Вероятность, статистика и прикладные исследования
в аграрном университете

Серия основана в 2012 году

Министерство сельского хозяйства Российской Федерации
ФГБОУ ВПО «Кубанский государственный аграрный университет»

Теория вероятностей и математическая статистика

Допущено
Министерством сельского хозяйства
Российской Федерации
в качестве учебного пособия для студентов
аграрных учебных заведений, обучающихся по направлениям 110400.62 «Агрономия», 110500.62 «Садоводство», 260100.62 «Продукты питания», 080100.62 «Экономика», 080500.62 «Бизнес-информатика», 230700.62 «Прикладная информатика», 230400.62 «Информационные системы и технологии», 270800.62 «Строительство», 080200. 62 «Менеджмент»

Серия: Вероятность, статистика и прикладные исследования в аграрном университете
Под редакцией: профессора А.И. Трубилина, профессора И.А. Кацко

Краснодар
2014

УДК 519.2
ББК 22.17я73
Б 811

Рецензенты:
кафедра математической статистики, эконометрики и актуарных расчётов
РГЭУ (РИНХ, г. Ростов н/Дону),
заведующая кафедрой – заслуженный деятель науки РФ,
доктор экономических наук,
профессор **Л.И. Ниворожкина**;

С.Г. Чефранов – доктор экономических наук, профессор, заведующий кафедрой организации и технологии защиты информации (Майкопский государственный технологический университет)

Б811 Бондаренко П.С. Теория вероятностей и математическая статистика: Учебное пособие для бакалавров / П.С. Бондаренко, Г.В. Горелова, И.А. Кацко. Краснодар: Кубанский ГАУ. – 2014. – 340 с., ил. (Серия: Вероятность, статистика и прикладные исследования в аграрном университете)

В учебном пособии рассмотрены основные вопросы теории вероятностей и математической статистики в соответствие с программой подготовки бакалавров.

Пособие отражает опыт преподавания авторами дисциплины в Кубанском государственном аграрном университете и Таганрогском государственном радиотехническом университете.

Предназначается для студентов, обучающихся по экономическим и агроинженерным специальностям, преподавателей, аспирантов и слушателей курсов повышения квалификаций, применяющих в своей практике вероятностные и статистические методы.

УДК 519.2
ББК 22.17я73

ISBN 978-5-94672-616-0

ОГЛАВЛЕНИЕ

Студенты всегда умнее своих учителей, поэтому мы попросили изучавших впервые этот материал откровенно поделиться своим мнением в форме «граффити»... Также цитируются знаменитые математики – подлинные слова, которыми они сопровождали свои работы... «граффити», несомненно, отражают реальные чувства..., что должно облегчить восприятие книги.

Р. Грэхем, Д. Кнут, О. Паташник.
Конкретная математика. Основания информатики, 1998

Замечательно, что науке, начинавшейся с рассмотрения азартных игр, суждено было стать важнейшим объектом человеческого знания...

П. Лаплас.
Аналитическая теория вероятностей, 1812.

При изучении наук примеры полезнее правил.

И. Ньютон

... истинной логикой этого мира является исчисление вероятностей, занимающееся нахождением величин вероятностей, которые учитывает или должен учитывать любой здравомыслящий человек.

Дж. Максвелл

Свои способности человек может узнать, только попытавшись приложить их.

Сенека

Есть люди, полагающие, что математика – это нудное занятие, которое всегда уныло и скучно; мы же находим математику развлечением и не стыдимся признаться в этом.

Р. Грэхем, Д. Кнут, О. Паташник.
Конкретная математика. Основания информатики, 1998

На книжной полке рядом стоят два тома Пушкина: первый и второй. Страницы каждого тома имеют вместе толщину 2 см, а обложка каждая 2мм. Червь прогрыз (перпендикулярно страницам) от первой страницы первого тома до последней страницы второго тома. Какой путь он прогрыз?

[Эта топологическая задача с невероятным ответом – 4 мм – совершенно недоступна академикам, но некоторые дошкольники легко справляются с ней.]

В. И. Арнольд. Задачи для детей от 5 до 15 лето, 2004

Конечно, мы будем учиться доказывать, но будем также учиться догадываться.

Д. Пойа

Введение

На современном этапе развития Российского общества, с переходом к рыночным отношениям, резко повысилась управленческая роль руководителя производства (предприятия). В связи с этим в нашей стране проводятся многочисленные исследования, перенимается и пропагандируется опыт зарубежных стран в области менеджмента и маркетинга. Одним из важнейших моментов в деятельности руководителя, менеджера, экономиста является принятие решений в условиях неопределённости. При этом наиболее разработанным инструментарием является математическая статистика, позволяющая решать *задачи принятия решений*[1] *в условиях вероятностной неопределённости* и имеющая достаточно распространённое программное обеспечение (например, в Excel).

В процессе всей своей жизни человек часто сталкивается с событиями и явлениями, исход которых заранее не определен. Например, студент не знает, какие именно вопросы задаст экзаменатор, служащий – сколько времени у него займет дорога на работу завтра (через неделю), инвестор – окупятся ли его инвестиции, страховщик – причину и размер выплаты страхового вознаграждения и т. д. Тем не менее, в подобных ситуациях, связанных с неопределенностью, человеку необходимо принимать решение.

Обычно принятию решений предшествует анализ известных данных (на основании предшествующего опыта, здравого смысла, интуиции и т.д.). Первые известные примеры обработки данных описаны в Ветхом Завете (Книга Чисел). Стремясь увидеть и обосновать закономерности в неопределенных процессах, человечество выработало целый арсенал методов, которые называются математической статистикой (прикладной статистикой или анализом данных).

Кратко рассмотрим основные направления методов математической статистики.

Выборочное наблюдение - решает задачу обобщения на всю совокупность результатов, полученных при изучении ее части, например, рейтинг политиков, анкетирование и т. д.

Проверка статистических гипотез – позволяет ответить на вопрос о достоверности принимаемого решения (например, обоснованность рейтинга популярности).

Дисперсионный анализ – изучает влияние факторных признаков на результативный (например, зависит ли производительность труда рабочего от стажа, возраста, стажа и возраста).

Корреляционно-регрессионный анализ – позволяет выявить связи и построить модели зависимости (например, какая зависимость существует между спросом на продукцию и курсом доллара).

Анализ временных рядов – рассматривает последовательности чисел, зависящих от времени и изучает их свойства (например, количество пятен на Солнце, характеризующее его активность по годам; курс доллара по дням и т. д.).

---

[1] О теории принятия решений см., например, учебник академика РАН О.И.Ларичева Теория и методы принятия решений, а также Хроника событий в Волшебных Странах.-М.:-Логос, 2000.-296с.:ил.

Последовательный анализ А. Вальда - один из первых разработанных методов, дающий возможность осуществлять контроль качества выпускаемой продукции.

Перечисленные выше (и другие) методы основываются на теоретических положениях теории вероятностей и являются классическими.

Существует ряд направлений, которые не используют предпосылки теории вероятностей. Наиболее известные из них указаны ниже.

Непараметрическая статистика, в отличие от классической (параметрической) не предполагает, что наблюдения подчиняются определенному закону распределения.

Бутстреп – способ обработки выборочных данных с помощью метода статистических испытаний (метода Монте-Карло), при котором выборку "размножают" и изучают устойчивость получаемых выводов.

Статистика интервальных данных. Реальные данные − это не числа, а интервалы (результат измерения плюс – минус погрешность), поэтому возникает необходимость в соответствующей статистике.

Статистика объектов нечисловой природы. Например: качественные оценки признака (машина – очень плохая, плохая, хорошая, очень хорошая), ранжировка (распределение студентов по росту), классификация (группировка студентов курса не по одному, а по ряду признаков, например: интересы, достаток, оценки, симпатии, антипатии) и т.д.

Указанные выше (и другие) методы прикладной статистики быстро входят в нашу жизнь посредством пакетов прикладных математических (Mathcad, Mat Lab), статистических (STATISTICA, STATGRAPHICS, STADIA, SAS, IBM SPSS, SYSTAT) и других программ, в которых предусматриваются средства обработки данных. Здесь даже у серьезных людей возникает вопрос «А зачем знать теоретические положения? Главное уметь кнопки нажимать!…». Ответом на этот вопрос может служить притча, приводимая Ф. Фишером.

В *N*–ской губернии разразилась эпидемия чумы. Крестьяне узнали, что можно «научно» выяснить причину болезни, для этого нужно посчитать коэффициент корреляции (Глава 4). Таким образом, они выяснили, что между количеством заболеваний в деревнях и количеством врачей существует прямая корреляционная зависимость. Поэтому крестьяне решили избавиться от врачей, посчитав их причиной болезни…

Притча, скорее всего, вызывает улыбку, однако, в менее очевидном случае мы зачастую поступаем аналогично.

Поэтому, несомненно, необходимо знать основные идеи и методы прикладной статистики, условия их применения, а затем «нажимать кнопки».

Методически более целесообразно изучать анализ данных на компьютере в Excel, а затем, по мере возникновения соответствующих вопросов, переходить к профессиональным программам.

Настоящее учебное пособие призвано помочь студентам в изучении основ теории вероятностей и математической статистики. Оно состоит из двух частей, разбитых на главы - две по теории вероятностей и две по математической статистике, предусматривающие сдачу четырех домашних контрольных работ (индивидуальные задания к главам 1-4). Первые две контрольные работы содержат 31 вариант двух уровней: уровень А - средней степени сложности, уровень В - повышенной сложности (рекомендуется для студентов, обучающихся по специальности «Прикладная информатика» и другим направлениям IT). Сочетание заданий по уровню сложности может варьироваться преподавателем.

Первая часть пособия посвящена теории вероятностей.

Теория вероятностей - это математическая дисциплина, изучающая закономерности, происходящие в массовых однородных случайных явлениях и процессах.

С возникновением теории вероятностей наука получила мощный аппарат исследования случайных явлений и процессов, до этого исследовались лишь детерминированные явления и опыты, в которых первоначальные условия однозначно позволяли определить исход. Между тем, случайные явления присутствуют во многих областях науки (биологии, генетике, агрономии, экономике, демографии, технике и т.д.), когда заранее невозможно предсказать результат опыта.

Исторически зарождение и развитие теории вероятностей связано с азартными играми, в которых требовалось обосновать то или иное решение. Классическим примером является задача рассматриваемая ниже.

Задача. Двое играют в безобидную игру (шансы выиграть у обоих одинаковы). Договариваются, что всю ставку забирает игрок, выигравший первым 6 партий. Как правильно разделить ставку, если игра остановилась при счете 5:3?

Решение. Для выигрыша первому игроку достаточно выиграть одну партию, второму игроку необходимо подряд выиграть три партии. Всего три партии предполагает $2\cdot2\cdot2=8$ исходов (каждая партия имеет два исхода: выиграл, проиграл). В пользу второго игрока только один исход из 8 возможных, а в пользу первого - 7 исходов. Поэтому справедливо разделить ставку пропорционально шансам выиграть, то есть 7:1.

Решение подобных задач Б. Паскалем, П. Ферма, Х. Гюйгенсом послужило появлению теории вероятностей. Дальнейшее ее развитие связано с именами Я. Бернулли, С. Пуассона, А. Муавра, П. Лапласа, К. Гаусса, П. Чебышева, А. Маркова, А. Ляпунова, А. Хинчина, А. Колмогорова и др.

Как и всякая математическая теория, с точки зрения аксиоматического подхода, теория вероятностей занимается изучением соотношений между неопределяемыми объектами (понятиями). В геометрии неопределяемые понятия это точка, прямая и плоскость, аналогично, в теории вероятностей неопределяемые понятия это элементарные события (исходы) (ω_i) и пространство элементарных событий ($\Omega=\{\omega_i\}$).

Пример В1. Монета подбрасывается один раз. Возможные элементарные исходы: выпал герб - ω_1, выпала решка - ω_2. Пространство элементарных событий $\Omega=\{\omega_1, \omega_2,\}$.

Пример В2. Игральная кость подбрасывается один раз. Элементарные события: ω_1- появление 1, ω_2 -2, ω_3 -3, ω_4 -4, ω_5 -5, ω_6 -6. Пространство элементарных событий $\Omega=\{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$.

Вероятность события – это число, имеющее ту же природу, что и расстояние в геометрии или масса в теоретической механике и всегда связанное с каким-либо пространством элементарных событий, природа которого не имеет значения. Понятие вероятности обычно строится на интуитивных соображениях (например, вероятность появления герба при подбрасывании симметричной монеты очевидно равна 1/2) и связано с понятием *статистической устойчивости* относительной частоты события при большом числе опытов (относительная частота события $\mu(A)=\frac{m}{n}$, где m - число появлений события A в серии из n опытов). При подбрасывании монеты достаточно большое число раз относительная частота появлений герба будет колебаться около 0,5, следовательно, можно говорить, что вероятность появления герба равна 0,5 (табл.1). Наличие устойчивости относительной частоты появления события позволяет судить о вероятности, как об объективной характеристике события в данном опыте, имеющей вполне определенное значение, независимо от того, будут проводиться опыты или нет. (То есть имеет место *вероятностная неопределённость*.)

Целью современной теории вероятностей является выявление общих закономерностей и зависимостей, а также описание физических явлений с помощью абстрактных моделей [4].

Вторая часть пособия посвящена математической статистике. Математическая статистика - это раздел математики, в котором изучаются математические методы систематизации, обработки и использования статистических данных для научных и практических выводов.

Математическая статистика использует математический аппарат и выводы теории вероятностей. Связующим звеном между теорией вероятностей и математической статистикой является закон больших чисел и так называемые предельные теоремы. В частности, закон больших чисел аргументирует применение средней арифметической в качестве оценки математического ожидания и относительной частоты появления события как оценки вероятности. Последнее обосновывает понятие статистической устойчивости**.**

Всю жизнь человек вынужден принимать решения: в личной сфере (в какой вуз поступать, с кем общаться, как учиться); в общественной (посещать вечера, театры, митинги, собрания, выборы); в производственной (определение факторов, существенно влияющих на урожайность, качество материалов и т.д.); научной (выдвижение и проверка научных гипотез).

Принятие решений обычно преследует одну из целей: прогнозирование будущего состояния процесса (объекта); управление (т.е. как следует изменять одни параметры объекта (процесса), чтобы другие параметры приняли желаемое значение); объяснение внутренней структуры объекта (процесса).

Одним из основных подходов к обоснованию и последующему принятию решений является статистический.

Статистические методы обработки данных можно классифицировать по нижеследующим признакам.

По способу получения экспериментальных данных:

- ✓ активный эксперимент;

- ✓ пассивный эксперимент (выборочное или сплошное наблюдение).

По цели обработки данных:

- ✓ описательные (получение и сравнение числовых характеристик экспериментальных данных) – анализ вариационных рядов, выборочный метод, и проверка статистических гипотез, и другие;
- ✓ аналитические (количественная оценка и анализ зависимостей описывающих изучаемые объекты (процессы) – дисперсионный анализ, регрессионный анализ, анализ рядов динамики, и другие).

В математической статистике предполагается, что результаты опытных данных и наблюдений являются реализацией различных случайных процессов, имеющих те или иные законы распределения (причем неизвестные заранее), а иногда и детерминированные составляющие (регрессионный анализ). Отсюда вытекают основные задачи математической статистики:

1) организация наблюдений;
2) нахождение по результатам выборочных наблюдений оценок числовых характеристик всей совокупности и исследование точности их приближения (выборочный метод);
3) решение вопроса согласования результатов оценивания с опытными данными (проверка статистических гипотез);
4) оценка существенности влияния факторных признаков на результативный (дисперсионный анализ);
5) выявление аналитической зависимости между наблюдениями факторных и результативных признаков (корреляционно-регрессионный анализ).

А. Вальд говорил, что "математическая статистика - это теория принятия решений в условиях неопределенности".

По существу математическая статистика дает единственный, математически обоснованный аппарат для решения задач управления и прогнозирования при отсутствии явных закономерностей (наличии случайностей) в изучаемых процессах.

В основу упражнений и контрольных работ положены сборники задач и учебные пособия, разработанные на кафедре статистики и прикладной математики КГАУ.

Часть I. Теория вероятностей

Предмет теории вероятностей и математической статистики настолько серьезен, что полезно не упускать случая сделать его немного занимательным.
Перефразированный Б. Паскаль

Глава 1. Случайные события

1.1. Алгебра событий

Одним из основных понятий теории вероятностей является опыт. Под опытом понимается выполнение комплекса условий, в результате которого происходят или не происходят определенные события (факты).

Простейшие неразложимые результаты опыта называются *элементарными событиями* (ω_i), а вся совокупность элементарных событий называется *пространством элементарных событий* $\Omega=\{\omega i\}$. С каждым опытом связано свое пространство элементарных событий Ω (примеры В1, В2).

Любое конечное или счетное[2] подмножество Ω называется *событием*. Различают три типа событий:

1) достоверные (Ω),
2) случайные,
3) невозможные (Ø или $\overline{\Omega}$).

События обычно обозначают первыми прописными буквами латинского алфавита: *A, B, C,…*. Например, в примере В2 событие $A=\{\omega_2, \omega_4, \omega_6\}$ – появление четного числа при подбрасывании игральной кости.

События *A* и *B* *несовместны*, если в результате одного опыта они не могут происходить одновременно, в противном случае - *совместны*. Например, при одном подбрасывании монеты не могут одновременно появиться герб и решетка.

Элементы последовательности событий $A_1, A_2, \ldots, A_n$ *попарно несовместны,* если любые два из них несовместны. Например, при подбрасывании игральной

---

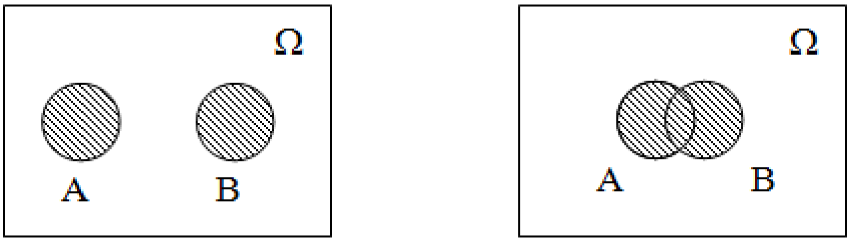
[2]Счетным называется множество, элементам которого можно поставить в соответствие ряд натуральных чисел, например, последовательность $\left\{\frac{1}{n}\right\}$, где $n \in N$ - счетное множество, так как можно установить соответствие между элементами последовательности и множеством натуральных чисел:

$$\begin{array}{cccccccc} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \ldots & \frac{1}{n} & \ldots \\ \updownarrow & \updownarrow & \updownarrow & \updownarrow & \updownarrow & \ldots & \updownarrow & \ldots \\ 1 & 2 & 3 & 4 & 5 & \ldots & n & \ldots \end{array}$$

кости никакие два элементарные исхода (появление цифр 1, 2, 3, 4, 5, 6) не могут произойти одновременно.

Несколько событий *равновозможны*, если ни одно из них не имеет объективного преимущества перед другими. Например, элементарные исходы при подбрасывании монеты, игральной кости. События $A_1,...,A_n$ образуют *полную группу*, если в результате опыта кроме этих событий ничего не может произойти. Обычно Ω изображают на плоскости в виде некоторой области, а ω_i в виде точек этой области, устанавливая, таким образом, соответствие между событиями и точечными множествами. Над событиями вводятся операции, совпадающие с операциями над множествами: сумма, произведение, отрицание.

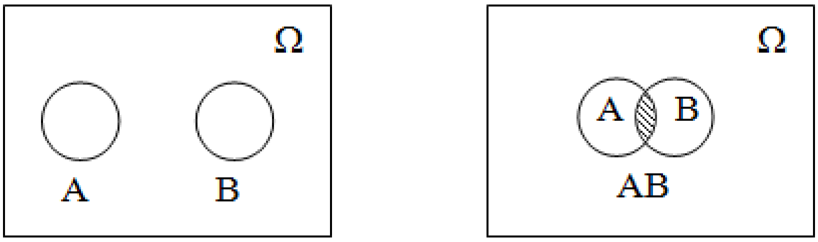
1. *Суммой* событий A и B называется такое третье событие $A+B$ (или $A\cup B$), которое заключается в наступлении хотя бы одного из событий или A, или B (рис. 1).



События несовместны   События совместны

Рис. 1 - Сумма событий

2. *Произведением* двух событий A и B называется такое третье событие AB (или $A\cap B$), которое заключается в наступлении событий A и B одновременно. Если события A и B несовместны, то $A\cdot B=\varnothing$ (рис.2).



События несовместны   События совместны

Рис. 2 - Произведение событий

3. *Отрицанием* события А называется событие $\overline{A}$ (не А), заключающееся в не наступлении события А ($A+\overline{A}=\Omega$, $A\cdot\overline{A}=\overline{\Omega}$). Причем, если в результате опыта может произойти событие А, то может произойти и обратное ему событие $\overline{A}$ (рис.3).

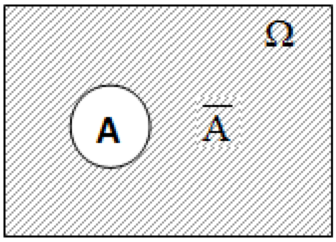


Рис. 3 - $\overline{A}$ -отрицание события A

Если наступление события A приводит к наступлению события B и наоборот (наступление B влечет наступление A), то события A и B *равны* (A=B).

Пусть S – множество всех подмножеств Ω, для которого выполняются следующие свойства:

1) если $A \in \mathrm{S}$ и $\mathrm{B} \in \mathrm{S}$, то $A + B = A \cup B \in S$,
2) если $A \in \mathrm{S}$ и $\mathrm{B} \in \mathrm{S}$, то $A \cdot B = A \cap B \in S$,
3) если $A \in \mathrm{S}$, то $\overline{A} \in S$,

тогда множество S называется *алгеброй событий.*

Замечание.
1. При более точном подходе достаточно одного из свойств 1) или 2), так как одно из них следует из другого.
2. При расширении операций сложения и умножения на случай счетного множества событий, алгебра событий S называется σ- алгеброй или борелевской алгеброй [5]:

$$\textit{если } A_i \in S, \mathrm{i} \in \mathrm{N}, \text{ тт}$$

$$\sum_{i=1}^{\infty} A_i \in S, \ \prod_{i=1}^{\infty} A_i \in S.$$

Контрольные задания 1.1

1. Являются ли несовместными следующие события:
   а) опыт - бросание двух монет. События:
   A_1 – появление двух гербов, A_2 – появление двух цифр;
   б) опыт – три выстрела по мишени. События:
   B_1 – хотя бы одно попадание, B_2 – хотя бы один промах;
   в) опыт – бросание двух игральных костей. События:
   C_1 – хотя бы на одной кости появилось три очка,
   C_2 – появление четного числа очков на каждой кости;
   г) опыт – извлечение двух шаров из урны, содержащей белые и черные шары. События:
   D_1 – взято два белых шара, D_2 – оба извлеченных шара одного цвета;
   д) опыт – покупка двух лотерейных билетов. События:
   E_1 – выиграют два билета, E_2 – выиграет хотя бы один билет,
   E_3 – выиграет только один лотерейный билет;
   е) опыт – лифт отправляется с 10 пассажирами и останавливается на пяти этажах. События:
   F_1 – на первых четырех остановках вышло не более 9 человек,
   F_2 – на последней остановке вышел хотя бы один человек.
2. Образуют ли полную группу следующие события:
   а) опыт – два выстрела по мишени. События:
   A_1 – два попадания в мишень, A_2 – хотя бы один промах по мишени;

б) опыт – бросание двух игральных костей. События:
B_1 – сумма очков на верхних гранях больше 3,
B_2 – сумма очков на верхних гранях равна 3;
в) опыт – посажено четыре зерна. События:
C_1 – взошло одно зерно, C_2 – взошло два зерна,
C_3 – взошло три зерна, C_4 – взошло четыре зерна.
г) покупатель посещает три магазина. События:
D_1 – покупатель купит товар хотя бы в одном магазине,
D_2 – покупатель не купит товар ни в одном магазине.

3. Являются ли равновозможными следующие события:
а) опыт – выстрел по мишени. События:
A_1 – попадание при выстреле, A_2 – промах при выстреле;
б) опыт – бросание двух игральных костей. События:
B_1 – произведение очков на верхних гранях равно 12,
B_2 – сумма очков на верхних гранях равна 9;
в) опыт - бросание двух монет. События:
C_1 – появление двух гербов, C_2 – появление двух цифр,
C_3 – появление одного герба и одной цифры.
Элементарные исходы: $Г_1$, $Р_1$, $Г_2$, $Р_2$ - появление гербов и решек при бросании первой и второй монеты соответственно.

Замечание. В данном случае (пункт в)) возможно несколько подходов, связанных с различными статистическими моделями физических частиц [4, 5].

1) Модель (статистика) Больцмана-Максвелла (частицы различимы, хотя известно, что таких частиц в природе не существует). Для опыта с монетами: $Г_1Г_2$, $Г_1Р_2$, $Г_2$ $Р_1$, $Р_1Р_2$ (монеты различимы).
2) Модель Бозе-Эйнштейна (частицы неразличимы, например: фотоны, атомные ядра, атомы с четным числом частиц). Для опыта с монетами: $Г_1Г_2$, $Г_1Р_2$, $Р_1Р_2$ (монеты не различимы).
3) Модель Ферми-Дирака (частицы не могут принимать одинаковые значения, например: электроны, нейтроны, протоны). Для опыта с монетами: $Г_1Р_2$, $Г_2$ $Р_1$.

Опыты с монетами показывают, что здесь выполняется статистика Больцмана-Максвелла, поэтому нельзя заранее, до проведения эксперимента, утверждать истинность той или иной модели.

4. Брошены 3 монеты.
Составить события, образующие полную группу.
Сколько равновозможных исходов образуют полную группу событий?
Укажите элементарные события, не образующие полной группы событий.
5. Приведите примеры:
а) трех событий, образующих полную группу событий;
б) трех событий, равновозможных и несовместных, но не образующих полной группы событий;
в) двух событий, несовместных и образующих полную группу событий, но не равновозможных.

1.2. Алгоритм, примеры и контрольные задания

Алгоритм. Пусть рассматривается опыт и связанное с ним сложное событие A, тогда A можно выразить через элементарные события, для этого:
1. Рассматриваются простейшие элементарные события, образующие для данного опыта полную группу: $A_1, A_2, \dots, A_n$;
2. Из элементарных событий с помощью операций сложения, умножения и отрицания формируют необходимое, для решения данной задачи сложное событие.

Пример 1.2.1. Стрелок произвел 3 выстрела по мишени, элементарные события:

A_1 - попал при 1-ом выстреле; $\overline{A_1}$ - не попал при 1-ом, выстреле;
A_2 - попал при 2-ом выстреле; $\overline{A_2}$ - не попал при 2-ом, выстреле;
A_3 - попал при 3-ем выстреле; $\overline{A_3}$ - не попал при 3-ем, выстреле.

Выразим через A_1, A_2, A_3 и их отрицания следующие события:

а) одно попадание: $A = A_1 \cdot \overline{A_2} \cdot \overline{A_3} + \overline{A_1} \cdot A_2 \cdot \overline{A_3} + \overline{A_1} \cdot \overline{A_2} \cdot A_3$,

б) три промаха: $B = \overline{A_1} \cdot \overline{A_2} \cdot \overline{A_3}$,

в) три попадания: $C = A_1 \cdot A_2 \cdot A_3$,

г) хотя бы один промах: $D = A + B + A_1 \cdot A_2 \cdot \overline{A_3} + A_1 \cdot \overline{A_2} \cdot A_3 + \overline{A_1} \cdot A_2 \cdot A_3$

или $D = \overline{A_1} + \overline{A_2} + \overline{A_3}$,

д) не менее 2-х попаданий: $E = A_1 \cdot A_2 \cdot \overline{A_3} + A_1 \cdot \overline{A_2} \cdot A_3 + \overline{A_1} \cdot A_2 \cdot A_3 + C$,

е) не больше 1-го попадания: $F = A + B$,

ж) попадание в мишень после 1-го выстрела: $G = \overline{A_1}\left(\overline{A_2} A_3 + A_2 \overline{A_3} + A_2 A_3\right)$

Контрольные задания 1.2

1. Производится три выстрела по мишени. Рассматриваются события: A_1 – попадание в цель первым выстрелом; A_2 – попадание в цель вторым выстрелом; A_3 – попадание в цель третьим выстрелом. Определить, каким событиям равносильны следующие события: 1) $A_1 + A_2 + A_3$; 2) $A_1 A_2 A_3$;
 3) $\overline{A_1 + A_2 + A_3}$; 4) $\overline{A_1}\ \overline{A_2}\ \overline{A_3}$; 5) $A_1 \overline{A_2}\ \overline{A_3} + \overline{A_1} A_2 \overline{A_3} + \overline{A_1}\ \overline{A_2} A_3$; 6) $\overline{\overline{A_1} + \overline{A_2} + \overline{A_3}}$; 7) $A_1 + \overline{A_1} A_2 + \overline{A_1}\ \overline{A_2} A_3$; 8) $(\overline{A_1} + \overline{A_2}) A_3$.
2. Монета подбрасывается три раза. Рассматриваются события A_i – появление герба при i – ом подбрасывании (i=1,2,3). Представить в виде сумм, произведений и сумм произведений событий A_i и $\overline{A_i}$ следующие события: A - появились все три герба; B - появились все три цифры; C - появился хотя бы один герб; D - появилась хотя бы одна цифра; E - появился только один герб; F - появилась только одна цифра.

1.3. Вероятность события

Существует несколько подходов к определению вероятности события. Рассмотрим основные из них.

Аксиоматическое определение вероятности.

Вероятность события - это численная мера объективной возможности его появления.

Аксиомы вероятности:

1. Каждому событию A ставится в соответствие неотрицательное число p, которое называется вероятностью события A:

$$P(A)=p \geq 0, \text{ г, где } \in S,\ S \subseteq \Omega.$$

2. Если события $A_1, A_2, ..., A_n$ несовместны, то верно равенство:

$$P(A_1 + A_2 + ... + A_n) = P(A_1) + P(A_2) + ... + P(A_n),$$

где $A_i \in S\ (i = 1,2,...,n),\ S \subseteq \Omega.$.

3. $P(\Omega) = 1$,

где Ω - истинное (достоверное) событие.

Пространство элементарных событий Ω с заданной в нем алгеброй S (или σ - алгеброй) и определенной на S вероятностью – неотрицательной мерой $P(A)$, $A \in S$ называется *вероятностным пространством* и обозначается (Ω, S, P). Вероятностное пространство служит математической моделью любого случайного явления в теории вероятностей.

Аксиоматический подход не указывает, как конкретно находить вероятность, поэтому для решения задач целесообразно использовать подходы к определению вероятности, которые перечислены ниже.

Классическое определение вероятности.

Пусть события $A_1, A_2, ..., A_n \in S$ образуют множество элементарных событий. Тогда события, которые приводят к наступлению события A, называются благоприятствующими исходами для события A, $m(A)$ - число благоприятствующих исходов.

Вероятностью события A называется отношение числа исходов благоприятствующих наступлению события A к числу всех возможных исходов

$$P(A) = \frac{m(A)}{n}. \tag{1.3.1}$$

В примере введения B1 вероятность появления герба $P(A) = \frac{1}{2}$; в примере B2 вероятность появления числа больше четырёх $P(A) = \frac{2}{6} = \frac{1}{3}$. Из классического определения следуют свойства вероятности:

1) $0 \leq P(A) \leq 1$,

2) P(Ω)=1,

3) $P(\overline{\Omega})$=0.
$A+\bar{A}=\Omega$ - достоверное событие, поэтому

$$P(A)+P(\bar{A})=1 \text{ или } P(\bar{A})=1-P(A).$$

Статистическое определение вероятности.

Пусть проводится серия опытов (n раз), в результате которой наступает или не наступает некоторое событие A (m раз), тогда отношение $\frac{m}{n}$, при $n\to\infty$, называется статистической вероятностью события A.

Таблица 1 - Опыты по подбрасыванию монеты

Опыт	Число опытов, n	Появление герба, m	$\frac{m}{n}$
Опыт Керриха	10000	5087	0,5087
Опыт Бюффона	4040	2048	0,5069
1Опыт Пирсона	12000	6019	0,5016
2Опыт Пирсона	24000	12012	0,5005

Из таблицы 1, описывающей опыт подбрасывания монеты, следует, что $\frac{m}{n}\to 0{,}5$, где $\frac{m}{n}$ - относительная частота или частость события A .

Иногда, при рассмотрении бесконечных множеств удобно рассматривать геометрическое определение вероятности.

Геометрическое определение вероятности.

Геометрической вероятностью события A называется отношение меры области, благоприятствующей появлению события A, к мере всей области.

Пример 1.3.1. Найти вероятность того, что точка случайным образом брошенная в квадрат $ABCD$ со стороной 4 попадет в квадрат $A_1E_1C_1D_1$ со стороной 3, находящийся внутри $ABCD$.

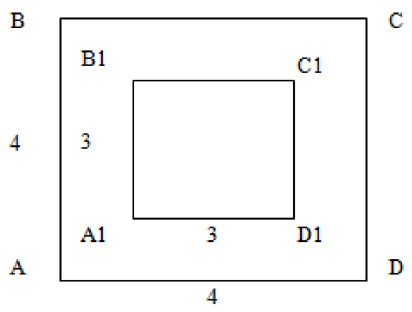


Рис. 4

Решение.

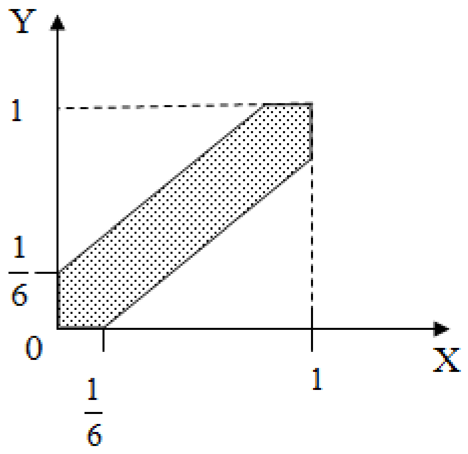
Вероятность события определяется как отношение меры части области (в данном случае площади), благоприятствующей событию A -$S_{A_1B_1C_1D_1}$, к мере всей области - S_{ABCD}.

$$P(A)=\frac{S_{A_1B_1C_1D_1}}{S_{ABCD}}=\frac{9}{16}$$

Пример 1.3.2. Два лица A и B договорились встретиться в определенном месте в промежутке времени от 9^{00} до 10^{00}. Каждый из них приходит наудачу, независимо от другого и ожидает 10 минут. Какова вероятность того, что они встретятся?

Решение. Рассмотрим прямоугольную систему координат XOY, в качестве единиц масштаба выберем часы. Пусть x и y – моменты прихода A и B соответственно. Необходимым и достаточным условием встречи является выполнение неравенства $|y\text{-}x| \le 1/6$ (или $x\text{-}1/6 \le y \le x\text{+}1/6$). Тогда все возможные исходы будут являться точками квадрата 1×1.Заштрихованной области квадрата, ограниченной сторонами квадрата, а также прямыми $y=x\text{-}1/6$ и $y=x\text{+}1/6$, соответствуют исходы благоприятствующие встрече (рис.5).

Искомая вероятность равна отношению площади заштрихованной фигуры к площади всего квадрата.



$$P = \frac{1^2 - \left(\frac{5}{6}\right)^2}{1^2} = 1 - \frac{25}{36} = \frac{11}{36}.$$

Рис. 5

Контрольные задания 1.3

1. Брошена игральная кость. Найти вероятность того, что на ее верхней грани появится: а) шесть очков; б) нечетное количество очков; в) не менее четырех очков; г) не более двух очков.
2. Набирая номер телефона, абонент забыл две цифры и, помня лишь, что они различны, набрал их наудачу. Найти вероятность того, что набраны нужные цифры.
3. Брошены две игральные кости. Найти вероятность того, что:
   а) на обеих костях появится одинаковое число очков;
   б) хотя бы на одной кости появится два очка;
   в) сумма выпавших очков равна пяти, а произведение шести;
   г) сумма очков не превосходит 6;
   д) произведение числа очков не превосходит 6;
   е) произведение очков делится на 6.
4. Из 100 посаженых семян проросло 78.
   а). Какова статистическая вероятность прорастания семян?
   б). Каков процент всхожести семян?
5. Относительная частота (частость) работников предприятия, имеющих высшее образование, равна 0,15. Определить число работников, имеющих высшее образование, если всего на предприятии работает 40 человек.

6. В отрезке AB длины 3 случайно появляется точка C. Определить вероятность того, что расстояние от точки C до B превосходит 1.
7. В круг радиуса 5 вписан треугольник наибольшей площади. Определить вероятность попадания в треугольник точки, случайно брошенной в круг.
8. Расстояние от пункта M до пункта N автобус проходит за 2 минуты, а пешеход за 15 минут. Интервал движения автобусов 25 минут. Пешеход в случайный момент времени отправляется из M в N пешком. Найти вероятность того, что его в пути догонит автобус.
9. Какой толщины должна быть монета радиуса r, чтобы вероятность падения на ребро была 1/3. *Указание*. Необходимо рассмотреть монету как прямой круговой цилиндр с радиусом основания r, вписанный в шар радиуса R. Монета бросается на клейкую поверхность. (Знаменитый математик фон Нейман, впервые услышав эту задачу, дал ответ с точностью трех знаков после запятой через 20 секунд в присутствии публики, которой потребовалось для решения значительно больше времени.)
10. *Задача Бюффона*. Игла длины l бросается на плоскость, разграфленную параллельными прямыми, разделенными расстояниями L ($L>l$). Все положения центра иглы и все ее направления одинаково вероятны. Найти вероятность того, что игла пересечет какую-нибудь из линий.
11. *Парадокс Бертрана.* Для некоторой окружности случайно выбирается хорда. Найти вероятность того, что эта хорда длиннее стороны правильно треугольника, вписанного в данную окружность, если:
    а) середина хорды равномерно распределена в круге;
    б) направление хорды задано, а ее середина равномерно распределена на диаметре, перпендикулярном направлению;
    в) один конец хорды закреплен, а другой равномерно распределен на окружности. (Парадокс заключается в том, что вероятности для а), б), в) различны.)

12. *Вероятность «черной пятницы».* Доказать, что 13 число месяца с большей вероятностью приходится на пятницу, чем на другие дни недели.

1.4. Комбинаторика

Нас в комбинаторике будет интересовать возможность определения количественно различных подмножеств конечных множеств, для вычисления вероятности классическим способом.

Комбинаторика (комбинаторный анализ) - раздел дискретной математики, посвященный решению задач выбора и расположения элементов некоторого, обычно конечного, множества в соответствии с заданными правилами[3]. Рождение комбинаторики связано с работами Б. Паскаля и П. Ферма по поводу азартных игр (Введение), большой вклад внесли Лейбниц, Бернулли, Эйлер. В настоящее время интерес к комбинаторике связан с развитием компьютеров.

Правило произведения. Пусть из некоторого конечного множества

1-й объект можно выбрать k_1 способами,
2-ой объект - k_2 способами,
…..…………………………..…, (1.4.1)
n-ый объект - k_n способами.

Тогда произвольный набор, перечисленных n объектов, из данного множества можно выбрать $k_1 \cdot k_2 \cdot \ldots \cdot k_n$ способами.

Пример 1.4.1. Сколько существует трехзначных чисел с разными цифрами.

Решение. В десятичной системе исчисления десять цифр: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9.

---
[3]Сачков В.Н. Комбинаторный анализ // Математическая энциклопедия.–М., 1979. – Т.2.-с.974.

На первом месте может стоять любая из девяти цифр (кроме нуля). На втором месте - любая из оставшихся 9 цифр, кроме выбранной. На последнем месте любая из оставшихся 8 цифр.

По правилу произведения, 9·9·8 = 648 трёхзначных чисел имеют разные цифры.

Правило суммы. При выполнении условий (1.4.1), любой из объектов можно выбрать $k_1 + k_2 + k_3 + \ldots + k_n$ способами.

Пример 1.4.2. Сколько существует способов выбора одного карандаша из коробки, содержащей 5 красных, 7 синих, 3 зеленых карандаша.

Решение. Один карандаш, по правилу суммы, можно выбрать 5+7+3=15 способами.

Обычно в комбинаторике рассматривается идеализированный эксперимент по выбору наудачу k элементов из n. При этом элементы: а) не возвращаются обратно (схема выбора без возвращений); б) возвращаются обратно (схема выбора с возвращением).

I. Схема выбора без возвращений. Размещением из n элементов по k называют любой упорядоченный набор из k элементов, принадлежащих n элементному множеству. Различные размещения отличны друг от друга или порядком элементов, или составом.

Число *размещений из n элементов по k* обозначается $A_n^к$ и вычисляется по формуле

$$A_n^k = \frac{n!}{(n-k)!}, \qquad (1.4.2)$$

где $n! = 1\cdot 2\cdot 3\cdot \ldots \cdot n$, 1!=1 , 0!=1.

Пример 1.4.3. В соревнованиях участвует 10 человек, трое из них займут 1 , 2 , 3 место. Сколько существует различных вариантов?

Решение. Число различных вариантов равно

$$\mathrm{A}_{10}^3 = \frac{10!}{(10-3)!} = \frac{7!\cdot 8\cdot 9\cdot 10}{7!} = 8\cdot 9\cdot 10 = 720\,.$$

Перестановкой из n элементов называют размещение из n элементов по n. Число перестановок из n элементов обозначают P_n и вычисляют по формуле

$$P_n = A_n^n = \frac{n!}{0!} = \frac{n!}{1} = n!\,. \qquad (1.4.3)$$

Пример 1.4.4. Сколько существует способов расстановки 10 книг на полке?

Решение. Общее число способов расстановки определяется как число перестановок из 10 элементов и равно P_{10} = 10! = 3 628 800.

Сочетанием из n элементов по к называется любой набор из k элементов, принадлежащих n элементному множеству. Различные сочетания отличаются друг от друга только составом.

Число сочетаний из n элементов по k обозначается C_n^k и вычисляется по формуле

$$C_n^k = \frac{A_n^k}{P_к} = \frac{n!}{(n-k)!\,k!}\,. \qquad (1.4.4)$$

Числа, которые определяются по формуле (1.4.4), называются *биномиальными коэффициентами*.

Справедливы тождества:

$$C_n^{\kappa} = C_n^{n-\kappa}, \quad C_{n+1}^{\kappa} = C_n^{\kappa} + C_n^{\kappa-1}, \quad C_n^0 = 1, \quad C_n^1 = n,$$

$$C_n^0 + C_n^1 + \ldots + C_n^n = 2^n, \sum_{\kappa=0}^{n} \left(C_n^{\kappa}\right)^2 = C_{2n}^n.$$

Пример 1.4.5. Сколько существует способов выбора трех человек из десяти.
Решение. В данном случае при выборе для нас важен только состав наборов по три человека, порядок выбора роли не играет, поэтому в отличие от примера 1.4.3 число способов выбора подсчитаем по формуле сочетаний

$$C_{10}^3 = \frac{10}{(10-3)!3!} = \frac{7!\cdot 8\cdot 9\cdot 10}{7!\cdot 1\cdot 2\cdot 3} = 120.$$

II.Схема выбора с возвращениями. Если при выборе *k* элементов из *n* - элементы возвращаются обратно и упорядочиваются, то говорят, что это *размещения с повторениями.*
Число размещений с повторениями

$$\overline{A_n^k} = n^k. \tag{1.4.5}$$

Пример 1.4.6. В гостинице 10 комнат, каждая из которых может разместить четырех человек. Сколько существует вариантов размещения, прибывших четырех гостей?
Решение. Каждый следующий гость из 4 может быть помещён в любую из 10 комнат, поэтому общее число размещений, по формуле размещений с повторениями, равно $\overline{A_{10}^4} = 10^4 = 10000.$

Если при выборе к элементов из n элементы возвращаются обратно без последующего упорядочивания, то говорят, что это *сочетания с повторениями.*
Число сочетаний с повторениями из *n* элементов по *k*

$$\overline{C_n^k} = C_{n+k-1}^k = \frac{(n+k-1)!}{(n-1)!k!}. \tag{1.4.6}$$

Пример 1.4.7. В магазине продается 10 видов тортов. Очередной покупатель выбил чек на три торта. Считая, что любой набор товаров равновозможен, определить число возможных заказов.
Решение. Число равновозможных заказов по формуле (1.4.6) равно

$$\overline{C_{10}^3} = C_{10+3-1}^3 = C_{12}^3 = \frac{12!}{(12-3)!3!} = \frac{9!\cdot 10\cdot 11\cdot 12}{9!3!} = 220.$$

III. Схема упорядоченных разбиений. Пусть $k_1, k_2, \ldots, k_r$ – целые числа, такие, что $k_1 + k\kappa_2 + \ldots + k_r = n$, $k_i \geq 0$ (i=1, 2,..., r).Число способов, которыми генеральную совокупность из *n* элементов можно разделить на *r* упорядоченных частей (*r* подмножеств или r групп), из которых первая содержит k_1 элементов, вторая - k_2 элементов и *r*-тая - k_r элементов обозначается C_n (k_1, k_2,..., k_r) и вычисляется по формуле

$$C_n(k_1, k_2, \ldots, k_r) = \frac{n!}{k_1!k_2!\ldots k_r!}. \tag{1.4.7}$$

Числа, которые определяются по формуле (1.4.7), называются *полиномиальными коэффициентами*.

Пример 1.4.8. Девять человек размещается в гостинице в четырехместный, трехместный и двухместный номера. Сколько существует способов их размещения?

Решение. Число способов размещения по формуле (1.4.7) равно

$$C_9(4,3,2)=\frac{9!}{4!3!2!}=1260\,.$$

***Замечание**. Формулы размещений и сочетаний обычно используют в двух классических задачах: 1) размещения k различимых и не различимых шаров по n урнам с запретом (в каждой урне может находиться не более одного элемента) или без запрета (в урне может находиться любое число элементов; 2) выбора k упорядоченных и не упорядоченных шаров из урны с n шарами с возвращениями или без возвращений. Указанные способы выбора соответствуют физическим статистикам, рассмотренным в контрольных заданиях 1.1(№ 3 в). Эти задачи сведены таблицу 2.*

Таблица 2

<table>
<tr><td colspan="3">Задача размещения k шаров по n урнам</td><td></td></tr>
<tr><td>Тип элементов
Размещение</td><td>Элементы различимы</td><td>Элементы не различимы</td><td></td></tr>
<tr><td>Без запрета</td><td>$\overline{A_n^k}$
(статистика Больцмана-Максвелла)</td><td>$\overline{C_n^k}$
(статистика Бозе-Эйнштейна)</td><td>С возвращением</td></tr>
<tr><td>С запретом</td><td>A_n^k</td><td>C_n^k
(статистика Ферми-Дирака)</td><td>Без возвращения</td></tr>
<tr><td></td><td>Упорядоченный</td><td>Неупорядоченный</td><td>Выбор
Набор</td></tr>
<tr><td></td><td colspan="3">Задача выбора *k* шаров из урны с n шарами</td></tr>
</table>

Контрольные задания 1.4.

1. В сельскохозяйственном эксперименте проверяют влияние на урожайность трех различных факторов (например, применение удобрений, орошение, температура). Факторы имеют соответственно k_1, k_2, k_3 уровней. Сколько существует комбинаций или способов воздействия?
2. Сколько можно образовать различных инициалов, если каждый человек имеет одну фамилию, имя, отчество?
3. В азбуке Морзе буквы представляются последовательностями тире и точек с возможными повторениями. Сколько букв можно составить из 5 и менее символов?
4. Девять запечатанных пакетов с предложениями цены на аренду участков для бурения нефтяных скважин поступили утром в специальное агентство утренней почтой. Сколько существует различных способов очерёдности вскрытия конвертов с предложениями цены?
5. Компания имеет четыре отдела: по производству продукции, отдел снабжения, занимающийся обеспечением сырья, а также отделы менеджмента и маркетинга. Количество людей в каждом из отделов 55, 30, 21 и 13 соответственно. Каждый отдел собирается послать одного

представителя на ежегодную встречу с директором компании. Сколько различных групп для встречи можно составить из числа работников компании?

6. Для доступа в компьютерную сеть оператору необходимо набрать пароль из 4 цифр. Оператор забыл или не знает необходимого кода. Сколько всевозможных комбинаций он может составить для набора пароля:

   а) если цифры в коде не повторяются;

   б) если повторяются?

7. Директор корпорации рассматривает заявления о приёме на работу 10 выпускников университета. На одном из предприятий корпорации имеются три различных вакансии. Сколькими способами директор может заполнить эти вакансии?

1.5. Алгоритм, примеры и контрольные задания

Алгоритм. Рассмотрение в задаче количественно различных подмножеств конечных множеств, предполагает привлечение комбинаторики.

1.Для вычисления вероятности классическим способом определяют с помощью описанных выше правил и формул: *m(A)* - число исходов, благоприятствующих наступлению события *A*, *n*- общее число возможных исходов.

2. Находят вероятность события *A*, согласно формуле классического определения вероятности: $P(A) = \dfrac{m(A)}{n}$.

Пример 1.5.1. Набирая номер телефона, абонент забыл последние 3 цифры и набрал их наудачу, помня, что они различны. Найти вероятность того, что набраны нужные цифры.

Решение. Событие A - номер набран верно

$$P(A) = \frac{m(A)}{n} = \frac{1}{720},$$

где $m(A) = 1$ - т.к. только один набор из 3-х цифр является нужным, всего таких наборов $n = A_{10}^{3} = \dfrac{10!}{(10-3)!} = \dfrac{7!\cdot 8\cdot 9\cdot 10}{7!} = 720$.

Пример 1.5.2. В ящике 15 деталей, среди которых 10 окрашены. Сборщик наудачу выбрал 3 детали. Найти вероятность того, что детали окрашены.

Решение. Событие A - 3 детали окрашены.

$m(A) = C_{10}^{3} = \dfrac{10!}{(10-3)!3!} = \dfrac{7!\cdot 8\cdot 9\cdot 10}{7!\cdot 1\cdot 2\cdot 3} = 120$ - число благоприятствующих исходов.

$n = C_{15}^{3} = \dfrac{15!}{(15-3)!3!} = \dfrac{12!\cdot 13\cdot 14\cdot 15}{12!\cdot 1\cdot 2\cdot 3} = 455$ - общее число возможных исходов.

Имеем $P(A) = \dfrac{m(A)}{n} = \dfrac{120}{455} = \dfrac{24}{91}$.

Пример 1.5.3. В группе 12 студентов, среди которых 8 отличников. По списку отбирают 9. Найти вероятность того, что отберут 5 отличников.

Решение. Событие A - отобрали 5 отличников. $P(A) = \dfrac{m(A)}{n} = \dfrac{56}{220} = \dfrac{14}{55}$, так как по правилу произведения $m(A) = \mathrm{C}_8^5$.

$m(A) = C_8^5 \cdot C_4^4 = 56 \cdot 1 = 56$, где C_8^5 - число возможных наборов из 8 отличников по 5,

C_4^4 - число возможных наборов по 4 из остальных студентов;

n-общее число способов выбора из 12 студентов 9 равно:

$$C_{12}^9 = \frac{12!}{(12-9)!9!} = \frac{9!\cdot 10\cdot 11\cdot 12}{3!\cdot 9!} = 220 .$$

Пример 1.5.4. В коробке 5 красных, 3 зеленых и 2 синих карандаша. Наудачу без возвращения извлекают 3 карандаша. Найти вероятность следующих событий:

A – все извлеченные карандаши разного цвета,

B – все извлеченные карандаши одного цвета,

C – среди извлеченных карандашей 1 синий,

D – среди извлеченных карандашей в точности 2 одного цвета.

Решение. Всего в коробке 5+3+2=10 карандашей.

1.По правилу произведения $m(A) = 5\cdot 3\cdot 2$ =30 – число исходов, благоприятствующих наступлению события *A*; общее число способов выбора из 10 карандашей 3 вычисляется как число сочетаний из 10 по 3

$$n = C_{10}^3 = \frac{10!}{(10-3)!3!} = \frac{7!\cdot 8\cdot 9\cdot 10}{7!\cdot 1\cdot 2\cdot 3} = 120 ,$$

отсюда $$P(A) = \frac{30}{120} = \frac{1}{4}.$$

2.Если все извлечённые карандаши одного цвета, то это либо 3 красных, либо 3 зелёных (3 синих не может быть, т.к. их в коробке всего 2).

Поэтому, по правилу суммы

$$m(B) = C_5^3 + C_3^3 = \frac{5!}{(5-3)!\cdot 3!} + 1 = \frac{3!\cdot 4\cdot 5}{2!\cdot 3!} = 10 + 1 = 11.$$

n=120, следовательно, $P(B) = \dfrac{m(B)}{n} = \dfrac{11}{120}.$

3. Из 10 карандашей (т.к. по условию: 5 + 3 + 2=10 карандашей) один синий можно выбрать $C_2^1 = \dfrac{2!}{1!\cdot 1!} = 2$ - способами , 2 из оставшихся 8 карандашей не синего цвета можно выбрать: $C_8^2 = 28$– способами. Отсюда, по правилу произведения, *m* (*C*)= $C_2^1 \cdot C_8^2 = 2\cdot 28$=56

$$P(C) = \frac{m(C)}{n} = \frac{56}{120} = \frac{7}{15}.$$

4. Событие *D* - 2 карандаша одного цвета − произойдет, если из трех карандашей вытащили: 2 красных + (1 зелёный или 1 синий) или 2 зелёных + (1 красный или 1 синий), или 2 синих + (1 красный или 1 зелёный).

По правилу произведения: $C_5^2 \cdot C_5^1$ - число способов выбора 2 красных карандашей и 1 другого цвета; $C_3^2 \cdot C_7^1$- число способов выбора 2 зелёных карандашей и 1 другого цвета; $C_2^2 \cdot C_8^1$- число способов выбора 2 синих карандашей и 1 другого цвета.

Общее число исходов благоприятствующих наступлению события *D*, по правилу суммы равно $m(D) = C_5^2 \cdot C_5^1 + C_3^2 \cdot C_7^1 + C_2^2 \cdot C_8^1 = 79$. Следовательно,

$$P(D) = \frac{m(D)}{n} = \frac{79}{120}.$$

Пример 1.5.5. Лифт начинает движение с четырьмя пассажирами и останавливается на 10 этаже. Какова вероятность, что никакие два пассажира не выйдут на одном этаже.

Решение. Пусть все возможные случаи выхода пассажиров равновероятны, тогда первый пассажир имеет 10 возможностей выхода на 10 этажах, второй - 9, на 9 оставшихся этажах, третий - 8 на 8 оставшихся этажах, четвертый - 7. По правилу произведения, общее число исходов, благоприятствующих событию A (никакие два пассажира не выйдут на одном этаже), $m(A)=10\cdot 9\cdot 8\cdot 7=A_{10}^4$. Общее число вариантов выхода четырех пассажиров на 10 этажах равно числу размещений с возвращениями из 10 элементов по 4, $\overline{A_{10}^4}$.

Отсюда, $$P(A) = \frac{m(A)}{n} = \frac{10\cdot 9\cdot 8\cdot 7}{10^4} = \frac{9\cdot 8\cdot 7}{10^3} = \frac{504}{1000} = 0{,}504.$$

Пример 1.5.6. На полке стоит n книг, какова вероятность, что k из них находятся рядом ($k<n$).

Решение. В качестве элементарных событий, условно можно рассматривать различные подмножества пространства элементарных событий, полагая, что они имеют соответствующие свойства (неразложимости, равновозможности).

I способ. Элементарные события – книги, которые могут стоять на n местах. В одном ряду на соседних местах k книг можно поставить (n-k+1) способами. Причем, по правилу произведения, m(А)=(n-k+1) k!(n-k)!, где k! – количество перестановок из k книг, стоящих рядом, (n-k)! – количество перестановок из (n-k) оставшихся книг.

Общее число перестановок из n книг равно n!.

Имеем, $P(A) = \dfrac{m(A)}{n} = \dfrac{(n-k+1)k!(n-k)!}{n!} = \dfrac{k!(n-k+1)!}{n!}$.

II способ. Рассмотрим в качестве элементарных событий все размещения из n книг по k, получим

$$P(A) = \frac{(n-k+1)\cdot P_к}{A_n^к} = \frac{(n-k+1)k!}{\dfrac{n!}{(n-k)!}} = \frac{(n-k+1)!k!}{n!},$$

где P_k – количество всех перестановок из k книг, стоящих рядом, A_n^k – количество всех размещений из n книг по k.

III способ. Рассмотрим в качестве пространства элементарных событий множество всех сочетаний из n книг по k, имеем

$$P(A) = \frac{(n-k+1)}{C_n^k} = \frac{(n-k+1)\cdot(n-k)!k!}{n!} = \frac{(n-k+1)!k!}{n!},$$

где C_n^k – количество всех сочетаний из n книг по k.

Контрольные задания 1.5.

1. Имеются две урны. В первой – 10 красных и 6 черных шаров. Во второй – 4 красных и 6 черных шаров. Из каждой урны вынимается по шару. Найти вероятность того, что: а) оба шара будут красными; б) из первой урны будет вынут красный шар, а из второй – черный; в) хотя бы один из вынутых шаров черный.
2. Из коробки, содержащей 8 пронумерованных жетонов, вынимают один за другим все находящиеся в ней жетоны и укладывают рядом. Найти вероятность того, что номера вынутых жетонов будут идти по порядку 1, 2, 3, 4, 5, 6, 7, 8.
3. Из пяти букв разрезной азбуки составлено слово «книга». Ребенок, не умеющий читать, рассыпал эти буквы, а затем собрал в произвольном порядке. Найти вероятность того, что у него снова получилось слово «книга».
4. В колоде 36 карт четырех мастей. После извлечения и возвращения одной карты, колода перемешивается и снова извлекается одна карта. Определить вероятность того, что обе извлеченные карты одной масти.
5. На отдельных одинаковых карточках написаны цифры: 1, 2, 3, 4, 5, 6, 7, 8, 9. Все девять карточек перемешивают, после чего наугад берут четыре карточки и раскладывают в ряд в порядке появления. Какова вероятность получить при этом: а) четное число? б) число 1234?
6. Какова вероятность, что на трех карточках, вынутых по одной и положенных в порядке их появления, получим число 325, если всего карточек было шесть с цифрами 1, 2, 3, 4, 5, 6?
7. Восемь различных книг расставляются наугад на полке. Найти вероятность того, что две определенные книги окажутся поставленными рядом.
8. Среди изготовленных 15 деталей имеется 5 нестандартных. Определить вероятность того, что взятые наугад три детали окажутся стандартными.
9. В партии готовой продукции из 10 изделий имеется 7 изделий повышенного качества. Наудачу отбираются шесть изделий. Какова вероятность того, что четыре из них будут повышенного качества?
10. Какова вероятность того, что два определенных студента будут посланы на практику в Лабинск, если предоставлено 6 мест в г. Лабинск, 10 – в г. Анапу и 4 – в г. Тимашевск?
11. В клетке содержится 18 кур. Из них 6 не вакцинированы. Партию делят на 2 равные части. Какова вероятность того, что не вакцинированные куры разделятся поровну?
12. Из 25 студентов группы 12 занимаются научной работой на кафедре бухгалтерского учета, 7 - экономического анализа, остальные – на кафедре статистики. Какова вероятность того, что два случайно отобранных студента занимаются научной работой на кафедре статистики?
13. Собрание, на котором присутствует 25 человек, в том числе 5 женщин, выбирает делегацию из трех человек. Найти вероятность того, что в делегацию войдут: а) две женщины и один мужчина; б) все женщины.
14. Среди 20 студентов группы, в которой 10 девушек, разыгрываются 5 билетов в театр. Определить вероятность того, что среди обладателей билетов окажутся три девушки.
15. Определить вероятность того, что участник лотереи «Спортлото – 5 из 36» угадает правильно: а) все 5 номеров; б) 3 номера.
16. Цифровой замок содержит на общей оси 4 диска, каждый из которых разделен на 6 секторов, отмеченных цифрами. Замок открывается в том случае, если диски установлены так, что цифры на них составляют определенное четырехзначное число. Какова вероятность того, что замок откроется, если установить произвольную комбинацию цифр?
17. Устройство состоит из пяти элементов, из которых два изношены. При включении устройства включается случайным образом два элемента. Найти вероятность того, что включенными окажутся неизношенные элементы.
18. В коробке пять одинаковых изделий, причем три из них окрашены. Наудачу извлечены два изделия. Найти вероятность того, что среди двух извлеченных изделий окажутся: а) одно окрашенное изделие; б) два окрашенных изделия; в) хотя бы одно окрашенное изделие.

1.6. Основные теоремы теории вероятностей

Теорема 1. Вероятность суммы двух несовместных событий A и B равна сумме их вероятностей:

$$P(A+B)=P(A)+P(B).$$

Следствие 1. Если $A_1, A_2, ..., A_n$ - попарно несовместные события, то вероятность их суммы равна сумме вероятностей этих событий:

$$P(A_1+A_2+...+A_n)=P(A_1)+P(A_2)+...+P(A_n).$$

Следствие 2. Вероятность суммы попарно несовместных событий $A_1, A_2, ..., A_n$, образующих полную группу, равна 1:

$$P(A_1+A_2+...+A_n)=P(A_1)+P(A_2)+...+P(A_n)=1.$$

Следствие 3. События A и $\overline{A}$ несовместны и образуют полную группу событий, поэтому

$$P(A+\overline{A})=P(A)+P(\overline{A})=1.$$

Отсюда,

$$P(\overline{A})=1-P(A).$$

Теорема 2. Вероятность суммы двух совместных событий A и B равна сумме вероятностей этих событий без вероятности их произведения:

$$P(A+B)=P(A)+P(B)-P(A\cdot B).$$

Введем понятие зависимых и независимых событий.

Два события A и B называются *независимыми*, если появление одного из них не влияет на вероятность появления другого (в противном случае события *зависимы*).

Теорема 3. Вероятность произведения двух независимых событий A и B равна произведению их вероятностей:

$$P(A\cdot B)=P(A)\cdot P(B).$$

Следствие. Вероятность произведения n независимых событий $A_1, A_2, ..., A_n$ равна произведению их вероятностей:

$$P(A_1\cdot A_2\cdot ...\cdot A_n)=P(A_1)\cdot P(A_2)\cdot ...\cdot P(A_n).$$

Условной вероятностью события B, при условии, что событие A уже произошло, называется число P(AB)/P(A), которое обозначается

$$P(AB)/P(A)=P(B/A)=P_A(B).$$

Аналогично, $P(AB)/P(B)=P(A/B)=P_B(A)$ - условная вероятность события A, при условии, что событие B уже произошло.

Теорема 4. Вероятность произведения 2-х зависимых событий A и B равна произведению вероятности наступления события A на условную вероятность события B при условии, что событие A уже произошло:

$$P(A\cdot B)=P(A)\cdot P(B/A).$$

Следствие. Если события A и B независимы, то из теоремы 4 следует теорема 3.

Событие B не зависит от события A, если P(B/A)=P(B).

Теорему 4 можно обобщить на n событий.

Теорема 5. Вероятность произведения n зависимых событий - $A_1, A_2,\ldots, A_n$ равна произведению последовательных условных вероятностей:

$$P(A_1 \cdot A_2 \cdot\ldots\cdot A_{n-1} \cdot A_n) = P(A_1)\cdot P(A_2 / A_1)\cdot\ldots\cdot P(A_n / A_1 \cdot A_2 \cdot\ldots\cdot A_{n-1}).$$

Теорема 6. Вероятность наступления хотя бы одного из событий $A_1, A_2,\ldots, A_n$ равна разности между единицей и вероятностью произведения отрицаний событий $A_1, A_2,\ldots, A_n$:

$$P(A) = 1 - P(\overline{A}_1 \cdot \overline{A}_2 \cdot\ldots\cdot \overline{A}_n) = 1 - P(\overline{A}_1)\cdot P(\overline{A}_2 / \overline{A}_1)\cdot\ldots\cdot P(\overline{A}_n / \overline{A}_1 \cdot \overline{A}_2 \cdot\ldots\cdot \overline{A}_{n-1}).$$

Следствие 1. Вероятность наступления хотя бы одного из событий $A_1, A_2, \ldots, A_n$, независимых в совокупности, равна разности между единицей и произведением вероятностей противоположных событий:

$$P(A)=1-P(\overline{\mathrm{A}_1})\cdot P(\overline{\mathrm{A}_2})\cdot\ldots\cdot P(\overline{\mathrm{A}_n}).$$

Следствие 2. Если события имеют одинаковую вероятность появиться ($P(A_i)=p$, $P(\overline{A}_i)=1-p=q$, где $i=1, 2,\ldots, n$), то вероятность появления хотя бы одного из них равна

$$P(A) = 1 - q^n.$$

Замечание. В теоремах 1-6 неявно предполагается, что все события, в рамках каждой теоремы, принадлежат одному пространству элементарных событий.

1.7. Алгоритм, примеры и контрольные задания

Алгоритм. Составить по алгоритму 1.2 формулу, выражающую событие, вероятность которого необходимо определить, через элементарные события, а затем применить теоремы 1-6.

Пример 1.7.1. В урне 10 шаров, из которых два белые, а остальные черные. Наудачу взято 2 шара. Найдем вероятность того, что оба шара черные.

Решение. Пусть элементарные события: A_1 - первый шар черный; A_2 - второй шар черный. Тогда событие $A=A_1\cdot A_2$ – оба шара черные. Вероятность того, что второй шар черный, будет зависеть от того какого цвета первый шар. Если первый шар черный, то вероятность того, что второй шар также черного цвета, равна условной вероятности $P(A_2/A_1)=7/9$, так как после наступления события A_1 –всего шаров останется 9, из них 7 черных. Отсюда,

$$P(A) = P(A_1 \cdot A_2) \overset{T4}{=} P(A_1)\cdot P(A_2 / A_1) = \frac{8}{10}\cdot\frac{7}{9} = \frac{28}{45}.$$

Пример 1.7.2. Два стрелка сделали по 1-му выстрелу в мишень, вероятность попадания первого 0,8 , а второго 0,6. Найти вероятность следующих событий:1) событие A - оба попали; 2) событие B - попал один ; 3) событие C - попал хотя бы один.

Решение. Пусть A_1, A_2 − события, обозначающие соответственно, что 1-й и 2-й стрелок попали в цель. По условию: $P(A_1) = 0{,}8$; $P(A_2) = 0{,}6$.

1) Событие A - оба стрелка попали в цель, наступит при одновременном попадании, поэтому: $A= A_1\cdot A_2$. Отсюда, в силу независимости событий A_1, A_2, по теореме 3 имеем:

$$P(A)=P(A_1)\cdot P(A_2)=0,48$$

2) B - попал один стрелок. $B=\overline{A}_1\cdot A_2+A_1\cdot\overline{A}_2$. Применим последовательно теоремы 1 и 3 ($T1$, $T3$):

$$P(B)=P(\overline{A}_1\cdot A_2+A_1\cdot\overline{A}_2)\overset{\text{T1}}{=}P(\overline{A}_1\cdot A_2)+P(A_1\cdot\overline{A}_2)\overset{\text{T3}}{=}$$
$$=P(\overline{A}_1)\cdot P(A_2)+P(A_1)\cdot P(\overline{A}_2)=$$

$$=(1-0,8)\cdot 0,6+0,8\cdot 1-0,6=0,12+0,32=0,44$$

3) Событие C - хотя бы один стрелок попал, $C=A_1+A_2$,

$$P(C)=P(A_1+A_2)\overset{\text{T2}}{=}P(A_1)+P(A_2)-P(A_1\cdot A_2)=0,8+0,6-0,8\cdot 0,6=0,92$$

или

$$P(\overline{C})=P(\overline{A}_1\cdot\overline{A}_2)=P(\overline{A}_1)\cdot P(\overline{A}_2)=(1-P(A_1))\cdot(1-P(A_2))=(1-0,8)\cdot(1-0,6)=0,08,$$

отсюда (по следствию 3 из $T1$), $P(C)=1-0,08=0,92$.

Пример 1.7.3. В урне 2 белых и 3 черных шара. Из урны вынимают подряд 2 шара. Найти вероятность того, что оба шара белые.

Решение. Событие A - оба шара белые,
элементарное событие A_1 – 1-ый вытащили белый шар,
элементарное событие A_2 – 2-ой вытащили белый шар.
Событие A наступит, если наступят одновременно и A_1 и A_2, $A=A_1\cdot A_2$, отсюда,

$$P(A)=P(A_1\cdot A_2)\overset{\text{T4}}{=}P(A_1)P(A_2/A_1)=\frac{2}{5}\cdot\frac{1}{4}=\frac{1}{10}$$

Пример 1.7.4. Условия примера 1.7.3, но, после 1-го извлечения, шар возвращается в урну.

Решение. В этом случае события A_1 и A_2- независимы.

$$A=A_1\cdot A_2,\quad P(A)=P(A_1\cdot A_2)\overset{\text{T3}}{=}P(A_1)\cdot P(A_2)=\frac{2}{5}\cdot\frac{2}{5}=0,16$$

Пример 1.7.5. Брошены 2 игральные кости. Найти вероятность того, что сумма очков, выпавших на гранях равна 7.

Решение. Согласно классическому определению вероятности:

$$P(A)=\frac{m(A)}{n}.$$

1+6 = 6+1 = 2+5 = 5+2 = 3+4 = 4+3 – все возможные варианты получения в сумме 7 очков при подбрасывании двух игральных костей, следовательно, $m(A)=6$.

Общее число возможных случаев $n=6\cdot 6=36$, поэтому

$$P(A)=\frac{6}{36}=\frac{1}{6}.$$

Пример 1.7.6. Вероятность одного попадания в цель при одном залпе из двух орудий равна 0,38. Найти вероятность поражения цели при одном выстреле, первого орудия, если известно, что для второго орудия эта вероятность равна 0,8.

Решение. Событие A - попадание одного орудия при одновременном залпе из 2-х орудий. Элементарные события:

A_1 - первое орудие попало,

A_2 - второе орудие попало.

По условию: $P(A) = 0{,}38$, $P(A_2) = 0{,}8$. Событие A наступит, если наступит A_1, но не наступит A_2 или наступит событие A_2, но не наступит событие A_1. Имеем:

$$A=\overline{A}_1A_2+A_1\overline{A}_2,\quad \mathrm{P(A)} = \mathrm{P}(\overline{A}_1A_2+A_1\overline{A}_2)\overset{\mathrm{T1}}{=}\mathrm{P}(\overline{A}_1A_2)+\mathrm{P}(A_1\overline{A}_2)$$

Так как A_1 и A_2 независимы, то

$$P(\overline{A}_1A_2)+P(A_1\overline{A}_2)\overset{\mathrm{T3}}{=}P(A_1)\cdot P(\overline{A}_2)+P(\overline{A}_1)\cdot P(A_2)=P(A_1)\cdot(1-P(A_2))+(1-P(A_1))\cdot P(A_2)=0{,}38$$

$$\Rightarrow P(A_1)\cdot(1-0{,}8)+(1-P(A_1))\cdot 0{,}8=0{,}38,$$

$$0{,}2P(A_1)+0{,}8-0{,}8P(A_1)=0{,}38$$

$$P(A_1)=0{,}7$$

Пример 1.7.7. Студент разыскивает нужную ему формулу в 3-х справочниках. Вероятность того, что формула содержится в 1,2 и 3 справочнике соответственно равна 0,6; 0,7; 0,8.

Найти вероятность того, что формула содержится:

а) только в одном справочнике (событие A);

б) только в двух справочниках (событие B);

в) во всех трех справочниках (событие C);

г) хотя бы в одном справочнике (событие D);

д) ни в одном справочнике (событие E).

Решение. Рассмотрим элементарные события и их вероятности:

A_1 - формула находится в 1-м справочнике, $P(A_1)$=0,6,

$$P(\overline{A}_1) = 1 - 0{,}6 = 0{,}4;$$

A_2 - формула находится во 2-м справочнике, $P(A_2)$=0,7,

$$P(\overline{A}_2) = 1-0{,}7 = 0{,}3;$$

A_3 - формула находится в 3-м справочнике, $P(A_3)$=0,8,

$$P(\overline{A}_3) = 1 - 0{,}8 = 0{,}2.$$

Выразим через элементарные события и их отрицания все события A-E, применим теоремы параграфа 1.6:

а) $A=A_1\cdot\overline{A}_2\cdot\overline{A}_3+\overline{A}_1\cdot A_2\cdot\overline{A}_3+\overline{A}_1\cdot\overline{A}_2\cdot A_3,$

$$P(A)=P(A_1\cdot\overline{A}_2\cdot\overline{A}_3)+P(\overline{A}_1\cdot A_2\cdot\overline{A}_3)+P(\overline{A}_1\cdot\overline{A}_2\cdot A_3)=$$

$$=P(A_1)P(\overline{A}_2)P(\overline{A}_3)+P(\overline{A}_1)P(A_2)P(\overline{A}_3)+P(\overline{A}_1)P(\overline{A}_2)P(A_3)=$$

$$=0{,}6\cdot 0{,}3\cdot 0{,}2+0{,}4\cdot 0{,}7\cdot 0{,}2+0{,}4\cdot 0{,}3\cdot 0{,}8=0{,}188$$

б) $B=A_1A_2\overline{A}_3+A_1\overline{A}_2A_3+\overline{A}_1A_2A_3$,

далее аналогично пункту а) получим, что $P(B)$=0,452;

в) $C= A_1.A_2.A_3,$

$$P(C) = P(A_1.A_2.A_3) \overset{\mathrm{T3}}{=} P(A_1)\,.P(A_2)\,.P(A_3) = 0{,}6.0{,}7.0{,}8=0{,}336;$$

г) $D= A_1+A_2+A_3$,

вероятность события D можно найти, обобщив теорему 2 для трёх событий:

$$P(D)= P(A_1+A_2+A_3) = P(A_1) + P(A_2) + P(A_3) - P(A_1 A_2) - P(A_1 A_3) - $$
$$- P(A_2 A_3) + P(A_1 A_2 A_3).$$

Но проще воспользоваться следствием к теореме 6:

$$P(D)= 1 - P(\overline{A}_1) P(\overline{A}_2) P(\overline{A}_3) = 1 – 0,4 \cdot 0,3 \cdot 0,2 = 0,976;$$

е) $E=\overline{A}_1\overline{A}_2\overline{A}_3$, $P(E)=0,024$

Пример 1.7.8. Два игрока поочередно бросают игральную кость. Выигрывает первый, у которого появится шесть очков. Найти вероятность выигрыша каждого игрока.

Решение. Пусть событие A_i – выиграл первый игрок, событие B_i – выиграл второй при i-ом подбрасывании (i=1,2…):

$$P(A_i)=P(B_i)=\frac{1}{6},\ P(\overline{A}_i) = P(\overline{B}_i) = \frac{5}{6}.$$

Выигрыш первого игрока до $(k+1)$ - го подбрасывания – событие A:

$$A = A_1 + \overline{A}_1\overline{B}_1 A_2 + \overline{A}_1\overline{B}_1\overline{A}_2\overline{B}_2 A_3 + ... + \overline{A}_1\overline{B}_1\overline{A}_2 ... \overline{A}_{k-1}\overline{B}_{k-1} A_k .$$

Событие A - сумма несовместных событий, каждый член которой, начиная со второго, является произведением независимых событий, поэтому:

$$P(A)=\frac{1}{6}+\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6}+\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6}+...+\left(\frac{5}{6}\right)^{(k-1)}\cdot\frac{1}{6}=\frac{1}{6}\cdot\frac{1-\left(\frac{25}{36}\right)^k}{1-\frac{25}{36}}=\frac{1}{6}\frac{1-\left(\frac{25}{36}\right)^k}{\frac{11}{36}}$$

как сумма k членов геометрической прогрессии с первым членом $\frac{1}{6}$ и знаменателем $\frac{25}{36}$.

При $k\to\infty$, $P(A)=\frac{\frac{1}{6}}{1-\frac{25}{36}}=\frac{\frac{1}{6}}{\frac{11}{36}}=\frac{6}{11}$ - по формуле бесконечной геометрической прогрессии, где $|q|<1$.

Аналогично, если выигрыш второго игрока до (k+1) - го подбрасывания – событие B, то $B = \overline{A}_1 B_1 + \overline{A}_1\overline{B}_1\overline{A}_2 B_2 + ... + \overline{A}_1\overline{B}_1\overline{A}_2\overline{B}_2 \cdot ... \cdot \overline{A}_k B_k$,

$$P(B) = \frac{5}{6}\cdot\frac{1}{6}+\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{5}{6}\cdot\frac{1}{6}+...+\left(\frac{5}{6}\right)^{2(k-1)}\cdot\frac{5}{6}\cdot\frac{1}{6}=\frac{5}{36}\cdot\frac{1-\left(\frac{25}{36}\right)^k}{1-\frac{25}{36}}.$$

При $k\to\infty$, $P(B)=\frac{\frac{5}{36}}{1-\frac{25}{36}}=\frac{\frac{5}{36}}{\frac{11}{36}}=\frac{5}{11}$.

Контрольные задания 1.7.

1. Круговая мишень состоит из трех зон. Вероятности попадания в эти зоны при одном выстреле соответственно равны 0,1; 0,35 и 0,4. Найти вероятность: а) попадания в первую или третью зоны; б) промаха по мишени.

2. Вероятность поражения первой мишени для данного стрелка равна 0,6. Если при первом выстреле зафиксировано попадание, то стрелок получает право на следующий выстрел по второй мишени. Вероятность поражения обеих мишеней при двух выстрелах равна 0,3. Определить вероятность поражения второй мишени.
3. В группе 25 студентов, из них 10 юношей и 15 девушек. Какова вероятность того, что из вызванных наудачу трех студентов: а) все три девушки; б) первые две девушки, третий - юноша; в) все три юноши?
4. Вероятность попадания в цель при одном выстреле равна 0,7. Выстрелы производятся по одному до первого попадания. Определить вероятность того, что придется производить четвертый выстрел.
5. Вероятность безотказной работы автомобиля равна 0,9. Автомобиль перед выходом на линию осматривается двумя механиками. Вероятность того, что первый механик обнаружит неисправность в автомобиле равна 0,8, а второй – 0,9. Если хотя бы один механик обнаружит неисправность, то автомобиль отправляется на ремонт. Найти вероятность того, что: а) автомобиль будет выпущен на линию; б) автомобиль не будет выпущен на линию.
6. Вероятность одного попадания в цель при одновременном залпе из двух орудий равно 0,44. Найти вероятность поражения цели при одном выстреле первым орудием, если для второго орудия эта вероятность равна 0,8.
7. Из 40 деталей в ящике 5 бракованных. Какова вероятность того, что взятые одновременно две детали не будут бракованными?
8. В коробке 12 карандашей трех цветов, по четыре карандаша каждого цвета. Наудачу вынимают три карандаша. Найти вероятность того, что все карандаши окажутся разного цвета. Решить задачу при условии: а) карандаши возвращают в коробку; б) карандаши не возвращают в коробку.
9. Из урны, содержащей четыре красных и шесть черных шаров, вынимают два шара (без возвращения первого). Какова вероятность того, что будут вынуты: а) два шара черного цвета; б) красный и черный в любой последовательности; в) второй шар будет черным; г) оба шара одного цвета?
10. Вероятность выигрыша по лотерейному билету равна 0,1. Приобретено три билета. Какова вероятность выиграть хотя бы по одному из них?
11. Вероятность попадания в цель при стрельбе из орудия равна 0,6. Производится по одному выстрелу одновременно из трех орудий. Цель будет поражена, если в нее попадут не менее двух орудий. Найти вероятность: а) поражения цели; б) промаха одним или двумя орудиями.
12. Слово «машина» составлено из букв разрезной азбуки. Какова вероятность того, что, перемешав все буквы и укладывая их в ряд по одной, получим слово: а) «машина»; б) «шина»; в) «маша»?
13. В магазин вошли три покупателя. Вероятность того, что каждый что-нибудь купит, равна 0,3. Найти вероятность того, что: а) два из них совершат покупки; б) все три совершат покупки; в) ни один не совершит покупки; г) по крайней мере, два совершат покупки; д) хотя бы один купит товар.
14. Вероятность получить высокие дивиденды по акциям на первом предприятии – 0,2, на втором – 0,35, на третьем – 0,15. Определить вероятность того, что акционер, имеющий акции всех предприятий, получит высокие дивиденды: а) на всех предприятиях; б) только на одном предприятии; в) хотя бы на одном предприятии.
15. Брошены две игральные кости. Предполагается, что все комбинации выпавших очков равновероятны. Найти условную вероятность того, что выпали две пятерки, если известно, что сумма выпавших очков делится на пять.
16. Два игрока поочередно бросают 2 игральные кости. Выигрывает тот, у которого первым появится в сумме двенадцать очков. Найти вероятность выигрыша для каждого игрока.
17. Через автобусную остановку проходят автобусы семи маршрутов с равной частотой. Пассажир ожидает автобус одного из маршрутов №1, №5, №7. Какова вероятность, что нужный ему автобус будет одним из первых трех подошедших к остановке?

18. Читатель в поисках нужной книги обходит три библиотеки. Вероятность того, что она имеется в очередной библиотеке равна 0,3. Что вероятнее – найдет читатель книгу или нет?
19. В денежно-вещевой лотерее на каждые 1000 билетов приходится 12 денежных и 8 вещевых выигрышей. Какова вероятность выигрыша хотя бы на один из трех приобретенных билетов?
20. В урне 10 красных, 5 зеленых и 3 черных шара. Определить вероятность того, что взятые наудачу два шара будут: а) одного цвета; б) разных цветов.
21. На базу поступило 40 ящиков овощей, из них 30 первого сорта. Наудачу для проверки берут два ящика. Какова вероятность, что: а) оба содержат овощи первого сорта; б) разного сорта; в) одного сорта?
22. Читатель разыскивает книгу в трех библиотеках. Одинаково вероятно, есть или нет в фонде очередной библиотеки книга и также одинаково вероятно, выдана она или нет. Чему равна вероятность того, что читатель найдет нужную книгу?
23. Три студента сдают экзамен. Вероятность того, что отдельный студент сдаст экзамен на «отлично» равна для первого студента 0,7, для второго - 0,6, для третьего – 0,2. Какова вероятность того, что экзамен будет сдан на «отлично»: а) только одним студентов; б) двумя студентами; в) хотя бы одним; г) ни одним?
24. Первый студент из 20 вопросов программы выучил 17, второй – 12. Каждому студенту задают по одному вопросу. Определить вероятность того, что: а) оба студента правильно ответят на вопрос; б) хотя бы один ответит верно; в) правильно ответит только первый студент.
25. Студент из 40 экзаменационных вопросов выучил только 30. Каким выгодней ему зайти на экзамен, первым или вторым?
26. В первой бригаде 6 тракторов, во второй – 9. В каждой бригаде один трактор требует ремонта. Из каждой бригады наудачу выбирают по одному трактору. Какова вероятность того, что: а) оба трактора исправны; б) один требует ремонта; в) трактор из второй бригады исправен.
27. На предприятии имеется три автомобиля. Вероятность безотказной работы первого из них равна 0,9, второго – 0,7, третьего – 0,8. Найти вероятности всех возможных значений числа автомобилей работающих безотказно в течение определенного времени.
28. Вероятность хотя бы одного попадания в мишень стрелком при трех выстрелах равна 0,784. Найти вероятность одного промаха при трех выстрелах.
29. В круг радиуса R вписан прямоугольник наибольшей площади. Чему равна вероятность того, что поставленные наудачу внутри круга две точки окажутся внутри заданного прямоугольника?
30. Сколько раз необходимо бросить игральную кость, чтобы с вероятностью 0,9 хотя бы один раз выпало не менее четырех очков?
31. Вероятность спортсменом взять в одной попытке высоту 1,8 м равна 0,6, высоту 2 м – 0,2, высоту 2 м 10 см – 0,1. Спортсмен, не взявший предыдущую высоту, выбывает из соревнований. Спортсмену на каждую высоту дается три попытки. Определить вероятность того, что спортсмен закончит соревнования, взяв высоту: а) 1,8 м; б) 2 м; в) 2 м 10 см.
32. В первой урне 5 красных, 3 белых и 2 черных шара. Во второй 3 белых и 2 черных шара. Из первой урны взято 2 шара, а из второй один. Определить вероятность того, что среди них: а) все шары одного цвета; б) все шары разного цвета.
33. Предположим, что 85% людей, которые интересуются возможными инвестициями (вложениями) в брокерскую фирму, не покупают акции, а 33 % не покупают облигации. Также известно, что 28% интересующихся прерывают покупку ценных бумаг – как акций, так и облигаций. Некто интересуется делами компании; чему равна вероятность, что он будет покупать либо облигации, либо акции, либо и то и другое?
34. Консультационная фирма получила приглашение для выполнения двух работ от двух международных корпораций. Руководство фирмы оценивает вероятность получения заказа от фирмы А (событие А) равной 0,45. Также, по мнению руководителей фирмы, в случае, если

фирма заключит договор с компанией А, то с вероятностью в 90 % компания В даст фирме консультационную работу. С какой вероятностью компания получит оба заказа?

35. Вероятность того, что покупатель, собирающийся приобрести компьютер и пакет прикладных программ, приобретёт только микрокомпьютер, равна 0,15. Вероятность, что покупатель купит только пакет программ, равна 0,1. Вероятность того, что будут куплены и компьютер, и пакет программ, равна 0,05.Чему равна вероятность того, что будут куплены или компьютер, или пакет программ, или компьютер и пакет программ вместе?

1.8. Формулы полной вероятности и вероятности гипотез

Пусть событие *A* может наступать только одновременно с одним несовместных событий $H_1, H_2, \ldots, H_n$, образующих полную группу (рис. 6).

.

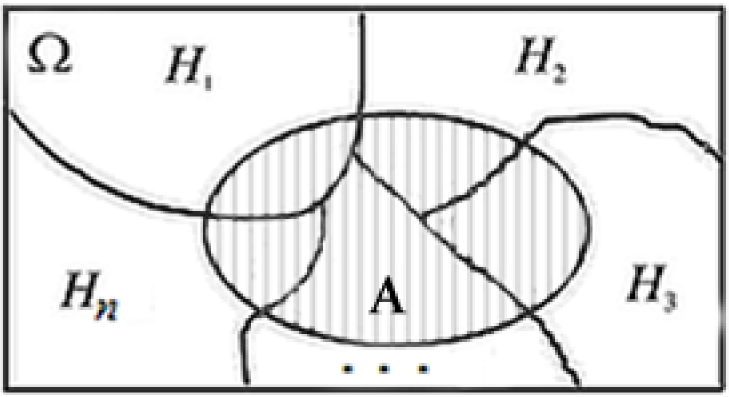


Рис. 6

Тогда вероятность события *A* определятся по *формуле полной вероятности*:

$$P(A)= P(H_1)\cdot P(A/H_1)+P(H_2)\cdot P(A/H_2)+\ldots+ P(H_n)\cdot P(A/H_n)$$

или

$$P(A)=\sum_{i=1}^{n} P(H_i)P(A/H_i),\quad i=1, 2, \ldots, n. \qquad (1.8.1)$$

где события $H_1, H_2, \ldots, H_n$, - гипотезы, а $P(A/H_i)$ - условная вероятность наступления события *A* при наступлении *i*-ой гипотезы (i=1, 2,…, n).

Условная вероятность гипотезы H_i, при условии того, что событие *A* произошло, определяется по формуле вероятности гипотез или *формуле Байеса* (она позволяет пересмотреть вероятности гипотез после наступления события *A*):

$$P(H_i / A)=\frac{P(H_i)\cdot P(A/H_i)}{P(A)} \qquad (1.8.2)$$

1.9. Алгоритм, примеры и контрольные задания

Алгоритм.

1. Исходя из условий задачи определяют, что некоторое событие *A* может наступить только при одновременном наступлении одного из попарно несовместных событий: $H_1, H_2, \ldots, H_n$.

2. По имеющимся данным определяют вероятности $P(H_i)$ и $P(A/H_i)$, где i=1, 2,…, n.

3. Для нахождения вероятности событий A и H_i/A применяют формулы (1.8.1) и (1.8.2).

Пример 1.9.1. Команда стрелков состоит из 5 человек 3-е из них попадают с вероятностью 0,8, а двое с вероятностью 0,6. Наудачу из команды берется стрелок и производит выстрел.

а) Какова вероятность того, что стрелок попадет.

б) Если стрелок попал в цель, то какова вероятность, что это один из 3-х (один из 2-х).

Решение. а). Событие А может произойти, если произойдёт одно из несовместных событий: H_1 – наудачу взятый стрелок один из трех, H_2 - наудачу взятый стрелок один из двух. Для определения вероятности события A воспользуемся формулой (1.8.1):

$$P(A) = P(H_1)\cdot P(A/H_1) + P(H_2)\cdot P(A/H_2) = \frac{3}{5}\cdot\frac{8}{10}+\frac{2}{5}\cdot\frac{6}{10}=\frac{18}{25},$$

т.к. $P(H_1)=\frac{3}{5}, \quad P(A/H_1)=\frac{8}{10}; P(H_2)=\frac{2}{5}, \quad P(A/H_2)=\frac{6}{10}.$

б) По формуле (1.8.2):

$$P(H_1/A)=\frac{\mathrm{P}(H_1)\cdot \mathrm{P}(A/H_1)}{P(A)}=\frac{\frac{24}{50}}{\frac{36}{50}}=\frac{2}{3},$$

$$P(H_2/A)=\frac{\mathrm{P}(H_2)\cdot \mathrm{P}(A/H_2)}{P(A)}=\frac{\frac{12}{50}}{\frac{36}{50}}=\frac{1}{3}$$

или $P(H_2/A) = 1 - \mathrm{P}(H_1/A) = 1-\frac{2}{3}=\frac{1}{3}.$

Пример 1.9.2. По предмету теория вероятностей и математическая статистика имеется 30 экзаменационных билетов. Студент Павлов выучил только 20. Каким выгоднее ему зайти на экзамен, первым или вторым?

Решение. Событие А – студент Павлов заходит первым,

$$P(A)=\frac{20}{30}=\frac{2}{3}.$$

Событие B – студент Павлов заходит вторым может произойти только с одним из попарно несовместных событий A_1, A_2, где:

событие A_1 – 1-й студент вытащит 1 из 20 билетов, которые Павлов знает; событие A_2 - 1-й студент вытащит 1 из 10 остальных билетов.

$$B=A_1B + A_2B,$$

$$P(B)=P(A_1)\cdot P(B/A_1) + P(A_2)\cdot P(B/A_2) =\frac{20}{30}\cdot\frac{19}{29}+\frac{10}{30}\cdot\frac{20}{29}=\frac{2}{3}.$$

То есть все равно, зайдет студент первым или вторым.

Графическое представление. Анализ логических возможностей при проведении опыта (эксперимента), анализа социально-экономического или технологического процесса обычно проводится с использованием графов. Граф – это множество вершин, соединенных ребрами.

С помощью графов можно иллюстрировать правила сложения и умножения вероятностей событий (рис. 7, 8):

1. $P(A{\cdot}B + C{\cdot}D) = P(A) \cdot P(B/A) + P(C) \cdot P(D/C)$, где $A{\cdot}B$ и $C{\cdot}D$ – несовместные события.

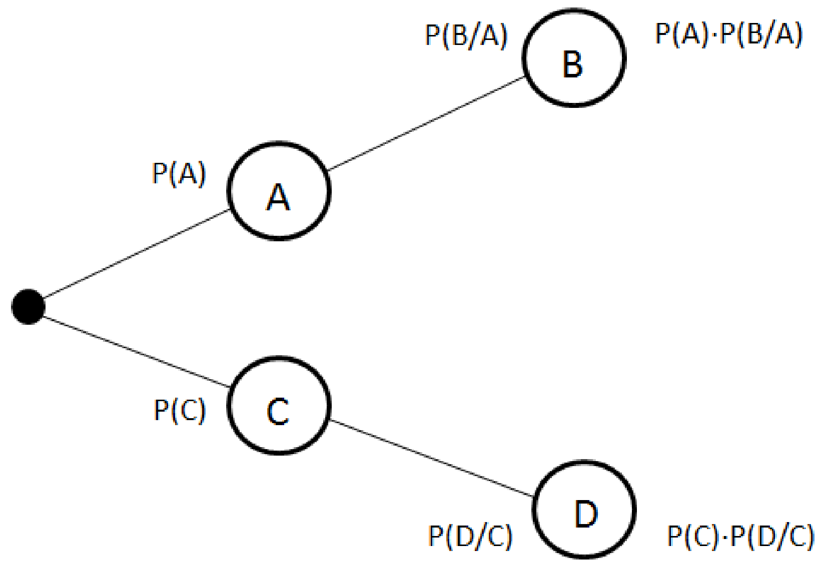


Рис. 7

2. $P(A_1 \cdot A_2 \cdot \ldots \cdot A_{n-1}{\cdot}A_n) = P(A_1) \cdot P(A_2/A_1) \cdot \ldots \cdot P(A_n/A_1 \cdot A_2 \cdot \ldots \cdot A_{n-1})$.

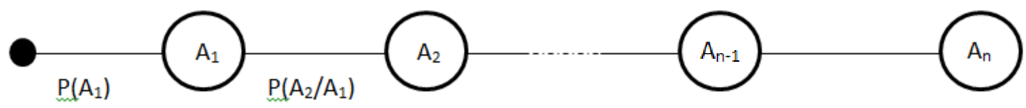


Рис. 8

Часто выделяется одна вершина (корень) графа, которая обозначает начало изучаемого процесса, тогда (при отсутствии циклов), такой граф называют деревом. Дерево представляет собой графический вариант классификации, когда одна группа детализируется по ярусам (уровням).

Пример 1.9.3. Имеется две урны, первая содержит два черных и один белый шар, а вторая один черный и два белых шара. Наудачу выбирается урна и из неё последовательно выбирается два шара. Какова вероятность того, что второй шар белый. Если шар белый, то какова вероятность, что выбрали 2 урну.

Решение. Событие A – выбранный шар белый; гипотеза H_1 – выбрана первая урна, H_2 – вторая урна. Логические возможности выбора представлены в таблице 3 и на рисунке 9, соответствующее вероятностное дерево на рисунке 10.

Таблица 3 – Логические возможности

Случай	*Урна*	*Первый шар*	*Второй шар*
1	1	Черный	Черный
2	1	Черный	Белый
3	1	Белый	Черный
4	2	Черный	Белый
5	2	Белый	Черный
6	2	Белый	Белый

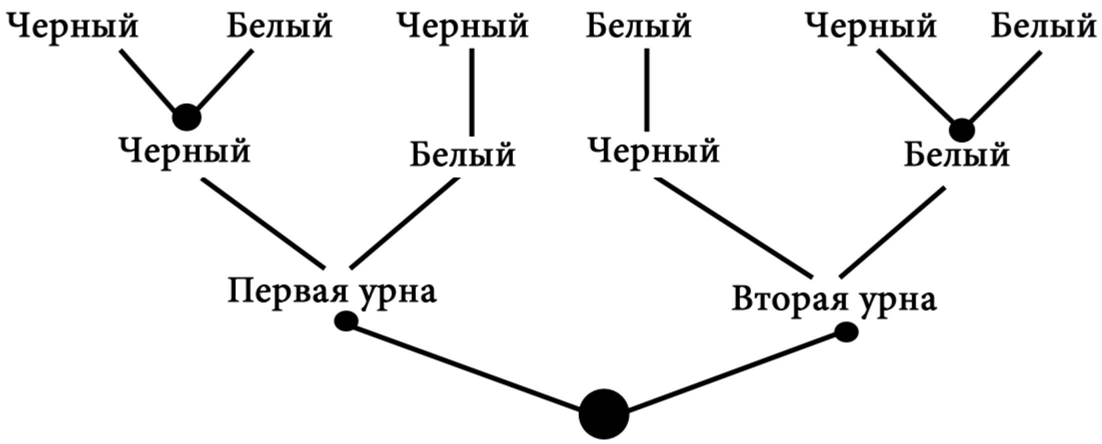


Рис. 9 – Дерево логических возможностей

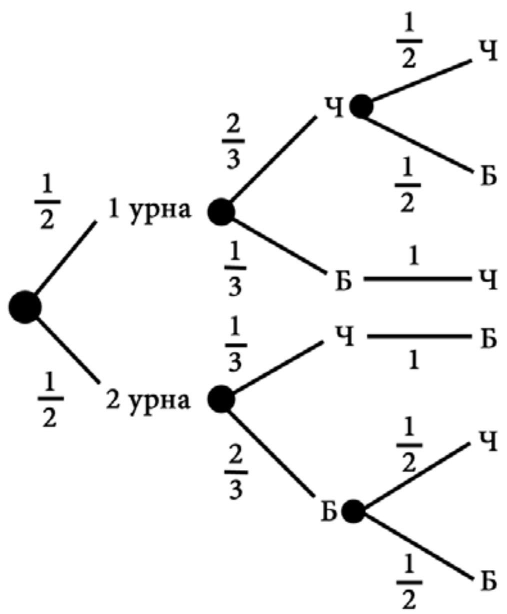


Рис.10 – Вероятностное дерево

Согласно формуле полной вероятности:

$$P(A) = P(H_1)P(A/H_1) + P(H_2)P(A/H_2),$$

$$P(A)=\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{1}{2}+\frac{1}{2}\cdot\frac{1}{3}\cdot 1+\frac{1}{2}\cdot\frac{2}{3}\cdot\frac{1}{2}=\frac{6}{12}=\frac{1}{2}$$

вероятность того, что второй шар белый равна 0,5.

Если второй выбранный шар белый, то согласно формуле Байеса вероятность того, что была выбрана вторая урна

$$P(H_2/A)=\frac{P(H_2)P(A/H_2)}{P(A)},$$

$$P(H_2/A)=\frac{4/12}{1/2}=\frac{2}{3}.$$

Современная экономика базируется на том, что два действующих лица – производитель и покупатель решают, что и по какой цене выгодно производить и покупать соответственно. На протяжении долгого времени считалось, что человек действуя в условиях неопределенности и имея знания о вероятностях тех или иных событий поступает рационально, то есть в явной или неявной форме пользуется так называемыми аксиомами рационального поведения. Предполагается, что существует некоторая функция полезности, которую рационально действующий экономический субъект в процессе выбора стратегии поведения старается максимизировать.

Пример 1.9.4.На рынке кофе машин покупатель хочет купить 1 кофе-машину, продавец продать 1 кофе-машину. Есть два типа кофе-машин: высокого и низкого качества. Вероятность отказа машин высокого качества 0,2, а низкого 0,75. Если кофе-машина работает без сбоев, то полезность для покупателя 400 у.е., сбой уменьшает полезность до 200 у.е. Известно, что 25% кофе-машин высокого качества. Какую кофе машину следует выбрать?

Решение. Пусть H_1, H_2 – выбор кофе машин высокого и низкого качества соответственно. Условия задачи можно представить в виде таблицы 4 и графически в виде дерева решений (рис.11).

Таблица 4.

Тип машины	Доля на рынке	Вероятность отказа машины	Выигрыш при работе без сбоя	Выигрыш при сбое
Высокого качества, H_1	0,25	0,2	400	200
Низкого качества, H_2	0,75	0,75	400	200

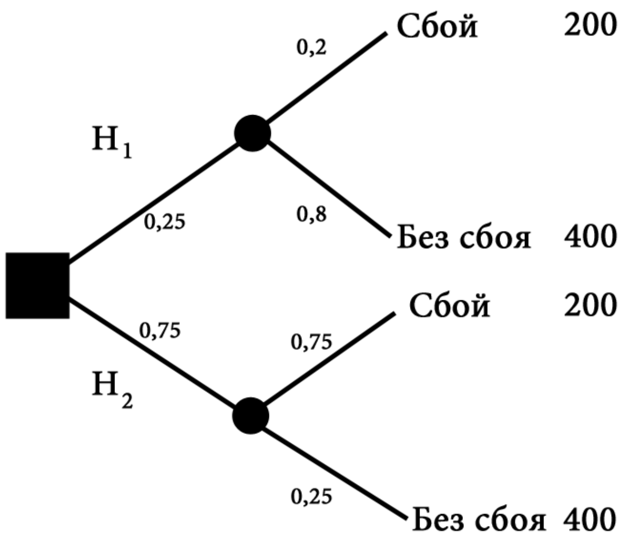


Рис. 11 – Дерево решений

Оценим среднюю ожидаемую полезность *U* (*utility*).

$U(H_1)$=0,25(0,2·200+0,8·400)=105 у.е.

$U(H_2)$=0,75(0,75·200+0,25·400)=187,5 у.е.

Если выбирать действие с максимальной ожидаемой полезностью, то рациональный человек должен выбрать действие H_2, а не H_1. Приведенный выше рисунок 11 называют деревом решений. Квадратик – место где человек принимает решение, а кружочек – место где все решает случай. На ветвях дерева написаны значения вероятностей, а справа у конечных ветвей значения исходов (результаты). Деревья решений используются для представления возможных действий.

Контрольные задания 1.9.

1. При исследовании жирности молока коров все стадо было разбито на три группы. В первой группе оказалось 70%, во второй 23% и в третьей 7% всех коров. Вероятность того, что молоко, полученное от отдельной коровы, имеет не менее 4% жирности, для каждой группы коров соответственно равна 0,6; 0,35 и 0,1. 1)Определить вероятность того, что для взятой наудачу коровы жирность молока составит не менее 4%.
   2)Взятая на удачу корова дает молоко жирностью не менее 4%. Найти вероятность того, что эта корова из первой группы.
2. В первой урне 10 деталей, из них 8 стандартных. Во второй 6 деталей, из которых 5 стандартных. Из второй урны переложили в первую одну деталь. Какова вероятность того, что деталь, извлеченная после этого из второй урны, нестандартная?
3. Имеются две урны. В первой – семь красных шаров и три черных, во второй – три красных и четыре черных. Из первой урны переложили во вторую один шар, затем, перемешав шары, из второй урны переложили в первую один шар. Найти вероятность того, что шар, извлеченный после этого из первой урны, окажется красным.
4. Перед посевом 90% всех семян было обработано ядохимикатами. Вероятность поражения вредителями для растений из обработанных семян равна 0,08, для растений из необработан-

ных семян − 0,4. Взятое наудачу растение оказалось пораженным. Какова вероятность того, что оно выращено из партии обработанных семян?

5. В районе 24 человека обучаются на заочном факультете института, из них шесть – на мехфаке, двенадцать – на агрофаке и шесть – на экономфаке. Вероятность успешно сдать все экзамены на предстоящей сессии для студентов мехфака равна 0,6, агрофака – 0,76 и экономфака – 0,8. Найти вероятность того, что наудачу взятый студент, сдавший успешно все экзамены, окажется студентом экономфака.
6. В первом ящике из 20 деталей 4 бракованных, во втором из 30 деталей 5 бракованных. Из первого во второй переложили две детали. Найти вероятность того, что деталь, извлеченная после этого из второго ящика, бракованная.
7. Стрелковое отделение получило 10 винтовок, из которых 8 пристрелянных, две нет. Вероятность попадания в цель из пристрелянной винтовки равна 0,6, а из не пристрелянной 0,4. 1)Какова вероятность, что стрелок из наудачу взятой винтовки попадет в цель при одном выстреле?
   2)Стрелок поразил цель. Какова вероятность, что он стрелял из пристрелянной винтовки?
8. Для посева заготовлены семена 4 сортов пшеницы. Причем, 20% всех семян 1-го сорта, 30% - 2-го сорта, 10% - 3-го сорта и 40% - 4-го сорта. Вероятность того, что из зерна вырастет колос, содержащий не менее 40 зерен, для первого сорта равна 0,5, для второго – 0,3, для третьего – 0,2, для четвертого – 0,1. Найти вероятность того, что наудачу взятое зерно даст колос, содержащий не менее 40 зерен.
9. Из 25 студентов группы 5 студентов знают все 30 вопросов программы, 10 студентов выучили по 25 вопросов, 7 студентов по 20 вопросов, трое по 10 вопросов. Случайно вызванный студент ответил на два заданных вопроса. Какова вероятность, что он из тех трех студентов, которые подготовили только по 10 вопросов.
10. Запасная деталь может находиться в одной из трех партий с вероятностями $p_1 = 0,2$; $p_2 = 0,5$; $p_3 = 0,3$. Вероятности того, что деталь проработает положенное время без ремонта, равны соответственно 0,9; 0,8 и 0,7. Определить вероятность того, что: а) взятая наудачу деталь проработает положенное время; б) деталь, проработавшая положенное время, взята из второй или третьей партии.
11. Имеется 5 урн. В первой, второй и третьей находится по 4 белых и 6 черных шаров, в четвертой и пятой урнах по 2 белых и 3 черных шара. Случайно выбирается урна и из нее извлекается шар. Какова вероятность того, что была выбрана четвертая или пятая урна, если извлеченный шар оказался белым?
12. В первой бригаде производится в три раза больше продукции, чем во второй. Вероятность того, что производимая продукция окажется стандартной для первой бригады равна 0,7, для второй – 0,8. 1)Определить вероятность того, что взятая наугад единица продукции будет стандартной. 2)Взятая наугад единица продукции оказалась стандартной. Какова вероятность, что она из второй бригады?
13. Покупатель с равной вероятностью посещает 3 магазина. Вероятность того, что он купит товар в первом магазине равна 0,4, во втором 0,3, в третьем 0,2. Определить вероятность того, что покупатель купит товар только в одном магазине, если каждый магазин он посетил дважды.
14. Вероятность того, что клиент банка не вернёт заём в период экономического роста, равна 0,04, а в период экономического кризиса – 0,13. Предположим, что вероятность того, что начнётся период экономического роста, равна 0,65. Чему равна вероятность того, что случайно выбранный клиент банка не вернёт полученный кредит?
15. Экономист-аналитик условно подразделяет экономическую ситуацию в стране на «хорошую», «посредственную» и «плохую» и оценивает их вероятности для данного момента времени как 0,15, 0,70 и 0,15 соответственно. Некоторый индекс экономического состояния возрастает с вероятностью 0,6, когда ситуация «хорошая»; с вероятностью 0,3, когда ситуация

«посредственная», и с вероятностью 0,1, когда ситуация «плохая». Пусть в настоящий момент индекс экономического состояния изменился. Какова вероятность того, что экономика страны на подъёме?

1.10. Повторные независимые испытания

1. *Постоянные условия опыта.* а) Пусть некоторый опыт повторяется в неизменных условиях n раз, причём каждый раз может либо наступить (успех), либо не наступить (неудача) некоторое событие A, где $P(A) = p -$ вероятность успеха, $P(\bar{A}) = 1 - p = q$ – вероятность неудачи. Тогда вероятность того, что в k случаях из n произойдёт событие A вычисляется по *формуле Бернулли*:

$$P_n(k) = C_n^k \cdot p^k \cdot q^{n-k}. \qquad (1.10.1)$$

Условия, приводящие к формуле Бернулли, называются *частной схемой повторных независимых испытаний* или *схемой Бернулли*. Так как вероятности $P_n(k)$ для различных значений k представляют собой слагаемые в разложении бинома Ньютона:
$(p+q)^n = C_n^0 \cdot p^0 \cdot q^n + C_n^1 \cdot p^1 \cdot q^{n-1} + \ldots C_n^k \cdot p^k \cdot q^{n-k} + \ldots + C_n^n \cdot p^n \cdot q^0$, то распределение вероятностей $P_n(k)$, где $0 \le k \le n$, называется *биномиальным*.

б) Если производится ряд независимых опытов в каждом из которых событие A появляется с вероятностью p, до получения k успехов ($k \ge 1$),то при этом вероятность m "неудачных" опытов ($P(\bar{A}) = 1 - p = q$) можно определить по формуле:

$$P_{m+\kappa}(m) = C_{m+\kappa-1}^{m} p^k q^m \quad (m = 0,1,2,...). \qquad (1.10.2)$$

Соответствующее распределение вероятностей называется *отрицательным биномиальным* (множество возможных случаев бесконечно).

2. *Переменные условия опыта*. Если в каждом из независимых испытаний вероятности наступления события A разные (*общая схема повторения опытов*), то вероятность наступления события A k раз в n опытах, определяется как коэффициент, при k-ой степени полинома

$$\varphi_n(Z) = \prod_{i=1}^{n} (q_i + p_i Z) = a_n Z^n + a_{n-1} Z^{n-1} + \ldots + a_1 Z_1 + a_0, \qquad (1.10.3)$$

$\varphi_n(Z)$ – *производящая функция.*

3. *Опыт с несколькими событиями*. Если в результате опыта может появиться одно из несовместных событий $A_1, A_2, \ldots, A_L$, образующих полную группу, где $P(A_1)=p_1$, $P(A_2)=p_2, \ldots, P(A_L)=p_L$ и $\sum p_i = 1$, то вероятность того, что в n опытах появится событие A_1-k_1 раз, $A_2 - k_2$ раз,…, $A_L - k_L$ раз ($\sum k_i = n$), определяется по формуле:

$$P(k_1, k_2, \ldots, k_L) = \frac{n!}{k_1!\, k_2! \ldots k_L!} p_1^{k_1} \cdot p_2^{k_2} \cdot \ldots \cdot p_L^{k_L}. \qquad (1.10.4)$$

Соответствующее распределение вероятностей называется п*олиномиальным*.

Пример 1.10.1. Производится три независимых выстрела по некоторой цели. Вероятности попадания при разных выстрелах различны и равны: p_1=0,7; p_2=0,8; p_3=0,9. Найти вероятность промаха, 1, 2, 3 попаданий.
Решение. Производящая функция:
$\varphi_3(Z)=(0{,}3+0{,}7Z)\cdot(0{,}2+0{,}8Z)\cdot(0{,}1+0{,}9Z)= 0{,}504Z^3+0{,}398Z^2+0{,}092Z+0{,}006$.
Отсюда, вероятность:

− промаха $P_3(0)=0{,}006$;
− одного попадания $P_3(1)=0{,}092$;
− двух попаданий $P_3(2)=0{,}398$;
− трёх попаданий $P_3(3)=0{,}504$.

1.11. Алгоритм, примеры и контрольные задания

Алгоритм.

1. Проверить, выполняются ли условия схемы повторных независимых испытаний.

2. Если опыт, описываемый в задаче, приводит к схеме повторных независимых испытаний, то из условий определить: $P(A) = p -$ вероятность успеха, $P(\bar{A}) = 1 - p = q$ – вероятность неудачи в каждом опыте.

3. При заданных числе опытов количестве успехов, воспользоваться формулами (1.10.1)-(1.10.4).

Часто применяемые формулы в схеме Бернулли.
Вероятность наступления события А:
а) менее k раз:
$P_n(0) + P_n(1) +\ldots+ P_n(k\text{-}1)$;
б) более k раз:
$P_n(k+1) + P_n(k+2) +\ldots+ P_n(n)$;
в) не менее k раз:
$P_n(k) + P_n(k+1) + \ldots+ P_n(n)$;
г) не более k раз:
$P_n(0) + P_n(1) +\ldots+ P_n(k)$;
д) хотя бы один раз:
$1\text{-}P_n(0)$.

Пример 1.11.1. Монета, подбрасывается 6 раз в неизменных условиях. Успехом считается герб (событие А); найти вероятность того, что герб появится 4 раза.
Решение. Условия проведения опыта соответствуют схеме Бернулли.
$P(A) = p = \frac{1}{2}$, $P(\bar{A}) = 1 - p = q = 1 - \frac{1}{2} = \frac{1}{2}$. Согласно условию n=6 , k=4. По формуле Бернулли имеем

$$P_6(4) = C_6^4 \cdot \mathrm{p}^4 \cdot q^{6-4} = \frac{6!}{(6-4)!4!}\left(\frac{1}{2}\right)^2 \cdot \left(\frac{1}{2}\right)^4 =$$

$$= \frac{4!\cdot 5 \cdot 6}{2!4!}\left(\frac{1}{2}\right)^6 = \frac{15}{64}.$$

Пример 1.11.2. Некоторый стрелок попадает в цель с вероятностью 0,6, он собирается произвести 10 выстрелов. Найти вероятность того, что он попадет в цель: а) три раза, б) хотя бы один раз.
Решение. p = 0,6, q =1- p= 1 – 0,6= 0,4, n=10.

а) $$\mathrm{P}_{10}(3) = C_{10}^{3} \cdot \mathrm{p}^{3} \cdot q^{10-3} = \frac{10!}{(10-3)!3!} \cdot 0{,}6^{3} \cdot 0{,}4^{7} =$$

$$= \frac{7! \cdot 8 \cdot 9 \cdot 10}{7! \cdot 3!} \cdot 0{,}6^{3} \cdot 0{,}4^{7} \approx 0{,}255;$$

б) $P_{10}(k \geq 10) = 1 - P_{10}(0) = C_{10}^{0} p^{0} q^{10-0} = 1 - 0{,}4^{10}$.

Наивероятнейшее число наступивших событий в схеме Бернулли - $к_о$ ($к_o \in$N), определяется из следующего неравенства:

$$np\text{-}\ q \leq k_o \leq np + p. \qquad (1.11.1)$$

В примере 1.11.1: $$6 \cdot \frac{1}{2} - \frac{1}{2} \leq k_о \leq 6 \cdot \frac{1}{2} + \frac{1}{2},$$
$$2{,}5 \leq k_о \leq 3{,}5,$$
$$k_о = 3.$$

В примере 1.11.2: $$10 \cdot 0{,}6 - 0{,}4 \leq k_о \leq 10 \cdot 0{,}6 + 0{,}6,$$
$$5{,}6 \leq k_о \leq 6{,}6,$$
$$k_о = 6.$$

Пример 1.11.3. Опыт состоит в подбрасывании 12 игральных костей. Какова вероятность, что выпадут 7 одинаковых цифр, а остальные по разу?
Решение. Для каждой игральной кости элементарные события $A_1, A_2, A_3, A_4, A_5, A_6$ соответствуют 6 граням, вероятность каждого события $P(A_i) = \frac{1}{6}$. Одинаковые цифры могут быть: 1, 2, 3, 4, 5, 6. Поэтому сначала найдем вероятность того, что появилось семь единиц. По формуле полиномиального распределения (1.11.3) имеем:

$$\frac{12!}{7! \cdot 1! \cdot 1! \cdot 1! \cdot 1! \cdot 1!} \cdot \left(\frac{1}{6}\right)^{7} \cdot \left(\frac{1}{6}\right)^{1} \cdot \left(\frac{1}{6}\right)^{1} \cdot \left(\frac{1}{6}\right)^{1} \cdot \left(\frac{1}{6}\right)^{1} \cdot \left(\frac{1}{6}\right)^{1} = \frac{12!}{7!} \cdot \left(\frac{1}{6}\right)^{12}.$$

Вероятность появления любых семи одинаковых цифр из 6 очевидно будет в 6 раз больше:

$$P(A) = 6 \cdot \frac{12!}{7!} \cdot \left(\frac{1}{6}\right)^{12} = \frac{12!}{7!}\left(\frac{1}{6}\right)^{11}.$$

Контрольные задания 1.11.

1. Найти вероятность того, что при четырех подбрасываниях игральной кости, 5 очков появится: а) два раза; б) хотя бы один раз.
2. Всхожесть семян некоторого растения составляет 80%. Найти вероятность того, что из пяти посеянных семян взойдут: а) пять семян; б) не менее четырех; в) не более одного.
3. Вероятность выбора отличника на факультете равна 1/7. Из 28 студентов группы наудачу вызываются три студента. Определить вероятности всех возможных значений числа отличников, которые могут оказаться среди вызванных трех студентов.

4. В семье 5 детей. Считая вероятности рождений мальчика и девочки одинаковыми, найти вероятность того, что среди этих детей: а) два мальчика; б) не более двух мальчиков; в) более двух мальчиков; г) не менее двух и не более трех мальчиков.
5. Всхожесть клубней картофеля равна 80%. Сколько нужно посадить клубней, чтобы наивероятнейшее число взошедших из них было равно 100?
6. Сколько раз нужно подбросить игральную кость, чтобы наивероятнейшее число выпадения 6 очков было равно 50?
7. Два равносильных противника играют в шахматы. Для каждого из них, что вероятнее выиграть: а) одну партию из двух или две из четырех; б) не менее двух партий из четырех или не менее трех партий из пяти. Ничьи во внимание не принимаются.
8. Бланк программированного опроса состоит из пяти вопросов. На каждый даны три ответа, среди которых один правильный. Какова вероятность, что методом угадывания студенту удастся выбрать, по крайней мере, четыре правильных ответа?
9. Вероятность появления события A в каждом из 6 независимых испытаний равна 0,7. Найти вероятность того, что событие A наступит хотя бы в одном испытании.
10. Событие A появится в случае, если событие B наступит не менее четырех раз. Найти вероятность наступления события A, если будет произведено 5 независимых испытаний, в каждом из которых вероятность наступления события B равна 0,8.
11. Два стрелка производят по n выстрелов, причём каждый стреляет по своей мишени. Определить вероятность того, что у них будет по одинаковому числу попаданий, если вероятность попадания при каждом выстреле постоянна и равна 0,5.
12. Торговый агент в среднем контактирует с восемью потенциальными покупателями в день. Из опыта ему известно, что вероятность того, что потенциальный покупатель совершит покупку, равна 0,1.
    а) Чему равна для агента вероятность двух продаж в течение одного дня?
    б) Чему равна вероятность того, что у агента будут хотя бы две продажи в течение дня?
    в) Чему равна вероятность того, что в течение одного дня не будет продаж?
13. Фирма предлагает в продажу со склада партию из 10 компьютеров, 4 из которых – с дефектами. Покупатель приобретает 5 из них, не зная о возможных дефектах. Чему равна вероятность что, что все 5 компьютеров окажутся без дефектов?

1.12. Приближенные формулы в схеме Бернулли

При большом числе опытов по схеме Бернулли удобнее пользоваться приближенными формулами.

1. *Локальная формула Муавра-Лапласа.*

Если $npq \geq 10$, то

$$P_n(k) \approx \frac{1}{\sqrt{npq}} \cdot \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}, \tag{1.12.1}$$

где вероятность p отлична от 0 и 1 ($p \to 0{,}5$), $x = \dfrac{\mathrm{k} - np}{\sqrt{npq}}$.

Для облегчения вычислений функция

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} \tag{1.12.2}$$

представлена в виде таблицы (прил.1).

$\varphi(x)$ – функция вероятности нормального распределения (рис. 12) имеет следующие свойства:

1) $\varphi(x)$ – четная;
2) точки перегиба $x=\pm 1$;
3) при $x\geq 4$, $\varphi(x) \to 0$, поэтому функция $\varphi(x)$ представлена в виде таблицы (затабулирована) для $0 \leq x \leq 4$ (прил.1).

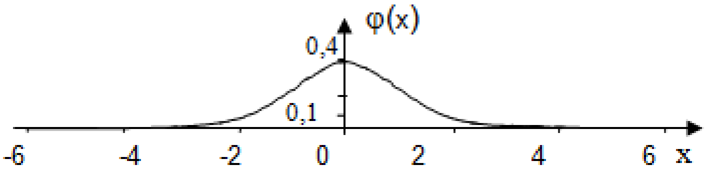


Рис. 12- Функция вероятности нормального распределения

2. *Формула Пуассона.*

Если npq<10 и p<0,1, то:

$$P_n(k) \approx \frac{\lambda^k}{k!} \cdot e^{-\lambda}, \qquad (1.12.3)$$

где $\lambda = np$.

Пример 1.12.1. Стрелок выполнил 400 выстрелов. Найти вероятность 325 попаданий, если вероятность попадания при каждом выстреле равна 0,8.

Решение. npq=400·0,8·0,2=64≥10, следовательно**,** по формуле (1.12.1):

$$P_n(k) \approx \frac{1}{\sqrt{npq}}\varphi(x), \text{ где } x = \frac{k-np}{\sqrt{npq}},$$

$$P_{400}(325) \approx \frac{1}{\sqrt{64}}\varphi(0{,}63) \approx \frac{1}{8} \cdot 0{,}3271 \approx 0{,}041.$$

Пример 1.12.2. Завод отправил 5000 доброкачественных изделий. Вероятность того, что в пути разбили одно изделие 0,0002. Найти вероятность того, что в пути будет повреждено:

а) 3 изделия;
б) 1 изделие;
в) не более 3-х изделий.

Решение. $npq = 5000 \cdot 0{,}0002 \cdot 0{,}9998 = 0{,}9998 < 10$ и $p < 0{,}1$, поэтому применяем формулу Пуассона (1.12.3), где $\lambda=np = 5000 \cdot 0{,}0002 = 1$.

а) при k=3: $P_{5000}(3) \approx \frac{1^3}{3!} \cdot e^{-1} = \frac{1}{6e} \approx 0{,}061$,

б) при k=1: $P_{5000}(1) \approx \frac{1^1}{1!} \cdot e^{-1} = \frac{1}{e} \approx \frac{1}{2{,}71828} \approx 0{,}368.$

в) $P_{5000}(0 \leq k \leq 3) = P_{5000}(0) + P_{5000}(1) + P_{5000}(2) + P_{5000}(3)=$

$$= \frac{1}{e} + \frac{1}{e} + \frac{1}{2e} + \frac{1}{6e} = \frac{16}{6e} = \frac{8}{3e} \approx 0{,}981.$$

3. При больших значениях n , для вычисления вероятности того, что произойдет от k_1 до k_2 событий по схеме Бернулли, используется *интегральная формула Муавра-Лапласса:*

$$P_n(k_1 \le k \le k_2) = \Phi(x_2) - \Phi(x_1), \qquad (1.12.4)$$

где $x_1 = \dfrac{k_1 - np}{\sqrt{npq}}$, $x_2 = \dfrac{k_2 - np}{\sqrt{npq}}$, Φ(x) - функция Лапласа (рис. 10).

$\Phi(x)$ имеет следующие свойства:

1) $\Phi(-x) = -\Phi(x)$ – функция нечетная, поэтому достаточно применять её для неотрицательных значений x:

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt; \qquad (1.12.5)$$

2) функция $\Phi(x)$ возрастает на всей числовой оси;

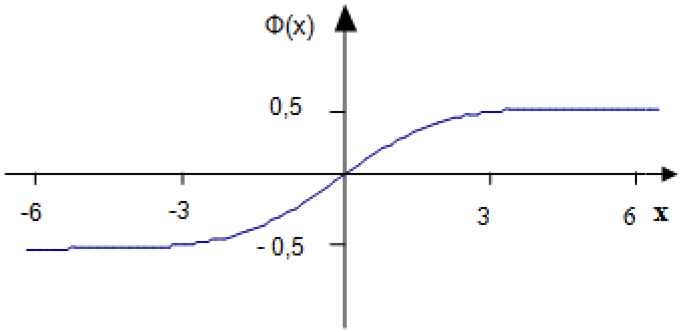


Рис. 13 - Функция Лапласа

3) при $x \ge 4$, $\Phi(x) \to \frac{1}{2}$ (y=0,5- горизонтальная асимптота при $x>0$), поэтому функция представлена в виде таблицы для $0 \le x \le 4$ (прил.1);

4) вероятность отклонения относительной частоты от постоянной вероятности в независимых испытаниях не более чем на некоторое число ε>0:

$$P_n\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = 2\Phi\left(\varepsilon \cdot \sqrt{\frac{n}{pq}}\right). \qquad (1.12.6)$$

Пример 1.12.3. Стрелок выполнил 400 выстрелов, вероятность одного попадания 0,8. Найти вероятность того, что он попадет от 310 до 325 раз.

Решение. $P_{400}(310 \le x \le 325) = \Phi(x_2) - \Phi(x_1)$, где

$$x_2 = \frac{k_2 - np}{\sqrt{npq}} = \frac{325 - 320}{\sqrt{64}} = \frac{5}{8} = 0{,}63, \quad x_1 = \frac{k_1 - np}{\sqrt{npq}} = \frac{310 - 320}{\sqrt{64}} = \frac{-10}{8} = -1{,}25,$$

$$P_{400}(310 \le k \le 325) \approx \Phi(0{,}63) - \Phi(-1{,}25) = \Phi(0{,}63) + \Phi(1{,}25) = 0{,}2357 + 0{,}3944 = 0{,}6301.$$

Пример 1.12.4. В каждом из 10000 независимых испытаний вероятность успеха p=0,75. Найти вероятность того, что относительная частота появления события отклонится от постоянной вероятности по абсолютной величине не более чем на 0,001.

Решение. n=10000, $p = 0{,}75$, $q = 1 - p = 1 - 0{,}75 = 0{,}25$, $\varepsilon = 0{,}001$, следовательно,

$$P_{10000}\left(\left|\frac{k}{n}-0{,}75\right|<0{,}001\right)=2\Phi\left(\varepsilon\cdot\sqrt{\frac{n}{pq}}\right)=$$

$$=2\Phi\left(0{,}001\cdot\sqrt{\frac{10000}{0{,}75\cdot 0{,}25}}\right)=2\Phi(0{,}23)=0{,}182.$$

Вероятность того, что отклонение относительной частоты от вероятности в независимых испытаниях не превысит 0,001, равна 0,182.

Пример 1.12.5. Вероятность появления события в каждом независимом испытании p=0,2. Найти, какое отклонение относительной частоты появления события от его вероятности может произойти с вероятностью 0,9128 , при 5 тысячах повторениях независимых испытаний по схеме Бернулли.

Решение. $P_{5000}\left(\left|\frac{k}{n}-p\right|<\varepsilon\right)$=0,9128 или $2\Phi\left(\varepsilon\cdot\sqrt{\frac{n}{pq}}\right)=0{,}9128$, отсюда:

$$\Phi\left(\varepsilon\cdot\sqrt{\frac{5000}{0{,}2\cdot 0{,}8}}\right)=0{,}4564,\quad \varepsilon\cdot\sqrt{\frac{5000}{0{,}16}}=1{,}71,\qquad \varepsilon=\frac{1{,}71}{\sqrt{5000/0{,}16}}\approx 0{,}00967\ .$$

Пример 1.12.6. Сколько раз нужно бросить монету, чтобы с вероятностью 0,6 можно было ожидать, что отклонение относительной частоты появления герба от вероятности p=0,5 окажется по абсолютной величине не больше 0,01.

Решение. По условию $2\Phi\left(\varepsilon\cdot\sqrt{\frac{n}{pq}}\right)=0{,}6$. Отсюда:

$$\Phi\left(\varepsilon\cdot\sqrt{\frac{n}{pq}}\right)=0{,}3,\quad \varepsilon\cdot\sqrt{\frac{n}{pq}}=0{,}84,\qquad \sqrt{\frac{n}{pq}}=0{,}84/\varepsilon,$$

$$n=\left(\frac{0{,}84}{\varepsilon}\right)^2 pq=\left(\frac{0{,}84}{0{,}01}\right)^2\cdot 0{,}5\cdot(1-0{,}5)=84^2\cdot 0{,}25=(84\cdot 0{,}5)^2=42^2=1764.$$

Контрольные задания 1.12.

1. Вероятность того, что при сортировке изделий одно из них будет разбито, равна 0,005. Найти вероятность того, что из 200 изделий окажутся разбитыми: а) три изделия; б) не более двух; в) не менее двух изделий.
2. Станок автомат делает детали. Вероятность того, что деталь окажется бракованной равна 0,01. Найти вероятность того, что среди 200 деталей окажется ровно 4 бракованных.
3. Установлено, что виноградник поражен вредителями в среднем на 10 %. Определить вероятность того, что из 10 проверенных кустов винограда один будет поражен. Вычислить вероятности по формулам Бернулли, Лапласа, Пуассона. Сравнить результаты, сделать выводы.
4. На факультете 900 студентов. Вероятность дня рождения каждого студента в данный день равна $\frac{1}{365}$. Найти вероятность того, что найдутся три студента с одним и тем же днем рождения.
5. Вероятность получения отличной оценки на экзамене равна 0,2. Найти наивероятнейшее число отличных оценок и вероятность этого числа, если сдают экзамен 50 студентов.
6. Вероятность того, что автомат при опускании одной монеты правильно сработает, равна 0,99. Найти наиболее вероятное число случаев неправильной работы автомата и вероятность этого числа случаев, если будет опущено 200 монет.

7. Численность работников предприятия составляет 500 человек. Вероятность невыхода на работу из-за болезни равна 0,01 для каждого работника предприятия. Определить вероятность того, что в ближайший день не выйдет на работу хотя бы один из работников.
8. В пчелиной семье 5000 пчел. Вероятность заболевания в течение дня равна 0,001 для каждой пчелы. Найти вероятность того, что в течение дня заболеет более чем одна пчела.
9. Известно, что 80 % специалистов в районе имеет высшее образование. Найти вероятность того, что из 100 наудачу отобранных человек высшее образование имеет: а) не менее 70; б) от 65 до 90 человек.
10. Всхожесть семян составляет 80 %. Какова вероятность того, что из 1000 посеянных семян взойдут от 650 до 760?
11. Найти такое число К, чтобы с вероятностью 0,9 можно было утверждать, что среди 900 новорожденных более К мальчиков, если вероятность рождения мальчика 0,515.
12. В автопарке 70 машин. Вероятность поломки машины равна 0,2. Найти наивероятнейшее число исправных автомобилей и вероятность этого числа.
13. Всхожесть зерна, хранящегося на складе равна 80%. Какова вероятность того, что среди 100 зерен: а) число всхожих составит от 68 до 90 шт.; б) доля (частость) всхожих зерен будет отличаться от вероятности 0,8 по абсолютной величине не более чем на 0,1?
14. Два стрелка одновременно делают выстрелы по мишени. Сколько нужно произвести залпов, если наивероятнейшее число залпов, при которых оба стрелка попадут в мишень, равно 8, причем вероятность попадания в мишень при одном выстреле для первого стрелка равна 0,5, а для второго – 0,8?
15. При проведении некоторого опыта вероятность появления ожидаемого результата равна 0,01. Сколько раз нужно провести опыт, чтобы с вероятностью 0,5 можно было бы ожидать хотя бы одного появления этого результата?
16. В автопарке имеется 400 автомобилей. Вероятность безотказной работы каждого из них равна 0,9. С вероятностью 0,95 определить границы, в которых будет находиться доля безотказно работавших машин в определенный момент времени.
17. Всхожесть зерна равна 90%. Определить вероятность того, что для отобранных случайным образом 100 зерен относительная частота всхожести по абсолютной величине будет отличаться от вероятности взойти $p = 0,9$ не более чем на 0,1.
18. Вероятность появления события в каждом из 400 независимых испытаний равна 0,8. Найти такое положительное число ε, чтобы с вероятностью 0,9876 абсолютная величина отклонения относительной частоты появления события от вероятности 0,8, не превысила ε.
19. Отдел контроля проверяет на стандартность 900 деталей. Вероятность того, что деталь стандартна, равна 0,9. С вероятностью 0,9544 найти границы, в которых будет заключено число стандартных деталей.
20. Известно, что 10% делянок под овощами плохо обработаны. Сколько нужно проверить делянок, чтобы с вероятностью 0,9973 можно было утверждать, что относительная частота засоренных делянок будет отличаться от вероятности засоренности по модулю не более чем на 0,01?
21. Для определения степени поражения винограда вредителями было обследовано 400 кустов. Вероятность поражения куста виноградника равна 0,03. Определить границы, в которых с вероятностью 0,9545 будет заключено число кустов, не пораженных вредителями.
22. Проверяется всхожесть кукурузы. Сколько семян необходимо посеять с вероятностью всхожести 0,99, чтобы частота всхожести отличалась от 0,95 меньше, чем на 0,01?
23. Вероятность того, что человек в период страхования будет травмирован, равна 0,006. Компанией застраховано 1000 человек. Годовой взнос с человека составляет 150 руб. В случае получения травмы застраховавшийся получает 12000 руб. Какова вероятность того, что выплата по страховкам превысит сумму страховых взносов?

Индивидуальные задания к главе 1

Задания уровня сложности А.

Для всех вариантов номера задач находятся в приложении 10.

Задания уровня сложности В[4].

Для всех вариантов задания 81-101 (исходные данные находятся в приложении 11) .

1. Из 20 вопросов, входящих в экзаменационный билет, студент подготовил 17. Найти вероятность того, что студент ответил правильно на экзаменационный билет, состоящий из 2-х вопросов.
2. Рабочий обслуживает 3 станка. Вероятность безотказной работы первого из них равна 0,75, второго 0,85, третьего 0,95. Найти вероятность того, что а) откажут два станка, б) все три станка будут работать безотказно, в) хотя бы один станок откажет в работе.
3. Из колоды содержащей 52 карты вынимается наугад 3. Найти вероятность, что это тройка, семерка и туз.
4. Найти вероятность того, что абонент наберет правильный двухзначный номер, если он знает, что данный номер не делится на 5.
5. Игральная кость подброшена два раза: а) найти вероятность того, что сумма очков на верхних гранях составит 7; б) найти вероятность того, что хотя бы два очка появится при одном подбрасывании.
6. В урне имеется 5 черных и 7 красных шаров. Последовательно (без возвращения) извлекается три шара. Найти вероятность того, что: а) все три шара будут красными; б) три шара будут красными или черными.
7. В группе из 15 человек 6 человек занимаются спортом. Найти вероятность того, что из случайно отобранных 7 человек 5 человек занимаются спортом.
8. Мышь может выбрать наугад один из 5 лабиринтов. Известно, что вероятности ее выхода из различных лабиринтов за три минуты равны 0,5; 0,6; 0,2; 0,1; 0,1. Пусть оказалось, что мышь вырвалась из лабиринта через три минуты. Какова вероятность того, что она выбрала: а) первый лабиринт? б) второй лабиринт?
9. Из 10 билетов выигрышными являются 2. Найти вероятность того, что из 5 случайно взятых билетов выигрышным является один.
10. В сентябре вероятность дождливого дня равна 0,3. Команда «Статистик» выигрывает в футбол в ясный день с вероятность 0,8, а в дождливый день эта вероятность равна 0,3. Известно, что в сентябре они выиграли некоторую игру. Какова вероятность, что в тот день: а) шел дождь?; б) был ясный день?
11. Вероятность попадания в цель первым стрелком равна 0,7, вторым – 0,5, третьим – 0,4. Найти вероятность того, что хотя бы один стрелок попадет в цель.
12. В первом ящике содержится 20 деталей, из них 10 стандартных, во втором 30 деталей, из них 25 стандартных, в третьем 10 деталей, из них 8 стандартных. Из случайно взятого ящика наудачу взята одна деталь, которая оказалась стандартной. Найти вероятность того, что она взята из второго ящика.
13. На каждой из пяти одинаковых карточек написана одна из следующих букв: А, Е, Н, С, Т. Карточки перемешаны. Определить вероятность того, что из вынутых и положенных в ряд карточек а) можно составить слово «СТЕНА», б) из трех карточек можно составить слово «НЕТ».
14. Для поражения цели достаточно попадания хотя бы одного снаряда. Произведено два залпа из двух орудий. Найти вероятность поражения цели, если вероятность попадания в цель при одном выстреле из первого орудия равна 0,46, второго 0,6.
15. Имеется 3 урны. В первой урне 6 черных и 4 белых шара, во второй 5 белых и 5 черных шаров, в третьей 7 белых и 3 черных шара. Случайно выбирается урна и из нее извлекается шар, который оказался белым. Найти вероятность того, что выбрана вторая урна.
16. Монета подбрасывается 3 раза. Найти вероятность того, что герб появится: а) все 3 раза, б) только один раз, в) хотя бы один раз.

[4] Задания уровня В заимствованы из типовых расчетов Чудесенко В.Ф. [12].

17. На отдельных карточках написаны цифры 0, 1, 2, 3, 4, 5, 6, 7, 8, 9. Все карточки перемешиваются, после чего наугад берут 5 карточек и раскладывают их в ряд. Определить вероятность того, что будет получено число 1 2 0 3 5. (Задачу решить, используя определение вероятности события и теоремы теории вероятностей).
18. Три известных экономиста предложили одновременно три экономические теории, которые считались равновероятными. После наблюдения над состоянием экономики оказалось, что вероятность того развития, которое она получила на самом деле, в соответствии с первой теорией была равна 0,5; со второй – 0,7; с третьей – 0,4. Каким образом это изменяет вероятности правильности трех теорий?
19. В магазине продается 4 магнитофона. Вероятность того, что они выдержат гарантийный срок, соответственно равны: 0,91; 0,9; 0,95; 0,94. Найти вероятность того, что взятый наудачу магнитофон выдержит гарантийный срок.
20. Игральная кость сделана так, что вероятность выпадения определенного числа пропорциональна числу очков. Какова вероятность выпадения трех очков, если известно, что выпало нечетное число очков.
21. Брошены 2 игральные кости. Какова вероятность того, что абсолютная величина разности выпавших очков равна 3?
22. Студент в поисках книги посещает 3 библиотеки. Вероятность того, что они есть в библиотеках равны 0,4, 0,5, 0,1, а того, а того что они выданы или нет - равновероятные события. Какова вероятность того, что нужная книга найдена?
23. Найти вероятность того, что дни рождения 12 человек придутся на разные месяцы года.
24. В урне имеется 10 белых, 5 черных и 15 красных шаров. Извлекается последовательно 2 шара. Рассматриваются 2 события: А – хотя бы один шар из двух вынутых красный; В – хотя бы один вынутый шар белый. Найти вероятность события С = А + В.
25. Наудачу набранный номер состоит из 5 цифр. Определить вероятность того, что все цифры в нем различны.
26. В магазин трикотажных изделий поступили носки, 60% которых получено от одной фабрики, 25% - другой и 15% - третьей. Найти вероятность того, что купленные покупателем носки изготовлены на второй или третьей фабрике.
27. Пассажир за получением билета может обратиться в одну из касс. Вероятность обращения в первую кассу составляет 0,4, вторую – 0,35 и третью – 0,25. Вероятность того, что к моменту прихода пассажира имеющиеся в кассе билеты будут проданы, равна для первой кассы 0,3, для второй 0,4, для третьей 0,6. Найти вероятность того, что пассажир купит билет.
28. Бросаются 4 игральные кости. Найти вероятность того, что: а) хотя бы на одной появится 2 очка, б) на них выпадет по одинаковому числу очков.
29. Из 9 жетонов, занумерованных разными однозначными цифрами, выбирается 3. Найти вероятность того, что последовательная запись их номеров покажет возрастание значений цифр.
30. Вероятность выигрыша по лотерейному билету равна 0,1. Какова вероятность того, что выиграет хотя бы один билет из трех купленных?
31. Из полной колоды карт (52 листа) вынимают сразу 4 карты. Найти вероятность того, что все эти карты будут разных мастей.
32. Имеется 3 урны. В первой из них 5 белых и 6 черных шаров, во второй 4 белых и 3 черных шара, в третьей 5 белых и 3 черных шара. Некто наугад выбирает одну из урн и вынимает из нее шар. Этот шар оказался белым. Найти вероятность того, что этот шар вынут из второй урны.
33. В магазине имеется в продаже 20 пар обуви, из которых 7 пар 42 размера. Найти вероятность того, что из 8 покупателей 3 выберут обувь 42 размера.
34. В мешке смешаны катушки с нитями трех цветов: 30% белых, 50% красных, остальные зеленые. Определить вероятность того, что при последовательном вытягивании наугад трех нитей окажется, что все они одного цвета.

35. В урне «a» белых и «b» черных шаров. Из урны вынули один шар и, не глядя, отложили в сторону. После этого из урны взяли еще один шар. Он оказался белым. Найти вероятность того, что первый шар, отложенный в сторону – тоже белый.
36. У рыбака имеется 2 места ловли рыбы, которые он посещает с одинаковой вероятностью. Если он закидывает удочку на первом месте, рыба клюет с вероятностью 0,6, на втором – с вероятностью 0,7. Рыбак, выйдя на ловлю в одно из мест, 2 раза закинул удочку. Найти вероятность того, что рыба клюнет только один раз.
37. На сборку поступило 50 деталей от первого станка, 100 от второго и 150 от третьего. Первый станок дает 2%, второй 1% и третий 2% брака. Найти вероятность того, что взятая наугад деталь окажется не бракованной. Задачу решить, используя: а) определение вероятности события; б) по формуле полной вероятности.
38. Найти вероятность того, что на 2 определенные карточки в Спортлото «5 из 36» будет получено по минимальному выигрышу (угадано ровно три числа).
39. Вероятность того, что стрелок попадет, хотя бы один раз при трех выстрелах равна, 0,992. Найти вероятность попадания в цель при одном выстреле, предполагая ее постоянной при каждом выстреле.
40. Пусть 3% всех мужчин и 0,5% всех женщин − дальтоники. Наугад выбранное человек оказался дальтоником. Какова вероятность, что это мужчина? (Считать, что количество мужчин и женщин одинаково.)
41. В группе из 25 человек 10 учится на «отлично», 8 на «хорошо» и 7 на «удовлетворительно». Найти вероятность того, что из взятых наугад 8 человек 3 человека учатся на «отлично».
42. Какова вероятность того, что наудачу вырванный листок из нового календаря соответствует первому числу месяца? (Год считается не високосным.)
43. В группе спортсменов 10 лыжников, 6 боксеров и 4 бегуна. Вероятность выполнить квалификационную норму для лыжников составляет 0,8, боксеров 0,7, бегунов 0,9. Найти вероятность того, что спортсмен, выбранный наудачу, выполнит квалификационную норму.
44. На одной полке наудачу расставляется 8 книг. Найти вероятность того, что определенные 3 книги окажутся поставленными рядом.
45. Монету бросают три раза. Какое из событий более вероятно: событие А – все три раза выпала цифра или событие В – два раза выпала цифра и один раз герб? Подсчитать вероятности этих событий.
46. К концу дня в магазине осталось 60 арбузов, из которых 50 спелых. Покупатель выбирает два арбуза. Какова вероятность, что оба арбуза спелые?
47. На один ряд из 7 мест, случайным образом садятся семь учеников. Найти вероятность того, что три определенных ученика окажутся рядом.
48. Известно, что при 10 – кратном бросании монеты 5 раз выпали гербы и 5 раз цифры. Какова вероятность того, что все гербы выпали при первых пяти бросаниях?
49. Из 15 строительных рабочих 10 – штукатуры, а 5 – маляры. Наудачу отбирается бригада из 5 рабочих. Какова вероятность того, что среди них будет 3 маляра и 2 штукатура?
50. Игральная кость подброшена 3 раза. Найти вероятность того, что: а) все 3 раза выпадет четное число очков; б) четное число очков выпадет только один раз; в) четное число очков выпадет хотя бы один раз.
51. Два автомата производят детали, которые поступают на общий конвейер. Производительность первого автомата в 3 раза больше производительности второго. Вероятность изготовления не бракованной детали первым автоматом равна 0,95, а вторым 0,8. Найти вероятность того, что взятая наугад деталь будет стандартной.
52. Какова вероятность получения 1 туза, туза и короля при сдаче 6 карт из колоды в 52 карты?

53. В соревнованиях по футболу участвуют 20 команд. Случайным образом они делятся на две группы по 10 команд. Какова вероятность того, что 2 наиболее сильные команды при этом окажутся в одной группе?
54. Гардеробщица одновременно выдала номерки пяти лицам, сдавшим в гардероб свои шляпы, и повесила их наугад. Найти вероятность того, что она каждому выдаст его собственную шляпу.
55. Несколько раз бросают игральную кость. Какова вероятность того, что одно очко появится впервые при третьем бросании?
56. 20 машин были доставлены на станцию технического обслуживания. При этом 5 из них имели неисправность в ходовой части, 8 имели неисправности в моторе, а 10 были полностью исправны. Какова вероятность, что машина с неисправной ходовой частью имеет также неисправный мотор?
57. Из 15 билетов выигрышными являются 2. Найти вероятность того, что из 10 билетов выигрышным является один.
58. Готовясь к вступительному экзамену по математике, абитуриент должен подготовить 20 вопросов по элементам математического анализа и 25 по геометрии. Однако он успел подготовить только 15 вопросов по элементам математического анализа и 20 по геометрии. Билет содержит 3 вопроса, 2 из которых по элементам математического анализа и 1 по геометрии. Какова вероятность, что: а) студент сдает экзамен на отлично (отвечает на все три вопроса); б) на хорошо (отвечает на любые два вопроса)?
59. На стеллаже 15 учебников, 5 из них в переплете. Наудачу выбирают 3 учебника. Какова вероятность, что хотя бы один из них будет в переплете?
60. Из пяти винтовок, из которых 3 снайперские и 2 обычные, наудачу выбирается одна, и из нее производится выстрел. Найти вероятность попадания в мишень, если вероятность попадания из снайперской винтовки – 0,95, а из обычной − 0,7.
61. Вероятность попадания в цель при одном выстреле равна 0,7. Произведено 3 выстрела. Какова вероятность, что будет: а) три попадания; б) один промах; в) хотя бы одно попадание.
62. На спортивных соревнованиях вероятность показать рекордный результат для первого спортсмена − 0,5, для второго − 0,3, для третьего − 0,1. Какова вероятность того, что: а) рекорд будет установлен одним спортсменом; б) рекорд будет установлен хотя бы одним спортсменом; в) рекорд не будет установлен.
63. В первой урне из 10 шаров, 6 черного и 4 белого цвета, во второй 3 черных и 7 белых шаров. Из каждой урны наудачу извлекается один шар. Какова вероятность того, что вынуты: а) 2 белых шара; б) хотя бы один шар черный; в) белый и черный в любой последовательности.
64. Вероятность того, что хотя бы один из трех покупателей купит определенный товар, равна 0,784. Вероятности покупки товара покупателями одинаковы. Определить вероятность того, что: а) два покупателя совершат покупки; б) три покупателя совершат покупки.
65. В коробке находятся жетоны с цифрами от 1 до 10. Наудачу извлекаются два жетона. Какова вероятность того, что будут вынуты: а) оба жетона с нечетными номерами; б) хотя бы один жетон с нечетным номером; в) один жетон с четным номером.
66. В двух группах обучается по 25 студентов. В первой группе сессию на «отлично» сдали 7 человек, во второй − 4 человека. Из каждой группы наудачу вызывают по одному студенту. Какова вероятность того, что: а) оба студента отличники; б) только один отличник; в) хотя бы один отличник.
67. В первой бригаде из 8 тракторов 2 требуют ремонта, во второй из 6 - тракторов 1 требует ремонта. Из каждой бригады наудачу выбирают по одному трактору. Определить вероятность того, что: а) оба трактора исправны; б) хотя бы один исправен; в) только один исправен.

68. В организации работают 12 мужчин и 8 женщин. Для них выделено 3 премии. Определить вероятность того, что премию получат: а) двое мужчин и одна женщина; б) только женщины; в) хотя бы один мужчина.
69. Из 25 работников, предприятия 10 имеют высшее образование: Определить вероятность того, что из случайно отобранных трех человек высшее образование имеют; а) три человека; б) один человек; в) хотя бы один человек.
70. На карточках написаны буквы «К», «А», «Р», «Т», «О», «Ч», «К», «А». Карточки перемешивают и кладут в порядке их вытаскивания. Какова вероятность того, что получится: а) слово «КАРТОЧКА»; б) слово «КАРТА»; в) слово «ТОК».
71. В коробке из 25 изделий 15 повышенного качества. Наудачу извлекается 3 изделия. Определить вероятность того, что: а) одно из них повышенного качества; б) все три изделия повышенного качества; в) хотя бы одно изделие повышенного качества.
72. Бросается три игральных кости. Какова вероятность того, что: а) хотя бы на одной из них появится 5 очков; б) на всех выпадут нечетные цифры; в) на всех костях выпадут одинаковые цифры.
73. В первом ящике из 6 шаров 4 красных и 2 черных, во втором ящике из 7 шаров 2 красных и 5 черных. Из первого ящика во второй переложили один шар, затем из второго в первый переложили один шар. Найти вероятность того, что шар, извлеченный после этого из первого ящика - черный.
74. Два предприятия выпускают однотипные изделия. Причем второе выпускает 55% изделий обоих предприятий. Вероятность выпуска нестандартного изделия первым предприятием равна 0,1, вторым − 0,15. а) Определить вероятность того, что: а) взятое наудачу изделие окажется не стандартным; б) взятое изделие оказалось не стандартным. Какова вероятность, что оно выпущено на втором предприятии?
75. Имеется три урны. В первой 3 белых и 2 черных шара, во второй и третьей по 4 белых и 3 черных шара. Из случайно выбранной урны извлекается шар. Он оказался белым. Какова вероятность того, что шар взят из третьей урны?
76. Семена для посева в хозяйство поступают из трех семеноводческих хозяйств. Причем первое и второе хозяйства присылают по 40 % всех семян. Всхожесть семян из первого хозяйства равна 90%, второго − 85%, третьего 95%. а) Определить вероятность того, что наудачу взятое семя не взойдет. б) Наудачу взятое семя не взошло. Какова вероятность, что оно получено от второго хозяйства?
77. Программа экзамена состоит из 30 вопросов. Из двадцати студентов группы 8 человек выучили все вопросы, 6 человек − по 25 вопросов, 5 человек − по 20 вопросов, а один человек − 10 вопросов. Определить вероятность того, что случайно вызванный студент ответит на два вопроса билета.
78. Перед посевом 95% семян обрабатываются специальным раствором. Всхожесть семян после обработки − 99%, необработанных − 85%. а) Какова вероятность того, что случайно взятое семя взойдет? б) Случайно взятое семя взошло. Какова вероятность того, что оно выращено из обработанного семени?
79. В магазин поступают телевизоры четырех заводов. Вероятность того, что в течение года телевизор не будет иметь неисправность, равна: для первого завода − 0,9, для второго − 0,8, для третьего − 0,8 и для четвертого − 0,99. Случайно выбранный телевизор в течение года вышел из - строя. Какова вероятность того, что он изготовлен на первом заводе?
80. Покупатель с равной вероятностью посещает каждый из трех магазинов. Вероятность того, что покупатель купит товар в первом магазине, равна 0,4, втором − 0,6 и третьем − 0,8. а) Определить вероятность того, что покупатель купит товар в каком-то магазине. б) Покупатель купил товар. Найти вероятность того, что он купил его во втором магазине.
81. Для проверки результатов геодезических работ назначена группа экспертов, состоящая из трех подгрупп. В первой подгруппе - n_1 человек, во второй − n_2 и в третьей − n_3. Эксперты первой подгруппы принимают верное решение с вероятностью p_1 эксперты второй

подгруппы − p_2 эксперты третьей подгруппы − p_3. Наудачу вызванный эксперт принимает k независимых решений. Найти вероятность того, что: а) ровно 3 решения приняты, верно; в) принимал решения эксперт из первой подгруппы, если 3 решения приняты, верно. (Исходные данные находятся в приложении 11).

82. Бросаются две игральные кости. Определить вероятность того, что: а) сумма числа очков не превосходит N; б) произведение числа очков не превосходит N; в) произведение числа очков делится на N.
83. Имеются изделия четырех сортов, причем число изделий i-го сорта равно n_i, i=1, 2, 3, 4. Для контроля наудачу берутся m изделий. Определить вероятность того, что среди них m_1 первосортных, m_2, m_3 и m_4 второго, третьего и четвертого сорта соответственно $\left(\sum_{i=1}^{4} m = m\right)$.
84. Среди n лотерейных билетов k выигрышных. Наудачу взяли m билетов. Определить вероятность того, что среди них l выигрышных.
85. В лифт k-этажного дома сели n пассажиров ($n < k$).Каждый независимо от других с одинаковой вероятностью может выйти на любом (начиная со второго) этаже. Определить вероятность того, что: а) все вышли на разных этажах; б) по крайней мере, двое сошли на одном этаже.
86. В отрезке единичной длины наудачу появляется точка. Определить вероятность того, что расстояние от точки до концов отрезка превосходит величину $1/k$.
87. Моменты начала двух событий наудачу распределены в промежутке времена от T_1 до T_2. Одно из событий длится 10 мин., другое - t мин. Определить вероятность того, что: а) события «перекрываются» по времени; б) «не перекрываются».
88. В круге радиуса R наудачу появляется точка. Определить вероятность того, что она попадает в одну из двух непересекающихся фигур, площади которых равны S_1 и S_2.
89. В двух партиях процент доброкачественных изделий k_1 и k_2 - соответственно. Наудачу выбирают по одному изделию из каждой партии. Какова вероятность обнаружить среди них: а) хотя бы одно бракованное; б) два бракованных; в) одно доброкачественное и одно бракованное?
90. Вероятность того, что цель поражена при одном выстреле первым стрелком равна p_1, вторым − p_2. Первый сделал n_1, второй − n_2 выстрелов. Определить вероятность того, что цель не поражена.
91. Два игрока A и B поочередно бросают монету. Выигравшим считается тот, у кого раньше выпадает герб. Первый бросок делает игрок A, второй − B, третий − A и т. д.
    1. Найти вероятность указанного ниже события.
        Варианты 1-8. Выиграл A до k-го броска.
        Варианты 9-15. Выиграл A не позднее k-го броска.
        Варианты 16-23. Выиграл B до k-го броска.
        Варианты 24—31. Выиграл B не позднее k-го броска.
    2. Каковы вероятности выигрыша для каждого игрока при сколь угодно длительной игре?
92. Урна содержит M занумерованных шаров с номерами от 1 до M. Шары, извлекаются по одному без возвращения. Рассматриваются следующие события: A - номера шаров в порядке поступления образуют последовательность 1,2,... M; B - хотя бы один раз совпадает номер шара и порядковый номер извлечения; C - нет ни одного совпадения номера шара и порядкового номера извлечения. Определить вероятности событий A, B, C. Найти предельные значения вероятностей при $M \to \infty$.
93. Из 1000 ламп n_i, принадлежат i-й партии, i=1, 2, 3, $\sum_{i=1}^{3} n_i = 1000$. В первой партии 6%, во второй 5%, в третьей 4% бракованных ламп. Наудачу выбирается одна лампа. Определить вероятность того, что выбранная лампа - бракованная.

94. В первой урне − N_1 белых и M_1 черных шаров, во второй − N_2 белых и M_2 черных. Из первой во вторую переложено K шаров, затем из второй урны извлечен один шар. Определить вероятность того, что выбранный из второй урны шар - белый.
95. В альбоме k чистых и l гашеных марок. Из них наудачу извлекаются m марок (среди которых могут быть и чистые, и гашеные), подвергаются спецгашению и возвращаются в альбом. После этого вновь наудачу извлекается n марок. Определить вероятность того, что все n марок чистые.
96. В магазин поступают однотипные изделия с трех заводов, причем i - й завод поставляет m_i% изделий (i=1, 2, 3). Среди изделий i-го завода n_i% первосортных. Куплено одно изделие. Оно оказалось первосортным. Определить вероятность того, что купленное изделие выпущено i-м заводом.
97. Монета бросается до тех пор, пока герб не выпадает n раз. Определить вероятность того, что цифра выпадает m раз.
98. Вероятность выигрыша в лотерею на один билет равна p. Куплено n билетов. Найти наивероятнейшее число выигравших билетов и соответствующую вероятность.
99. На каждый лотерейный билет с вероятностью p_1 может выпасть крупный выигрыш, с вероятностью p_2 - мелкий выигрыш и с вероятностью p_3 билет может оказаться без выигрыша. Куплено n билетов. Определить вероятность получения n_1 крупных выигрышей и n_2 мелких.
100. Вероятность «сбоя» в работе телефонной станции при каждом вызове равна p. Поступило n вызовов. Определить вероятность m «сбоев».
101. Вероятность наступления некоторого события в каждом из n независимых испытаний равна p. Определить вероятность того, что число m наступлений события удовлетворяет следующему неравенству.

   Варианты 1-11: $k_1 \le m \le k_2$.

   Варианты 12—21: $k_1 \le m$.

   Варианты 22—31: $m \le k_2$

Глава 2. Случайные величины

2.1. Закон распределения дискретных случайных величин

Случайной величиной (СВ) называют такую величину, которая в результате опыта может принимать те или иные значения, причем до опыта мы не можем сказать какое именно значение она примет. (Более точно, СВ – это действительная функция, определенная на пространстве элементарных событий Ω).

Случайные величины обозначаются последними буквами латинского алфавита – *X,Y,Z*. Случайные величины могут быть трех типов:

- дискретные,
- непрерывные,
- смешанные (дискретно-непрерывные).

Дискретная случайная величина (ДСВ) может принимать конечное или бесконечное счетное число значений. Например, подбрасываем монету 5 раз. Случайная величина X - число появлений герба: 0, 1, 2, 3, 4, 5.

Непрерывная случайная величина (НСВ) в отличие от ДСВ принимает бесконечное несчетное число значений. Например, мишень имеет форму круга радиуса R. По этой мишени произвели выстрел с обязательным попаданием. Обозначим через Y расстояние от центра до точки попадания в мишень, $Y \in [0; R]$. Y - непрерывная случайная величина, так как она принимает бесконечное несчетное число значений.

Пусть X - дискретная случайная величина, которая принимает значения: x_1, x_2, …,x_n,… с некоторой вероятностью p_i, где $i = 1, 2, …, n,…$ Тогда можно говорить о вероятности того, что случайная величина X приняла значение x_i: $p_i=P(X=x_i)$.

Значения x_i и соответствующие p_i представляют в виде таблицы:

x_i	x_1	x_2	x_3	…	x_n	…
p_i	p_1	p_2	p_3	…	p_n	…

Эта таблица является одной из форм задания ДСВ.

Обычно случайные величины располагаются в возрастающем порядке.

Основное свойство таблицы заключено в том, что сумма вероятностей равна 1:

$$\sum_{i=1}^{\infty} p_i = p_1 + p_2 + p_3 + ... + p_n + … = 1. \qquad (2.1.1)$$

Пример 2.1.1. Монета бросается 5 раз. Представим закон распределения ДСВ X - числа появлений герба, в виде таблицы.

ДСВ X может принимать значения: 0, 1, 2, 3, 4, 5. Вероятность появления герба в одном опыте $p=\frac{1}{2}$, не появления $q=\frac{1}{2}$, $n = 5$.Таким образом, выполняются условия применения формулы Бернулли. Имеем:

$$P_5(X=0) = C_5^0 \mathrm{p}^0 q^{5-0} = 1 \cdot \left(\frac{1}{2}\right)^0 \cdot \left(\frac{1}{2}\right)^{5-0} = \frac{1}{32},$$

$$P_5(X=1)=C_5^1\,p^1 q^{5-1}=5\cdot\left(\frac{1}{2}\right)^1\cdot\left(\frac{1}{2}\right)^{5-1}=\frac{5}{32},$$

$$P_5(X=2)=C_5^2\,p^2 q^{5-2}=10\cdot\left(\frac{1}{2}\right)^2\cdot\left(\frac{1}{2}\right)^{5-2}=\frac{10}{32},$$

$$P_5(X=3)=C_5^3\,p^3 q^{5-3}=10\cdot\left(\frac{1}{2}\right)^3\cdot\left(\frac{1}{2}\right)^{5-3}=\frac{10}{32},$$

$$P_5(X=4)=C_5^4\,p^4 q^{5-4}=5\cdot\left(\frac{1}{2}\right)^4\cdot\left(\frac{1}{2}\right)^{5-4}=\frac{5}{32},$$

$$P_5(X=5)=C_5^5\,p^5 q^{5-5}=1\cdot\left(\frac{1}{2}\right)^5\cdot\left(\frac{1}{2}\right)^{5-5}=\frac{1}{32}.$$

Полученные данные представим в виде таблицы распределения:

x_i	0	1	2	3	4	5
p_i	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$

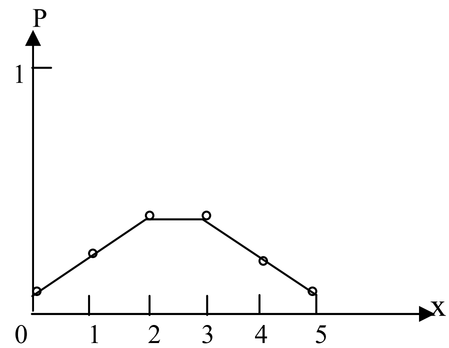


Рис. 14 - Многоугольник распределения

Дискретная случайная величина может быть представлена так же   в виде многоугольника распределения – фигуры состоящей из точек (x_i, p_i), соединённых отрезками (рис.14).

Над случайными величинами устанавливаются операции сложения и умножения.

1. Суммой двух случайных величин X и Y называется случайная величина, которая получается в результате сложения всех значений случайной величины X и всех значений случайной величины Y, соответствующие вероятности перемножаются.
2. Произведением двух случайных величин X и Y называется случайная величина, которая получается в результате перемножения всех значений случайной величины X и всех значений случайной величины Y, соответствующие вероятности перемножаются.

Пример 2.1.2.

X

x_i	0	1	2	3
p_i	0,1	0,4	0,3	0,2

Y

y_i	-1	0	1
p_i	0,2	0,3	0,5

Найти: 1) $X+C$, где C=2; 2)$X+Y$.

1. $Z = X + C$, C=2.

z_i	0+2	1+2	2+2	3+2
p_i	0,1	0,4	0,3	0,2

z_i	2	3	4	5
p_i	0,1	0,4	0,3	0,2

2. $Z = X + Y$.

z_i	0-1	0+0	0+1	1-1	1+0	1+1	2-1	2+0	2+1	3-1	3+0	3+1
p_i	0,02	0,03	0,05	0,08	0,12	0,2	0,06	0,09	0,15	0,04	0,06	0,1

Одинаковые значения СВ можно записать один раз, предварительно сложив соответствующие вероятности:

z_i	- 1	0	1	2	3	4
p_i	0,02	0,11	0,23	0,33	0,21	0,1

2.2. Часто встречающиеся распределения дискретных случайных величин

1. ***Закон распределения Бернулли***. Случайная величина X, распределенная по закону Бернулли (индикаторная случайная величина) принимает значения 1- успех или 0 - неудача, с вероятностями p и q соответственно ($p+q$=1).

x_i	0	1
p_i	q	p

2. ***Биномиальный закон распределения***. Случайная величина X принимает значения: 0, 1, 2, 3, 4, 5,…, n , с вероятностью, определяемой по формуле Бернулли (1.10.1):

x_i	0	1	2	…	k	...	n
p_i	$C_n^0 \cdot p^0 \cdot q^n$	$C_n^1 \cdot p^1 \cdot q^{n-1}$	$C_n^2 \cdot p^2 \cdot q^{n-2}$	…	$C_n^k \cdot p^k \cdot q^{n-k}$	…	$C_n^n \cdot p^n \cdot q^0$

3. ***Закон распределения Пуассона***. Случайная величина X принимает бесконечное счетное число значений: 0, 1, 2, 3, 4, 5,…, k ,… , с вероятностью, определяющейся по формуле Пуассона:

$$P(X=k)=\frac{a^k}{k!}e^{-a}, \qquad (2.2.1)$$

где a>0 - параметр распределения Пуассона.

x_i	0	1	2	…	k	…
p_i	$e^{-\alpha}$	e^{-a}	$\frac{a^2}{2!}\cdot e^{-a}$	…	$\frac{a^k}{k!}\cdot e^{-a}$	…

При $n\to\infty$ и $p\to 0$ биномиальный закон приближается к закону распределения Пуассона, где $a=np$.

4. ***Геометрический закон распределения***. Пусть $P(A)=p$- вероятность наступления события A в каждом опыте, соответственно, $q=1-p$ - вероятность не наступления события A (схема Бернулли).
Вероятность появления m-неудач до первого наступления события А определяется по формуле:

$$P(X=m)=p\cdot q^m. \tag{2.2.2}$$

Случайная величина X - распределенная по геометрическому закону принимает значения: 0, 1,2,…,m,… , с вероятностью, определяемой по формуле (2.2.2):

x_i	0	1	2	...	m	…
p_i	p	pq	pq^2	...	pq^m	…

5. ***Геометрический закон распределения*** **сдвинутый на единицу.**
Вероятность наступления события A в m-ом опыте определяется по формуле:

$$P(X=m)=p\cdot q^{m-1}. \tag{2.2.3}$$

Случайная величина X - распределенная по геометрическому закону, сдвинутому на 1 (геометрический закон +1), означает число опытов до первого появления события А и принимает значения: 1, 2,…,m,… , с вероятностью, определяемой по формуле (2.2.3):

x_i	1	2	3	…	m	…
p_i	p	pq	pq^2	...	pq^{m-1}	…

Пример 2.2.1. Из орудия производили выстрел по цели до первого попадания. Вероятность попадания в цель 0,6. Найти вероятность того, что попадание произойдет при втором, третьем, k-ом выстреле.

Решение.
$P\ (X=1)$=p=0,6;
$P\ (X=2)=pq$= 0,6 · 0,4 = 0,24;

$P(X=3)=pq^2$=0,6 · 0,4 · 0,4 = 0,096;
…………………………………….
$\mathrm{P}(\mathrm{X}=k)=pq^{k-1}$.

x_i	1	2	3	…	k	…
p_i	0,6	0,24	0,096	…	pq^{k-1}	…

Сумма вероятностей, как и для других законов, равна единице:

$S = \dfrac{p}{1-q} = \dfrac{0{,}6}{1-0{,}4} = \dfrac{0{,}6}{0{,}6} = 1$ - согласно формуле суммы бесконечной геометрической прогрессии, со знаменателем q меньше единицы.

6. **Отрицательное биномиальное распределение**. Если производится ряд независимых опытов в каждом из которых событие А появляется с вероятностью p, до получения k успехов (k = 1,2, 3, …),то при этом вероятность X=m "*неудачных*" опытов можно определить по формуле:

$$P_{m+k} = C_{m+к-1}^{k-1}\, p^k\, q^m \quad (m = 0,1,2,...).$$

Вероятность появления m-неудач, до получения k-успехов совпадает с k-ым членом разложения выражения $q^m(1\text{-}p)^{-m}$ по степеням p, т.е. отрицательного бинома (отсюда и название):

$$1 = q^m(1-p)^{-m} = q^m \sum_{k=1}^{\infty} C_{m+к-1}^{k-1} p^k \quad .$$

Распределение определяется двумя параметрами «k» и «p».

7. ***Гипергеометрический закон распределения***. Пусть в урне N-шаров, из них M белых, а остальные (N – M) черные. Найдем вероятность того, что из извлеченных n шаров m белых и (n-m) черных.

$N_{\text{шаров}} = M_{\text{белых}} + (N\text{-}M)_{\text{черных}}$;

$n_{\text{шаров}} = m_{\text{белых}} + (n–m)_{\text{черных}}$;

C_M^m - число способов выбора m белых шаров из M;

C_{N-M}^{n-m}- число способов выбора (n–m) черных шаров из (N-M).

Всего возможных наборов из m-белых и (n–m) черных, по правилу произведения, равно $C_M^m\ C_{N-M}^{n-m}$; C_N^n - общее число способов выбора из N шаров n.

Отсюда, по формуле классического определения вероятности

$$P(A) = \frac{C_M^m\, C_{N-M}^{n-m}}{C_N^n}. \tag{2.2.4}$$

Ограничения на параметры: $M \le N,\ m \le n;\ m = m_0,\ m_0+1,\ m_0+2, ...,\ \min(M,n)$, где $m_0 = \max\{O,\ n\text{-}(N\text{-}M)\}$. Случайная величина X=m, распределенная по гипергеометрическому закону распределения (при m=0,1,2,3,...,M) имеет вид:

x_i	0	1	2	…	m	...	M
p_i	$\dfrac{C_M^0\, C^n}{C_N^n}$	$\dfrac{C_M^1\, C_{N-M}^{n-1}}{C_N^n}$	$\dfrac{C_M^2\, C_{N-M}^{n-2}}{C_N^n}$	…	$\dfrac{C_M^m\, C_{N-M}^{n-m}}{C_N^n}$	...	$\dfrac{C_M^M\, C_{N-M}^{n-M}}{C_N^n}$

Гипергеометрический закон определяется тремя параметрами *N, M, n*. При $n<0,1N$ этот закон стремится к биномиальному.

Замечание.

1. В теории вероятностей различают две основные схемы: выбора элементов с возвращением каждый раз обратно и выбора без возвращения, которые описываются соответственно биномиальным и гипергеометрическим законами.
2. Геометрический закон описывает схему повторения опытов (в каждом из которых может наступить или не наступить событие A: $P(A)=p$, $q=1-p$), до первого появления события А, то есть фактически это отрицательное биномиальное распределение при $m=1$.
3. Закон распределения Пуассона обычно используют при изучении событий, вероятность которых близка к нулю (маловероятных событий), его иногда называют законом редких событий.

2.3. Числовые характеристики дискретных случайных величин

На практике нет необходимости характеризовать величину полностью. Обычно достаточно указать только отдельные числовые параметры распределения. Такие числовые параметры принято называть числовыми характеристиками распределения. Прежде всего, это характеристики положения ряда распределения: математическое ожидание, медиана, мода; характеристики рассеяния: дисперсия, среднее квадратическое отклонение.

Математическим ожиданием $M(X)$ ДСВ X называется среднее значение случайной величины:

$$M(x)=\frac{x_1p_1+x_2p_2+\ldots+x_np_n}{p_1+p_2+\ldots+p_n}=\frac{\sum_{i=1}^{n}x_ip_i}{\sum_{i=1}^{n}p_i}=\frac{\sum_{i=1}^{n}x_ip_i}{1}=\sum_{i=1}^{n}x_ip_i\,.$$

Или иначе, $M(X)$ - это сумма парных произведений случайной величины на соответствующую вероятность:

$$M(x)=\sum_{i=1}^{n}x_ip_i\,. \qquad (2.3.1)$$

Мода $M_o(X)$ распределения – это значение СВ, имеющее наиболее вероятное значение.

Медиана $M_e(X)$ – это значение случайной величины, которое делит таблицу распределения на две части таким образом, что вероятность попадания в одну из них равна 0,5. Медиана обычно не определяется для ДСВ.

Пример 2.3.1.

x_i	-1	0	1	2
p_i	0,1	0,2	0,1	0,6

$M(X) = 0{,}1\cdot(-1)+0\cdot 0{,}2+1\cdot 0{,}1+2\cdot 0{,}6=1{,}2$; $M_o(X) =2$.

Свойства математического ожидания:

1) $M(C) = C$, где C =const;
2) $M(CX) = CM(X)$;
3) $M(X\pm Y) = M(X) \pm M(Y)$;
4) Если случайные величины X и Y, независимы, то $M(XY) = M(X)\cdot M(Y)$.

Для распределения Бернулли $M(X)=p$;
для биномиального распределения: $M(X)=np$;
для геометрического закона: $M(X)=q/p$;
для геометрического закона +1: $M(X)= 1/p$;
для отрицательного биномиального распределения: $M(X)=(kq)/p$;
для распределения Пуассона: $M(X)=a$;
для гипергеометрического распределения: $M(X)=n(M/N)$.

Дисперсия ДСВ и ее свойства.

Дисперсия служит для характеристики рассеяния СВ относительно ее математического ожидания и характеризует форму кривой распределения.

Она является более полной оценки ДСВ. Пусть заданы СВ X и Y:

x_i	-1	1
p_i	0,5	0,5

y_i	-100	100
p_i	0,5	0,5

$M(X)= -1\cdot 0{,}5+1\cdot 0{,}5=0$, $M(Y)= -100\cdot 0{,}5+100\cdot 0{,}5=0$ – хотя математические ожидания равны, однако, случайные величины X иY явно различны, поэтому для характеристики случайной величины одного математического ожидания недостаточно и необходимо ввести другие характеристики, одна из них дисперсия.

Дисперсией ДСВ X называется математическое ожидание квадрата отклонения СВ от ее математического ожидания:

$$D(X)=M((x-M(X))^2)=(x_1-M(X))^2p_1+(x_2-M(X))^2p_2+....+(x_n-M(X))^2p_n. \quad (2.3.2)$$

Свойства дисперсии:

1) $D(C) = 0$, где C=const;
2) $D(CX)=C^2D(X)$;
3) $D(X)=M(X^2)-(M(X))^2$,

где $M(X^2)=x_1^2p_1+x_2^2p_2+...+x_n^2p_n$;

4) Если СВ X и Y независимы, то:

$D(X\pm Y)= D(X) + D(Y)$;

5) $D(C+X)= D(X)$;
6) Для любых СВ X и Y, $D(X\pm Y)=D(X)+D(Y)\pm 2cov(X,Y)$,

где $cov(X,Y)=M((X-m_x)(Y-m_y))$ - ковариация случайных величин X и Y ($M(X)= m_x$, $M(Y)= m_y$).

Дисперсия характеризует средний квадрат отклонения ДСВ, поэтому на практике часто используют в качестве характеристики разброса *среднее квадратическое отклонение* $\sigma(X)=\sqrt{D(X)}$, которое имеет ту же размерность, что и СВ X.

Например, для рассмотренных выше СВ X и Y:
$D(X)=M(X^2)-(M(X))^2=1-0^2=1$, $\sigma(X)=\sqrt{D(X)}=1$,
так как $M(X^2)=(-1)^2\cdot 0{,}5+(1)^2\cdot 0{,}5=1$;
$D(X)=M(Y^2)-(M(Y))^2=10000-0^2=10000$, $\sigma(Y)=\sqrt{10000}=100$,
так как $M(Y^2)=(-100)^2\cdot 0{,}5+(100)^2\cdot 0{,}5=10000$.

Для распределения Бернулли**:** $D(X)=pq$;

для биноминального закона: $D(X)=npq$, $\sigma(X)=\sqrt{npq}$;

для геометрического закона и для геометрического закона+1: $D(X)=\dfrac{q}{p^2}$;

для отрицательного биномиального распределения: $D(X)=\dfrac{kq}{(p^2)}$;

для гипергеометрического: $D(X)=n\dfrac{M}{N}\left(1-\dfrac{M}{N}\right)\dfrac{N-n}{N-1}$;

для распределения Пуассона: $D(X)=a$.

Только для распределения Пуассона $M(X)=D(X)=a$

2.4. Одинаково распределённые, взаимонезависимые дискретные случайные величины

СВ называют *одинаково распределенными*, если они имеют одинаковые законы распределения. Поэтому у них совпадают числовые характеристики: математическое ожидание, дисперсия, среднее квадратическое отклонение.

Пусть $X_1, X_2, \ldots, X_n$ одинаково распределенные, взаимонезависимые ДСВ, тогда: $M(X_1)=M(X_2)=\ldots=M(X_n)=M(X)$, $D(X_1)=D(X_2)=\ldots=D(X_n)=D(X)$.

Рассмотрим характеристики их средней арифметической $\overline{X}=\dfrac{X_1+X_2+\ldots+X_n}{n}$:

1) $M(\overline{X})=M\left(\dfrac{X_1+X_2+\ldots+X_n}{n}\right)=\dfrac{1}{n}(M(X_1)+M(X_2)+\ldots+M(X_n))=\dfrac{1}{n}nM(X)=M(X)$;

2) $D(\overline{X})=D\left(\dfrac{X_1+X_2+\ldots+X_n}{n}\right)=\dfrac{1}{n^2}(D(X_1)+D(X_2)+\ldots+D(X_n))=\dfrac{1}{n^2}nD(X)=D(X)/n$;

3) $\sigma=(\overline{X})=\sqrt{D(\overline{x})}=\sqrt{\dfrac{D(X)}{n}}=\dfrac{\sigma(X)}{\sqrt{n}}$ - стандартное отклонение СВ X.

Дисперсия относительной частоты $(\dfrac{m}{n})$ появления события A в n независимых испытаниях (в каждом из которых событие A появляется с вероятностью равной p, и не появляется с вероятностью q=1-p; m - число появлений события A в серии из n испытаний), равна

$$D\left(\frac{m}{n}\right)=\frac{pq}{n}.$$

Контрольные задания 2.4.

1. Вероятность работы каждого из четырех комбайнов без поломок в течение определенного времени равна 0,9. Составить закон распределения случайной величины X – числа комбайнов, работавших безотказно. Построить график распределения вероятностей. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.
2. Вероятность рождения в семье мальчика равна 0,515. Составить закон распределения случайной величины X – числа мальчиков в семьях, имеющих четырех детей. Построить график распределения вероятностей. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение.
3. Вероятность того, что покупатель совершит покупку в магазине 0,4. Составить закон распределения случайной величины X – числа покупателей, совершивших покупку, если магазин посетило 3 покупателя. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.
4. В группе из 10 спортсменов 6 мастеров спорта. Отбирают (по схеме без возвращения) 3−х спортсменов. Составить закон распределения случайной величины X – числа мастеров спорта из отобранных спортсменов. Найти математическое ожидание случайной величины X.
5. В группе, состоящей из ($2N$ + 1) студентов, N девушек. Составить закон распределения случайной величины X – числа девушек из случайно отобранных трех студентов (N – номер студента в группе).
6. В партии из (N+5) изделий (N+1) изделие высокого качества. Случайно отбирается 3 изделия. Составить закон распределения случайной величины X – числа изделий высокого качества среди отобранных.
7. Стрелок производит выстрелы по цели до первого попадания. Составить закон распределения случайной величины X – числа выстрелов, сделанных стрелком. Вероятность попадания в цель при каждом выстреле составляет 0,7. Найти наивероятнейшее число выданных стрелку патронов.
8. Покупатель посещает магазины для приобретения нужного товара. Вероятность того, что товар имеется в определенном магазине составляет 0,4. Составить закон распределения случайной величины X – числа магазинов, которые посетит покупатель из четырех возможных. Построить график распределения. Найти наиболее вероятное число магазинов, которые посетит покупатель.
9. Игрок поочередно покупает билеты двух разных лотерей до первого выигрыша. Вероятность выигрыша по одному билету первой лотереи составляет 0,2, а второй 0,3. Игрок вначале покупает билет первой лотереи. Составить закон распределения и найти математическое ожидание случайной величины X – числа купленных билетов, если он имеет возможность купить: а) только 5 билетов; б) неограниченное число билетов.
10. На конноспортивных соревнованиях необходимо преодолеть четыре препятствия с вероятностями, равными соответственно 0,9; 0,8; 0,7; 0,6. При первой неудаче спортсмен в дальнейших состязаниях не участвует. Составить закон распределения случайной величины X – числа взятых препятствий. Найти математическое ожидание случайной величины X.
11. В игре спортивной лотереи угадывается 5 номеров из 36. Игрок получает выигрыш, если угадает 5, 4 или 3 номера. За 5 угаданных номеров выигрыш составляет 10 тыс. руб. Сумма выигрыша по одной карточке за 4 правильно угаданных номера в 10 раз больше, чем за 3. Составить закон распределения случайной величины X – числа правильно угаданных номеров. Определить среднюю величину выигрыша, если известно, что карточек было выпущено 1 млн. шт. Стоимость одной карточки 1 руб. Выигрыши составляют 50 % общей суммы тиража.

12. Вероятность попадания в цель первым стрелком равна 0,9, вторым − 0,8 и третьим − 0,7. Составить закон распределения случайной величины X – числа попаданий в цель, если каждый стрелок производит по одному выстрелу. Определить математическое ожидание случайной величины X.
13. Вероятность успешной сдачи экзамена первым студентом составляет 0,7, а вторым − 0,8. Составить закон распределения случайной величины X – числа студентов, успешно сдавших экзамен, если каждый из них может пересдать один раз экзамен, если он его первый раз не сдал. Найти математическое ожидание случайной величины X.
14. Дискретная случайная величина X задана законом распределения:

X	Значения вероятностей по вариантам											
	1	2	3	4	5	6	7	8	9	10	11	12
1	0,1	0,2	0,05	0,15	0,1	0,2	0,25	0,1	0,4	0,05	0,15	0,12
3	0,2	0,25	0,15	0,2	0,3	0,4	0,3	0,15	0,3	0,1	0,25	0,23
5	0,4	0,3	0,2	0,25	0,3	0,3	0,2	0,25	0,2	0,15	0,25	0,46
7	0,2	0,15	0,4	0,25	0,2	0,05	0,15	0,35	0,08	0,25	0,2	0,14
9	0,1	0,1	0,2	0,15	0,1	0,05	0,1	0,15	0,02	0,45	0,15	0,05

По одному из 12 вариантов построить многоугольник распределения. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.
15. Предприниматель рассматривает возможность покупки акций трех предприятий, по каждой из которых известна доходность, как отношение величины получаемого дохода за период времени к цене акции и вероятности возможных значений доходности. Акции какого предприятия следует считать более доходными, если руководствоваться средним значением (математическим ожиданием) доходности?

Предприятие 1		Предприятие 2		Предприятие 3	
Доходность (%),X	Вероятность, P_x	Доходность (%),Y	Вероятность, P_x	Доходность (%),Z	Вероятность, P_z
5	0,2	3	0,1	1	0,1
7	0,3	7	0,4	6	0,4
9	0,4	10	0,3	10	0,25
11	0,1	15	0,2	20	0,25

Акции какого предприятия являются менее рискованными (считается, что чем выше колеблемость доходности акций, тем больше их рискованность)?
16. Бросают 12 игральных костей. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X – суммы числа очков, которые могут появиться на всех выпавших гранях.
17. Математическое ожидание случайной величины X равно 8. Найти математическое ожидание случайных величин: а) X-4; б) X+6; в) 3X-4; г) 4X+3.
18. Дисперсия случайной величины X равна 8. Найти дисперсию следующих величин: а) X-2; б) X+6; в) 3X-2; г) 2X+7.
19. Найти математическое ожидание и дисперсию случайных величин:
а) $Z=4X-2Y$; б) $Z=2X-4Y$; в) $Z=3X+5Y$; если $M(X)=5$, $M(Y)=3$, $D(X)=4$,
$D(Y)=6$. Случайные величины X и Y независимы.
20. Случайные величины X и Y независимы. Найти математическое ожидание и дисперсию случайных величин: а) $Z=4X+2Y$; б) $Z= 5X-3Y$;
в) $Z=3X-Y$, если $M(X) = 6$, $M(Y)=5$, $D(X)=5$, $D(Y)=4$.

21.Вероятность изготовления бракованной детали автоматом равна 0,002. Найти математическое ожидание и среднее квадратическое отклонение случайной величины X - числа бракованных деталей, если деталей изготовлено 1000. Определить вероятность того, что из 1000 деталей будет изготовлено: а) не более двух бракованных; б) хотя бы одна бракованная.

22.Независимые случайные величины X и Y имеют следующие распределения:

x_i	2	4	6
p_i	0,3	0,5	0,2

y_j	3	4
p_j	0,4	0,6

Составить закон распределения случайных величин: а) $Z=X+Y$; б) $V=XY$. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайных величин Z и V.

23.В бригаде имеется два звена тракторов. В первом звене − 3 трактора, причем вероятность безотказной работы каждого из них в течение смены равна 0,9. Во втором звене − 2 трактора, вероятность безотказной работы первого из них равна 0,8, а второго − 0,7. Составить закон распределения случайной величины X – числа тракторов, работавших безотказно в бригаде. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.

24.Два стрелка производят по одному выстрелу по мишени. Вероятность поражения мишени первым стрелком равна N: (N+5), вторым N: (N+2). Составить закон распределения случайной величины $Z=X+Y$, где X - число поражений мишени первым стрелком, Y – число поражений мишени вторым стрелком. Найти числовые характеристики случайной величины Z.

25.Случайные величины X – площадь посева овощей на хозяйство (га) и Y – урожайность овощей с 1 га (т) имеют следующие распределения:

x_i	1	2	3
p_i	0,1	0,6	0,3

y_j	10	15	20
p_j	0,2	0,5	0,3

Определить средний валовой сбор овощей на хозяйство, дисперсию и среднее квадратическое отклонение валового сбора овощей.

26.Дискретная случайная величина X принимает три возможных значения: $x_1=1$ с вероятностью $p_1=0,2$; $x_3=5$ с вероятностью 0,3 и x_2 с вероятностью p_2. Найти x_2 и p_2, если известно, что $M(X)=3$.

27.Вероятность сдать экзамен студентом на «отлично» равна 0,3, на «хорошо» – 0,4. Определить вероятности получения других оценок (2; 3), если известно, что $M(X)=3,9$.

28. Вероятность выигрыша по лотерейному билету составляет 0,02. Найти $M(X)$ и $\sigma(X)$ числа выигравших билетов, если их было приобретено 100.

29.По одному тиражу лотереи куплено 100 билетов. Среднее квадратическое отклонение числа выигранных билетов равно трем. Найти вероятность выигрыша по одному билету лотереи.

30.Подброшены две игральные кости. Найти $M(X)$, где X - случайная величина – сумма числа очков, которые могут появиться на двух выпавших гранях.

31.Хозяйство продает крупный рогатый скот живым весом x_1 и x_2 ($x_2>x_1$). Вероятность того, что крупный рогатый скот будет продан весом x_1 равна 0,4. Найти закон распределения случайной величины X – веса крупного рогатого скота, если математическое ожидание составило 4,60 ц, а дисперсия 0,24.

32.Совокупность семей имеет следующее распределение по числу детей:

x_i	x_1	x_2	2	3
p_i	0,1	p_2	0,4	0,35

Определить x_1, x_2, p_2, если известно, что $M(X)=2$, $D(X)=0,9$.

33.Дискретная случайная величина X задана законом распределения:

x_i	1	x_2	x_3	8
p_i	0,1	p_2	0,5	0,1

Найти x_2, x_3, p_2, если известно, что $M(X)=4$, $M(X^2)=20,2$.

34.Совокупность студентов имеет следующее распределение по результатам сдачи сессии:

x_i	2	3	4	5
p_i	0,1	p_2	p_3	p_4

Найти вероятности получения удовлетворительных, хороших и отличных оценок, если известно, что математическое ожидание (среднее значение) результатов сдачи экзаменов составило 3,7, а среднее квадратическое отклонение 0,9.

35.По данным задачи 14 определить модальное и медианное значения случайной величины X.

36. Для того чтобы проверить точность своих финансовых счетов, компания регулярно пользуется услугами аудиторов для проверки в бухгалтерских проводках счетов. Предположим, что служащие компании при обработке 100 входящих счетов допускают примерно 5 ошибок. Аудитор случайно отбирает 3 входящих документа.

а) Найти закон распределения случайной величины X – числа ошибок, выявленных аудитором.

б) Построить функцию распределения и её график (вероятностную гистограмму).

в) Определить вероятность того, что аудитор обнаружит более чем одну ошибку.

37. Фирма предлагает в продажу со склада партию из 10 компьютеров, 4 из которых – с дефектами. Покупатель приобретает 5 из них, не зная о возможных дефектах. Чему равна вероятность что, что все 5 компьютеров окажутся без дефектов? Ремонт одной дефектной машины будет стоить 50$. Найдите математическое ожидание общей средней стоимости ремонта и его дисперсию.

2.5. Интегральная функция распределения и ее свойства

Для непрерывной случайной величины X вероятность $P(X = x_i) \to 0$, поэтому для НСВ удобнее использовать вероятность того, что СВ $X < x$, где $x = x_i$ - текущее значение переменной. Эта вероятность $P(X < x) = F(x)$ называется *интегральной функцией* распределения. Интегральная функция является универсальным способом задания СВ (как для ДСВ, так и для НСВ).

Свойства интегральной функции распределения:

1) $F(x)$ не убывает (если $x_2 > x_1$, то $F(x_1) \geq F(x_2)$;

2) $F(-\infty) = 0$;

3) $F(+\infty) = 1$;

4) вероятность попадания СВ X в интервал $a<X<b$: $P(a \leq X < b) = F(b) - F(a)$.

Замечание. Обычно для определённости левую границу включают в интервал, а правую нет. Вообще для НСВ верно: $P(a<X<b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$.

2.6. Дифференциальная функция распределения и ее свойства

СВ X непрерывна, если ее интегральная функция непрерывна на всей числовой оси. СВ X непрерывна и имеет дифференциальную функцию, если ее интегральная функция непрерывна и дифференцируема всюду, за исключением конечного числа точек на любом конечном промежутке.

Дифференциальной функцией (функцией плотности вероятности) СВ X называется производная ее функции распределения:

$$f(x) = F'(x). \tag{2.6.1}$$

С помощью дифференциальной функции можно получить формулу вероятности попадания СВ X в заданный интервал:

$$P(a \le X < b) = P(a \le X < b) = \int_a^b f(x)dx = F(b) - F(a). \qquad (2.6.2)$$

Свойства дифференциальной функции:

1) $f(x) \ge 0$;

2) $\int_{-\infty}^{+\infty} f(x)dx = 1$;

3) $F(x) = \int_{-\infty}^{x} f(t)dt$.

2.7. Числовые характеристики непрерывных случайных величин

1) Математическое ожидание НСВ X определяется по формуле:

$$M(X) = \int_{-\infty}^{+\infty} xf(x)dx. \qquad (2.7.1)$$

Если НСВ X определена на интервале ($a; b$), то:

$$M(X) = \int_a^b xf(x)dx. \qquad (2.7.2)$$

2) Мода НСВ X будет определяться как максимум ее дифференциальной функции:

$$M_0(X) = \max_{(-\infty;+\infty)} f(x). \qquad (2.7.3)$$

3) Медиана определяется как значение случайной величины, которое делит площадь под дифференциальной функцией на две равные части.

$$M_e(X) = P(x < M_e(X)) = P(x > M_e(X)) = \frac{1}{2}. \qquad (2.7.4)$$

4) Дисперсия НСВ:

$$D(X) = \int_{-\infty}^{+\infty} (x - M(X))^2 f(x)dx = \int_{-\infty}^{+\infty} x^2 f(x)dx - (M(X))^2. \qquad (2.7.5)$$

Все свойства дисперсии и математического ожидания, установленные для ДСВ, сохраняются для НСВ.

Замечание: Если распределение симметрично, то его мода, медиана и математическое ожидание совпадают.

5) Моменты случайных величин.

Кроме характеристик положения и рассеяния существует ряд других числовых характеристик распределения, например, моменты.

Начальным моментом порядка s называется математическое ожидание степени *s* СВ *X*:

$$\alpha_s = M(X^s). \quad (2.7.6)$$

Для ДСВ: $\alpha_s = \sum_{i=1}^{n} x_i^s p_i = x_1^s p_1 + x_1^s p_1 + \ldots + x_n^s p_n$.

Для НСВ: $\alpha_s = \int_{-\infty}^{+\infty} x^s f(x)dx$.

При s=1: $\alpha_1 = M(X) = m_x$, то есть, первый начальный момент – это математическое ожидание СВ.

Отклонение СВ от ее математического ожидания называется *центрированной* СВ $\dot{X}$: $\dot{X} = X - m_x$.

Центральным моментом порядка s СВ X называется математическое ожидание степени s, соответствующей центрированной СВ:

$$\mu_S = \mu(\dot{X}^S) = M((x - m_x)^s). \quad (2.7.7)$$

Для ДСВ: $\mu_s = \sum_{i=1}^{n} (x_i - m_x)^s p_i = (x_1 - m_x)^s p_1 + \ldots + (x_n - m_x)^s p_n$.

Для НСВ: $\mu_S = \int_{-\infty}^{+\infty} (x - m_X)^S f(x)dx$.

При вычислении центральных моментов пользуются формулами связи между центральными и начальными моментами:

$$\begin{aligned}
&\mu_1 = 0, \\
&\mu_2 = \alpha_2 - m_x^2, \\
&\mu_3 = \alpha_3 - 3m_x\alpha_2 + 2m_x^3, \\
&\mu_4 = \alpha_4 - 4m_x\alpha_3 + 6m_x^2\alpha_2 - 3m_x^4.
\end{aligned} \quad (2.7.8)$$

Обычно рассматривают первые четыре центральных момента:

1) $\mu_1 = M(x - m_x) = 0$- математическое ожидание центрированной СВ равно нулю;

2) $\mu_2 = M(x - m_x)^2 = D(X)$ -второй центральный момент это дисперсия;

3) $\mu_3 = M(x - m_x)^3$-третий центральный момент может служить для характеристики асимметрии (скошенности распределения), обычно рассматривают безразмерный коэффициент асимметрии:

$$Sk = \frac{\mu_3}{\sigma^3}. \quad (2.7.9)$$

4) Четвёртый центральный момент $\mu_4 = M(x - m_x)^4$, может служить для характеристики «крутости» или островершинности распределения, описывающейся с помощью эксцесса:

$$E_x = \frac{\mu_4}{\sigma^4} - 3. \qquad (2.7.10)$$

Основным моментом порядка s называется нормированный центральный момент порядка S:

$$r_s = \frac{\mu_s}{\sigma^s}, \qquad (2.7.11)$$

то есть $Sk=r_3$, $Ex=r_4$-3.

Замечание. 1). Sk=0 - распределение симметрично $M_o(X)=M_e(X)=M(X)$,
Sk>0 - распределение имеет положительную асимметрию $M_o(X)<M(X)$,
Sk<0 - распределение имеет отрицательную асимметрию $M_o(X)>M(X)$.
2). Распределение имеет вершину:
при Ex=0 - типа φ(x) (рис.12);
при Ex>0 - более заостренную, чем φ(x);
при Ex<0 - более плоскую, чем φ(x).
3). Фактически начальные и центральные моменты служат, для вычисления основных моментов, представляющих вполне определённые численные характеристики различных свойств случайных величин.

Пример 2.7.1. Непрерывная случайная величина X задана интегральной функцией:

$$F(x) = \begin{cases} 0 \quad при \quad x \le 0, \\ \dfrac{x^3}{125} \quad при \quad 0 < x \le 5, \\ 1 \quad при \quad x > 5. \end{cases}$$

Определить: а) вероятность попадания случайной величины в интервал (2;3); б) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X; в) функции распределения изобразить графически.
Решение. По четвертому свойству интегральной функции:

$$P(2 \le X < 3) = F(3) - F(2) = \left.\frac{x^3}{125}\right|_{x=3} - \left.\frac{x^3}{125}\right|_{x=2} = \frac{27}{125} - \frac{8}{125} = \frac{19}{125} = 0{,}152.$$

Найдем функцию плотности вероятности (дифференциальную функцию) (2.6.1):

$$f(x) = F'(x) = \begin{cases} 0 \quad при \quad x \le 0\,, \\ \dfrac{3x^2}{125} \quad при \quad 0 < x \le 5\,, \\ 0 \quad при \quad x > 5\,. \end{cases}$$

Вероятность попадания случайной величины в интервал (2, 3) также можно найти, зная функцию плотности вероятности (2.6.2):

$$P(2 < X < 3) = \int_2^3 \frac{3x^2}{125}dx = \frac{3}{125}\left.\frac{x^3}{3}\right|_2^3 = \frac{27}{125} - \frac{8}{125} = \frac{19}{125} = 0{,}152\,.$$

Найдем числовые характеристики непрерывной случайной величины *X*. Следует обратить внимание, что случайная величина задана на интервале (0;5).

$$M(X)=\int_0^5 x\cdot\frac{3}{125}x^2dx=\frac{3}{125}\int_0^5 x^3dx=\frac{3}{125}\cdot\frac{x^4}{4}\bigg|_0^5=\frac{3}{5^3}\cdot\frac{5^4}{4}=\frac{15}{4}=3\frac{3}{4}=3{,}75\ .$$

$$D(X)=\int_0^5 x^2\cdot\frac{3x^2}{125}dx-3{,}75^2=\frac{3}{125}\int_0^5 x^4\,dx-14{,}0625=\frac{3}{125}\cdot\frac{x^5}{5}\bigg|_0^5-14{,}0625=15-14{,}0625=0{,}9375\ .$$

$\sigma(X)=\sqrt{D(x)}=\sqrt{0{,}9375}=0{,}9682$. Построим графики функций *F(x)* и *f(x)*.

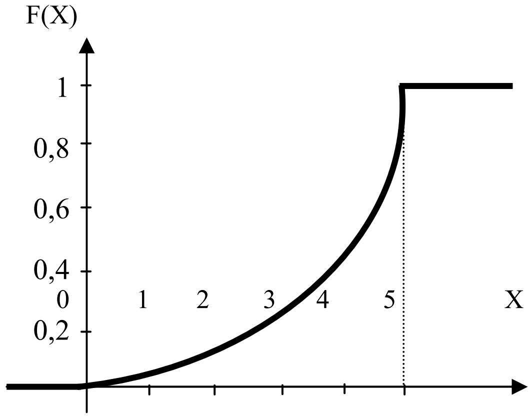


Рис.15 – Интегральная функция случайной величины *X*

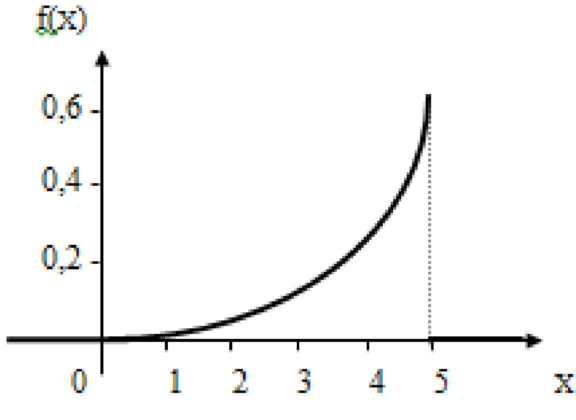


Рис.16 - Дифференциальная функция случайной величины *X*

Контрольные задания 2.7.

1. Даны законы распределения дискретной случайной величины:

а)

X	1	4	6	8
p	0,1	0,3	0,4	0,2

б)

X	-2	5	7	9
p	0,4	0,3	0,2	0,1

Найти интегральную функцию случайной величины *X* и построить ее график.

2. По данным задачи 5, контрольных заданий 2.4 составить интегральную функцию случайной величины X и построить ее график.

3. По данным задачи 6, контрольных заданий 2.4 составить интегральную функцию случайной величины *X* и начертить ее график

4. По одному варианту задачи 14, контрольных заданий 2.4 составить интегральную функцию случайной величины *X* и начертить ее график.

5. Найти интегральную функцию распределения случайной величины *X* – числа попаданий в цель, если произведено три выстрела с вероятностью попадания в цель при каждом выстреле 0,8.

6. Вероятность сдачи первого экзамена студентом составляет 0,7, второго 0,6 и третьего 0,8. Найти интегральную функцию случайной величины *X* – числа экзаменов, сданных студентом. Определить $M(X)$.

7. Случайная величина *X* задана интегральной функцией:

$$F(x)=\begin{cases}0 & \text{при } x\le -2,\\ \dfrac{x}{4}+\dfrac{1}{2} & \text{при } -2<x\le 2,\\ 1 & \text{при } x>2.\end{cases}$$

Найти вероятность того, что в результате испытания случайная величина *X* примет значение: а) меньше 0; б) меньше 1; в) не меньше 1; г) заключенное в интервале (0;2).

8. Дана интегральная функция случайной величины *X*:

$$F(x)=\begin{cases}0 & \text{при } x\le 0,\\ \dfrac{x^6}{4} & \text{при } 0<\text{x}\le\sqrt[3]{2},\\ 1 & \text{при } x>\sqrt[3]{2}.\end{cases}$$

Найти вероятность того, что в результате шести испытаний случайная величина *X* два раза примет значение, принадлежащее интервалу (0;1).

9. Случайная величина задана интегральной функцией:

$$F(x)=\begin{cases}0 & \textit{при } x\le 1,\\ \dfrac{x^2}{8}-\dfrac{1}{8} & \textit{при } 1<x\le 3,\\ 1 & \textit{при } x>3.\end{cases}$$

Найти: а) дифференциальную функцию случайной величины *X*; б) математическое ожидание, дисперсию, среднее квадратическое отклонение случайной величины *X*; в) вероятность попадания случайной величины в интервал (1;2).

10. Дана функция распределения случайной величины *X*:

$$F(x)=\begin{cases}0 & \text{при } x\le -2a),\\ \dfrac{x}{4}+\dfrac{a}{2} & \text{при } -2a<x\le(4-2a),\\ 1 & \text{при } x>(4-2a).\end{cases}$$

а) Определить вероятность попадания случайной величины в интервал
(*-a;a*). б) Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины *X*.

11. Случайная величина *X* задана интегральной функцией:

$$F(x)=\begin{cases}0 & \textit{при } x\le A,\\ \dfrac{x^3}{8} & \textit{при } A<x\le B,\\ 1 & \textit{при } x>B.\end{cases}$$

Найти значения A и B, математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.

12. Случайная величина X задана интегральной функцией:

$$F(x)=\begin{cases}0 & \text{при } x\le 1,\\ \dfrac{x^2}{2}-\dfrac{x}{2} & \text{при } 1<x\le 2,\\ 1 & \text{при } x>2.\end{cases}$$

Найти: а) дифференциальную функцию случайной величины X; б) вероятность того, что в результате четырех независимых испытаний случайная величина X хотя бы один раз примет значение, принадлежащее интервалу $(1;1{,}5)$; в) начертить графики функций.

13. Случайная величина X задана интегральной функцией:

$$F(x)=\begin{cases}0 & \text{\textit{при} } x\le 2,\\ \dfrac{x^3-8}{19} & \text{\textit{при} } 2<x\le 3,\\ 1 & \text{\textit{при} } x>3.\end{cases}$$

Найти: а) дифференциальную функцию; б) вероятность попадания случайной величины в интервал (2,5; 3); в) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X; г) моду и медиану величины X. Построить графики функций.

14. Случайная величина X задана дифференциальной функцией:

$$f(x)=\begin{cases}0 & \text{\textit{при} } x\le 0,\\ \dfrac{4a-2x}{3a^2} & \text{\textit{при} } 0<x\le a,\\ 0 & \text{\textit{при} } x>a.\end{cases}$$

Найти: а) интегральную функцию; б) вероятность попадания случайной величины в интервал $(\frac{a}{6};\frac{a}{3})$.

15. Случайная величина X задана дифференциальной функцией:

$$f(x)=\begin{cases}0 & \text{при } x\le 0,\\ x^3+x & \text{при } 0<x\le\sqrt{\sqrt{5}-1},\\ 0 & \text{при } x>\sqrt{\sqrt{5}-1}.\end{cases}$$

Определить: а) интегральную функцию случайной величины X; б) вероятность попадания случайной величины в интервал (1;1,1). Начертить графики интегральной и дифференциальной функций.

16. Случайная величина X задана дифференциальной функцией:

$$\text{f(x)}=\begin{cases}0 & \text{при } x\le 1,\\ \dfrac{1}{16} & \text{при } 1<x\le 17,\\ 0 & \text{при } x>17.\end{cases}$$

Определить: а) интегральную функцию случайной величины X; б) вероятность попадания случайной величины в интервал (9;12); в) начертить графики интегральной и дифференциальной функции;

17. Случайная величина X задана дифференциальной функцией:

$$f(x)=\begin{cases}0 & \text{при } x\le 0,\\ \dfrac{4}{a^2}x^3 & \text{при } 0<x<\sqrt{a},\\ 0 & \text{при } x>\sqrt{a}.\end{cases}$$

Найти: а) интегральную функцию; б) вероятность того, что в результате испытания случайная величина X примет значение, заключенное в интервале $(\sqrt{0{,}25a};\sqrt{0{,}5a})$; в) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.

18. Случайная величина X задана дифференциальной функцией:

$$f(x)=\begin{cases}0 & \text{при } x<0,\\ 3e^{-3x} & \text{при } x\ge 0.\end{cases}$$

Найти: а) интегральную функцию случайной величины X и начертить её график; б) вероятность попадания случайной величины X в интервал $(-1;\frac{1}{3})$; в) математическое ожидание, дисперсию и среднее квадратическое отклонение.

19. Случайная величина X задана интегральной функцией:

$$F(x)=\begin{cases}0 & \text{при } x\le 1,\\ a(x-1)^2 & \text{при } 1<x\le 3,\\ 1 & \text{при } x>3.\end{cases}$$

Определить: а) значение а; б) математическое ожидание; в) вероятность попадания случайной величины в интервал (1;2); г) построить графики функций F(x) и f(x).

20. Дана интегральная функция:

$$F(x)=\begin{cases}0 & \text{при } x\le -\dfrac{\pi}{2},\\ 0{,}5(1+\sin x) & \text{при } -\dfrac{\pi}{2}<x\le\dfrac{\pi}{2},\\ 1 & \text{при } x>\dfrac{\pi}{2}.\end{cases}$$

Построить графики интегральной и дифференциальной функций.

Найти математическое ожидание и дисперсию случайной величины X

21. Случайная величина задана дифференциальной функцией:

$$f(x)=\begin{cases}0 & \text{при } x\le 1,\\ \dfrac{3x^2-2x}{c} & \text{при } 1<x\le 4,\\ 0 & \text{при } x>4.\end{cases}$$

Найти: а) постоянную c; б) математическое ожидание, дисперсию и среднее квадратическое отклонение.

22. Случайная величина X задана дифференциальной функцией $f(x)=\dfrac{2c}{e^x+e^{-x}}$,

при $-\infty<x<+\infty$. Найти постоянную c.

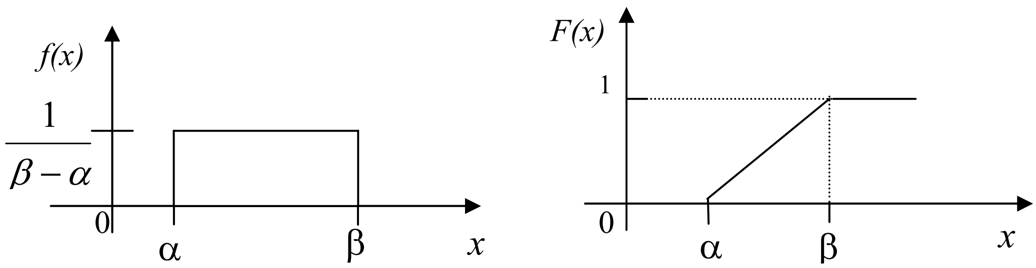
2.8. Основные законы распределения непрерывных случайных величин

Равномерный закон распределения. СВ X распределена по *равномерному (прямоугольному) закону*, если все значения СВ лежат внутри некоторого интервала и все они равновероятны (точнее обладают одной плотностью вероятности).Например, если весы имеют точность 1г и полученное значение округляется до ближайшего целого числа k, то точный вес можно считать равномерно распределенной СВ на интервале (k-0,5; k+0,5).
Дифференциальная функция равномерного закона на интервале (α, β) (рис.17):

$$f(x)=\begin{cases}0 \text{ при } x\le\alpha,\\ \dfrac{1}{\beta-\alpha} \text{ при } \alpha< x\le\beta,\\ 0 \text{ при } x>\beta.\end{cases} \qquad (2.8.1)$$

Интегральная функция равномерного закона на интервале (α, β) (рис.17):

$$F(x)=\begin{cases}0 \text{ при } x\le\alpha,\\ \dfrac{x-\alpha}{\beta-\alpha} \text{ при } \alpha< x\le\beta,\\ 1 \text{ при } x>\beta.\end{cases} \qquad (2.8.2)$$



Дифференциальная функция Интегральная функция
Рис.17 - Равномерный закон распределения

Основные числовые характеристики равномерного закона:
1. Математическое ожидание:

$$M(X)=\int_{\alpha}^{\beta}xf(x)dx=\int_{\alpha}^{\beta}x\frac{1}{\beta-\alpha}dx=\frac{1}{\beta-\alpha}\frac{x^2}{2}\Big|_{\alpha}^{\beta}=\frac{\beta+\alpha}{2}.$$

Математическое ожидание совпадает, в силу симметрии распределения, с медианой.

2. Моды равномерное распределение не имеет.

3. Дисперсия: $D(X)=\int\limits_{\alpha}^{\beta} x^2 f(x)dx-(M(X))^2=\int\limits_{\alpha}^{\beta} x^2\,\frac{1}{\beta-\alpha}\,dx-\left(\frac{\beta+\alpha}{2}\right)^2=$

$=\frac{1}{\beta-\alpha}\,\frac{x^3}{3}\,\Big|_{\alpha}^{\beta}-\left(\frac{\beta+\alpha}{2}\right)^2=\frac{1}{\beta-\alpha}\,\frac{1}{3}\,\left(\beta^3-\alpha^3\right)\text{-}\left(\frac{\beta+\alpha}{4}\right)^2=\frac{1}{3}\,\left(\beta^2+\alpha\beta-\alpha^2\right)\text{-}$

$\text{-}\left(\frac{\beta^2+2\alpha\beta+\alpha^2}{4}\right)=\left(\frac{\beta^2-2\alpha\beta+\alpha^2}{12}\right)=\frac{(\beta-\alpha)^2}{12}.$

Отсюда, среднее квадратическое отклонение $\sigma(X)=\sqrt{D(X)}=\frac{\beta-\alpha}{2\sqrt{3}}$.

4. Третий центральный момент: $\mu_3=$ $M\left(\left(x-\frac{\beta+\alpha}{2}\right)^3\right)=0\Rightarrow Sk=0$, поэтому распределение симметрично относительно $M(X)$.

5. Четвёртый центральный момент:

$$\mu_4=\frac{1}{\beta-\alpha}\int\limits_{\alpha}^{\beta}\left(x-\frac{\beta+\alpha}{2}\right)^4 dx=\frac{(\beta-\alpha)^4}{80}\Rightarrow E_x=\frac{\mu_4}{\sigma_4}\text{-}3=\text{-}1{,}2.$$

6. Вероятность попадания СВ в заданный интервал ($a;b$).
Пусть СВ X распределена по равномерному закону, тогда (рис.18):

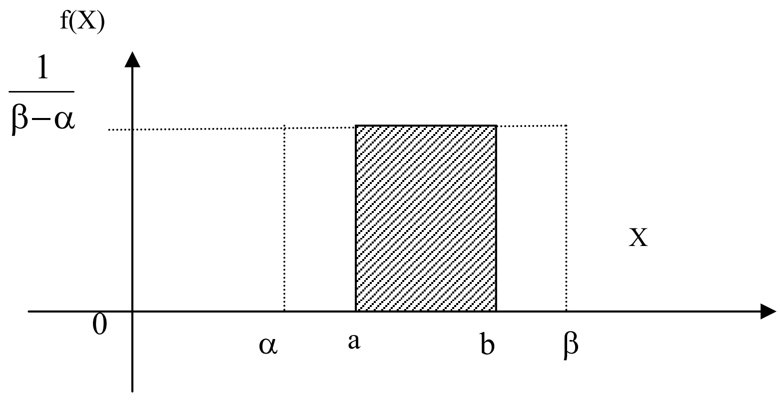


Рис.18 - Вероятность попадания равномерно распределенной СВ X в ($a; b$)

$$P(a\le X<b)=\int\limits_a^b f(x)dx=\int\limits_a^b\frac{1}{\beta-\alpha}dx=\frac{1}{\beta-\alpha}\int\limits_a^b 1dx=\frac{1}{\beta-\alpha}\,x\Big|_a^b=\frac{b-a}{\beta-\alpha}.$$

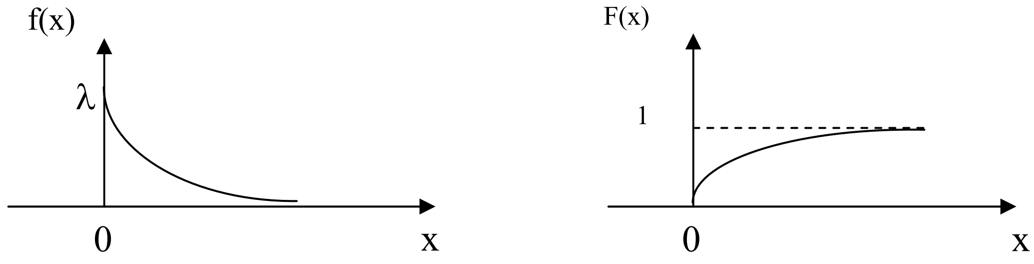
Показательное распределение. НСВ X, принимающая неотрицательные значения, имеет показательное распределение, если ее дифференциальная функция имеет вид (рис.13):

$$f(x)==\begin{cases}\lambda\cdot e^{-\lambda x}, x\ge 0,\\ 0, x<0.\end{cases}\tag{2.8.3}$$

где λ =const, λ >0.

Интегральная функция показательного закона (рис.19):

$$F(x)=\begin{cases}1-e^{-\lambda x}, x\geq 0,\\ 0, x<0.\end{cases} \qquad (2.8.4)$$



Дифференциальная функция Интегральная функция

Рис. 19 - Показательный закон

Числовые характеристики показательного закона:

1. Математическое ожидание: $M(X)=\frac{1}{\lambda}$;

2. Дисперсия: $D(X)=\frac{1}{\lambda^2}$,

среднее квадратическое отклонение: $\sigma(x)=\sqrt{D}=\frac{1}{\lambda}$.

3. Вероятность попадания СВ X в заданный интервал: $P(a\leq x<b)=e^{-\lambda a}-e^{-\lambda b}$.

Замечание. Показательное распределение играет большую роль в теории массового обслуживания (ТМО), теории надежности. В ТМО λ - среднее число событий приходящихся на единицу времени. При определенных условиях, число событий, произошедших за промежуток времени τ, распределено по закону Пуассона с математическим ожиданием $a = \lambda\tau$. Длина промежутка t, между произвольными двумя соседними событиями, подчиняется показательному закону: $P(T<t)=F(t)=1-e^{\lambda t}$ [1].

Нормальный закон распределения (рис. 20) играет исключительную роль в теории вероятностей. Это наиболее часто встречающийся закон распределения, главной особенностью которого является то, что он является предельным законом, к которому, при определённых условиях, приближаются другие законы распределения.

Дифференциальная функция нормального закона имеет вид (рис.20):

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\cdot e^{-\frac{(x-a)^2}{2\sigma^2}}. \qquad (2.8.5)$$

Числовые характеристики нормального закона:

1. Математическое ожидание характеризует центр распределения:

$$M(x)=\int_{-\infty}^{\infty} x\cdot f(x)dx=\int_{-\infty}^{\infty}\frac{x}{\sigma\sqrt{2\pi}}\cdot\exp\left(-\frac{(x-a)^2}{2\sigma^2}\right)dx=a, \text{ где } e^x=\exp(x).$$

2. Дисперсия характеризует форму распределения:

$$D(x)=\int_{-\infty}^{\infty} x^2\cdot f(x)dx-(M(x))^2=\sigma^2.$$

Свойства дифференциальной функции нормального закона:

1. Область определения: $D_f = R$;
2. Ось $0X$ – горизонтальная асимптота;
3. $x = a \pm \sigma$ - две точки перегиба;
4. Максимум в точке с координатами: $(a;\ \frac{1}{\sigma\sqrt{2\pi}})$;
5. График симметричен относительно прямой $x=a$;
6. Моменты:

$\mu_1= \mu_3 =\ldots= \mu_{2к+1}= \ldots= 0$,

$\mu_2=\sigma^2$, $\mu_4= 3\sigma^4$,

$Sk=\frac{\mu^3}{\sigma^3}=0;\ Ex=\frac{\mu_4}{\sigma^4}-3=0.$

7. Вероятность попадания нормально распределенной случайной величины в заданный интервал определяется, по свойству интегральной функции:

$P(\alpha < x < \beta) = \Phi^*\left(\frac{\beta-a}{\sigma}\right)-\Phi^*\left(\frac{\alpha-a}{\sigma}\right)=\Phi\left(\frac{\beta-a}{\sigma}\right)-\Phi\left(\frac{\alpha-a}{\sigma}\right)$, где

$$\Phi^*(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}\exp\left(-\frac{(t-a)^2}{2\sigma^2}\right)dt \quad - \qquad (2.8.6)$$

интегральная функция нормального закона (рис.20);

$\Phi(x)$ - функция Лапласа (1.12.5).

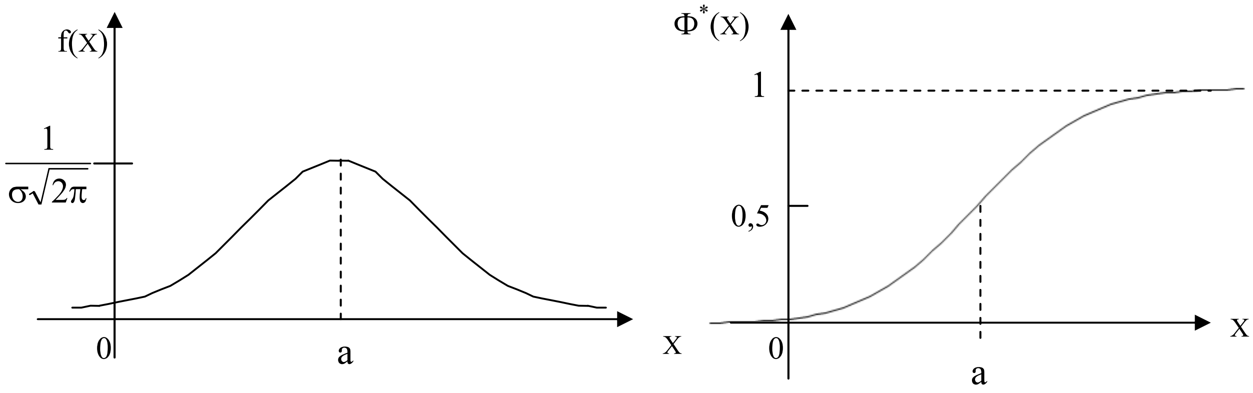


Рис. 20 - Нормальный закон распределения

Свойства интегральной функции нормального закона:

1. $\Phi^*(-\infty)=0$;
2. $\Phi^*(+\infty)=1$;

3. $Ф^*(x)=\frac{1}{2}+Ф(x)$;

4. $Ф^*(-x)=1-Ф^*(x)$.

Вероятность заданного отклонения. Правило трех сигм. Найдем вероятность того, что случайная величина X, распределённая по нормальному закону, отклонится от математического ожидания $M(X)=a$ не более чем на величину $\varepsilon>0$.

$$P(|x-a|<\varepsilon)=P(-\varepsilon<x-a<+\varepsilon)=P(a-\varepsilon<x<a+\varepsilon)=$$

$$=Ф^*\left(\frac{a+\varepsilon-a}{\sigma}\right)-Ф^*\left(\frac{a-\varepsilon-a}{\sigma}\right)=Ф^*\left(\frac{\varepsilon}{\sigma}\right)-(1-Ф^*\left(\frac{\varepsilon}{\sigma}\right))=2Ф^*\left(\frac{\varepsilon}{\sigma}\right)-1$$

или используя функцию Лапласа:

$$P(|X-a|<\varepsilon)=2Ф\left(\frac{\varepsilon}{\sigma}\right).$$

Найдём вероятность того, что нормально распределённая случайная величина X отклонится от $M(X)$ = а на σ, 2σ, 3σ:

$$P(|x-a|<\sigma)=2Ф\left(\frac{\sigma}{\sigma}\right)=2\,Ф(1)=2\cdot0{,}3413=0{,}6826,$$

$$P(|x-a|<2\sigma)=2\,Ф\left(\frac{2\sigma}{\sigma}\right)=2\,Ф(2)=2\cdot0{,}4772=0{,}9544,$$

$$P(|x-a|<3\sigma)=2\,Ф\left(\frac{3\sigma}{\sigma}\right)=2\,Ф(3)=2\cdot0{,}49865=0{,}9973.$$

Отсюда следует *правило* 3σ: если случайная величина X имеет нормальное распределение, то отклонение этой случайной величины от ее математического ожидания по абсолютной величине не превышает утроенное среднее квадратическое отклонение (3σ).

Замечание. Если СВ X распределена нормально с $M(X)=a$ и $D(X)=\sigma^2$, то это обозначают: $X\in N(a,\ \sigma^2)$.

Контрольные задания 2.8.

1. Случайная величина X равномерно распределена в интервале (-2; N). Найти: а) дифференциальную функцию случайной величины X; б) интегральную функцию; в) вероятность попадания случайной величины в интервал $(-1;\frac{N}{2})$; г) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.
2. Найти математическое ожидание и дисперсию случайной величины, равномерно распределенной в интервале: а) (5; 11); б) (-3; 5). Начертить графики этих функций.
3. Равномерно распределенная случайная величина X задана плотностью распределения $f(x)=$ 0,125 в интервале (a-4; a+4), вне него $f(x)=0$. Найти $M(X)$, $D(X)$, $\sigma(X)$.
4. Случайная величина X распределена по закону прямоугольного треугольника (рис. 21) в интервале (0; a). Найти: а) дифференциальную функцию случайной величины X; б) интегральную функцию; в) вероятность попадания случайной величины в интервал (0,25a; 0,5a); г) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.

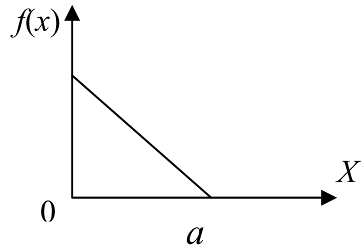


Рис. 21 - Дифференциальная функция закона прямоугольного треугольника

5. Случайная величина X распределена по закону Симпсона («закону равнобедренного треугольника») (рис. 22) на интервале $(-a\,;a)$. Найти: а) дифференциальную функцию распределения вероятностей случайной величины X; б) интегральную функцию и построить ее график; в) вероятность попадания случайной величины в интервал $(-0{,}5a\,;0{,}5a)$; г) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.

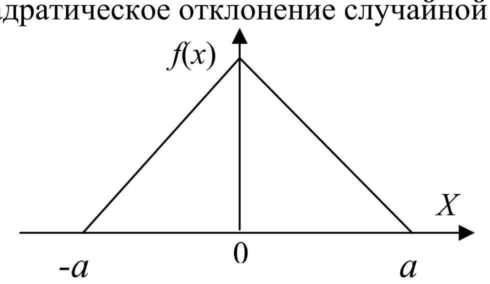


Рис. 22 - Дифференциальная функция распределения Симпсона

6. Для исследования продуктивности определенной породы домашней птицы измеряют диаметр яиц. Наибольший поперечный диаметр яиц представляет собой случайную величину, распределенную по нормальному закону со средним значением 5 см и средним квадратическим отклонением 0,3 см. Найти вероятность того, что: а) диаметр взятого наудачу яйца будет заключен в границах от 4,7 до 6,2 см; б) отклонение диаметра от среднего не превзойдет по абсолютной величине 0,6 см.
7. Вес вылавливаемых в пруду рыб подчиняется нормальному закону распределения со средним квадратическим отклонением 150 г и математическим ожиданием а = 1000 г. Найти вероятность того, что вес пойманной рыбы будет: а) от 900 до 1300 г; б) не более 1500 г; в) не менее 800 г; г) отличаться от среднего веса по модулю не более чем на 200 г; д) начертить график дифференциальной функции случайной величины X.
8. Урожайность озимой пшеницы по совокупности участков распределяется по нормальному закону с параметрами: a = 50 ц/га, σ = 10 ц/га. Определить: а) какой процент участков будет иметь урожайность свыше 40 ц/га; б) процент участков с урожайностью от 45 до 60 ц/га.
9. Выборочным методом измеряется засоренность зерна, случайные ошибки измерения подчинены нормальному закону распределения со средним квадратическим отклонением 0,2 г и математическим ожиданием a = 0. Найти вероятность того, что из четырех независимых измерений ошибка хотя бы одного из них не превзойдет по абсолютной величине 0,3 г.
10. Количество зерна, собранного с каждой делянки опытного поля, есть нормально распределенная случайная величина X, имеющая математическое ожидание m = 60 кг и среднее квадратическое отклонение равное 1,5 кг. Найти симметричный относительно m интервал, в котором с вероятностью 0,9906 будет заключена величина X. Написать дифференциальную функцию этой случайной величины.
11. С вероятностью 0,9973 было установлено, что абсолютное отклонение живого веса случайно взятой особи крупного рогатого скота от среднего веса животного по всему стаду не превосходит

30 кг. Найти среднее квадратическое отклонение живого веса скота, считая, что распределение скота по живому весу подчиняется нормальному закону.

12. Урожайность овощей по участкам является нормально распределенной случайной величиной с математическим ожиданием 300 ц/га и средним квадратическим отклонением 30 ц/га. С вероятностью 0,9545 определить границы, в которых будет находиться средняя урожайность овощей на участках.

13. Нормально распределенная случайная величина *X* задана дифференциальной функцией:

$$f(x) = \frac{1}{4\sqrt{2\pi}} \cdot e^{-\frac{(x-5)^2}{32}}.$$

Определить: а) вероятность попадания случайной величины в интервал (3; 9); б) моду и медиану случайной величины *X*.

14. Нормально распределенная случайная величина *X* задана дифференциальной функцией:

$$f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}}.$$

Найти интервал, в который с вероятностью 0,9545 попадет случайная величина *X* в результате испытаний.

15. Случайная величина распределена по нормальному закону с математическим ожиданием а = 20. Вероятность попадания ее в интервал (20; 30) равна 0,4772. Определить вероятность попадания случайной величины в интервал (10; 25).

16. Написать плотность и функцию распределения показательного закона, если: а) параметр λ=2; б) λ=5; в) λ=0,5.

17. Случайная величина *X* распределена по показательному закону, причем λ=2. Найти вероятность попадания случайной величины *X* в интервал: а) (0; 1); б) (2; 4).

18. Найти *M*(*X*), *D*(*X*), σ(*X*) показательного закона распределения случайной величины X заданной функцией:

$$F(x) = \begin{cases} 0 & при \quad x < 0, \\ 1 - e^{-\lambda x} & при \quad x \geq 0. \end{cases}$$

если: а) λ=0,4; б) λ=3; в) λ=4.

19. Испытываются два независимо работающих элемента. Длительность безотказной работы первого имеет показательнее распределение $F_1(t) = 1 - e^{-0,1t}$, второго $F_2(t) = 1 - e^{-0,05t}$. Найти вероятность того, что за время длительностью 20 часов: а) оба элемента будут работать; б) откажет только один элемент; в) откажет хотя бы один элемент; г) оба элемента откажут.

20. Вероятность того, что оба независимых элемента будут работать в течение 10 суток равна 0,64. Определить функцию надежности для каждого элемента, если функции одинаковы.

21. Среднее число ошибок, которые делает оператор в течение часа работы равно 2. Найти вероятность того, что за 3 часа работы оператор сделает: а) 4 ошибки; б) не менее двух ошибок; в) хотя бы одну ошибку.

22. Среднее число вызовов, поступающих на АТС в одну минуту, равно трем. Найти вероятность того, что за 2 минуты поступит: а) 4 вызова; б) не менее трех вызовов.

23. Случайная величина *X* распределена по закону Коши:

$$f(x) = \frac{1}{\pi(1 + x^2)}.$$

Определить: а) интегральную функцию случайной величины *X*; б) вероятность попадания случайной величины в интервал ($-\sqrt{3}$; $\sqrt{3}$).

24. Случайная величина *X* распределена по закону Релея:

$$F(x)=\begin{cases}0 & при \quad x<0,\\ 1-e^{-\frac{x^2}{2a^2}} & при \quad x\ge 0.\end{cases}$$

Найти: а) дифференциальную функцию случайной величины X; б) вероятность попадания случайной величины в интервал (1; 2), при a=1; в) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X.

25. Функция распределения годовых доходов лиц, облагаемых налогом, имеет вид закона Парето:

$$F(x)=\begin{cases}0 & при \quad x<a,\\ 1-\left(\dfrac{a}{x}\right)^{\beta} & при \quad x\ge a,\ \beta>0.\end{cases}$$

Определить: а) размер годового дохода, который для случайно взятого лица будет превышен с вероятностью 0,8; б) дифференциальную функцию случайной величины X; в) математическое ожидание случайной величины X при $\beta > 1$.

26. Фирма, занимающаяся продажей товаров по каталогу, ежемесячно получает по почте заказы. Число этих заказов – есть нормально распределённая случайная величина со средним квадратическим отклонением $\sigma = 560$ и неизвестным математическим ожиданием a. В 90% случаев число ежемесячных заказов превышает 12 439. Найдите среднее число заказов, получаемых фирмой за месяц.

27. Масса товаров, помещаемых в контейнер определённого размера, - нормально распределённая случайная величина. Известно, что 65% контейнеров имеют чистую массу меньшую, чем 4,2 т. Найдите среднюю и среднее квадратическое отклонение чистой массы контейнера.

28. Служащий рекламного агентства утверждает, что время, в течение которого телезрители помнят содержание коммерческого рекламного ролика, подчиняется экспоненциальному закону с $\lambda = 0{,}25$ дня. Найдите долю зрителей, способных вспомнить рекламу спустя 7 дней.

2.9. Система двух случайных величин

В практических задачах приходится сталкиваться со случаями, когда результат описывается двумя и более случайными величинами, образующими систему случайных величин (случайный вектор) ($x_1, x_2, \dots, x_n$). Например, точка попадания снаряда имеет две координаты: x и y, которые можно принять за систему случайных величин, (x, y) определенных на одном и том же пространстве элементарных событий Ω.

Закон распределения *дискретной двумерной случайной* величины можно представить в виде таблицы, характеризующей собой совокупность всех значений случайных величин и соответствующих вероятностей:

	x_1	x_2	…	x_n	$\sum P(y_j)$
y_1	$P(x_1,y_1)$	$P(x_2,y_2)$	…	$P(x_n,y_1)$	$P(y_1)$
y_2	$P(x_1,y_2)$	$P(x_2,y_2)$	…	$P(x_n,y_2)$	$P(y_2)$
…	…	…	…	…	…
y_m	$P(x_1,y_m)$	$P(x_2,y_m)$	…	$P(x_n,y_m)$	$P(y_m)$
$\sum Px_i$	$P(x_1)$	$P(x_2)$	…	$P(x_n)$	1

В общем случае двумерная случайная величина задается в виде интегральной функции: $F(x, y) = P(X<x,\ Y<y)$, которая означает вероятность попадания двумерной случайной величины в квадрант левее и ниже точки с координатами (x, y).

Свойства интегральной функции:

1. F- не убывает и непрерывна слева по каждому аргументу.
2. $F(-\infty, y) = F(x, -\infty) = F(-\infty, \infty) = 0$
3. $F(+\infty, y) = F_2(y)$ – функция распределения случайной величины Y.
   $F(x, +\infty\ ) = F_1(x)$– функция распределения случайной величины X.
4. $F(+\infty, +\infty) = 1$.

Вероятность попадания двумерной случайной величины в прямоугольник определяется, исходя из определения интегральной функции двумерной случайной величины (рис.23):

$$P\big((x, y) \in D\big) = F(\beta, \delta) - F(\alpha, \delta) + F(\alpha, \gamma). \qquad (2.9.1)$$

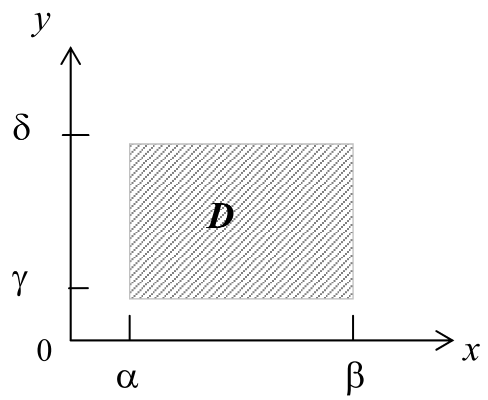


Рис. 23 - Вероятность попадания точки (x, y) в прямоугольник *D*

Случайные величины *X, Y независимы*, если $F(x, y) = F_1(x)\ \ F_2(y)$.
Дифференциальная функция системы двух непрерывных случайных величин определяется как вторая смешанная производная функции распределения:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x \partial y} = F''_{xy}(x, y). \qquad (2.9.2)$$

Свойства дифференциальной функции:

1. $f(x, y) > 0$;
2. $\int\limits_{-\infty}^{+\infty}\int\limits_{-\infty}^{+\infty} f(x, y)\,dx\,dy = 1$;
3. $F(x, y) = \int\limits_{-\infty}^{x}\int\limits_{-\infty}^{y} f(x, y)\,dx\,dy$.

Геометрически свойство 2 означает, что объем тела ограниченного поверхностью $f(x, y)$ и плоскостью XOY равен 1.

Если случайные величины x и y независимы, то

$$f(x, y) = f_1(x)\ f_2(y), \qquad (2.9.3)$$

где $f_1(x) = F_1'(x), f_2(y) = F_2'(y)$, – безусловные законы распределения.

В противном случае:

$$f(x, y) = f_1(x)\ f(y/x) \quad \text{или} \quad f(x,y) = f_2(y)\, f(x/y), \text{ где}$$

$$f(y/x) = \frac{f(x,y)}{f_1(x)} \text{ -} \qquad (2.9.4)$$

условная дифференциальная функция СВ Y при заданном значении $X = x$,

$$f(x/y) = \frac{f(x,y)}{f_2(y)} \text{ -} \qquad (2.9.5)$$

условная дифференциальная функция СВ X при заданном значении Y= y;

$$f_1(x) = \int_{-\infty}^{+\infty} f(x,y)dy \quad и \quad f_2(y) = \int_{-\infty}^{+\infty} f(x,y)dx -$$

дифференциальные функции отдельных величин *X* и *Y*, входящих в систему.

2.10. Числовые характеристики системы двух случайных величин. Корреляционный момент. Коэффициент корреляции.

Начальным моментом порядка s,h системы двух случайных величин *X, Y* называется математическое ожидание произведения степени s случайной величины *X* и степени *h* случайной величины *Y*:

$$\alpha_{s,h} = M(X^s Y^h). \qquad (2.10.1)$$

Центральным моментом порядка s, h системы СВ (*X, Y*) называется математическое ожидание произведения степеней *s, h* соответствующих центрированных случайных величин:

$$\mu_{s,h} = M(\dot{X}^s \dot{Y}^h), \qquad (2.10.2.)$$

где $\dot{X} = X - M(X), \dot{Y} = Y - M(Y)$-центрированные случайные величины *X* и *Y*.

Основным моментом порядка *s, h* системы СВ (*X,Y*) называется нормированный центральный момент порядка *s, h*:

$$r_{s,h} = \frac{\mu_{s,h}}{\sigma_x^s \sigma_y^h}. \qquad (2.10.3)$$

Начальные моменты $\alpha_{1,0}$, $\alpha_{0,1}$:

$\alpha_{1,0} = M(X^1 Y^0) = M(X)$; $\alpha_{0,1} = M(X^0 Y^1) = M(Y)$.

Вторые центральные моменты:

$\mu_{2,0} = M(\dot{X}^2 \dot{Y}^0) = M(x - M(X))^2 = D(X),$ – характеризует рассеяние случайных величин в направлении оси *0X*.

$\mu_{0,2} = M(\dot{X}^0 \dot{Y}^2) = M(y - M(Y))^2 = D(Y),$ – характеризует рассеяние случайных величин в направлении оси *0Y*.

Особую роль в качестве характеристики совместной вариации случайных величин X и Y играет второй смешанный центральный момент, который называется корреляционным моментом (ковариацией):

$$\mu_{1,1}=M(\dot{X}^{1}\dot{Y}^{1})=K(X,Y)=cov(X,\ Y)=M(XY)\ -\ M(X)M(Y). \qquad (2.10.3)$$

Корреляционный момент является мерой связи случайных величин. Если случайные величины X и Y независимы, то математическое ожидание равно произведению их математических ожиданий:

$M\ (XY)=M\ (X)M\ (Y)$, отсюда $cov\ (X,Y)=0$

Если ковариация случайных величин не равна нулю, то говорят, что случайные величины коррелированны. Ковариация может принимать значения на всей числовой оси, поэтому в качестве меры связи используют основной момент порядка s=1, h=1,который называют *коэффициентом корреляции*:

$$r_{xy}=\frac{cov(x,y)}{\sigma(x)\sigma(y)}, \qquad (2.10.4)$$

где $\sigma(X)=\sqrt{D(X)}, \sigma(Y)=\sqrt{D(Y)}$.

Пример 2.10.1. Докажем, что если случайные величины X и Y линейно зависимы, то коэффициент корреляции равен ± 1. Пусть между случайными величинами X и Y имеет место зависимость $Y=AX+B$, где $M\ (X)=a$, $D(X)=\sigma^2$.Тогда имеем:

$$M\ (Y)=M\ (AX+B)=A\ M\ (X)+B=Aa+B,$$

$$D(Y)=D(AX+B)=D(AX)+D(B)=A^2\ D(X)=A^2\sigma^2,$$

следовательно,

$$\sigma(Y)=\sqrt{A^2\sigma^2}=|A|\sigma\ ,$$

$$cov\ (X,Y)=\mathrm{M}\ ((X-a)(Y-A\ a\ -\ B))=M\ ((X-a)(AX+B-A\ a\ -\ B))=AM\ ((X-a)^2)=A\ D(X)=A\sigma^2.$$

Отсюда,

$$r_{xy}=\frac{A\sigma^2}{\sigma(|\mathrm{A}|\sigma)}=\frac{A}{|A|}=\pm 1,$$ что и требовалось доказать.

Таким образом, мы показали, что если между случайными величинами X и Y существует линейная связь, то коэффициент корреляции равен ± 1. Коэффициент корреляции служит мерой *линейной зависимости* между случайными величинами.

Свойства коэффициента корреляции:

1. $-1\le r_{xy}\le 1$.
2. Если $r_{xy}=\pm 1$, то случайные величины линейно зависимы;
3. Если $r_{xy}=0$, то случайные величины не коррелированны, что не означает их независимости вообще.

Замечание. Если случайные величины X и Y подчиняются нормальному закону распределения, то некоррелированность СВ X и Y означает их независимость.

Первые моменты:

а) для дискретных СВ:

$$M(X)=\sum_{i=1}^{n}\sum_{j=1}^{m}x_i p_{ij},$$

$$M(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}y_j p_{ij},$$

$$D(X)=\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - M(X))^2 p_{ij},$$

$$D(Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}(y_j - M(Y))^2 p_{ij},$$

$$K(X,Y)=\sum_{i=1}^{n}\sum_{j=1}^{m}(x_i - M(X))(y_j - M(Y))p_{ij};$$

б) для непрерывных СВ:

$$M(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x\, f(x,y)dxdy,$$

$$M(Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} y\, f(x,y)dxdy,$$

$$D(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - M(X))^2 f(x,y)dxdy,$$

$$D(X)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(y - M(Y))^2 f(x,y)dxdy,$$

$$K(X,Y)=\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - M(X))(y - M(Y))f(x,y)dxdy.$$

Контрольные задания 2.10.

1. Задана двумерная дискретная случайная величина:

Y \ *X*	2	4
0	0,1	0,3
5	0,2	0,15
10	0,15	0,1

Найти законы распределения составляющих случайных величин.

2. Задана двумерная дискретная случайная величина:

Y \ *X*	0	5	20
0	0,15	0,2	0,10
10	0,10	0,3	0,15

Найти математическое ожидание и дисперсию составляющих случайных величин *X* и *Y*.

3. Задана функция распределения двумерной случайной величины:

$$F(x,y)=\begin{cases}1-3^{-x}-3^{-y}+3^{-x-y} & \text{при} \quad x\ge 0,\, y\ge 0,\\ 0 & \text{при} \quad x<0,\, y<0.\end{cases}$$

Найти двумерную плотность вероятности системы (*X*, *Y*).

4. Функция распределения случайной двумерной величины задана в задании 3. Найти вероятность попадания случайной точки (*X*, *Y*) в прямоугольник, ограниченный прямыми $x = 0$, $x = 2$, $y = 1$, $y = 5$.

5. Найти вероятность попадания случайной точки (X, Y) в прямоугольник, ограниченный прямыми $x = 1$, $x = 2$, $y = 3$, $y = 5$, если известна функция распределения:

$$F(x,y)=\begin{cases}1-2^{-x}-2^{-y}+2^{-x-y} & \text{при } x\ge 0, y\ge 0,\\ 0 & \text{при } x<0, y<0.\end{cases}$$

6. Задана функция распределения двумерной случайной величины:

$$F(x,y)=\begin{cases}(1-e^{-4x})(1-e^{-2y}) & \text{при } x\ge 0, y\ge 0,\\ 0 & \text{при } x<0, y<0.\end{cases}$$

Найти двумерную плотность вероятности системы (*X, Y*).

7. Распределение 100 студентов по количеству пропущенных часов занятий и экзаменационной оценке представлено в следующей таблице. Найти безусловные и условные законы распределения случайных величин: количества пропущенных часов (*X*) и экзаменационной оценки (*Y*).

Количество пропущенных часов	Оценка на экзамене			
	2	3	4	5
0	0	5	10	10
4	5	15	20	15
10	10	5	5	0

8. Распределение хозяйств по дозам внесения удобрений и урожайности озимой пшеницы приведено в следующей таблице:

Дозы удобрений на 1 га, ц д.в.	Урожайность, ц с 1 га			
	до 35	30-35	35-40	Свыше 40
до 1	18	a	5	-
1-2	a	15	20	10
свыше 2	-	a	12	20

Найти безусловные и условные законы распределения случайных величин урожайности (*X*) и доз внесения удобрений (*Y*).

9. Система случайных величин (*X,Y*) подчинена закону распределения с плотностью $f(x,y)=a\cdot\sin(x+y)$ в квадрате $0\le x\le\frac{\pi}{2}$; $0\le y\le\frac{\pi}{2}$ и $f(x,y)=0$, вне квадрата.

Определить: а) коэффициент *a*; б) $M(X)$, $M(Y)$; в) $D(X)$, $D(Y)$.

10. Дана дискретная двумерная величина (*X, Y*):

а)

Y \ X	2	4
10	0,15	0,10
15	0,3	0,05
20	0,15	0,25

б)

Y \ X	100	200
0	0,1	0,25
5	0,05	0,2
10	0,1	0,3

Найти: а) условный закон распределения *X* при условии, что *y*=20; б) условный закон распределения *Y* при условии, что *x*=200.

11.Плотность совместного распределения непрерывной двумерной случайной величины $f(x,y)=\cos x\cdot\cos y$ в квадрате $0\le x\le\frac{\pi}{2};\;\; 0\le y\le\frac{\pi}{2};$ и $f(x,y)=0$, вне квадрата. Доказать, что составляющие *X* и *Y* независимы.

12. Непрерывная двумерная случайная величина (*X, Y*) распределена равномерно внутри треугольника с вершинами O (0,0), *A* (0,6) и *B* (6,0). Найти: а) двумерную плотность вероятности системы; б) плотности и условные плотности составляющих системы.

13.Система случайных величин (*X, Y*) распределена равномерно внутри квадрата со стороной *a*, диагонали которого совпадают с осями координат. Найти: а) двумерную плотность вероятности системы; б) плотности и условные плотности составляющих системы.

14.Задана плотность совместного распределения непрерывной двумерной случайной величины (*X, Y*):

$$f(x,y)=\begin{cases}4xye^{-(x^2+y^2)}\textit{при } x>0,\ y>0;\\ 0 \qquad\qquad \textit{при } x\le 0,\ y\le 0.\end{cases}$$

Найти математические ожидания и дисперсии составляющих.

15.Система случайных величин (*X, Y*) равномерно распределена в треугольнике, ограниченном прямыми $x=0$, $y=0$, $x+y=a$ ($a>0$). Определить: а) математические ожидания и дисперсии случайных величин *X* и *Y*, б) корреляционный момент.

16.Заданы плотности распределения независимых составляющих непрерывной случайной величины (*X, Y*):

$$f(x)=\begin{cases}0 \qquad \textit{при } x\le 0,\\ 2e^{-2x}\textit{при } x>0,\end{cases} \qquad\qquad f(y)=\begin{cases}0 \qquad \text{при } \ y\le 0,\\ 5e^{-5y} \ \text{при } \ y>0.\end{cases}$$

Найти: а) плотность совместного распределения системы; б) функцию распределения системы.

2.11. Функции случайных величин

Закон распределения функции случайных величин.

Пусть имеется непрерывная случайная величина *X* с функцией плотности вероятности f(x). Другая случайная величина *Y* связана со случайной величиной X функциональной зависимостью: $Y=\varphi(X)$. Случайная точка (*X, Y*) может находиться только на кривой $y=\varphi(x)$.

Дифференциальная функция случайной величины *Y* определяется при условии, что φ(x) - монотонна на интервале (*a, b*), тогда для функции $\varphi(x)$ существует обратная функция: $\varphi^{-1}=\Psi$, $x=\Psi(y)$.

Обычно числовая прямая разбивается на n промежутков монотонности и обратная функция находится на каждом из них, поэтому:

$$g(y)=\sum_{i=1}^{n} f(\psi_i(y))\cdot\left|\psi_i'(y)\right|, \tag{2.11.1}$$

g(*y*) - дифференциальная функция СВ *Y*.

Замечание.

Математическое ожидание и дисперсию СВ *Y* – функции случайной величины $X(Y=\varphi(X))$, имеющей дифференциальную функцию *f(x)*, можно определить по формулам:

$$M(Y)=M(\varphi(x))=\int_{-\infty}^{\infty}\varphi(x)f(x)dx, \qquad (2.11.2)$$

$$D(Y)=D(\varphi(x))=\int_{-\infty}^{\infty}\varphi^2(x)f(x)dx-M^2(Y). \qquad (2.11.2^1)$$

Пример 2.11.1. Случайная величина X подчиняется нормальному закону распределения с математическим ожиданием m и дисперсией σ^2, то есть дифференциальная функция имеет вид:

$$f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-a)^2}{2\sigma^2}}.$$

Найти дифференциальную функцию случайной величины $Y=X^2$

Решение. На $(0;\infty)$, для $y=x^2$, обратная функция $x=\sqrt{y}=\psi_1$;

на $(-\infty;0)$ - обратная функция $x=-\sqrt{y}=\psi_2$. По формуле (2.11.1):

$g(y)=f(\psi_1(y))\cdot\left|\psi_1^{/}(y)\right|+f(\psi_2(y))\cdot\left|\psi_2^{/}(y)\right|=$

$$=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(\sqrt{y}-a)^2}{2\sigma^2}}\cdot\frac{1}{2\sqrt{y}}+\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(-\sqrt{y}-a)^2}{2\sigma^2}}\cdot\frac{1}{2\sqrt{y}}.$$

При a=0 и σ=1: $g(y)=\frac{1}{\sqrt{2\pi y}}e^{-\frac{y}{2}}$.

Пример 2.11.2. Случайная величина X равномерно распределяется на интервале $(-\frac{\pi}{2};\frac{\pi}{2})$. Найти дифференциальную функцию g(y),где $Y=sinX$.

Решение. Дифференциальная функция равномерного закона на $(-\frac{\pi}{2};\frac{\pi}{2})$ имеет вид:

$$f(x)=\begin{cases}0 \ \text{при}\ x>-\frac{\pi}{2},\\ \frac{1}{\pi}\ при -\frac{\pi}{2}<x<\frac{\pi}{2},\\ 0\ при\ \ x>\frac{\pi}{2}.\end{cases}$$

На интервале $-\frac{\pi}{2}<x<\frac{\pi}{2}$ функция sin монотонна, поэтому имеет обратную функцию:

$x=arcsin\, y$, $x'=\frac{1}{\sqrt{1-y^2}}$. По формуле (2.11.1):

$$g(y)=f(\arcsin y)\cdot\frac{1}{\sqrt{1-y^2}}=\frac{1}{\pi}\frac{1}{\sqrt{1-y^2}},=\frac{1}{\pi}\cdot\frac{1}{\sqrt{1-y^2}},$$

где $-1<y<1(-\frac{\pi}{2}<x<\frac{\pi}{2})$.

Пример 2.11.3. Два человека договорились встретиться в определенном месте в промежутке времени от 9^{00} до 10^{00}. Каждый из них приходит на место встречи независимо от другого, с постоянной дифференциальной функцией (функцией плотности вероятности) в любой момент назначенного промежутка времени. Пришедший раньше ожидает другого. Найти распределение вероятностей времени ожидания и вероятность того, что ожидать придется менее десяти минут.

Решение. Пусть моменты прихода двух лиц T_1 и T_2 соответственно. За начало отчета времени выберем девять часов. Тогда каждая из независимых случайных величин равномерно распределена в промежутке (0;1). Случайная величина T - время ожидания: $T=|T_1-T_2|$. Найдем функцию распределения $F(t)$ этой величины. Рассмотрим на плоскости $t_1 O t_2$ область $D(t)$, в которой $|t_1-t_2|<t$ (рис. 24).

Интегральная функция распределения F(t) равна площади области $D(t)$:

$F(t)=1 - (1-t)^2 = (2-t)t.$

Найдем дифференциальную функцию времени ожидания:

$f(t) = F'(t) = 2-2t$, при $0 < t < 1$.

Вероятность того, что ожидать придется менее десяти минут определим как:

$P(T < \frac{10}{60}) = P(T < \frac{1}{6}) = F(\frac{1}{6}) = (2 - \frac{1}{6})\frac{1}{6} = \frac{11}{36}$. (Сравните с примером 1.3.2.)

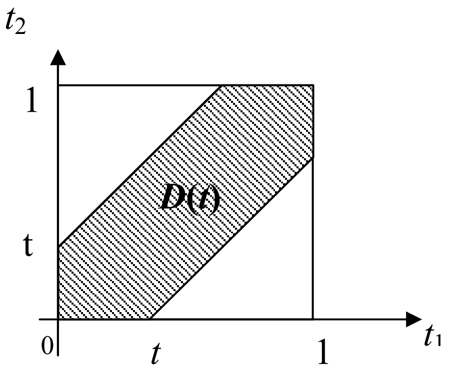


Рис. 24 - Область $D(t)$

Пример 2.11.4. Сторона квадрата равномерно распределена на отрезке [3; 4] оси *Y*. Найти закон распределения площади *S* прямоугольника со сторонами (*x, y*), расположенного внутри квадрата, причём две стороны прямоугольника лежат на сторонах квадрата.

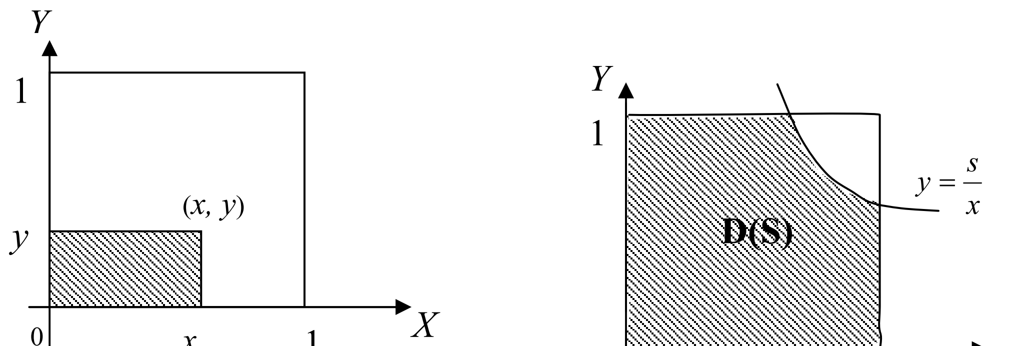


Рис. 25 - Закон распределения площади прямоугольника, определённого системой двух СВ (*X,Y*), равномерно распределённых в единичном квадрате

Решение. Выберем в качестве точки отсчета точку с координатами (0; 3). Тогда площадь прямоугольника со сторонами *x,y*: $S=xy$ (рис.25.а)). Выделим на плоскости

XOY область $D(s)$ в которой $xy<s$ (заштрихованная область на рис. 25.б)). Функция распределения $H(s)$ равна площади этой области:

$$H(s)=\iint\limits_{D(s)} dxdy = 1-\int_s^1 dx\int_{\frac{s}{x}}^1 dy = 1-\int_s^1 (1-\frac{s}{x})dx = 1-(x-s\ln x)\Big|_s^1 = 1-((1-\ln 1)-(s-s\ln s)) = s(1-\ln s).$$

Отсюда, $h(s) = H'(s) = -\ln s$, где $0<s<1$.

Рассмотрим *функции случайных величин в пространстве*. При выполнении исходных условий параграфа 2.11. СВ X - случайный вектор размерности n ($X=(X_1, X_2,\ldots,X_n)$), Y - случайный вектор размерности m ($Y=(Y_1, Y_2,\ldots,Y_m)$), $Y=\varphi(X)$.

Пусть $\varphi(X)$ имеет кусочно-непрерывные первые производные по всем координатам вектора X и не постоянна, ни на каком множестве значений аргумента X, имеющем отличную от нуля вероятность.

Тогда - дифференциальная функция СВ $Y=\varphi(X)$ при $n=m$:

$$g(y)=f(\varphi^{-1}(y))|J(y)|, \tag{2.11.3}$$

где $|J(y)|$ - абсолютная величина якобиана координат[1] вектора $X=\varphi^{-1}(Y)$ по координатам вектора Y:

$$J(y)=\frac{\partial(\varphi_1^{-1},..,\varphi_n^{-1})}{\partial(y_1,..,y_n)}=\begin{vmatrix} \frac{\partial\varphi_1^{-1}}{\partial y_1}\cdots\frac{\partial\varphi_1^{-1}}{\partial y_n} \\ \cdots\cdots\cdots \\ \frac{\partial\varphi_n^{-1}}{\partial y_1}\cdots\frac{\partial\varphi_n^{-1}}{\partial y_n}\end{vmatrix} \tag{2.11.4}$$

или

$$J(y)=\frac{\partial(x_1,x_2,\ldots,x_n)}{\partial(y_1,y_2,\ldots,y_n)}=\begin{vmatrix} \frac{\partial x_1}{\partial y_1}\frac{\partial x_1}{\partial y_2}\cdots\frac{\partial x_1}{\partial y_n} \\ \frac{\partial x_2}{\partial y_1}\frac{\partial x_2}{\partial y_2}\cdots\frac{\partial x_2}{\partial y_n} \\ \cdots\cdots\cdots \\ \frac{\partial x_n}{\partial y_1}\frac{\partial x_n}{\partial y_2}\cdots\frac{\partial x_n}{\partial y_n}\end{vmatrix}$$

(или $J(y)$ это производная функции $\varphi^{-1}(Y)$, в случае скалярных X и Y).

Отсюда, (2.11.3) можно записать в виде интегральной функции Y:

$$G(y)=\int_D f(\varphi^{-1}(y))\,|J(y)|\,dy, \tag{2.11.5}$$

где D - область задания Y.

Пример 2.11.5. По заданной дифференциальной функции распределения $f(x_1, x_2)$ двумерной случайной величины (X_1, X_2) найти плотность распределения $g(Y_1,Y_2)$

---

[1] $|J(y)|$- коэффициент искажения или коэффициент растяжения пространства (в данной точке), задаваемого вектором X, при преобразовании его в пространство, задаваемое вектором Y (в двумерном случае плоскости x_1x_2 в плоскость y_1y_2).

двумерной случайной величины (Y_1, Y_2), связанной взаимно однозначно с (X_1, X_2) указанными ниже соотношениями:

$$f(x_1, x_2) = \frac{1}{2\pi \cdot 3 \cdot 4} \cdot e^{-\frac{1}{2}\left(\frac{X_1^2}{3^2} + \frac{X_2^2}{4^2}\right)},$$

$$X_1 = 3Y_1 \cos 2Y_2,\ X_2 = 4Y_1 \sin 2Y_2,\ 0 \le Y_1 < \infty, \qquad 0 \le Y_2 < \pi.$$

Решение. Из (2.11.3) следует, что дифференциальная функция

$g(y) = f(x_1, x_2)\,|J(y)|$,

определена в области D, заданной неравенствами: $0 \le Y_1 < \infty$, $0 \le Y_2 < \pi$.

Имеем: $J(y) = \begin{vmatrix} 3\cos 2y_2 & -6y_1 \sin 2y_2 \\ 4\sin 2y_2 & 8y_1 \cos 2y_2 \end{vmatrix} = 24y_1 \cos^2 2y_2 + 24y_1 \sin^2 2y_2 = 24y_1$.

Отсюда,

$$g(y_1, y_2) = \frac{1}{2\pi \cdot 3 \cdot 4} \cdot e^{-\frac{1}{2}\left(\frac{3^2 y_1^2 \cos^2 2y_2}{3^2} + \frac{4^2 y_1^2 \sin^2 2y_2}{4^2}\right)} \cdot 24y_1,$$

$$g(y_1, y_2) = y_1 \cdot e^{-\frac{y_1^2}{2}} \cdot \frac{1}{\pi}$$

где $g(y_1) = y_1 \cdot e^{-\frac{y_1^2}{2}}$ - дифференциальная функция закона Рэлея;

$g_2(y_2) = \begin{cases} \frac{1}{\pi}, & 0 \le y_2 < \pi, \\ 0, & y_2 < 0 \text{ или } y_2 > \pi \end{cases}$ - дифференциальная функция равномерного закона на $[0, \pi)$.

2.12. Композиция законов распределения

В приложениях часто рассматривается вопрос о распределении суммы нескольких случайных величин. Например, пусть Z=X+Y, тогда G(z) -интегральную функцию СВ Z можно определить по формуле:

$$G(z) = \iint\limits_{(D)} f(x,y)dxdy = \int_{-\infty}^{+\infty} dy \int_{-\infty}^{z-x} f(x, z-x)dx = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{z-y} f(z-y, y)dy, \qquad (2.12.1)$$

где $f(x,y)$-дифференциальная функция системы случайных величин (X,Y);

область D – полуплоскость, ограниченная сверху прямой $y=z-x$.

Отсюда, $g(z) = G'(z) = \int\limits_{-\infty}^{+\infty} f(x, z-x)dx$.

Если X и Y независимы, то говорят о композиции законов распределения случайных величин и дифференциальная функция СВ Z определяется как

$$\mathrm{g}(z)=\int_{-\infty}^{+\infty} f_1(x) f_2(z-x)dx=\int_{-\infty}^{+\infty} f_1(z-y) f_2(y)\mathrm{d}x, \qquad (2.12.2)$$

где $f_1(x)$ и $f_2(y)$ дифференциальные функции СВ X и Y соответственно.

Если возможные значения аргументов неотрицательны, то дифференциальную функцию СВ Z определяют по формуле:

$$\mathrm{g}(z)=\int_0^z f_1(x) f_2(z-x)\mathrm{d}x \qquad (2.12.3)$$

или

$$\mathrm{g}(z)=\int_0^z f_1(z-y) f_2(y)\mathrm{d}y. \qquad (2.12.4)$$

Пример 2.12.1. Случайные величины X и Y распределены равномерно:

СВ X на (0;5),

СВ Y на (0;2).

Найти интегральную и дифференциальную функции СВ $Z = X + Y$.

Решение. Найдем функцию распределения $G(Z)$ величины $Z = X + Y$.

Система двух СВ (X, Y) равномерно распределена в прямоугольнике со сторонами 2 и 5 (рис.26.)

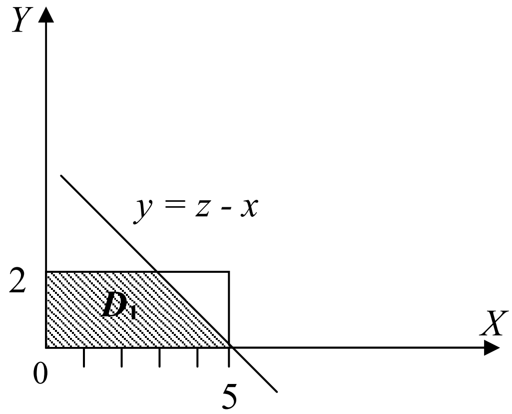


Рис. 26 - Область D_1

$$G(z)=\iint_{D_1} f(x,y)dxdy=\iint_{D_1} f_1(x) f_2(y)dxdy,$$

где D_1 – часть прямоугольника, лежащая левее и ниже прямой $y = z–x$; $f(x, y)$ - дифференциальная функция системы двух СВ (X,Y), f_1- дифференциальная функция СВ X, f_2 - дифференциальная функция СВ Y. В силу независимости СВ X и Y: $f(x,y)=f_1(x)f_2(y)=\frac{1}{5}\cdot\frac{1}{2}=\frac{1}{10}$.

То есть $G(z) = S(D_1)\cdot\frac{1}{10}$, где $S(D_1)$ – площадь области D_1, $\frac{1}{10}$ - плотность вероятности на единицу площади прямоугольника.

Имеем:

1) $z \le 0$, $G(z) = 0$;

2) $0 < z \le 2,\ \mathrm{G(z)} = \frac{z^2}{2} \cdot \frac{1}{10} = \frac{z^2}{20};$

3) $2 < z \le 5,\ \mathrm{G(z)} = \frac{1}{10}(2(z-1)) = \frac{1}{5}(z-1);$

4) $5 < z \le 7,\ G(z) = \frac{1}{10}(10 - \frac{(7-z)^2}{2}) = 1 - \frac{(7-z)^2}{20};$

5) z >7, $G(z)$=1.

Отсюда, $$g(z) = G'(z) = \begin{cases} 0 & \text{при } z \le 0, \\ \frac{z}{10} & \text{при } 0 < z \le 2, \\ \frac{1}{5} & \text{при } 2 < z \le 5, \\ \frac{7-z}{10} & \text{при } 5 < z \le 7, \\ 0 & \text{при } z > 7. \end{cases}$$

Замечание. Если рассматривается композиция нормальных законов: $X = \sum_{i=1}^{n} X_i$, где СВ X_i независимы (i=1,2,...,n) $X_i \in N(a_i, \sigma_i)$,то $X \in N(a, \sigma)$, где $M(X) = a = \sum_{i=1}^{n} a_i, D(X) = \sigma^2 = \sum_{i=1}^{n} \sigma_i^2.$

2.13. Специальные законы распределения

1. χ^2 *-распределение Пирсона*. Пусть $X_1, X_2, ..., X_n$ одинаково распределенные по нормальному закону случайные величины, являющиеся взаимно-независимыми, для которых математическое ожидание равно нулю, а среднеквадратическое отклонение 1, тогда сумма квадратов этих случайных величин носит название случайной величины χ^2 - хи-квадрат с $\nu=n$ степенями свободы:

$$\chi_n^2 = \sum_{i=1}^{n} X_i^2 \qquad (2.13.1)$$

При ν=1 (учитывая пример 2.11.1) дифференциальная функция χ^2:

$$g(c) = \frac{1}{2\sqrt{\chi^2}\sqrt{2\pi}} \cdot \left(e^{-\frac{(\sqrt{\chi^2})^2}{2}} + e^{-\frac{(-\sqrt{\chi^2})^2}{2}} \right) = \frac{1}{\sqrt{2\pi\chi^2}} \cdot e^{-\chi^2/2}.$$

Дифференциальная функция распределения χ^2 с ν=n степенями свободы задаётся формулой

$$f(\chi^2)=\frac{1}{2^{n/2}\cdot\Gamma(\frac{n}{2})}\cdot(\chi^2)^{(n/2-1)}\cdot e^{-\frac{\chi^2}{2}}, \qquad (2.13.2)$$

где $\Gamma(x)$ -гамма, функция Эйлера.

$$\Gamma(x)=\int_0^{+\infty} t^{x-1}e^{-t}dt \text{, при } R_+\text{; если } n\in Z\text{, то } \Gamma(n+1)=n!$$

С возрастанием числа степеней свободы $\nu = n$, распределение χ^2 медленно приближается к нормальному закону распределения. На практике используют обычно не плотность вероятности, а квантили распределения (прил. 2).

Квантилью χ^2_n распределения, отвечающей заданному уровню значимости α (альфа): $\chi^2_{\alpha,\nu}$, называется такое значение $\chi^2=\chi^2_{\alpha,\nu}$, при котором вероятность того, что χ^2 превысит значение $\chi^2_{\alpha,\nu}$ равна α (рис.27):

$$P(\chi^2>\chi^2_{\alpha,\nu})=\int_{\chi_{\alpha,\nu}}^{+\infty} f(\chi^2)d(\chi^2)=\alpha. \qquad (2.13.3)$$

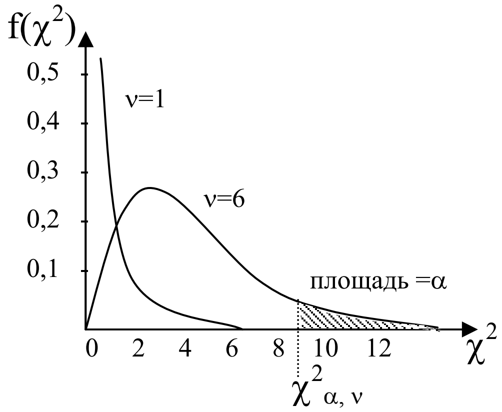


Рис. 27 -Дифференциальная функция распределения χ^2 с ν степенями свободы

С геометрической точки зрения нахождение квантили $\chi^2_{\alpha,\nu}$ заключается в выборе такого значения $\lambda^2=\chi^2_{\alpha,\nu}$, при котором площадь криволинейной трапеции ограниченной дифференциальной функцией была бы равна α. Значение квантилей затабулированны (прил.2). При n>30 распределение практически не отличается от нормального.

Замечание. Квантиль СВХ порядка α - это такое значение СВХ, что $F(x_\alpha)=1-\alpha$, где $F(x)=P(X<x)$. Например, медиана – это квантиль $x_{0,5}$.

2. *t– распределение Стьюдента*. Это распределение имеет важное значение при статистических вычислениях, связанных с нормальным законом распределения, где σ – неизвестный параметр распределения и подлежит определению из опытных данных, например, при статистической обработке наблюдений с неизвестной точностью.

Пусть $X, X_1, X_2,...,X_k$ − независимые нормально распределённые случайные величины с нулевыми математическими ожиданиями и одинаковыми дисперсиями. Безразмерная величина t:

$$t = \frac{X}{\sqrt{\frac{1}{k}\cdot\sum_{i=1}^{k} X_i^2}} = \frac{X}{\sqrt{\frac{\chi_k^2}{k}}}, \tag{2.13.4}$$

называется дробью Стьюдента.

Ее распределение не зависит от σ в силу ее безразмерности. Дифференциальная функция t-распределения с $\nu=k$ степенями свободы имеет вид:

$$f(x) = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{\pi k}\,\Gamma\left(\frac{k}{2}\right)}\left(1+\frac{t^2}{k}\right)^{-\frac{k+1}{2}}. \tag{2.13.5}$$

t-распределение Стьюдента, которое быстрее, чем χ^2 стремится к нормальному.

На практике используют квантили распределения в зависимости от числа степеней свободы и уровня значимости α. С геометрической точки зрения нахождение квантилей (для двусторонней области) заключается в выборе такого значения t, при котором суммарная площадь криволинейной трапеции была бы равна α, в силу симметрии распределения (рис. 28):

$$P(|t| > t_{\alpha/2,\nu}) = 2\int_{t_{\alpha/2,\nu}}^{+\infty} f(t)dt = \alpha. \tag{2.13.6}$$

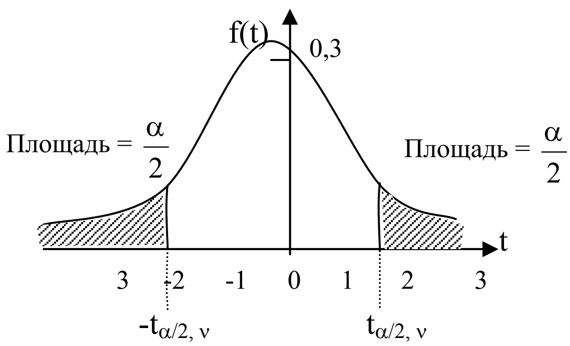


Рис. 28 - Дифференциальная функция t-распределения Стьюдента с $\nu=k \le 30$ степенями свободы

3. *F- распределение Фишера-Снедекора.*

Пусть $X_1, X_2, ...,X_m$ и $Y_1, Y_2, ...,Y_n$ одинаково распределенные по нормальному закону случайные величины, являющиеся взаимно-независимыми, для которых математическое ожидание равно нулю, а среднеквадратическое отклонение равно единице.

Рассмотрим дробь Фишера:

$$F(m, n) = \frac{\chi_m^2}{m} \Big/ \frac{\chi_n^2}{n}, \tag{2.13.7}$$

она имеет F - распределение с $\nu_1=m$ - числом степеней свободы числителя, и $\nu_2=n$ - числом степеней свободы знаменателя ((m, n) степенями свободы), которое называется распределением Фишера-Снедекора. Обычно используют квантили распределения в зависимости от числа степеней свободы (m, n) и уровня значимости α. (рис. 29):

$$P(F > F_\alpha(\nu_1, \nu_2)) = \int\limits_{F_\alpha(\nu_1,\nu_2)}^{+\infty} f(F)dF. \qquad (2.13.8)$$

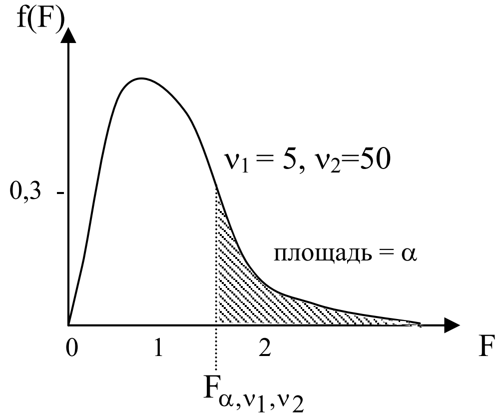


Рис. 29 - Дифференциальная функция F распределения Фишера - Снедекора с ν_1=5, ν_2=50 степенями свободы

Для квантилей распределения Фишера - Снедекора геометрический смысл аналогичен другим распределениям (рис.23). Имеет место равенство:

$$F_{1-\alpha}(m,n) = \frac{1}{F_\alpha(n,m)}. \qquad (2.13.9)$$

Распределения χ^2 - Пирсона, t - Стьюдента, F - Фишера-Снедекора нашли широкое применение в математической статистике, в частности при проверке статистических гипотез и в дисперсионном анализе.

Контрольные задания 2.13.

1. Дискретная случайная величина задана законом распределения:

X	0	1	4
p	0,3	0,5	0,2

Найти закон распределения случайной величины Y, где: а) $Y=2X-1$;
б) $Y=X+5$; в) $Y=X^2-2$; г) $Y=\sqrt{X}$.

2. Дискретная случайная величина X задана законом распределения:
3.

X	-2	-1	0	1
p	0,2	0,4	0,1	0,3

Найти закон распределения случайной величины Y, где: а) $Y=2X+1$; б) $Y=X^3-1$; в) $Y=X^2$; г) $Y=\sqrt{X+2}$.

4. Дискретная случайная величина X задана законом распределения:

X	1	2	4	5
p	0,1	0,3	0,2	0,4

Найти математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины Y, если: а) $Y=4X-4$; б) $Y=X^2$.

4. Дискретная случайная величина X задана законом распределения:

X	$\frac{\pi}{6}$	$\frac{\pi}{4}$	$\frac{\pi}{3}$
p	0,2	0,7	0,1

Найти: а) закон распределения случайной величины $Y=sin^2X$; б) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины Y.

5. Случайная величина X равномерно распределена в интервале $(-\frac{\pi}{2}; \frac{\pi}{2})$. Найти дифференциальную функцию случайной величины:
а) $Y=sin\,X$; б) $Y=cos\,X$.

6. Случайная величина X распределена нормально с параметрами $a = 2$, $\sigma=1$. Найти дифференциальную функцию случайной величины:
а) $Y=2X+6$; б) $Y=X^3$.

7. Сторона квадрата X имеет равномерное распределение на отрезке [1;2]. Найти функцию плотности вероятности площади квадрата.

8. Случайная величина X распределена по закону Коши:

$$f(x) = \frac{1}{\pi\,(1+x^2)} \,.$$

Найти дифференциальную функцию случайной величины: а) $Y=X^3$;
б) $Y=3X$.

9. Независимые случайные величины X и Y распределены равномерно. Случайная величина X распределена в интервале (0; 2), а случайная величина Y в интервале (0; 10). Найти интегральную и дифференциальную функции случайной величины $Z=X+Y$. Построить графики интегральной и дифференциальной функций случайной величины Z.

10. Случайная величина X равномерно распределена в интервале (-4; 1), а случайная величина Y равномерно распределена в интервале (1; 6). Найти дифференциальную функцию случайной величины $Z=X+Y$ и начертить ее график.

11. Независимые случайные величины X и Y заданы дифференциальными функциями:

$$f_1(x) = e^{-x}, \qquad при \quad 0 \le x < \infty,$$

$$f_2(y) = 0{,}5e^{-0{,}5y}, \quad при \quad 0 \le y < \infty.$$

Найти дифференциальную функцию случайной величины $Z=X+Y$.

12. Независимые случайные величины X и Y распределены по нормальному закону:

$$f_1(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \qquad f_2(y) = \frac{1}{\sqrt{2\pi}} e^{-\frac{y^2}{2}}.$$

Найти дифференциальную функцию случайной величины $Z=X+Y$. Показать, что случайная величина Z распределяется по нормальному закону.

13. Натуральный логарифм некоторой случайной величины X распределен по нормальному закону с центром рассеивания a и средним квадратическим отклонением σ. Найти плотность распределения случайной величины X.

2.14. Закон больших чисел

> - Законы статистики везде одинаковы, - продолжал Николай Петрович солидно. Утром, например, гостей бывает меньше, потому что публика ещё исправна; но чем больше солнце поднимается к зениту, тем наплыв делается сильнее. И, наконец ночью, по выходе из театров – это почти целая оргия!
> - И заметьте, - пояснил Семён Иванович, каждый день, в одни и те же промежутки времени, цифры всегда одинаковые. Колебаний – никаких! Такова незыблемость законов статистики!
>
> М.Е.Салтыков-Щедрин. «За рубежом»

Под законом больших чисел в теории вероятностей понимается совокупность теорем, в которых устанавливается связь между средним арифметическим достаточно большого числа случайных величин и средним арифметическим их математических ожиданий.

На протяжении веков математическая наука, основываясь на логике Аристотеля, развивалась под влиянием основной потребности – исключить неточность и неопределённость в человеческих суждениях о природе наблюдаемых социально-экономических, технических и других процессов и явлений.

В повседневной жизни, бизнесе, научных исследованиях мы постоянно сталкиваемся с событиями и явлениями с неопределённым исходом. Например, торговец не знает сколько посетителей придёт к нему в магазин, бизнесмен не знает курс доллара через 1 день или год; банкир – вернут ли ему заём в срок; страховые компании – когда и кому придётся выплачивать страховое вознаграждение.

Развитие любой науки предполагает установление основных закономерностей и причинно-следственных связей в виде определений, правил, аксиом, теорем. Однако, более углублённое рассмотрение указанных закономерностей показывает неточность первоначальных утверждений. Последнее связано с открытием в 1927 г. Гейзенбергом «принципа неопределённости», заключающегося в ограниченности «измерительного познания» окружающего нас мира.

Таким образом, неопределённость может служить философским и методологическим объяснением ограниченности традиционных подходов к моделированию экономических, биологических и др. процессов. Математическая статистика, бурно развивавшаяся последние 200 лет, позволяет учитывать случайность и неопределённость при анализе статистических данных для принятия решений. Это послужило толчком к развитию: теории ошибок наблюдений, теории связи, радиотехники, теории автоматического управления, социологии, экономики, психологии и др. наук.

Основой для математической статистики служит математический аппарат и выводы теории вероятностей, изучающей закономерности, происходящие в массовых, однородных случайных явлениях и процессах.

Связующим звеном между теорией вероятностей и математической статистикой являются так называемые предельные теоремы, к которым относится закон больших чисел. Закон больших чисел устанавливает условия при которых совокупное воздействие множества факторов приводит к результату, не зависящего от случая. В самом общем виде закон больших чисел сформулировал П.Л.Чебышев. Большой вклад в изучение закона больших чисел внесли А.Н. Колмогоров, А.Я. Хинчин, Б.В. Гнеденко, В.И. Гливенко.

К предельным теоремам относится также так называемая Центральная предельная теорема А.Ляпунова, устанавливающая условия при которых сумма случайных величин будет стремиться к случайной величине с нормальным законом распределения. Эта теорема позволяет обосновать методы проверки статистических гипотез, корреляционно-регрессионный анализ и др. методы классической статистики. Дальнейшее развитие центральной предельной теоремы связано с именами Линденберга, С.Н. Бернштейна, А.Я. Хинчина, П. Леви.

Замечание.
1) Теория вероятностей и классическая математическая статистика, получившая развитие в XIX и первой половине XX века трактуют понятие неопределённости только с точки зрения вероятности (вероятностная неопределённость). Между тем вероятность имеет место на практике (равно как и законы распределения вероятностей) только при наличии устойчивой частоты появления события, стремящейся к некоторому числу (об этом писали С.Н.Бернштейн, В.Н.Тутубалин, С.А.Айвазян, А.И. Орлов, С.А.Айвазян и др.). В других случаях говорить о вероятностной неопределённости нельзя (говорят о неопределённости II рода, которая является предметом теории игр, теории возможностей и теории нечётких множеств). А.Эйнштейн вообще не верил в существование вероятности, считая, что все процессы детерминированны. С другой стороны многие учёные (физики-теоретики) считают вероятность неотъемлемой частью нашей жизни.
2) практическое применение методов теории вероятностей и математической статистики основано на двух принципах, фактически основывающихся на предельных теоремах:
✓ Принцип невозможности наступления маловероятного события.
✓ Принцип достаточной уверенности в наступлении события, вероятность которого близка к 1.
3) следует отметить, что уже в конце XX века была известна ограниченность предельных теорем, в силу того, что выборки имеющие место на практике – конечны.

Неравенство и теорема Чебышева.

Рассмотрим закон больших чисел в форме Чебышева.

Лемма Чебышева (Маркова). Если случайная величина X принимает только неотрицательные значения и имеет математическое ожидание $M(X)$, то для любого $\alpha > 0$ имеет место неравенство:

$$P(X \geq \alpha) \leq \frac{M(X)}{\alpha}. \tag{2.14.1}$$

Неравенство Чебышева. Если случайная величина X имеет математическое ожидание $M(X)$ и дисперсию $D(X)$, то для любого $\varepsilon > 0$ имеет место неравенство:

$$P(|x - M(x) < \varepsilon|) \geq 1 - \frac{D(X)}{\varepsilon^2} \qquad (2.14.2)$$

Неравенство Чебышева является в теории вероятностей общим фактом и позволяет оценить нижнюю границу вероятности.

Пример 2.14.1. Если произведено n независимых испытаний по схеме Бернулли, где p – вероятность успеха, q - вероятность неудачи, n - число опытов, m - число успехов, то для случайной величины m имеет место неравенство:

$$P(|k - np| < \varepsilon) \geq 1 - \frac{npq}{\varepsilon^2}. \qquad (2.14.3)$$

Для относительной частоты появления события $\frac{k}{n}$ аналогичное неравенство имеет вид:

$$P\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) \geq 1 - \frac{pq}{n\varepsilon^2}. \qquad (2.14.4)$$

Теорема. *Закон больших чисел Чебышева.* Пусть $X_1, X_2, \ldots, X_n$ - последовательность попарно независимых случайных величин, имеющих конечные математические ожидания и дисперсии, ограниченные сверху постоянной $C = \mathrm{const}$ ($D(X_i) \leq C$ (i=1, 2,…,n)). Тогда для любого $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - \frac{1}{n}\sum_{i=1}^{n} M(X_i)\right| < \varepsilon\right) = 1. \qquad (2.14.5)$$

Теорема показывает, что среднее арифметическое большого числа случайных величин с вероятностью сколь угодно близкой к 1 будет мало отклоняться от среднего арифметического математических ожиданий.

Следствие 1. Если вероятность наступления события A в каждом из n независимых испытаний равна p, m - число наступлений события A в серии из n независимых испытаний, то, каково бы ни было число $\varepsilon > 0$, имеет место предел:

$$\lim_{n\to\infty} \mathrm{P}\left(\left|\frac{k}{n} - p\right| < \varepsilon\right) = 1. \qquad (2.14.7)$$

Таким образом устанавливается связь между относительной частотой появления события A и постоянной вероятностью p в серии из n независимых испытаний.

Следствие 2. *Теорема Пуассона.* Если в последовательности независимых испытаний вероятность появления события A в r-ом испытании равна p_r, то

$$\lim_{n\to\infty} P\left(\left|\frac{k}{n} - \frac{p_1 + p_2 + \ldots + p_n}{n}\right| < \varepsilon\right) = 1, \qquad (2.14.7)$$

где к - число появлений события А в серии из n испытаний.

Следствие 3. *Теорема Бернулли.* Если $X_1, X_2, \ldots, X_n$ - последовательность независимых случайных величин таких, что

$M(X_1) = M(X_2) = \ldots = M(X_n) = a$,

$D(X_1) < C, D(X_2) < C, \ldots, D(X_n) < C$, где $C = const$,

то, каково бы ни было постоянное число $\varepsilon > 0$, имеет место предел:

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i - a\right| < \varepsilon\right) = 1. \qquad (2.14.8)$$

Этот частный случай закона больших чисел позволяет обосновать правило средней арифметической.

Законы больших чисел не позволяют уменьшить неопределённость в каждом конкретном случае, они утверждают лишь о существовании закономерности при достаточно большом числе опытов. Например, если при подбрасывании монеты 10 раз появился герб, то это не означает, что в 11 раз появится цифра.

Пример 2.14.2. Оценить с помощью неравенства Чебышева вероятность того, что при подбрасывании 12 игральных костей сумма очков (СВ X) отклонится от математического ожидания меньше, чем на 15. СВ X_i – число очков на i-й кости (i = 1, 2, …, 12).

Решение. СВ $X=X_1+\ldots X_{12}$,

где $M(X_1)=M(X_2)=\ldots=M(X_{12})$, $D(X_1)=D(X_2)=\ldots=D(X_{12})$.

$$M(X_1)=\frac{1}{6}(1+2+3+4+5+6)=\frac{21}{6}=\frac{7}{2}=3{,}5,$$

$$M(X_1^2)=\frac{1}{6}(1+2^2+3^2+4^2+5^2+6^2)=\frac{91}{6},$$

$$M(X)=3{,}5\cdot 12=42,$$

$$D(X_1)=\frac{91}{6}-\frac{49}{4}=\frac{182-147}{12}=\frac{35}{12},$$

$$D(X)=(35/12)\cdot 12=35.$$

Согласно неравенству Чебышева имеем

$$P(|X-42|<15)\geq 1-\frac{35}{225},$$

$$P(|X-42|<15)\geq 0{,}844.$$

2.15. Понятие о центральной предельной теореме

В теории вероятностей и математической статистике большое значение имеет центральная предельная теорема Ляпунова, в которой утверждается, что если сложить большое число случайных величин, имеющих один или различные законы распределения, то случайная величина, являющаяся результатом суммы, при некоторых условиях будет иметь нормальный закон распределения.

Примером центральной предельной теоремы (для последовательности независимых случайных величин) является интегральная теорема Муавра-Лапласа.

Теорема 1. Пусть производится n независимых опытов в каждом из которых вероятность наступления события A равна p (не наступления q=1-p, $p\neq 0$, $p\neq 1$). Если K - число появлений события А в серии из n испытаний, то при достаточно больших n СВ K можно считать нормально распределенной ($M(K)=np$, $\sigma(K)=\sqrt{D(K)}=\sqrt{npq}$):

$$\mathrm{P}(K<k)\ \to \mathrm{P}(\mathrm{X}<\mathrm{x}_0)=\int_{-\infty}^{\mathrm{x}_0}\varphi(\mathrm{x})d\mathrm{x}=\frac{1}{2}+\Phi(\mathrm{x}_0), \qquad (2.15.1)$$

где $x_0 = \frac{k-np}{\sqrt{npp}}$, $\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$, $\Phi(x_0)$ - функция Лапласа.

В более общем случае верна следующая теорема.

Теорема 2. Если случайные величины $X_1, X_2, \ldots X_n$ независимы, одинаково распределены и имеют конечную дисперсию, то при $n \to \infty$:

$$P\left(\frac{X_1 + X_2 + \ldots + X_n - na}{\sigma\sqrt{n}} < t \right) \to \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-\frac{u^2}{2}} du , \qquad (2.15.2)$$

где $M(X_i)=a,\ \sigma^2=D(X_i);$

U - нормально распределенная случайная величина, $M(U)=0$, $D(U)=1$.

Пример 2.15.1. На отрезке [0; 1] случайным образом выбрано 100 чисел, точнее рассматриваются 100 независимых средних $X_1, X_2, \ldots X_n$, равномерно распределенных на отрезке [0; 1]. Найти вероятность того, что их сумма заключена между 51 и 60, т.е. $P(51<\sum X_i <60)$.

Решение. В силу Теоремы 2:

$$\frac{\sum X_i - na}{\sigma\sqrt{n}} = U ,$$

$$\sum(X_i) - na = \sigma\sqrt{n}U .$$

Из условия, в силу равномерности СВ X_i , следует, что

$$M(X_i) = \frac{1+0}{2} = \frac{1}{2},\ \sigma^2 = \frac{(1-0)^2}{12}.$$

Имеем:

$$M(\sum X_i) = M(na + \sigma\sqrt{n}U) = na = 100\frac{1}{2} = 50,$$

$$D(\sum X_i) = D(na + \sigma\sqrt{n}U) = n\sigma^2 = 100\frac{1}{12} = \frac{100}{12} .$$

Итак, $\sum X_i \in N(na, \sqrt{n}\sigma)$ - сумма, нормально распределённая случайная величина с математическим ожиданием na=50 и дисперсией $n\sigma^2$=100/12.

Отсюда,

$$P(51 \le \sum X_i \le 60) = \Phi\left(\frac{60-na}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{51-na}{\sqrt{n}\sigma}\right) =$$

$$= \Phi\left(\frac{60-50}{\frac{10}{\sqrt{12}}}\right) - \Phi\left(\frac{51-50}{\frac{10}{\sqrt{12}}}\right) = \Phi\left(\sqrt{12}\right) - \Phi\left(\frac{\sqrt{12}}{10}\right) = \Phi(3{,}464) - \Phi(0{,}3464) \approx$$

$$\approx 0{,}49971 - 0{,}1353 = 0{,}3644.$$

То есть вероятность того, что сумма 100 независимых средних $X_1, X_2, \ldots, X_n$, равномерно распределенных на отрезке [0; 1], заключена между 51 и 60 равна 0,3644.

Контрольные задания 2.15.

1. Количество кормов, расходуемых на ферме крупного рогатого скота в сутки, является случайной величиной, математическое ожидание которой равно 6 т. Оценить вероятность того, что в ближайшие сутки расход кормов на ферме превысит 10 т.
2. Количество электроэнергии, потребляемой поселком в течение суток, является случайной величиной, математическое ожидание которой равно 4 тыс. квт. ч. Оценить вероятность того, что в ближайшие сутки потребление энергии: а) превысит 8 тыс. квт. ч.; б) не превысит 6 тыс. квт. ч.
3. Пользуясь неравенством Чебышева, оценить вероятность того, что из посеянных 5000 семян число взошедших окажется от 3750 до 4250, если известно, что M(X) = 4000. Определить вероятность попадания случайной величины в заданный интервал.
4. Вероятность вызревания семян овощной культуры в данной местности составляет 0,8. С помощью неравенства Чебышева оценить вероятность того, что из 1000 растений число растений с вызревшими семенами составит от 750 до 850. Определить вероятность попадания случайной величины в заданный интервал.
5. В хозяйстве имеется 100 автомобилей. Вероятность безотказной работы каждого из них в течение определенного периода составляет 0,9. С помощью неравенства Чебышева оценить вероятность того, что отклонение числа безотказно работавших автомобилей за определенный период от его математического ожидания не превзойдет по модулю 5.

6. Дискретная случайная величина *X* задана законом распределения:

X	2	3	6	9
p	0,1	0,4	0,3	0,2

Используя неравенство Чебышева, оценить вероятность того, что $|X - M(X)| > 3$.

7. Величина *X* задана законом распределения:

X	-1	0	1	3	5
p	0,1	0,2	0,4	0,2	0,1

Используя неравенство Чебышева, оценить вероятность того, что $|X - M(X)| < 2{,}5$.

8. Всхожесть семян некоторого растения составляет 90 %. Используя неравенство Чебышева, оценить вероятность того, что из посеянных 5000 семян: а) отклонение доли взошедших семян от постоянной вероятности взойти каждому из них не превзойдет по модулю 0,03; б) отклонение числа взошедших семян от математического ожидания не превзойдет по модулю 100.
9. Случайная величина задана интегральной функцией:

$$F(x) = \begin{cases} 0 & \text{при} \quad x \le a, \\ \dfrac{(x-a)^2}{a^2} & \text{при} \; a < x \le 2a, \\ 1 & \text{при} \quad x > 2a. \end{cases}$$

а) с помощью неравенства Чебышева, оценить вероятность того, что $|X - M(X)| < \frac{a}{2}$; б) определить вероятность того, что $|X - M(X)| < \frac{a}{2}$.

10. Случайная величина задана интегральной функцией:

$$F(x)=\begin{cases}0 & \text{при} \quad x\le 0,\\ \dfrac{x^2}{4a^2} & \text{при} \quad 0<x\le 2a,\\ 1 & \text{при} \quad x>2a.\end{cases}$$

а) используя неравенство Чебышева, оценить вероятность того, что $|X-M(X)|$<а; б) определить вероятность того, что $|X-M(X)|$<а.

11. Выборочным способом определяют вес колосьев ячменя. Сколько необходимо отобрать колосьев, чтобы с вероятностью не меньшей 0,99, можно было утверждать, что средний вес случайно отобранных колосьев будет отличаться от среднего веса колосьев во всей партии (принимаемого за математическое ожидание) не более чем на 0,1 г? Установлено, что среднее квадратическое отклонение веса не превышает 0,2 г.

12. Сколько человек необходимо отобрать для определения удельного веса лиц со специальным образованием, чтобы с вероятностью 0,95 можно было утверждать, что отклонение относительной частоты лиц со специальным образованием от их доли, принимаемой за постоянную вероятность, не превышало по модулю 0,04?

13. В результате анализа торговой деятельности некоторого магазина установлено, что среднемесячные издержки обращения составляют 300 усл. ден. ед. Оцените вероятность того, что в очередном месяце издержки не выйдут за пределы 280-320 денежных единиц. Известно, что дисперсия издержек равна 16 ден. ед.

14. Для определения средней урожайности на площади 100000 га взято на выборку по одному гектару от каждого участка размером 100 га. Определите вероятность того, что средняя выборочная вероятность будет отличаться от действительной средней по всей площади не более чем на 0,5 ц, если дисперсия урожайности на отдельных участках (по 100 га) не превышает 2 ц.

2.16. Цепи Маркова[1]

Результаты различных экспериментов или испытаний могут быть представлены в виде определенной последовательности букв или цифр, которые рассматриваются как процессы перехода от одного состояния в другое. Можно оценить вероятность такого перехода. Модель случайного процесса объединяет множество состояний и множество вероятностей перехода из одного состояния в другое.

Пусть происходит случайный процесс в системе S с дискретными состояниями $s_1, s_2, \ldots, s_n$, которые она может принимать в последовательности шагов с номерами $0,1,2,\ldots k,\ldots$

Случайный процесс представляет собой последовательность («цепь») событий вида $\{S(k)=s_i\}$ (i=0,1,2,…,n; k=0,1,2,…). Цепь, в которой условные вероятности состояний в будущем зависят только от состояния на данном, последнем, шаге и не зависят от предыдущих, называют простой цепью Маркова. Наиболее важной ее характеристикой являются безусловные вероятности нахождения системы S на любом (k-м) шаге в состоянии s_i; обозначим эту вероятность $p_i(k)$:

$$p_i(k)=\mathrm{P}\{S(k)=s_i\} \quad (i=0,1,2,\ldots,n;\ k=0,1,2,\ldots). \qquad (2.16.1)$$

Для нахождения этих вероятностей необходимо знать условные вероятности перехода системы S на k-м шаге в состояние s_j, если известно, что на предыдущем (k-1)-м шаге она была в состоянии s_i. Обозначим эту вероятность

[1] Разработано ст. преподавателем Т.В. Соловьевой

$$P_{ij}(k)=\mathrm{P}\{\mathrm{S}(k)=s_j \mid \mathrm{S}(k\text{-}1)=s_i\}(i,j=0,1,2,\ldots,n). \quad (2.16.2)$$

Вероятности $P_{ij}(k)$ называются переходными вероятностями марковской цепи на k-м шаге. Вероятность $P_{ii}(k)$ есть вероятность того, что на k-м шаге система задержится (останется) в состоянии s_i.

Цепь называется однородной, если переходные вероятности $P_{ij}(k)$ не зависят от номера шага k: $P_{ij}(k)=P_{ij}$.

Переходные вероятности P_{ij} можно записать в виде квадратной матрицы размерности $n\times n$:

$$\|P_{\mathrm{ij}}\|=\begin{Vmatrix} P_{11} & P_{12} & \ldots & P_{1\mathrm{j}} & \ldots & P_{1\mathrm{n}} \\ P_{21} & P_{22} & \ldots & P_{2\mathrm{j}} & \ldots & P_{2\mathrm{n}} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ P_{\mathrm{i}1} & P_{\mathrm{i}2} & \ldots & P_{\mathrm{ij}} & \ldots & P_{\mathrm{in}} \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ P_{\mathrm{n}1} & P_{\mathrm{n}2} & \ldots & P_{\mathrm{nj}} & \ldots & P_{\mathrm{nn}} \end{Vmatrix}. \quad (2.16.3)$$

Так как на каждом шаге система S может находиться только в одном из взаимно исключающих состояний, то для любой i-й строки матрицы (2.16.3) сумма всех стоящих в ней вероятностей P_{ij} равна единице: $\sum_{j=1}^{n} P_{ij}=1(i=0,1,2,\ldots,n)$.

Основной задачей исследования марковской цепи является нахождение безусловных вероятностей $p_i(k)$ нахождения системы S на любом (k-м) шаге в состоянии s_i ($i=0,1,2,\ldots,n$; $k=0,1,2,\ldots$), если задана матрица переходных вероятностей и начальное распределение вероятностей

$$p_i(0) \quad (i=0,1,2,\ldots,n), \quad \sum_{i=1}^{n} p_i(0)=1. \quad (2.16.4)$$

Сформулируем гипотезу, что система S в начальный момент ($k=0$) находилась в состоянии s_i . Вероятность этой гипотезы известна из (2.16.4). Тогда условная вероятность того, что система S на первом шаге будет в состоянии s_j , равна переходной вероятности

$$P_{ij}=P\{S(1)=s_j \mid S(0)=s_i\}.$$

По формуле полной вероятности получим

$$p_j(1)=\sum_{i=1}^{n} p_i(0)P_{ij} \quad (j=1,2,\ldots,n). \quad (2.16.5)$$

После второго шага по формуле полной вероятности находим

$$p_j(2)=\sum_{i=1}^{n} p_i(1)P_{ij} \quad (j=1,2,\ldots,n). \quad (2.16.6)$$

Переходя таким же способом от k=2 к k=3 и т.д., получим рекуррентную формулу:

$$p_j(k)=\sum_{i=1}^{n} p_i(k-1)P_{ij} \quad (k=1,2,\ldots;\ j=1,2,\ldots,n). \quad (2.16.7)$$

Замечание. Вероятность перехода процесса из состояния s_i в состояние s_j за n шагов обозначается $P_{ij}^{(n)}$, а квадратная матрица всех этих вероятностей обозначается P_n. Вероятность перехода процесса из состояния i в состояние j за два шага вычисляется по формуле полной вероятности.

$P_{ij}^{(2)} = \sum_{z=1}^{m} P_{iz} \cdot P_{zj}$ или в матричной записи $P_2 = P_1 \cdot P_1 = P_1^2$.

Аналогично получается матрица перехода за три шага $P_3 = P_1^3$ и за n – шагов $P_n = P_1^n$.

$P_1 = \|P_{ij}\|$, где P_{ij} - элемент (вероятность), стоящий на пересечении i-ой строки и j-го столбца.

$\sum_{i=1}^{k} P_{ij} = 1, i = 1,2,3,......k, j = 1,2,....,k.$

Пример. По некоторой цели производятся выстрелы в моменты времени t_1, t_2, t_3,…Возможные состояния системы (цели):

s_1 – цель невредима,

s_2 – цель повреждена (но может функционировать),

s_1 – цель полностью поражена (не может функционировать).

Граф состояний представлен на рисунке 30. Матрица переходных вероятностей

$$\|P_{ij}\| = \begin{Vmatrix} P_{11} & P_{12} & P_{13} \\ P_{21} & P_{22} & P_{23} \\ P_{31} & P_{32} & P_{33} \end{Vmatrix} = \begin{Vmatrix} 0,1 & 0,7 & 0,2 \\ 0 & 0,6 & 0,4 \\ 0 & 0 & 1 \end{Vmatrix} \qquad (*)$$

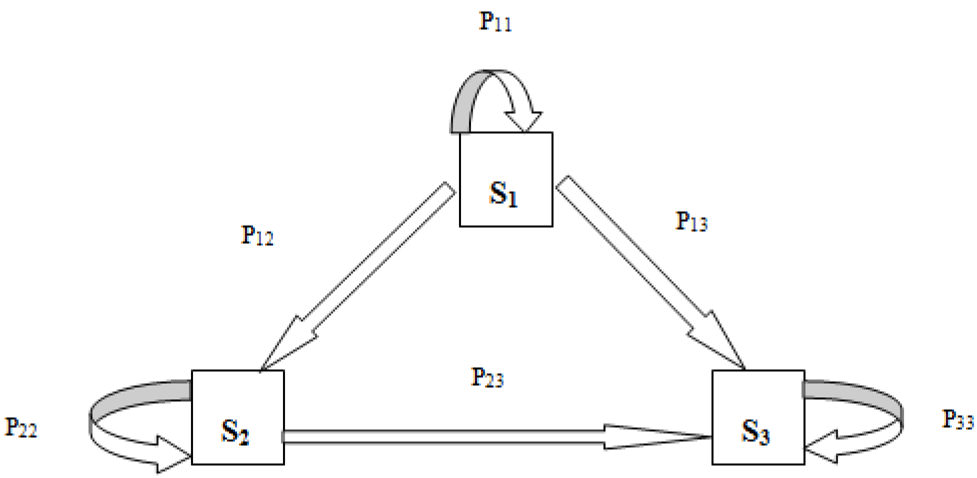


Рис. 30

Сколько надо сделать выстрелов, чтобы вероятность поражения цели оказалась не менее 0,6?

До того, как начались выстрелы, система (цель) находилась в состоянии s_1. После первого выстрела (после первого шага) вероятности состояний системы определяются из первой строки матрицы (*):

$$p_1(1)=0,1, \quad p_2(1)=0,7, \quad p_3(1)=0,2 \qquad (**)$$

Таким образом, после первого выстрела вероятность полного поражения цели составляет 0,2.

Делаем второй шаг (второй выстрел). Используя (2.16.6), (*) и (**), находим:

$$p_1(2) = p_1(1)P_{11} + p_2(1)P_{21} + p_3(1)P_{31} = 0{,}1\cdot 0{,}1 + 0{,}7\cdot 0 + 0{,}2\cdot 0 = 0{,}01;$$

$$p_2(2) = p_1(1)P_{12} + p_2(1)P_{22} + p_3(1)P_{32} = 0{,}1\cdot 0{,}7 + 0{,}7\cdot 0{,}6 + 0{,}2\cdot 0 = 0{,}49; \qquad (***)$$

$$p_3(2) = p_1(1)P_{13} + p_2(1)P_{23} + p_3(1)P_{33} = 0{,}1\cdot 0{,}2 + 0{,}7\cdot 0{,}4 + 0{,}2\cdot 1 = 0{,}5.$$

Видим, что вероятность поражения цели равна теперь 0,5.

Делаем третий шаг (третий выстрел). Используя (2.16.7), (*), (***), находим:

$$p_1(3) = p_1(2)P_{11} + p_2(2)P_{21} + p_3(2)P_{31} = 0{,}01\cdot 0{,}1 + 0{,}49\cdot 0 + 0{,}5\cdot 0 = 0{,}001;$$

$$p_2(3) = p_1(2)P_{12} + p_2(2)P_{22} + p_3(2)P_{32} = 0{,}01\cdot 0{,}7 + 0{,}49\cdot 0{,}6 + 0{,}5\cdot 0 = 0{,}301;$$

$$p_3(3) = p_1(2)P_{13} + p_2(2)P_{23} + p_3(2)P_{33} = 0{,}01\cdot 0{,}2 + 0{,}49\cdot 0{,}4 + 0{,}5\cdot 1 = 0{,}698.$$

Мы видим, что вероятность поражения цели равна теперь 0,698,т.е. больше 0,6. Значит, требуется сделать три выстрела, чтобы вероятность поражения цели превысила 0,6.

Контрольные задания 2.16.

1. Вероятности перехода за один шаг в цепи Маркова задаются матрицей:

а) $P = \begin{bmatrix} 0 & 0 & \frac{1}{4} & \frac{3}{4} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ \frac{1}{2} & \frac{1}{4} & 0 & \frac{1}{4} \end{bmatrix}$ б) $P = \begin{bmatrix} \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{bmatrix}.$

Найти число состояний системы. Построить граф, соответствующий матрице P.

2. Задана матрица вероятностей переходов

$$\begin{bmatrix} p & 1-p \\ 1-S & S \end{bmatrix}$$

Каковы пределы изменений p и S.

3. В урне имеется 5 белых и черных шаров. Из урны случайно извлекаются один шар, а обратно в урну возвращается один шар другого цвета. Опыт повторяется неоднократно. Найти матрицу переходных вероятностей, состояниями которой является количество белых шаров в урне. Найти вероятности перехода за два шага.

4. Игральная кость перекладывается многократно с равной вероятностью случайным образом с одной грани на любую из соседних четырех граней, независимо от исхода предыдущего испытания. К какому пределу стремится при $t \to \infty$ вероятность того, что в момент времени t игральная кость лежит на грани «5», если в момент времени t=0, она находится в этом же положении?
5. Имеется пять стульев, расположенных один после другого. Человек пересаживается с одного стула на рядом стоящий, причем эти перемещения определяются бросанием правильной игральной кости. Стулья обозначены буквами *A, B, C, D, E.* Вначале он сидит на среднем стуле *C*. Если человек сидит на крайнем стуле, то: возвращается на стул *C*, когда выпадет четное число очков; остается на том же месте при выпадении нечетного числа очков.

Если он сидит не на крайнем стуле, то: перемещается налево при выпадении одного или двух очков; перемещается направо при выпадении трех или четырех очков; остается на том же месте при выпадении пяти или шести очков. Найти: а) матрицу вероятностей переходов за один шаг; б) вероятности следующих последовательностей: *C, D, E, C, D, A, C; C, B, D, E, E, A; C, B, A, A, C, D; C, D, E ,C, E, C; A, A, C, D, E, E*.
6. Студент, для получения профессионального образования, обучается в колледже в течение трех лет. Ежегодно он сдает комплексный экзамен. Если студент успешно сдаст экзамен, то он переводится на следующий курс или заканчивает колледж с дипломом специалиста. Если студент экзамен не сдает, то он остается на соответствующем курсе второй год. Вероятность успешной сдачи экзамена на первом году обучения составляет 0,7; втором – 0,8; третьем – 0,9. а) Указать подходящее число состояний системы. б) Найти матрицу вероятностей перехода за один шаг для ежегодных передвижений студента по курсам (первый, второй, третий год обучения, окончание колледжа). в) Определить вероятность, что студент будет обучаться на третьем курсе после сдачи второго экзамена. г) Определить среднее число лет, которые студент проводит в колледже.

2.17 Индивидуальные задания к главе 2

Задания уровня сложности А.
Для всех вариантов номера задач находятся в приложении 12.
Задания уровня сложности В[1].
Для всех вариантов задания 91-100 (исходные данные находятся в приложении 13).
1. Имеется 6 ключей, из которых только один подходит к замку. Найти закон распределения СВ X, равной числу проб при открывании замка, если испробованный ключ в последующих пробах не участвует. Построить многоугольник распределения.
2. Вероятность того, что стрелок попадет в мишень при одном выстреле, равна 0,9. Стрелку выдаются патроны до тех пор, пока он не промахнется. Требуется: а) составить закон распределения

[1] Задания уровня В заимствованы из типовых расчетов Чудесенко В.Ф. [12].

дискретной случайной величины X - числа патронов, выданных стрелку; б) найти наивероятнейшее число выданных стрелку патронов.

3. НСВ задана дифференциальной функцией $f(x)$:

$$f(x)=\begin{cases}0 \text{ при } \ x<-\dfrac{\pi}{2} \text{ или } x>0,\\ \cos x \ \text{ при } -\dfrac{\pi}{2}<x<0.\end{cases}$$

а). Найти функцию распределения СВ X: $F(x)$.

б). Построить графики $F(x)$ и $f(x)$.

в). Найти вероятность попадания СВ X в интервал (-π/3;-π/4).

4. Дана интегральная функция распределения:

$$F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{t^2}{2}}dt\,.$$

Найти: дифференциальную функцию $f(x)$, $M(X)$, $\sigma(X)$, $D(X)$.

5. Из двух орудий поочередно ведется стрельба по цели до первого попадания одним из орудий. Вероятность попадания в цель первым орудием равна 0,4 вторым - 0,6. Начинает стрельбу первое орудие. Составить законы распределения дискретных случайных величин X и Y - числа израсходованных снарядов соответственно первым и вторым орудием.

6. Производится три независимых испытания, в каждом из которых вероятность появления события A равна 0,4. Составить закон распределения дискретной случайной величины X – числа появлений события A в указанных испытаниях. Найти математическое ожидание, дисперсию и среднее квадратическое отклонение X.

7. В партии из 10 деталей имеется 8 стандартных. Из этой партии наудачу взято 2 детали. Найти закон распределения случайной величины X, равный числу стандартных деталей в выборке. Построить многоугольник распределения.

8. НСВ X - задана функцией распределения $F(x)$:

$$F(x)=\begin{cases}0 & \text{при } x\le 0,\\ (1/\pi)(x-0{,}5\sin 2x) & \text{при } 0<x\le\pi,\\ 1 & \text{при } x>\pi.\end{cases}$$

а) Найти плотность вероятности СВ X.

б) Построить графики $f(x)$, $F(x)$.

в) Найти вероятность попадания СВ X в интервал (0; π/2).

9. Найти: $M(X)$ НСВ X, распределенной равномерно в интервале (2;8); функцию распределения $F(x)$ и функцию плотности вероятности f(x); вероятность попадания НСВ X в интервал (3; 6).

10. Устройство состоит из 1000 элементов, работающих независимо один от другого. Вероятность отказа любого элемента в течение времени Т равна 0,001. Найти вероятность того, что за время T откажут ровно 3 элемента. Определить закон распределения СВ Х и её числовые характеристики.

11. В коробке 7 карандашей, из которых 4 красных. Из этой коробки наудачу извлекается 3 карандаша.

а). Найти закон распределения случайной величины X равной числу красных карандашей в выборке.

б). Построить многоугольник распределения.

в). Найти вероятность события: $0<X\le 2$.

12.Станок-автомат штампует детали. Вероятность того, что изготовленная деталь окажется бракованной, равна 0,001 Найти вероятность того, что среди 250деталей окажется ровно 5 бракованных. Определить закон распределения СВ X и её числовые характеристики.

13.Устройство состоит из большего числа независимо работающих элементов с одинаковой (очень малой) вероятностью отказа каждого элемента за время *T*. Найти среднее число отказавших за время Т элементов, если вероятность того, что за это время не откажет хотя бы один элемент, равна 0,99.
14.НСВ *X* на всей числовой оси *OX* задана интегральной функцией:

$$F(x)= (1/2)+(1/\pi)arctg(x).$$

Найти вероятность того, что в результате двух испытаний случайная величина примет значение, заключенное в интервале (0;1).
15.Дана дифференциальная функция непрерывной СВ *X*:

$$f(x)=\begin{cases} 0, \ x\le\pi/6, \\ C\sin 3x, \ \pi/6<x\le\pi/3, \\ 0, \ x>\pi/3. \end{cases}$$

Найти: постоянную С, интегральную функцию *F(x).*
16.Из 25 контрольных работ, среди которых 5 оценены на «отлично», наугад извлекаются 3 работы. Найти закон распределения ДСВ *X*, если *X*-число работ оцененных на «отлично» среди извлеченных. Построить многоугольник распределения. Чему равна вероятность событий *X*>0.

17. Найти среднее число λ бракованных изделий в партии изделий, если вероятность того, что в этой партии содержится хотя бы одно бракованное изделие, равна 0,95. Предполагается, что число бракованных изделий в рассматриваемой партии распределено по закону Пуассона.
18. В урне 5 белых и 20 черных шаров. Вынули 3 шара. Случайная величина *X* - число вынутых белых шаров. Построить ряд распределения величины *X*.
19. Дискретная СВ задана законом распределения:

x_i	3	4	7	10
p_i	0,2	0,1	0,4	0,3

Найти интегральную функцию и построить ее график.
20. Дана дифференциальная функция непрерывной случайной величины *X*:

$$f(x)=\begin{cases} 0, \ x\le 1, \\ C(x^2-x), \ 1<x\le 2, \\ 0, \ x>2. \end{cases}$$

Найти: постоянную С, интегральную функцию *F(x).* Вероятность попадания СВ *X* в интервал (1/2; 3/2).
21.С вероятностью попадания при одном выстреле 0,9 охотник стреляет по дичи до первого попадания, но успевает сделать не более 4-х выстрелов. ДСВ *X*-число промахов:
а). Найти закон распределения *X*.
б). Построить многоугольник распределения.
в). Найти вероятность событий: $X<2$, $X\le 3$, $1<X\le 3$.
22. Бросают три монеты. Требуется: а) задать случайную величину *X*, равную числу выпавших "решеток"; б) построить ряд распределения.
23.НСВ X имеет плотность вероятности (закон Коши):
$f(x)= C/(1+x^2)$.
Найти: а) постоянную *C=const*;
б) функцию распределения *F(x);*
в) вероятность попадания в интервал $-1<X<1$;
г) построить графики *f(x), F(X).*
24.Найти *M(X)* и *σ(X)* НСВ, имеющей плотность вероятности:

$$f(x)= 1/(3\sqrt{2\pi})\exp(-(x+2)^2/18)$$

Указать интервал, симметричный относительно *M*(*X*) в который попадает случайная величина *X* с вероятностью р=0,9973.

25. Построить ряд распределения числа попаданий мячом в корзину при четырёх бросках, если вероятность попадания равна 0,7.
26. Два стрелка делают по одному выстрелу в мишень. Вероятность попадания
для первого стрелка при одном выстреле 0,5, для второго 0,4. ДСВ X - число попаданий в мишень.
а) Найти закон распределения X.
б) Построить многоугольник распределения.
в) Найти вероятность $X \geq 1$.
27. Из партии в 20 изделий, среди которых имеются 4 бракованных, выбраны случайным образом 3 изделия для проверки их качества. Построить ряд распределения случайного числа X бракованных изделий, содержащихся в выборке.
28.НСВ X задана функцией распределения:

$$F(x) = \begin{cases} e^x, & x < 0, \\ 1, & x > 0. \end{cases}$$

а) Найти плотность вероятности СВ X.
б) Построить графики $f(x)$, $F(x)$.
в) Найти вероятность попадания НСВ X в интервал (0; 1).
29.$M(X)$ и $\sigma(X)$ нормального распределения СВ X соответственно равны 10 и 2.
Найти вероятность того, что в результате испытания X примет значение, заключенное в интервале (12,14).
30.Случайная величина X задана функцией распределения:

$$F(x) = \begin{cases} 0, x \leq 2, \\ (x-2)^2, 2 < x \leq 3, \\ 1, \ x > 3. \end{cases}$$

а) Найти плотность вероятности СВ X.
б) Построить графики $f(x)$, $F(x)$.
в) Найти вероятность попадания НСВ X в интервал (2; 2,5).
31.Три стрелка независимо друг от друга сделали по одному выстрелу по мишени. Вероятность попадания для первого стрелка 0,9, для второго 0,8, для третьего – 0,7. Найти закон распределения величины X – числа попадания в мишень. Построить многоугольник распределения. Чему равна вероятность получения не менее двух попаданий?
32.Случайная величина X распределена равномерно интервале $(0;\pi)$. Найти закон распределения случайной величины $Y=\cos X$.
33.Случайная величина X равномерно распределена на отрезке [1;3]. Найти плотность вероятности случайной величины $Y=X^2$.
34.Дифференциальная функция НСВ X задана на всей числовой оси OX:

$$f(x)= 4C/ (1+x^2).$$

Найти постоянный параметр C.
35.НСВ X задана интегральной функцией

$$F(x)= \begin{cases} 0, \ x \leq -2, \\ (1/2)+(1/\pi) \arcsin (x/2), \ -2<x\leq 2, \\ 1, \ x>2. \end{cases}$$

Найти вероятность того, что в результате трех испытаний X примет значение в интервале (-1;1).
36.В первой урне 5 шаров – 2 белых и 3 черных. Во второй 3 шара – 1 белый и 2 черных. Из первой урны наудачу переложили во вторую 2 шара, после чего, из второй в первую переложили 1 шар. Найти закон распределения случайной величины X – числа белых шаров в первой урне, по-

сле всех перекладываний шаров. Какова вероятность того, что число белых шаров не больше, чем первоначально? Построить многоугольник распределения.
37.Случайную величину *X* умножили на *k*. Как от этого изменяются ее характеристики: 1) математическое ожидание; 2) дисперсия; 3) среднее квадратическое отклонение; 4) второй начальный момент?
38.Функция распределения случайной величины *X* задана формулой

$$F(x)=A+B\operatorname{arctg} x \ (-\infty<X+\infty).$$

Найти: а) постоянные *A* и *B*;
б) плотность вероятности *f(x);*
в) вероятность того, что величина *X* попадет в отрезок [-1; 1].
39. Случайная величина *X* задана интегральной функцией

$$F(x)=\begin{cases}0 & при\, x\le 2,\\ 0{,}5x-1 & при 2<x\le 4,\\ 1 & при\, x>4\end{cases}$$

Найти вероятность того, что в результате испытания X примет значение:
а) меньше 2, б) меньше 3, в) не меньше 3, г) не меньше 5.

40. Дана интегральная функция НСВ *X*:

$$F(x)=\begin{cases}0 & при \quad x\le 0,\\ \sin 2x & при \quad 0<x\le \dfrac{\pi}{4},\\ 1 & при \quad x>\dfrac{\pi}{4}.\end{cases}$$

Найти дифференциальную функцию и вероятность попадания СВ на интервал (π/16;π/8).
41. Вероятность изготовления стандартной детали – 0,98. Для контроля на удачу взято 100 деталей. Найти закон распределения СВ *X*, равный числу нестандартных деталей в выборке. Построить многоугольник распределения. Найти вероятность событий:
а) в выборке 2 стандартных детали;
б) в выборке более 2 стандартных деталей.
42. Найти *M*(*X*) числа лотерейных билетов, на которые выпадут выигрыши, если приобретено 50 билетов, причем вероятность выигрыша равна 0,01.

43. НСВ *X* задана дифференциальной функцией:

$$f(x)=\frac{1}{\pi\sqrt{c^2-x^2}}$$

в *интервале (-c; c), вне этого интервала f(x)=0. Найти вероятность попадания СВ X в интервал (-c*/2;c/2) и функцию распределения *F(x).*
44.НСВ *X* распределена нормально с математическим ожиданием *a*=10. Вероятность попадания СВ *X* в интервал (10; 20) равна 0,3. Чему равна вероятность попадания НСВ *X* в интервал (0; 10)?
45. Производятся 20 независимых испытаний, в каждом из которых вероятность появления успеха равна 0,2. Найти дисперсию числа появления успеха в этих испытаниях.
46.ДСВ *X* – число мальчиков в семьях с 5-тью детьми. Предполагают равновероятное рождение мальчика и девочки. Найти закон распределения случайной величины *X*. Построить многоугольник распределения.
Найти вероятность событий: а) в семье 2-3 мальчика,

б) не более 3-х мальчиков,
в) более 1 мальчика.
47. При 10 000 бросании монеты "герб" выпал 6400 раз. Следует ли считать, что монета несимметрична?
48. Устройство состоит из 10 независимо работающих элементов. Вероятность отказа каждого элемента за время t равна 0,01, Используя неравенство Чебышева, оценить вероятность того, что абсолютная величина разности между числом отказавших элементов и средним числом (математическим ожиданием) отказов за время t окажется меньше двух.
49.НСВ X задана дифференциальной функцией:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{при } x \geq 0 \; (\lambda > 0), \\ 0 & \text{при } x < 0 \end{cases}.$$

Найти вероятность того, что X примет значение, принадлежащее интервалу (2;3).
50.Случайная величина X задана дифференциальной функцией:

$$f(x) = \frac{1}{2} e^{-|x|}$$

(распределение Лапласа). Найти математическое ожидание величины X.

51–70. Случайные величины X и Y заданы законами распределений. Определить математическое ожидание, дисперсию и среднее квадратическое отклонение случайных величин X и Y. Составить законы распределений случайных величин $Z = X+Y$, $V=XY$. Построить многоугольник распределения вероятностей случайной величины Z. Найти математическое ожидание и дисперсию случайной величины $W=2X-4Y$.

51.

x_i	-1	3	4
p_i	0,2	p_2	0,6

y_j	2	5
q_j	0,4	0,6

52.

x_i	2	7	9
p_i	p_1	0,3	0,2

y_j	0	1
q_j	0,7	0,3

53.

x_i	4	6	9
p_i	0,1	0,5	p_3

y_j	3	5
q_j	0,4	0,6

54.

x_i	1	2	5
p_i	p_1	0,1	0,8

y_j	1	3
q_j	0,4	0,6

55.

x_i	-2	4
p_i	0,4	0,6

y_j	0	5	10
q_j	0,3	q_2	0,3

56.

x_i	0	5	10
p_i	0,3	0,1	p_3

y_j	-2	4
q_j	0,3	0,7

57.

x_i	*-2*	*0*	*3*
p_i	p_1	*0,5*	*0,2*

y_j	*4*	*6*
q_j	*0,5*	*0,5*

58.

x_i	-5	0	10
p_i	0,2	0,2	0,6

y_j	1	6
q_j	q_1	0,4

59.

x_i	-1	2	4
p_i	0,4	0,2	p_3

y_j	-3	1
q_j	0,4	0,6

60.

x_i	4	7	10
p_i	0,3	0,2	p_3

y_j	1	5
q_j	0,1	0,9

61.

x_i	-4	-2	1
p_i	0,1	0,6	0,3

y_j	0	4
q_j	q_1	0,2

62.

x_i	-10	-6	-1
p_i	0,4	p_2	0,2

y_j	-1	2
q_j	0,2	0,8

63.

x_i	-1	0	3
p_i	0,6	0,2	0,2

y_j	2	4
q_j	q_1	0,2

64.

x_i	-2	-1	1
p_i	0,3	0,2	p_3

y_j	4	5
q_j	0,2	0,8

65.

x_i	3	7	10
p_i	p_1	0,1	0,6

y_j	-4	4
q_j	0,3	0,7

66.

x_i	-6	-2	-1
p_i	0,2	p_2	0,2

y_j	1	4
q_j	0,2	0,8

67.

x_i	2	5
p_i	0,4	p_2

y_j	-1	3	7
q_j	0,1	0,3	0,6

68.

x_i	0	10	20
p_i	0,4	p_2	0,4

y_j	-2	-1
q_j	0,3	0,7

69.

x_i	-10	0	5
p_i	0,3	0,4	0,3

y_j	1	4
q_j	0,8	q_2

70.

x_i	-2	1
p_i	0,1	p_2

y_j	-6	-1	2
q_j	0,2	0,3	0,5

71–90 непрерывная случайная величина задана интегральной функцией (функцией распределения) $F(x)$. Найти: а) вероятность попадания случайной величины X в интервал (a, b); б) дифференциальную функцию (функцию плотности вероятностей) $f(x)$; в) математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X; г) построить графики функций $F(x)$ и $f(x)$.

71. $$F(x)=\begin{cases} 0 & при \quad x\le 0, \\ \dfrac{x^2}{\pi^2} & при \quad 0<x\le \pi, \\ 1 & при \quad x>\pi. \end{cases}$$

$a=1,\quad b=2.$

72. $$F(x)=\begin{cases} 0 & при \quad x\le 0, \\ \dfrac{64}{81}x^2 & при \quad 0<x\le \dfrac{9}{8}, \\ 1 & при \quad x>\dfrac{9}{8}. \end{cases}$$

$a=0{,}5,\quad b=0{,}9.$

73. $F(x)=\begin{cases}0 & при \quad x\le 2,\\ (x-2)^2 & при \quad 2<x\le 3,\\ 1 & при \quad x>3.\end{cases}$

$a = 2{,}5$, $b = 3$.

74. $F(x)=\begin{cases}0 & при \quad x\le 1,\\ \frac{1}{6}(x^2-x) & при \quad 1<x\le 3,\\ 1 & при \quad x>3.\end{cases}$

$a = 1$, $b = 2$.

75. $F(x)=\begin{cases}0 & при \quad x\le -\frac{\pi}{2},\\ 0{,}5\cdot(\sin x+1) & при \quad -\frac{\pi}{2}<x\le \frac{\pi}{2},\\ 1 & при \quad x>\frac{\pi}{2}.\end{cases}$

$a = 0$, $b = \frac{\pi}{6}$.

76. $F(x)=\begin{cases}0 & при \quad x\le 0,\\ \frac{16}{25}x^2 & при \quad 0<x\le \frac{5}{4},\\ 1 & при \quad x>\frac{5}{4}.\end{cases}$

$a = 0{,}5$, $b = 1$.

77. $F(x)=\begin{cases}e^x & при \quad x\le 0,\\ 1 & при \quad x>0.\end{cases}$

$a = -2$, $b = 0$.

78. $F(x)=\begin{cases}0 & при \quad x\le 1,\\ \ln x & при \quad 1<x\le e,\\ 1 & при \quad x>e.\end{cases}$

$a = 2$, $b = e$.

79. $F(x)=\begin{cases}0 & при \quad x\le 100,\\ 1-\left(\frac{100}{x}\right)^3 & при \quad x>100.\end{cases}$

$a = 110$, $b = 120$.

80. $F(x)=\begin{cases}0 & при \quad x\le 0,\\ 1-e^{-2x} & при \quad x>0.\end{cases}$

$a = 0$, $b = 2$.

81. $F(x)=\begin{cases}0 & при \quad x\le 0,\\ \frac{x^2}{e^2} & при \quad 0<x\le e,\\ 1 & при \quad x>e.\end{cases}$

$a = 1$, $b = 2$.

82. $F(x)=\begin{cases}0 & при \quad x\le 3,\\ \frac{x^4-81}{175} & при \quad 3<x\le 4,\\ 1 & при \quad x>4.\end{cases}$

$a = 3{,}2$, $b = 3{,}5$.

83. $F(x)=\begin{cases}0 & при \quad x\le 0,\\ \frac{x^4}{16} & при \quad 0<x\le 2,\\ 1 & при \quad x>2.\end{cases}$

$a = 1$, $b = 1{,}5$.

84. $F(x)=\begin{cases}0 & при \quad x\le 1,\\ \frac{(x-1)^2}{25} & при \quad 1<x\le 6,\\ 1 & при \quad x>6.\end{cases}$

$a = 2$, $b = 4$.

85. $F(x)=\begin{cases}0 & при \quad x\le -2,\\ \frac{(x+2)^3}{216} & при \quad -2<x\le 4,\\ 1 & при \quad x>4.\end{cases}$

$a = -1$, $b = 3$.

86. $F(x)=\begin{cases}0 & при \quad x\le 1,\\ \frac{x^3-x}{60} & при \quad 1<x\le 4,\\ 1 & при \quad x>4.\end{cases}$

$a = 1$, $b = 2$.

87. $F(x)=\begin{cases}0 & при \quad x\le 0, \\ \dfrac{64}{49}x^2 & при \quad 0<x\le\dfrac{7}{8}, \\ 1 & при \quad x>\dfrac{7}{8}.\end{cases}$

$a=0{,}5, \quad b=1.$

88. $F(x)=\begin{cases}0 & при \quad x\le -2, \\ \dfrac{x^3+8}{16} & при \quad -2<x\le 2, \\ 1 & при \quad x>2.\end{cases}$

$a=-1, \quad b=1.$

89. $F(x)=\begin{cases}0 & при \quad x\le 1, \\ \dfrac{x^3-x^2}{48} & при \quad 1<x\le 4, \\ 1 & при \quad x>4.\end{cases}$

$a=2, \quad b=3.$

90. $F(x)=\begin{cases}0 & при \quad x\le\sqrt{2}, \\ \dfrac{x^6-x^4-4}{96} & при \quad \sqrt{2}<x\le\sqrt{5}, \\ 1 & при \quad x<\sqrt{5}.\end{cases}$

$a=1{,}5, \quad b=2.$

91.Дана плотность распределения $f(x)$ случайной величины X. Найти параметр c, математическое ожидание $M(X)$, дисперсию $D(X)$, функцию распределения случайной величины X, вероятность выполнения неравенства $x_1<X<x_2$.

Варианты 1-8: $f(x)=\begin{cases}\dfrac{1}{c-a}, x\in[a,b], \\ 0, x\notin[a,b].\end{cases}$

Варианты 9—16: $f(x)=\begin{cases}a, x\in[c,b], \\ 0, x\notin[c,b].\end{cases}$

Варианты 17-24: $f(x)=\begin{cases}c, x\in[a,b], \\ 0, x\notin[a,b].\end{cases}$

Варианты 25—31: $f(x)=\begin{cases}a, x\in[\dfrac{b-c}{2},\dfrac{b+c}{2}], \\ 0, x\notin[\dfrac{b-c}{2},\dfrac{b+c}{2}].\end{cases}$

92. По данному закону распределения случайной величины найти характеристическую функцию g(t), математическое ожидание $M(X)$, дисперсию $D(X)$ случайной величины X.

Варианты 1-10. Биномиальный закон:

$$P(X=k)=C_n^k p^k(1-p)^{n-k},\ 0<p<1,\ k=0, 1,\ldots, n.$$

Варианты 11-20. Закон Паскаля (отрицательное биноминальное распределение):

$$P(X=k)=\frac{a^k}{(1+a)^{k+1}},\ a>0,\ k=0, 1, 2,...$$

Варианты 21-31. Закон Пуассона:

$$P(X=k)=\frac{a^k}{k!}e^{-a},\ a>0,\ k=0, 1, 2,\ldots$$

Замечание [1].

1. $g(t)=M[e^{itX}]$ - характеристическая функция случайной величины X ($i=\sqrt{-1}$ – мнимая единица).

Для ДСВ X: $g(t)=\sum_{k=1}^{n} e^{itx_k} p_k$, где $p_k=P(X=x_k)$ ($k=1,2,...,n$).

2. Характеристическая функция $g(t)$ - один из способов задания СВ, наряду с дифференциальной и интегральной функциями, которой удобнее пользоваться, чем законами распределения. Метод характеристических функций часто используется для нахождения композиции законов распределения.

3. Метод характеристических функций был создан и использован Н. М. Ляпуновым в 1900 г. для доказательства одной из наиболее общих форм центральной предельной теоремы.

4. Основные свойства характеристических функций:

а) Если СВ X и Y связаны соотношениями $Y=aX$, то $g_y(t)=g_x(at)$.

б) Характеристическая функция суммы независимых случайных величин равна произведению характеристических функций слагаемых. То есть, если даны независимые СВ $X_1, X_2,...,X_n$ с характеристическими функциями $g_{x_1}(t), g_{x_2}(t),..., g_{x_n}(t)$, то

$$\text{для СВ } Y=\sum_{\kappa=1}^{n} X_i : \ g_y(t)=\prod_{\kappa=1}^{n} g_{x_\kappa}(t)$$

5. Если $it=u$, (u – параметр, вообще говоря, комплексный), то $g_x(u)=M(e^{ux})$ – производящая функция. Наиболее важное свойство $g_x(u)$ состоит в том, что она содержит в себе все сведения о начальных моментах:

$$g_x'(0)=M(X)=\alpha_1,$$
$$g_x''(0)=M(X^2)=\alpha_2,$$
$$\ldots\ldots\ldots\ldots\ldots\ldots$$
$$g_x^{(\kappa)}(0)=M(X^\kappa)=\alpha_\kappa.$$

93. Дана плотность распределения p (x) случайной величины X. Найти плотность распределения $p(y)$, математическое ожидание $M(Y)$ и дисперсию $D(Y)$ случайной величины Y, которая представляет собой площадь одной из указанных ниже геометрических фигур.

$$\text{Варианты 1-15: } p(x)=\begin{cases}1/(b-a), x\in[a,b],\\ 0, x\notin[a,b];\end{cases}$$

в вариантах 1-5 Y - площадь равностороннего треугольника со стороной X, в вариантах 6-10 Y - площадь круга радиуса X, в вариантах 11-15 Y - площадь квадрата со стороной X.

Варианты 16-31:

$$p(x)=\begin{cases}\dfrac{2(x-a)}{(b-a)^2}, x\in[a,b],\\ 0, x\notin[a,b].\end{cases}$$

В вариантах 16-20 Y - площадь равностороннего треугольника со стороной X, в вариантах 21-25 Y - площадь круга радиуса X, в вариантах 26-31 Y - площадь квадрата со стороной X.

94. Случайная величина X имеет плотность распределения $p(x)$, указанную в задаче 93. Другая случайная величина Y связана с X функциональной зависимостью $Y=2X^m+1$. Определить математическое ожидание $M(Y)$ и дисперсию $D(Y)$ случайной величины Y.

95.Случайная величина X имеет плотность распределения вероятностей $p(x)$. Найти плотность распределения вероятностей $p(y)$ случайной величины $Y=\varphi(X)$.

96. По заданной плотности распределения $f(x_1, x_2)$ двумерной случайной величины (X_1, X_2) найти плотность распределения $f(y_1, y_2)$ двумерной случайной величины (Y_1, Y_2), связанной взаимно однозначно с (X_1, X_2) указанными ниже соотношениями.

Варианты 1—15: $f(x_1,x_2)=\frac{1}{2\pi ab}e^{-\frac{1}{2}\left(\frac{x_1^2}{a^2}+\frac{x_2^2}{b^2}\right)}$

$$X_1=aY_1\cos nY_2,\ X_2=bY_1\sin nY_2,\ 0\le Y_1<\infty,\ 0\le Y_2<\frac{2\pi}{n}.$$

Варианты 16-31: $f(x_1,x_2)=\dfrac{ab}{\pi^2(x_1^2+a^2)(x_2^2+b^2)}$;

$$X_1=\mathrm{atg}\ nY_1,\ X_2=btg\ nY_2,\ |Y_1|<\frac{\pi}{2n},\ |Y_2|<\frac{\pi}{2n}.$$

97. Двумерная случайная величина (*X, Y*) имеет равномерное распределение вероятностей в треугольной области *ABC*, т. е.

$$f(x,y)=\begin{cases}1/S,\text{если}(x,y)\in ABC,\\ 0,\text{если}(x,y)\notin ABC,\end{cases}$$

где S - площадь треугольника ABC. Определить: условные плотности распределения *f(x/y) и f(y/x);* математические ожидания *M(X), M(Y)*; дисперсии *D(X), D(Y);* коэффициент корреляции *r*. Являются ли случайные величины *X и Y* независимыми? (В исходных данных указаны декартовы координаты вершин треугольника *ABC*).

98. Используя неравенство Чебышева, оценить вероятность того, что случайная величина *X* отклонится от своего математического ожидания *M*(*X*) менее чем на $N\sigma$, где $\sigma=\sqrt{D(X)}$ - среднее квадратическое отклонение случайной величины *X; N* - номер варианта.

99. Случайная величина X_i с одинаковой вероятностью может принимать одно из двух значений: i^α или - i^α. Выяснить, удовлетворяет ли последовательность $X_1, X_2,..., X_i,...$ попарно независимых случайных величин закону больших чисел

$$\lim_{n\to\infty} P\left(\left|\frac{1}{n}\sum_{i=1}^{n} X_i-\frac{1}{n}\sum_{i=1}^{n} M(X)_i\right|<\varepsilon\right)=1,\varepsilon>0.$$

Решить задачу для двух значений параметра α: α_1 и α_2.

100. На отрезке [0, α] случайным образом выбраны n чисел, точнее, рассматриваются n независимых случайных величин $X_1, X_2,..., X_n$, равномерно распределенных на отрезке [0, α]. Найти вероятность того, что их сумма заключена между x_1 и x_2, т.е. $P\left(x_1<\sum_{i=1}^{n}X_i<x_2\right)$.

Литература
I.I. Учебники

1. Вентцель Е.С. Теория вероятностей: Учеб. для вузов. 5–е изд. стер. – М.: Высш. шк., 1998. – 576 с., ил.

Один из лучших учебников по теории вероятностей. На доступном уровне изложены основы теории вероятностей, а также важные для приложений разделы теории вероятностей: теория случайных процессов, теория информации, теория массового обслуживания.

2. Гнеденко Б.В. Курс теории вероятностей: Учебник – Изд. 6-е, перераб. и доп. – М.: Наука, 1988. – 448 с.
В настоящее время классический учебник, сочетающий большое количество примеров и исторических справок со строгим изложением на современном уровне.

3. Гмурман В.Е. Теория вероятностей и математическая статистика. Изд. 6 - е, стер. – М.: Высш. шк., 1997. – 479 с.
Учебник содержит основы теории вероятностей и математической статистики. На протяжении многих лет рекомендуется для студентов экономических специальностей вузов.

4. Феллер В. Введение в теорию вероятностей и ее приложения. В 2-х томах. Т.1: Пер. с англ.. –М.: Мир, 1984. – 528 с., ил.
Улучшению этой книги автор посвятил около четверти века, отличительной особенностью является многочисленность рассматриваемых приложений и в то же время ясное, математически строгое изложение. По полноте и широте охвата труд Феллера не имеет равных.

5. Чистяков В.П. Курс теории вероятностей. – 4-е издание – М.: Агар, 1996. – 256 с.
Дается аксиоматическое построение теории вероятностей, элементы математической статистики и теории случайных процессов. Большое количество интересных примеров и задач.

I.II. Задачники и учебные пособия

6. Агапов Г.И. Задачник по теории вероятностей: Учебное пособие для вузов. – 2-е изд., доп. – М.: Высш. шк., 1994. – 112 с.: ил.

7. Андрухаев Х. М. Сборник задач по теории вероятностей. – М.: Просвещение, 1985.- 160 с.

8. Бондаренко П.С., Кацко И.А., Чуприна И.В. Сборник задач по теории вероятностей и математической статистике. – Краснодар.: КГАУ, 1999. – 61 с.

9. Вентцель Е.С., Овчаров Л.А. Задачи и упражнения по теории вероятностей.– Учеб. пособие для втузов.- 3-е изд., стер.- М.:Высш.шк., 2000.– 366с.: ил.

10. Гмурман Д.Е. Руководство к решению задач по теории вероятностей и математической статистике: Учебное пособие. – 4-е изд., стереотип. – М.: Высш. шк., 1998. – 400 с.

11. Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей: Учеб. пособие для вузов. 2-е изд., испр. и доп.- М.: Наука.Гл. ред. физ.-мат. лит. – 1989.- 320с.

12. Чудесенко В.Ф. Сборник задач по специальным курсам высшей математики (типовые расчеты):Учебное пособие для вузов.– М.: Высш. шк., 1983.- 242 с.

13. Сборник задач по теории вероятностей, математической статистике и теории случайных функций, под редакцией А.А.Свешникова. Изд. «Наука», Главная редакция физико-математической литературы, 1970.- 656 с.

14. Теория статистики с основами теории вероятностей: Учеб. пособие для вузов/И.И. Елисеева, В.С. Князевский, Л.И. Ниворожкина, З.А. Морозова; Под ред. И,И, Елисеевой. – М.: ЮНИТИ-ДАНА, 2001. – 446 с.

2.18.Приложения теории вероятностей в компьютерных науках (computer science)[1]

АЛГОРИТМ ЧТЕНИЯ ЛЕКЦИИ

ШАГ1. Прочитайте введение без эпиграфа (и оцените необходимость применения теории вероятностей в компьютерных науках…)
ШАГ2. Перейдите к 1.
ШАГ3. Перейдите к примерам 1,2, если непонятно (или неинтересно), вернитесь к шагу 2.
ЩАГ4. Если на шагах 1-2 произошло «зацикливание» прочитайте эпиграф к введению и оставьте примеры 1,2 – главное, Вы поняли, что такие задачи есть и в принципе они разрешимы, просто этим нужно заниматься (например, Дональд Кнут занимается этим уже 40 лет). Перейдите к.2.
ШАГ5. Если 2 неинтересен, перейдите к 3-5, в противном случае продолжить и перейти к .3-5.
ШАГ6. Если п. 3-5 неинтересен – закончить чтение. В противном случае, окинув взором вопросы 1, 2, 3, 4, 5 Вы придете к заключению о важности и применимости теории вероятностей в компьютерных науках.

[Lector]

Лектор пользуется известными привилегиями,
в пользе которых, надеюсь, нет никаких оснований
сомневаться. Так, встретив у меня непонятное
место, читатель должен предположить, что под ним
кроется нечто весьма полезное и глубокомысленное
(1704г., Дж. Свифт (в интерпретации Lectora))

Приложения теории вероятностей в компьютерных науках.

Говорят, что XXI век – век генетики, информатики и т.д. Не секрет, что наша страна утратила большинство из ведущих позиций в мире за исключением космических технологий (наши спутники все еще остаются лучшими, за счет результатов отечественной науки). Профессиональные пользователи – математики, физики и астрофизики, генетики и метеорологи, военные и контрразведчики давно пришли к выводу, что вычислительные возможности современного компьютера ограничены, все чаще они сталкиваются с задачами, для решения которых требуется больше ресурсов, чем имеется в наличии на всем земном шаре. Как панацея от неотвратимо надвигающегося мрака вычислительной беспомощности, на пороге XXI века появились слова «су-

[1] Читается для студентов, обучающихся по специальности «Прикладная информатика»

перкомпьютер» и «petaflops» (петафлопс – миллион миллиардов операций с плавающей запятой в секунду).

Задачи, эффективное решение которых под силу исключительно суперкомпьютеру с производительностью порядка одного петафлопса, распадаются на два класса: задачи с преобладанием целочисленных вычислений и задачи с преобладанием вычислений с плавающей запятой. В каждом классе, в свою очередь, легко выделить подклассы военно-прикладных и научно-практических приложений. К первому классу относятся криптография (например, взламывание кодов) и создание полноценного искусственного интеллекта, ко второму – моделирование ядерных взрывов, долгосрочный прогноз погоды и вычислительные задачи гидродинамики, оптимизация металлургических процессов (в частности производства алюминия…) и т.д.

В связи с указанными выше проблемами возникает ряд задач, которые требуют оценки времени работы алгоритмов и поиска путей их ускорения на базе компьютеров со стандартной архитектурой, что дает неограниченное поле деятельности для приложения теории вероятностей…

1. Производящие функции.

Производящая функция является устройством, отчасти напоминающим мешок. Вместо того чтобы нести много предметов, что могло бы оказаться затруднительным, мы собираем их вместе, и тогда нам нужно нести лишь один предмет-мешок.

Д. Пойа

1). Пусть имеется некоторая (в общем случае бесконечная) последовательность $\{a_0, a_{1,}a_{2,\dots,}a_{n,\dots}\} = \{a_n\}$.Эту последовательность часто удобно выразить в виде степенного ряда по степеням вспомогательной переменной Z:

$$\text{A(Z)} = a_0 + a_1Z + a_2Z^2 + a_3Z^3 + \ldots = \sum_{k\geq 0} a_k Z^k. \qquad (2.18.1)$$

Теперь можно заняться изучением свойств функции A (Z), которая представляет всю последовательность. Это особенно важно, если последовательность получена по индукции с использованием рекуррентных выражений.

A (Z) называется производящей функцией для последовательности $\{a_n\}$.

Замечание 1. В настоящее время самый мощный метод работы с последовательностями чисел - это преобразование бесконечных рядов (то есть производящих функций), которые «порождают» эти последовательности. Впервые метод производящих функций для решения линейных рекуррентных соотношений ввел в начале 18 в Де Муавр, затем в 1730г. Дж. Стирлинг применил производящие функции для решения более сложных задач и показал, как применять при этом дифференцирование и интегрирование. Дальнейшее развитие метод производящих функций нашел в работах Л. Эйлера (1741-1750гг.) и Пьера Лапласа (1812.). Самое интересное заключается в том, что большинство операций, выполняемых над производящими функциями, можно обосновать, не затрагивая вопроса о сходимости соответствующего бесконечного ряда (в свое время это был один из опорных пунктов в критике Л.Эйлера).

Пусть для последовательности$\{b_0, b_{1,}b_{2,\dots,}b_{m,\dots}\} = \{b_m\}$ производящая функция

$$B\ (Z) = b_0 + b_1 Z + b_2 Z^2 + \ b_3 Z^3 + \ldots = \sum_{m \geq 0} b_m Z^m. \qquad (2.18.2)$$

Тогда рассмотрим произведение производящих функций A(Z)B(Z)=C(Z).
A(Z)B(Z)=$a_0 b_0 + (a_0 b_1 + a_1 b_0)Z + (a_0 b_2 + a_1 b_1 + a_2 b_0)Z^2 + . \ldots,$ (2.18.3)
следовательно,

$$\text{A(Z)B(Z)} = \ \sum_{k \geq 0} \cdot \sum_{m \geq 0} a_k\, Z^k b_m Z^m = C(Z) = \ c_0 + c_1 Z + \ c_2 Z^2 + \ldots \qquad (2.18.4)$$

Приравняем коэффициенты приZ^nи получим формулу

$$c_n = \ \sum_{k=0}^{n} a_k b_{n-k}. \qquad (2.18.5)$$

Последовательность, полученная по правилу (2.17.5) называется сверткой последовательностей $\{a_n\}$ и $\{b_m\}$.
Пример1. Пусть φ_1и φ_2две производящие функции:

$$\varphi_1 (1 + \mathrm{Z})^{\mathrm{m}} = \sum_{\mathrm{k} \geq 0} \mathrm{C}_{\mathrm{m}}^{\mathrm{k}} \mathrm{Z}^{\mathrm{k}},$$

$$\varphi_2(Z) = (1 + Z)^i = \sum_{k \geq 0} C_i^k Z^k.$$

Рассмотрим их произведение

$$\varphi_1(Z)\varphi_2(Z) = (1 + Z)^m (1 + Z)^i = (1 + Z)^{m+i}$$

и приравняем коэффициенты при Z^n:

$$\sum_{k=0}^{n} C_m^k C_i^{n-k} = C_{m+i}^n. \qquad (2.18.6)$$

Полученная формула носит название сверки Вандермонда(Александр Вандермонд написал по этому поводу в 1772г. статью, хотя формула была известна еще в 1303г. Чжу Ши-Цзе из Китая). Из сверки Вандермонда легко получить целый ряд частных случаев, например, пусть m=i=n, тогда

$$\sum_{k=0}^{n} C_n^k C_n^{n-k} = C_{2n}^n,$$

но по свойству биноминальных коэффициентов $C_n^k = C_n^{n-k}$, следовательно, имеем

$$\sum_{k=0}^{n} (C_n^k)^2 = C_{2n}^n. \qquad (2.18.7)$$

Пример2. Два стрелка сделали по n выстрелов, по разным мишеням. Какова вероятность одинакового числа попаданий, если вероятность попадания каждого стрелка равна 0,5.

Решение. Пусть ДСВ X_1- число попаданий первым стрелком в цель, ДСВ X_2- число попаданий вторым стрелком в цель, P_k-вероятность того, что оба стрелка попали k-раз. Тогда учитывая, что вероятность k попаданий из n находится по формуле Бернулли, получим, что

$$P_k = P(X_1 = k)P(X_2 = k) = C_n^k p^k q^{n-k} \cdot C_n^k p^k q^{n-k}.$$

По условию p=q= $\frac{1}{2}$,

следовательно, $P_k = P(X_1 = k)P(X_2 = k)$=$C_n^k (\frac{1}{2})^n \cdot C_n^k (\frac{1}{2})^n = (\frac{1}{4})^n (C_n^k)^2$.

Суммируя P_k от нуля до n получим вероятность одинакового числа попаданий:

$$\text{P}=\sum_{k=0}^n P_k = \sum_{k=0}^n (\tfrac{1}{4})^n (C_n^k)^2 = (\tfrac{1}{4})^n \sum_{k=0}^n (C_n^k)^2.$$

Учитывая формулу (5) получим выражение для вероятности одинакового числа попаданий в замкнутом виде(т.е. не в виде суммы, а в виде конечной формулы)

$$P=(\tfrac{1}{4})^n \sum_{k=0}^n (C_n^k)^2 = (\tfrac{1}{4})^n C_{2n}^n.$$

Ответ: $(\frac{1}{4})^n C_{2n}^n$.

2). Одним из важнейших в дискретной математике является понятие *производящей функции* неотрицательной целочисленной *случайной величины* X (ПФСВ) с вероятностями $P_0, P_1, P_2, \dots, P_k \dots$.

Если СВ *X* принимает только целые неотрицательные значения, то распределение ее вероятностей можно представить как степенной ряд (в частном случае многочлен) по степеням Z:

$$\varphi(Z) = \sum_{k\geq 0} P(X = k) \cdot Z^k = M(Z^x), \text{где } P(X = k) = P_k.$$

$\varphi(Z)$- производящая функция неотрицательной целочисленной случайной величины *X*.

Коэффициенты φ неотрицательны и их сумма равна 1($\sum_{k\geq 0} P(X = k) = 1$), следовательно, $\varphi(1) = 1$.

И обратно, любой степенной ряд $\varphi(Z)$с неотрицательными коэффициентами и свойством $\varphi(1) = 1$ является ПФСВ. Он содержит всю информацию о СВХ.

*Свойства производящей функции неотрицательной целочисленной случайной величин*ы-$\varphi_x(Z)$:

1) $M(x) = \sum_{k\geq 0} P(X = k) \cdot k \cdot Z^{k-1} = \varphi_x(1)$ (2.18.8)

2) $M(x^2) = \sum_{k\geq 0} P(X = k) \cdot k(k-1) \cdot Z^{k-2} + \sum_{k\geq 0} P(X = k) \cdot kZ^{k-1} = \varphi''_x(1) + +\varphi'_x(1)$, следовательно,

3) $D(x) = \varphi''_x(1) + \varphi'_x(1) - [\varphi'_x(1)]^2$. (2.18.9)

4) Если СВ *X* и *Y* независимы и принимают только целые неотрицательные значения, то распределение их суммы будет определяться сверсткой последовательностей

$P(X+Y=n)=\sum_k P(X=k, Y=n-k)=\sum_k P(X=k)P(Y=n-k)$ – формула полной вероятности.
Свертке этих последовательностей отвечает произведение производящих функций

$$\varphi_{X+Y}(Z)=\varphi_x(Z)\cdot\varphi_Y(Z).$$

Замечание. а) Из свойства 4) следует, что, для поиска числовых характеристик результата суммы независимых случайных величин $Y=X_1+X_2+X_3+\cdots X_n$ необходимо представить производящую функцию СВ Zв виде произведения производящих функций слагаемых $X_1, X_2, X_3, .. X_n$.

$$\varphi_y=\prod_{i=1}^{n}\varphi_{xi}(Z) \text{ и найти } \varphi'_y(1),$$

или найдя, с помощью производящих функций числовые характеристики $X_1, X_2, X_3, .. X_n$ (свойства 1, 2) легко перейти к числовым характеристикам их суммы, по свойствам математического ожидания и дисперсии, то есть просто сложить числовые характеристики СВ $X_1, X_2, X_3, .. X_n$.

б) Математическое ожидание и дисперсия- это лишь первые члены из бесконечного ряда так называемых кумулянтов (или семиинвариантов), введенных в 1903 г. датским астрономом Т.Н. Тиеле :

$$k_{1,}\, k_{2,} \dots;$$

$$k_1=M(X), k_2=D(X);$$

кумулянты более высоких порядков выражают более тонкие свойствараспределения.
Если $\varphi(Z)-$ ПФСВ, то

$$\ln\varphi(e^t)=\frac{k_1}{1!}t+\frac{k_2}{2!}t^2+\frac{k_3}{3!}t^3+\frac{k_4}{4!}t^4+\dots$$

в) ПФСВ для часто встречающихся законов распределения ДСВ:

1) Закон распределения Бернулли: $\varphi(Z)=q+PZ$
2) Биномиальный закон распределения: $\varphi(Z)=(q+PZ)^n$
3) Геометрический закон распределения: $\varphi(Z)=\sum_{m\geq0}Pq^m Z^m=\frac{P}{1-qZ}$
4) Геометрический закон сдвинутый на единицу: $\varphi(Z)=PZ+Pq^1Z^2+\cdots=$
$$=\frac{PZ}{1-qZ}$$
5) Закон отрицательного биномиального распределения:
$$\varphi(Z)=(P/(1-qZ))^k=\sum_m C^m{}_{k+m-1}P^k\cdot q^mZ^m$$
6) Закон Пуассона: $\varphi(Z)=\sum_{k\geq0}\frac{\lambda^k}{k!}e^{-}Z^k=e^{(Z-1)}$
7) Гипергеометрический закон: $\varphi(Z)=\sum_k\frac{C_M^kC_{N-M}^{n-k}}{C_N^n}Z^k.$

Пример 3. Игрок поочередно покупает билеты двух разных лотерей до первого выигрыша. Вероятность выигрыша по одному билету первой лотереи составляет 0,2, а второй 0,3. Игрок вначале покупает билет первой лотереи. Составить закон распределения и найти математическое ожидание случайной величины X- числа купленных билетов, если он имеет возможность купить: а) только 5 билетов; б) неограниченное число билетов.

Решение. Пусть событие A_i– выигрыш i-го билета первой из лотерей, B_i- выигрыш i-го билета второй из лотерей. По условию известно, что P(A_i)=0,2 и P(B_i)=0,3.

Рассмотрим ДСВ *X*- число билетов купленных до выигрыша: *X*={1,2,3,4,5}.

а) Представим события *X=k* (*k*=1,2,3,4,5) и вероятности в виде таблицы:

X	1	2	3	4	5
Событие *X=k*	A_1	$\bar{A}_1 B_1$	$\bar{A}_1\bar{B}_1A_2$	$\bar{A}_1\bar{B}_1\bar{A}_2B_2$	$\bar{A}_1\bar{B}_1\bar{A}_2\bar{B}_2A_3 + \bar{A}_1\bar{B}_1\bar{A}_2\bar{B}_2\bar{A}_3$
Формула Вероятности: *P(X=k)*	0,2	0,8·0,3	0,8·0,7·0,2	$0{,}8^2\cdot 0{,}7\cdot 0{,}3$	$0{,}8^2\cdot 0{,}7^2\cdot 0{,}2 + 0{,}8^3\cdot 0{,}7^2$
Значение вероятности: *P(X=k*)	0,2	0,24	0,112	0,1344	0,3136

M(*X*)=1·0,2+2·0,24+3·0,112+4·0,1344+5·0,3136=3,1216- математическое ожидание числа купленных билетов до первого выигрыша.

б) Рассмотрим ПФСВ *X*- числа купленных билетов до 1 выигрыша при возможности покупки неограниченного числа билетов. Учитывая результаты подсчета вероятностей в пункте а), легко получить производящую функцию СВ *X*. Она будет состоять из двух сумм $\sum_{k\geq 0} P(x=2k+1)Z^{2k+1}$ и $\sum_{k\geq 1}(x=2k)2^{2k}$ или P(x=2k+1)=$0{,}8^k\cdot 0{,}7^k\cdot 0{,}2$; P(x=2k)=$0{,}8^k\cdot 0{,}7^{k-1}\cdot 0{,}7$, соответствующих выигрышу на нечетном (X=2k+1) или четном (X=2k) билете:

$$\varphi_x(Z) = \sum_{k\geq 0} P(X=2k+1)Z^{2k+1} ++ \sum_{k\geq 1} P(X=2k)Z^{2k} =$$

$$= \sum_{k\geq 0} 0{,}8^k\cdot 0{,}7^k\cdot 0{,}2\cdot Z^{2k+1} + \sum_{k\geq 1} 0{,}8^k\cdot 0{,}7^{k-1}\cdot 0{,}3\cdot Z^{2k} =$$

$$=0{,}2Z\cdot \sum_{k\geq 0}(0{,}56\cdot Z^2)^k + \frac{3}{7}\sum_{k\geq 1}(0{,}56\cdot Z^2)^k = 0{,}2Z\cdot\frac{1}{1-0{,}56Z^2} + \frac{3}{7}\cdot\frac{0{,}56Z^2}{1-0{,}56^2} =$$

$$= \frac{0{,}24Z^2+0{,}2Z}{1-0{,}56Z^2}.$$

Таким образом, ПСФВ *X*- числа купленных билетов до первого выигрыша имеет вид $\varphi_x(Z) = \frac{0{,}24Z^2+0{,}2Z}{1-0{,}56Z^2}$. Найдем математическое ожидание, пользуясь свойством 1) производящей функции:

$$\varphi_x'(Z) = \frac{(0{,}48Z+0{,}2)\cdot(1-0{,}56Z^2) + 2\cdot 0{,}56Z(0{,}24Z^2+0{,}2Z)}{(1-0{,}56Z^2)^2}$$

M(*X*)=$\varphi'(1) = \frac{0{,}68\cdot 0{,}44+1{,}12\cdot 0{,}44}{0{,}44^2} = \frac{0{,}68+1{,}12}{0{,}44} = \frac{1{,}8}{0{,}44} = \frac{180}{44} = \frac{45}{11} \approx 4{,}091$

Ответ: а) 3,1216; б) 4,091.

2. Вероятностный анализ скорости выполнения алгоритмов.

Алгоритм – это последовательность инструкций для выполнения отдельного задания. При этом предполагается, что субъект, выполняющий алгоритм, следует инструкциям и умеет их выполнять.

Компьютер имеет ограниченный словарь (язык программирования), поэтому для написания алгоритмов используют формализованный стиль.

300 лет до н.э. Евклид изложил алгоритм решения целого ряда геометрических задач. Сначала излагается «словарь» - неопределяемые понятия и аксиомы, а затем на их основе построены алгоритмы решения сложных задач (алгоритмы деления угла (отрезка) в данном отношении, равенства (подобия) фигур и т.д.). Формализованные алгоритмы такого типа обычно используются для проверки истинности определённых положений или возможности выполнения каких-либо действий, скорость же работы алгоритма не важна.

При решении реальных задач на компьютере, например, сортировки записей о миллионе покупателей, эффективность становится основным критерием оценки алгоритма. Оценка сложности алгоритмов складывается из:

а) скорости алгоритма (например: 1 алгоритм 1000 записей сортирует за 1 с.; 1 млн. – 10 с, 2 алгоритм 1000 записей сортирует за 2 с.; 1 млн. – 5 с.) т.е. 1 алгоритм лучше для малых списков, 2 алгоритм лучше для больших списков;

б) требований к размеру памяти…(от быстрого алгоритма нет толка, если мало памяти);

в) свободному месту на диске.

Производительность алгоритма можно оценивать по порядку величины N – размерности исходных данных. Алгоритм имеет сложность порядка $O(f(N))$, если с увеличением размерности исходных данных N время выполнения алгоритма растет пропорционально функции $f(N)$.

Например, для алгоритма содержащего вложенный цикл:

```
……………
            For I=1 moN
For J=1 moN
…………..
Next J
Next I.
…………….
```

сложность алгоритма будет порядка $O(N^2)$.

Часто встречающиеся функции оценки порядка сложности алгоритма $(c = const, c > 1)$:

а) работают с достаточной скоростью:

$$f(N) = C,$$

$$f(N) = \log_2(\log_2 N),$$

$$f(N) = \log_2 N,$$

$$f(N) = NC\ ,$$
$$f(N) = N\ ,$$
$$f(N) = N^c;$$

б) пригодны только для решения задач с небольшими N:

$$f(N) = C^N,$$
$$f(N) = N!$$

Время выполнения сложных алгоритмов на компьютере со скоростью 1млн. операций в секунду (1MFLOPS).

	N=10	N=20	N=30	N=40	N=50
N^3	0,001с	0,008с	0,027с	0,064с	0,125с
2^N	0,001с	1,05с	17,9 мин.	1,27 дня	35,7 лет
3^N	0,059с	58,1 мин.	6,53 года	3,68 10^5 лет	8,20 10^{10} лет
$N!$	3,63с	7,71 10^4 лет	8,41 10^{18} лет	2,59 10^{34} лет	9,64 10^{50} лет

Для решения задачи $O(N!)$ при $N = 24$ требуется больше времени, чем существует вселенная!!!

Поэтому алгоритмы со сложностью порядка $O(C^N)$ *и* $O(N!)$ *пригодны только при малых значениях* N. *В этом случае существует два пути: 1) совершенствование аппаратной базы – совершенствование стандартных компьютеров с архитектурой фон Неймана или создание принципиально новых компьютеров с параллельной обработкой информации, 2) разработка рандомизированных алгоритмов содержащих элемент случайности (см. ниже «хеширование»).*

Анализ сложности алгоритмов – необходимая часть разработки приложения, часто это достигается тестированием (с использованием генератора случайных чисел). Важным фактором также является частота обращения к файлу подкачки, т.е. необходимо экономно расходовать оперативную память (например, задавая тип переменных).

При решении практических задач возникают следующие задачи:

1. *Задача построения больших массивов данных и их переупорядочивания* (базы данных, списки, бинарные и другие деревья, связи, структуры).

2. *Рекурсия* – вызов функции или подпрограммы самой себя.

3. *Моделирование реальных задач с помощью дерева решений*. Поиск наилучшего решения соответствует поиску наилучшего пути на дереве (метод полного перебора, ветвей и границ, эвристические методы, случайный поиск и последовательное приближение).

4. *Сортировка* (вставкой, выбором, пузырьковая, слиянием, пирамидальная, подсчетом, большая сортировка).

5. *Поиск* (полный перебор, двоичный поиск, интерполяционный поиск).

Сложность алгоритма оценивается по затратам времени и ресурсов, обычно находится компромисс между ними. Проведем анализ следующего алгоритма.

Пример 2.16.1. *Алгоритм М* (*нахождение максимума*).

Для данных n элементов $x[1], x[2], \ldots, x[n]$ необходимо найти такие величины m и j, что $m = x(j) = \max\limits_{1 \le i \le n} x[i]$, где j – наибольший индекс, удовлетворяющий этому соотношению (символ $\leftarrow$ означает операцию присвоения, $j \leftarrow n$ означает, что переменной j присвоено значение n).

M 1. (Инициализация). Положим $j \leftarrow n,\ k \leftarrow n-1,\ m \leftarrow x[n]$.
(Во время выполнения алгоритма будем иметь $m = [j] = \max\limits_{k \le i \le n} x[i]$).

M 2. (Все проверено?) Если k=0, то работа алгоритма заканчивается.

M 3. (Сравнение). Если $x[k] \le m$, перейти к шагу М 5.

M 4. (Замена m). Положим $j \leftarrow k,\ m \leftarrow x[k]$. (Это значение m является новым текущим максимумом).

M 5. (Уменьшение k). Уменьшим k на единицу и вернемся к шагу М 2.

Для выполнения алгоритма *M* требуется фиксированный объем памяти, поэтому будем анализировать время, необходимое для его выполнения. Для этого подсчитаем, сколько раз выполняется каждый шаг.

№ шага	Количество выполнений
M 1	1
M 2	n
M 3	n-1
M 4	A
M 5	n-1

Зная сколько раз выполняется каждый шаг можно оценить время выполнения алгоритма на конкретном компьютере.

В приведенной таблице значение A неизвестно (A определяет, сколько раз необходимо изменить значение текущего max). Для проведения анализа требуется оценить значение A и его σ(А).

$A \to \min$ (для оптимистов) $= 0 : x[n] = \max\limits_{1 \le k \le n} x[k]$,

$A \to \max$ (для пессимистов) $= (n-1) : x[1] > x[2] > \ldots > x[n]$,

$A \to M(X)$ к среднему – для специалистов по теории вероятностей и математической статистике.

Пусть все значения $x[k]$ - различны и все значения их n ! перестановок равновероятны.

Например, если $n = 3$, то следующие 6 возможностей равновероятны

Ситуация	Значение A
$x[1] < x[2] < x[3]$	0
$x[1] < x[3] < x[2]$	1
$x[2] < x[1] < x[3]$	0
$x[2] < x[3] < x[1]$	1
$x[3] < x[1] < x[2]$	1
$x[3] < x[2] < x[1]$	2

$$\overline{A} = \frac{0+1+0+1+1+2}{6} = \frac{5}{6}.$$

Вероятность того, что А имеет значение k *, равна*

$$P_n(A=k) = \frac{\textit{число перестанов ок, для которых } A=k}{n!}$$

при $n=3$: $\quad P_3(A=0)=\frac{2}{6}=\frac{1}{3},\ P_3(A=1)=\frac{3}{6}=\frac{1}{2},\ P_3(A=2)=\frac{1}{6}.$

Чтобы определить поведение А найдем вероятности $P_n(A=k)$ по индукции. Обозначим число перестановок, для которых $A=k$ через P_{nk}. Оно равно $P_{nk}=n!\cdot P_n(A=k)$.

Рассмотрим перестановки $x_1, x_2, \ldots x_n$ элементов $\{1, 2, \ldots, n\}$. Если $x_1=n$, то значение A на единицу больше, чем значение для перестановки $x_2, x_3, \ldots x_n$. Если $x_1 \neq n$, то значение A точно такое, как для перестановки $x_2, x_3, \ldots x_n$. Следовательно, по индукции получим, что

$$P_{nk} = P_{(n-1)(k-1)} + (n-1)\cdot P_{(n-1)k},$$

$$P_n(A=k) = \frac{P_{n-1}(A=k-1)}{n} + \frac{n-1}{n}\cdot P_{n-1}(A=k).$$

Пусть $G_n(Z)=\sum_k P_n(A=k)Z^k - \textit{ПФСВ}$, тогда

$$G_n(Z) = \frac{Z}{n}G_{n-1}(Z) + \frac{n-1}{n}G_{n-1}(Z) = \frac{Z+n-1}{n}G_{n-1}(Z).$$

$$G_n'(Z) = \frac{1}{n}G_{n-1}(Z) + \frac{Z+n-1}{n}G_{n-1}'(Z).$$

$$G_n'(1) = \frac{1}{n} + G_{n-1}'(1).$$

Отсюда найдем, что математическое ожидание A – это сумма ряда:

$M(A)=G_n'(1)=\frac{1}{n}+\frac{1}{n-1}+\frac{1}{n-2}+\ldots+\frac{1}{3}+\frac{1}{2}+\frac{1}{1}-1$, учитывая, что $G_1'(1)=0$.

Это и есть среднее число выполнений шага M 4.

При $n\to\infty$, $\quad M(A)\approx \ln n$.

6. *Хеширование.* Ряд важных алгоритмов хранения и выборки информации на ЭВМ основаны на методе, который называется «хеширование». «Анализ алгоритмов» - раздел информатики, занимающийся выводом количественной информации об эффективности компьютерных методов, его основателем является известный американский математик Д. Кнут, который занимается им с 1962г. «Вероятностный анализ алгоритмов» - это изучение времени работы алгоритмов, рассматриваемого как случайная величина, зависящая от предполагаемых характеристик исходных данных. Хеширование особенно подходит для вероятностного анализа, т.к. метод хеширования исключительно эффективен в среднем, хотя наихудший вариант ужасен, когда все ключи имеют одинаковый хеш-код. Общая задача заключается в хранении некоторого множества записей, каждая из которых содержит значения «ключа» K и некоторые данные об этом ключе $D(K)$. Цель находить $D(K)$ по заданному K.

Хеширование реализует подход аналогичный интерполяционному поиску, создающему функцию соответствия между искомым значением и индексом позиции, где он должен находиться.

Алгоритм хеширования использует *функцию, определяющую вероятное положение элемента в таблице* на основе значения искомого элемента. Хеш-таблицы обычно используют, если требуется часто вставлять или удалять элементы.

1. Пусть требуется заполнить несколько значений с ключом от 1 до 100, тогда формируют массив со 100 ячейками и устанавливают соответствие.
2. Реально диапазон ключа выше. Например, при использовании в качестве ключа идентификационные номера социального страхования, состоящего из 9 цифр (1 млрд.). Если одна запись 1 Кб., то весь массив – 1 млн. Мб (1 Гигабайт). При штате менее 10 млн., 99 % массива не востребовано (пустые значения).
3. Поэтому схемы хеширования, отображают большое количество возможных ключей, отображают на достаточно компактную хеш-таблицу. Однако, если номер социального страхования 1 млрд. записей, то для таблицы с 1000 позиций, в среднем одной ячейке будет соответствовать млн. записей!!!

Поэтому для решения этой проблемы схема хеширования предполагает наличие алгоритма разрешения конфликтов.

Для реализации хеширования необходимо:

а) Хеш-таблица для хранения данных;

б) функция хеширования (элемент$\leftrightarrow$место);

с) алгоритм решения конфликтов, определяющих последовательность действий, если несколько ключей соответствуют одной ячейке таблицы.

Общая задача: хранение некоторого множества записей, каждая из которых содержит значение «ключа» K и некоторые данные $D(K)$ об этом ключе.

Одно из решений. Хранятся две таблицы, $KEY(j)$ – для ключей и $DATA\ (j)$ для данных$1\leq j\leq N$, где N - общее число записей, которые могут быть размещены, n – обозначает фактическое число записей.

Поиск ключа K можно осуществить последовательно просматривая таблицу.

S1. Установить j=1 (уже просмотрены позиции $< j$).

S2. Если $j > n$, остановиться (безуспешный поиск).

S3. Если $KEY(j) = K$, остановиться (успешный поиск).

S4. Увеличить j на 1 и вернуться к шагу 2 (следующая попытка).

Метод работает правильно, но его работа может быть ужасающе медленной; при безуспешном поиске шаг $S2$ приходится повторять (n+1) раз, а n может быть большим.

Замечание. 1. Обычно заранее не известно какие ключи будут в таблице, но часто существует возможность выбора хеш-функции h, чтобы значения $h(K)$ можно было считать СВ равномерно распределенной в интервале от 1 до m и независимой от хеш-кодов других присутствующих ключей. В этом случае вычисления хеш-функции подобно бросанию кубика с m гранями.

2. Все записи могут попасть в один список, точно так же, как на кубике может выпадать цифра 6. Однако закон больших чисел говорит о том, что списки почти всегда будут хорошо сбалансированы.

Анализ хеширования.

Хеширование было придумано для ускорения поиска в базах и хранилищах данных. Например, используется m отдельных списков вместо одного огромного.

«Хеш-функция» преобразует любой возможный ключ K в номер списка $h(K)$, лежащий в диапазоне от 1 до m.
Используются вспомогательные таблицы:
1) *First* [i] – содержит для каждого i, $1 \le i \le m$ указатель на первый элемент в списке i.
2) *NEXT* [j], $1 \le j \le N$, указывает на запись, следующую за записью в списке, которому эта запись принадлежит
First [i] = -1, если список i пуст; *NEXT* [j] = 0, если j – последняя запись в своём списке; n – общее количество записей. Переменная n содержит информацию о количестве записей.

Пример 2.16.2. Пусть ключи – имена, и имеется m=4 списка, разделяемые по первой букве имени:

$$h(\text{имя}) = \begin{cases} 1 & \text{при} \quad А - Ж, \\ 2 & \text{при} \quad З - О, \\ 3 & \text{при} \quad П - Ц, \\ 4 & \text{при} \quad Ч - Я. \end{cases}$$

Вначале имеется 4 пустых списка и n=0. Если ключ первой записи будет, скажем, имя Даниил, то $h(\text{Даниил}) = 1$ и поэтому «Даниил» станет ключом первого элемента в списке 1. Если следующими именами окажутся Инна и Оксана, то они попадут в список 2.

Таблица памяти выглядит так:
First [1]=1, *First* [2]=2, *First* [3]= -1, *First* [4]= -1
Ключи: *KEY* [1]= Даниил, *NEXT* [1]=0,
KEY [2]= Инна, *NEXT* [2]=2,
KEY [3]= Оксана, *NEXT* [3]=0, n=3.
(Значения *DATA* (K) содержат секретную информацию и здесь не приводятся). После вставки 16 имен списки могли бы содержать следующие записи:

Список 1	Список 2	Список 3	Список 4
Даниил	Инна	Петр	Эльвира
Елена	Оксана	Роман	Юлия
Василий	Ирина	Светлана	
Владимир	Ольга	Тамара	
	Кристина		
	Леонид		

В массиве *KEY* те же имена записаны вперемешку, но значения *NEXT* позволяют разделить списки. Если нужно найти имя Кристина, то нужно просмотреть 6 имен списка 2, но это не идет в сравнение с просмотром 16 имен.

Алгоритм поиска ключа в соответствии с описанной схемой:
Н 1. Установить i:=h(K) и j:=*FIRST*[i].
Н 2. Если $j \le 0$, остановиться (безуспешный поиск). (*)
Н 3. Если *KEY* [j]=K, остановиться (успешный поиск).
Н 4. Установить i:=j, затем j:=*NEXT*[i] и вернуться к шагу Н2.

Например, при поиске имени Кристина установим на шаге *Н*1 i:=2, j:=2; на шаге *Н*3 найдем, что Инна ≠ Кристина; на шаге *Н*4 установим j: =3 и на шаге *Н*3 найдем, что Оксана ≠ Кристина; выполнив шаги *Н*4 и *Н*3 еще 3 раза мы найдем Кристину в таблице.

После успешного поиска данные *D*(*K*) содержатся в *DATA* [*j*].

Рассмотрим *вероятностный анализ хеширования*. Пусть *r* – число «проб» в таблице, или иначе – количество выполнений шага *Н*3 при работе алгоритма (*).

Шаг	Безуспешный поиск	Успешный поиск
1	1	1
2	*r+1*	*r*
3	*r*	*r*
4	*r*	*r-1*

Таким образом, главная характеристика, определяющая время работы процедуры поиска, есть число проб *r*. Например, существует некоторая книга в которой на каждой странице 1 запись, на обложке – номер станицы первой записи в каждом из *m* списков; каждому ключу *К* соответствует список *h*(*K*), которому тот принадлежит. На каждой странице книги содержится ссылка на следующую страницу в том списке, к которому эта страница относится. Число проб – это число страниц, которые придется просмотреть.

Случай 1. Ключ отсутствует.

Безуспешный поиск требует по одной пробе для каждого элемента списка h_{n+1}. Вероятность того, что $h_j = h_{n+1}$ равна 1/m для $1 \le j \le N$:

$$P(h_j = h_{n+1}) = \frac{1}{m}.$$

Математическое ожидание числа «проб r»:

$$M(r) = \frac{n}{m},$$

т.е. среднее число проб в m раз меньше, чем без хеширования.

Процесс поиска можно описать биноминальным распределением:

$p = \dfrac{1}{m}, \quad q = 1 - p = \dfrac{m-1}{m}, \quad npq = \dfrac{n(m-1)}{m^2}$ - дисперсия.

$P(Z) = \left(\dfrac{m-1+Z}{m}\right)^n$ - производящая функция общего числа проб в безуспешном поиске.

Итак, анализ алгоритма в случае безуспешного поиска дает следующие результаты:

min r = 0 (минимальное число проб r = 0);

max r = n (максимальное число проб r = n – числу записей в списке);

$M(r) = \dfrac{n}{m}$ (математическое ожидание числа проб);

$D(r) = \dfrac{n(m-1)}{m^2}$ (дисперсия числа проб);

$$\text{при } m \to \infty; \ Д(r) \approx \frac{n}{m}, \ \sigma(r) = \sqrt{n/m}.$$

Случай 2. Ключ присутствует.

Пусть S_j – вероятность того, что ищется j-ый ключ. Тогда можно показать [4], что

$G_r(Z)=\left(\dfrac{m-1+Z}{m}\right)\sum_{\kappa=1}^{n} S_\kappa Z^\kappa - ПФСВ$ числа проб в случае успешного поиска.

$$M(r)=\frac{n-1}{2m}+1, \text{ если } m\approx\frac{n}{\ln n}, \quad (n\to\infty), \text{ то } M(r)=\frac{1}{2}\ln n .$$

$$\sigma(r)=\ln n/\sqrt{12}.$$

В обоих случаях проведенный анализ позволяет спать спокойно пессимистам, которые опасаются наихудшего случая. Из неравенства Чебышева следует, что списки будут хорошими, за исключением крайне редких случаев, для любого ε >0:

1) $P(\left|r-\dfrac{n}{m}\right|<\varepsilon)\geq 1-\dfrac{n(n-1)}{m^2\varepsilon^2}$,

2) $P(\left|r-\dfrac{1}{2}\ln n\right|<\varepsilon)\geq 1-\dfrac{(\ln n)^2}{12\varepsilon^2}$.

3. Случайные числа, генераторы случайных чисел

Defindit numerus (В числах ты найдёшь покой) – это истина дураков;
Deperdit numerus (В числах ты найдёшь погибель) – истина мудрых.
[1820г.,Ч.Колтон]

Числа, которые выбираются случайным образом, находят множество полезных применений.

1. **Моделирование**. (Имитационное моделирование). Компьютер используется для моделирования естественных явлений или состояния различных систем, когда на вход подаются случайные числа, подчиняющиеся тому или иному закону распределения и изучаются параметры выходных переменных. Например, исследования в ядерной физике (первое применение случайных чисел на компьютере связано с именем Д. фон Неймана, который в годы второй мировой войны предложил использовать их при исследовании проблемы создания атомного оружия); теории массового обслуживания и т.д. Классическим примером является модель «ядерной зимы», полученная в 70-е годы на ВЦ АН СССР под руководством академика Н.Н. Моисеева. В современном пакете расширений Matlab – Simylink имеется прекрасная возможность построения имитационных моделей в режиме языка визуального программирования.

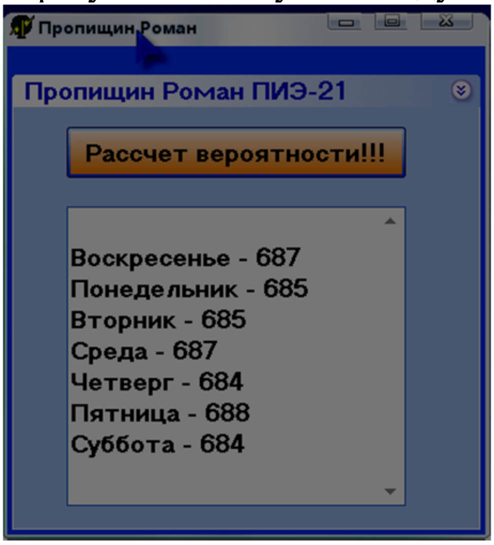
Рассмотрим задачу 12 из п. 1.3. Вероятность «черной пятницы». Доказать, что 13 число месяца с большей вероятностью приходится на пятницу, чем на другие дни недели.

Решение. Известно, что календарь повторяется с периодичностью 400 лет, учитывая это, студент Р. Пропищин предложил программу на языке turbo pascal:

```
procedure TForm1.PDJXPButton1Click(Sender: TObject);
//  В любых идущих подряд 400 годов
// имеется 400/4-4+1 високосных годов, т.е. 97
// значит дней будет 365*400+97=146097
// достаточно вычислить вероятность попадания 13 числа
// на пятницу в любые 400 лет идущие подряд
```

```
var
// номер дня недели (от 1 до 7)
Den:byte;
// переменная в которую сохраняется 13 + месяц + год
DatePR:TDate;
// считаем в массиве выпадание 13 на день недели
mas:array[1..7] of integer;
// используем для цикла
i:integer;
// год, месяц
year,month:word;
begin
// обнуляем счетчики
for i:=1 to 7 do mas[i]:=0;
// цикл по году (400 лет)
for year:=3000 to 3399 do
// цикл по месяцу
for month:=1 to 12 do
begin
// заносим в переменную дату
DatePR:=EncodeDate(year, month, 13);
// функция DayOfWeek возвращает от 1 (воскресенье) до 7 (суббота)
Den:=DayOfWeek(DatePR);
// увеличиваем счетчик
inc(mas[Den]);
end;
// выводим данные
Memo1.Lines.Add('Воскресенье'+' - '+inttostr(mas[1]));
Memo1.Lines.Add('Понедельник'+' - '+inttostr(mas[2]));
Memo1.Lines.Add('Вторник'+' - '+inttostr(mas[3]));
Memo1.Lines.Add('Среда'+' - '+inttostr(mas[4]));
Memo1.Lines.Add('Четверг'+' - '+inttostr(mas[5]));
Memo1.Lines.Add('Пятница'+' - '+inttostr(mas[6]));
Memo1.Lines.Add('Суббота'+' - '+inttostr(mas[7]));
end;
```

В результате получим следующие результаты:



Таким образом, частота попадания 13 числа на пятницу является наибольшей и равна 688, следовательно, 13 число месяца с большей вероятностью приходится на пятницу, чем на другие дни недели

2. **Выборочный метод**. Часто невозможно или нецелесообразно исследовать всю совокупность данных, например, базу данных сети предприятий или информацию в сети интернет. Но случайный отбор элементов совокупности позволяет изучить её «типичные свойства».

3.**Численный анализ**. Для решения сложных задач не всегда подходят точные методы, которые часто и разработать не возможно. Поэтому используют различные приближенные методы. Причем наиболее эффективными из них являются методы, использующие случайные числа.

Например, в пакете анализа Excel, опция **генерация случайных чисел** заполняет диапазон случайными числами, заданными по одному из законов: равномерному; нормальному; Бернулли (СВ X принимает значение 1 с вероятностью p и 0 с вероятностью (1-p) (индикаторная СВ)); биномиальному; Пуассона; модельному (позволяющему генерировать последовательности случайных чисел от a до b с шагом c, и возможностью повторения каждого числа и последовательности); дискретному (решающему задачу, получения по имеющемуся распределению, новых значений того же распределения).

Замечание. Инструмент генерации случайных чисел позволяет решать целый ряд задач: численных методов (например, приближённого вычисления определённых интегралов методом статистических испытаний - методом Монте-Карло, имитационного моделирования изучаемых процессов и т.д.).

4. **Компьютерное программирование**. С помощью случайных чисел часто тестируют эффективность компьютерных алгоритмов, а также создают рандомизированные (случайные) алгоритмы (например, хеширование), которые часто превосходят свои детерминированные аналоги.

5. **Принятие решений**. Иногда важно принять полностью беспристрастное решение, тогда говорят, что некоторые преподаватели кафедры статистики бросают монеты или игральные кости…

Сслучайность – важная часть оптимальных стратегий в теории матричных игр.

6. **Эстетика**. Небольшая добавка случайности в живопись, музыку часто их оживляет.

7. **Развлечения**. Люди любят проводить время, играя в карты, нарды, вращая рулетку и т.д. Такие традиционные способы использования случайных чисел получили название метод Монте-Карло. Это общее название всех алгоритмов, использующих случайные числа.

Каковы источники случайных чисел?

1) Вытаскивание шаров из урны, бросание костей, вращение рулетки и т.д.

2) Существуют *таблицы случайных чисел*. Первая, содержащая 40 000 значений взятых наудачу из отчетов о переписи опубликована в 1927 г. Типпетом.

3) *Механические генераторы случайных чисел.* Первая такая машина была использована в 1939 г. Кендаллом и Бабингтоном-Смитом для построения таблицы, содержащей 100 000 случайных цифр.

4) *Встроенные программы*. Первая запущенная в 1951 г. использовала резисторный генератор шума, поставляющий 20 случайных битов на сумматор (Тьюринг). В 1955 г. были опубликованы таблицы с 1 000 000 случайных чисел.

Использование таблиц было ограничено, но в 90-е годы интерес к ним вернулся. Д. Марсалья в 1995 году подготовил демо-диск с 650 Мбайт случайных чисел, при генерировании которых запись шума диодной цепи сочеталась с музыкой в стиле «рэп» (он назвал это «черным и белым шумом»).

Несовершенство первых механических методов побудило интерес к получению случайных чисел с помощью арифметических операций, заложенных в компьютер.

Джон фон Нейман в 1946 году первым предложил такой алгоритм. Идея – возводим в квадрат k-значное число и берем среднее k цифр и т.д.

Например, для десятизначного числа 5772156649;
возводим в квадрат 33317792380594909201;
значит следующее число 7923805949.
Однако! При небольших значениях k алгоритм части приводит к циклу (!), т.е. числа не являются случайными! «Каждый кто использует арифметические методы генерации случайных чисел, безусловно, грешит» – Дж. фон Нейман. Действительно, если их можно рассчитать, то они не случайны. Поэтому генераторы случайных чисел часто называют генераторами «псевдослучайных» чисел. В языке программирования ПАСКАЛЬ в основе генераторов случайных чисел лежат встроенные функции RANDOM и RANDOMIZE генерирующие равномерно распределенные числа на [0;1]. На их основе строятся случайные числа, подчиняющиеся другим законам распределения. Специалисты рекомендуют использовать для сравнения разные источники случайных чисел – это будет указывать на стабильность результатов. По высказыванию Дж. Морсалья «генератор случайных чисел похож на sex. Если он хорош, то это прекрасно. Если плох, то все равно приятно».

4. Вероятностный подход к понятию информации.

Что наша жизнь(?) — ИГРА…!!!???

На учетно-финансовом факультете КГАУ, благодаря Н.В. Чуприна, известна игра, когда один из участников отгадывает, что загадал второй. Вопрос допускает только два ответа «да» и «нет» (да-1, нет-0). Игра называется Бар-Кохба. Согласно легенде, Бар-Кохба («сын звезды») был предводителем восстания 135 г. в Иудее против владычества римлян. Превосходящее по силам войско осадило крепость, которую оборонял гарнизон под командованием Бар-Кохбы. Согласно легенде он послал в стан врага лазутчика, которого римляне схватили и отрезали язык. Несчастный бежал из стана врагов, пришел к Бар-Кохбе, но не смог рассказать, что увидел у противника. Тогда Бар-Кохба стал задавать ему вопросы, на которые можно было ответить только либо «да» либо «нет» - и получил необходимую информацию.

Таким образом, в известной мере, игру Бар-Кохба нужно считать предтечей теории информации. Видимо, ещё в древности была известна возможность закодировать любую информацию в виде последовательности двух символов «0» и «1».

Записав последовательность полученных ответов в виде «0» и «1», мы получим число в двоичной системе исчисления.

1 бит (bit – binary digit – двоичная единица) – единица информации, которую можно закодировать «0» и «1».

С примерами кодирования и декодирования информации мы встречаемся постоянно.

Например, пишем $\rightarrow$ кодируем, читаем $\rightarrow$ декодируем, радио-сигнал - кодирование $\rightarrow$ антенна $\rightarrow$ изображение на экране телевизора – декодирование;

пластинки – механическое кодирование,

кассеты – магнитное кодирование,

лазерные диски – оптическое кодирование.

Рассмотрим количество информации на примере игры Бар-Кохба. В списке 32 человека. Сколько вопросов нужно задать, чтобы отгадать человека, которого задумали? Разделим список на две части по 16 человек, зададим вопрос – в какой части загаданный человек? и т.д. 5 раз:

$$\begin{array}{ccccccccccc} 32 & \rightarrow & 16 & \rightarrow & 8 & \rightarrow & 4 & \rightarrow & 2 & \rightarrow & 1 \\ & 1в & & 2в & & 3в & & 4в & & 5в & \end{array}$$

Таким образом, необходимо задать 5 вопросов или получить информацию пять бит. А если список из N = 48 человек, то 1 из них отгадать можно $[\log_2 N]+1$ вопросами. Действительно, если N = 64, то нужно 6 бит информации (6 вопросов).

Имеем при N = 48:

$$\log_2 32 < \log_2 N < \log_2 64,$$

$$5 < \log_2 48 < 6.$$

Определение 1. Если в заданном множестве *H*, содержащем *N* элементов выделен элемент *X* ($X \in H$), то чтобы найти *X*, необходимо получить количество информации, равное

$$H(X) = \log_2 N \quad \text{битам.}$$

Это формула Хартли.

Определение 2. Если не все элементы из *N* равновероятны:

X_1	X_2	...	X_n
P_1	P_2	...	P_n

то количество информации, необходимое для поиска элемента *X* равно

$$H(X) = p_1 \log_2 \frac{1}{p_1} + p_2 \log_2 \frac{1}{p_2} + ... + p_n \log_2 \frac{1}{p_n}$$

Это формула Шеннона.

(Партнер загадывает x_i с вероятностями p_i и сыграно много партий)

при $p_1 = p_2 = ... = p_n = \dfrac{1}{N}$, $H(X) = \log_2 N$ - формула Хартли.

Например, бросаем две монеты. СВ *X* – число орлов $X = \{0, 1, 2\}$.

X:	0	1	2
P:	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

Среднее число знаков для составления кода одного бросания по формуле Шеннона

$$H(X)=\frac{1}{4}\log_2\frac{1}{\frac{1}{4}}+\frac{1}{2}\log_2\frac{1}{\frac{1}{2}}+\frac{1}{4}\log_2\frac{1}{\frac{1}{4}}=\frac{1}{4}\cdot 2+\frac{1}{2}+\frac{1}{4}\cdot 2=\frac{3}{2}=1{,}5.$$

Действительно: $X = 0$: 00_2 (2 цифры)

$X = 1$: 1_2 (1 цифра)

$X = 2$: 10_2 (2 цифры).

$$M(X)=2\cdot\frac{1}{4}+1\cdot\frac{1}{2}+2\cdot\frac{1}{4}=1{,}5$$ -

математическое ожидание длины кодового слова.

В переводе на язык игры Бар-Кохба нам следует спросить у партнёра – не выпал ли у него один герб? В случае утвердительного ответа мы будем знать все, в противном случае следует задать вопрос – не выпало ли у него 2 герба? Таким образом, в игре Бар-Кохба по возможности следует задавать вопросы так, чтобы утвердительные и отрицательные ответы были равновероятны (или близки к ним).

Формулу, названную впоследствии формулой Шеннона, в XIX веке вывел Больцман для решения другой задачи. Он показал, что если в газе, состоящем из множества молекул, вероятности состояний соседних из них $p_1, p_2, \dots p_n$, то энтропия системы определяется так

$$H=c\left(p_1\ln\frac{1}{p_1}+p_2\ln\frac{1}{p_2}+\dots+p_n\ln\frac{1}{p_n}\right),\text{ где } c=const.$$

Энтропия служит мерой неупорядоченности системы или иначе − это *мера неопределенности* системы. Таким образом, неопределенность – это недостача информации или отрицательная информация.

Замечание. 1. Можно оценить среднее количество вопросов, необходимых для ответа. Пусть каждый человек обладает словарным запасом из n слов, тогда он сможет разбить свой словарный запас на m групп так, что вероятности нахождения нужного (ключевого слова) для каждой группы одинаковы и равны

$$p=\frac{1}{m},$$

то вероятность неудачи

$$q=1-p=\frac{m-1}{m}.$$

Таким образом, процесс поиска можно описать биноминальным распределением.

I случай – Хеширование – безуспешный поиск. Итак, если X – число вопросов для нахождения ответа, то

$\min\ x=0$ (сразу отгадал, не задавая вопросы)

$\max x=n$ (практически $n\to\infty\dots$)

$$M(x)=\frac{n}{m},$$

$$D(x)=\frac{n(m-1)}{m^2}$$ - дисперсия числа вопросов.

$$\text{при } m \to \infty\text{: } D(x) = \frac{n}{m},\ \sigma(x) = \sqrt{n/m}.$$

II случай – Хеширование успешный поиск.
В «базе данных» студента есть информация о ключе (или иначе о заданном вопросе), тогда

$$\min\ x = 0,$$

$$\max x = n,$$

$$M(x) = \frac{n-1}{2m} + 1,$$

$$\text{если } m \approx \frac{n}{\ln n} \text{ при } n \to \infty\text{, то } M(x) = \frac{1}{2}\ln n \text{ и } \sigma(x) = \ln n/\sqrt{12},$$

что в какой-то степени ускоряет процесс поиска ответа.
2. *Задача для исследования*. Оцените возможное значение n и приведите численную оценку среднего количества вопросов и времени, если современные исследования психологов показали, что человек может держать в сознании (памяти) 7 ± 2 чанка (бита) информации.

5. Байесовские сети

Теорема умножения предполагает, что

$$P(AB) = P\ (A/B)P(B) = P(B/A)P(A)$$

Отсюда

$$P(A/B) = \frac{P\ (B/A)P(A)}{P(B)}$$

это известная формула Байеса.

Используя теоретико-множественный смысл события можно записать

$$B = (B \cap A) \cup (B \cap \overline{A})$$

тогда в силу независимости

$$P(B) = P(B \cap A) + P\ (B\ \cap \bar{A})$$

или

$$P(B) = P(B/A)P(A) + P(B/\bar{A})P(\bar{A})$$

Теперь формулу Байеса можно переписать в виде

$$P(A/B) = \frac{P(B/A)P(A)}{P(B/A)P(A) + P\left(B/\overline{A}\right)P(\overline{A})}$$

В современных информационных технологиях эта формула используется как основа для управления неопределенностью – «делать выводы вперед и назад».

Основной современного логического вывода считается «если..., то...» правила. В нашем случае «если событие *H* истинно, то событие *E* будет наблюдаться с вероятностью *P*», а если событие *E* уже произошло, то какова вероятность истинности *H*?

H – событие, заключающееся в том, что данная гипотеза верна

E – событие, заключающееся в том, что наступило определенное доказательство (свидетельство), которое может подтвердить правильность указанной гипотезы.

Формула Байеса примет вид

$$P(H/E) = \frac{P(E/H)P(H)}{P(E/H)P(H) + P(E/\bar{H})P(\bar{H})}$$

Она устанавливает связь гипотезы H и свидетельства E и в то же время свидетельства с неподтвержденной гипотезой. $P(H)$ – априорная вероятность гипотезы, известная до наступления события E.

Предполагается, что $P(H)$ и $P(E/H)$ находятся опытным или экспериментальным путем.

Формулу Байеса обобщают на случай множества гипотез ($H_1, H_2, \ldots H_m$) и множества свидетельств ($E_1, E_2, \ldots E_n$).

Вероятности каждой из гипотез можно определить по формуле

$$P(H_i \,/\, E_1E_2 \ldots E_n) = \frac{P(E_1E_2 \ldots E_n/\, H_i\,)P(H_i\,)}{\sum_{k=1}^{m} P(E_1E_2 \ldots E_n/\, H_k\,)P(H_k)}$$

$i = 1, m$

Сложность формулы заключается в необходимости знать все условные вероятности знаменателя, поэтому часто делается довольно сильное предположение о независимости свидетельств (подход называют наивный Байес – naïve Bayes). Тогда формула приобретает вид

$$P(H_i \,/\, E_1E_2 \ldots E_n) = \frac{P(E_1E_2 \ldots E_n/\, H_i\,)P(H_i\,)}{\sum_{k=1}^{m} P(E_1/\, H_i\,)P(E_2/\, H_i\,) \ldots P(E_n/\, H_k\,)P(H_k)}$$

Пример1. Имеется три взаимно независимые фирмы. Гипотезы

H_1 – «средняя надежность фирмы»

H_2 – «высокая надежность фирмы»

H_3 – «низкая надежность фирмы».

Имеется два условно независимых свидетельства, подтверждающих в разной степени исходные гипотезы.

P(i)	1	2	3
$P(H_I)$	0,6	0,4	0,1
$P(E_I/H_I)$	0,3	0,7	0,2
$P(E_2/H_2)$	0,6	0,8	0,0

Условно независимые свидетельства, поддерживающие исходные гипотезы:

E_1 – «наличие прибыли у фирмы»,

E_2 – «своевременный расчет с бюджетом».

Новые факты, получаемые в процессе сбора, будут повышать или понижать вероятности гипотез.

Пусть с вероятностью 1 наступило событие E_2, тогда апостериорные вероятности для гипотез согласно формуле Байеса для одного свидетельства:

$$P(H_i \,/\, E_2) = \frac{P(E_2/\, H_i\,)P(H_i\,)}{\sum_{k=1}^{3} P(E_2/\, H_k\,)P(H_k)}$$

i = 1,2,3

Имеем $P(H_1/\, E_2) = \frac{0{,}6 \cdot 0{,}6}{0{,}6 \cdot 0{,}6 + 0{,}4 \cdot 0{,}8 + 0{,}1 \cdot 0{,}0} = \frac{0{,}36}{0{,}68} = 0{,}53$

$$P(H_2 \,/\, E_2) = \frac{0{,}4 \cdot 0{,}8}{0{,}6 * 0{,}6 + 0{,}4 \cdot 0{,}8 + 0{,}1 \cdot 0{,}0} = \frac{0{,}32}{0{,}68} = 0{,}47$$

$$P(H_3 \,/\, E_2) = \frac{0{,}1 * 0{,}0}{0{,}6 \cdot 0{,}6 + 0{,}4 \cdot 0{,}8 + 0{,}1 \cdot 0{,}0} = 0$$

После того как событие E_2 произошло доверие к гипотезе H_1 и H_3 понизилось, а доверие к H_2 возросло. Если есть факты, подтверждающие и событие E_1, и событие E_2, то при условии их независимости формула Байеса будет выглядеть в следующем виде:

$$P(H_i \,/\, E_1 E_2) = \frac{P(E_1/\, H_i)P(E_2/\, H_i)P(H_i)}{\sum_{k=1}^{3} P(E_1/\, H_k)P(E_2/\, H_k)P(H_k)}$$

$i = 1, 2, 3.$

Таким образом

$$P(H_1 \,/\, E_1 E_2) = \frac{0{,}3 \cdot 0{,}6 \cdot 0{,}6}{0{,}3 \cdot 0{,}6 \cdot 0{,}6 + 0{,}7 \cdot 0{,}8 \cdot 0{,}4 + 0{,}2 \cdot 0{,}0 \cdot 0{,}1} = \frac{0{,}108}{0{,}332} = 0{,}325$$

$$P(H_2 \,/\, E_1 E_2) = \frac{0{,}7 \cdot 0{,}8 \cdot 0{,}4}{0{,}3 \cdot 0{,}6 \cdot 0{,}6 + 0{,}7 \cdot 0{,}8 \cdot 0{,}4 + 0{,}2 \cdot 0{,}0 \cdot 0{,}1} = \frac{0{,}224}{0{,}332} = 0{,}675$$

$$P(H_3 \,/\, E_1 E_2) = \frac{0{,}2 \cdot 0{,}0 \cdot 0{,}1}{0{,}3 \cdot 0{,}6 \cdot 0{,}6 + 0{,}7 \cdot 0{,}8 \cdot 0{,}4 + 0{,}2 \cdot 0{,}0 \cdot 0{,}1} = 0.$$

После получения свидетельств E_1 и E_2 осталось только две гипотеза H_1 и H_2. при этом H_2 более вероятно, чем H_1.

Данный пример иллюстрирует процесс распространения вероятностей по элементам экспертной системы (ЭС), основанной на Байесовских сетях, при поступлении в нее новых свидетельств. Можно показать что последовательное поступление свидетельств приводит к результатам аналогичным применению формулы Байеса для одновременно поступающих свидетельств.

Байесовская сеть доверия (БСД) – это направленный ациклический граф со следующими свойствами:

- каждая вершина – событие, описывающееся случайной величиной,
- вершины, связанные с «родительскими» определяются таблицей или функцией условных вероятностей,
- вероятности состояний вершин без «родителей» являются безусловными.

Таким образом, в байесовских сетях доверия вершины – случайные величины, дуги – вероятностные зависимости, определяющиеся таблицей или функцией условных вероятностей.

Пример 2. *Построение байесовской сети доверия*.

Фирма обанкротилась. Основные факторы: надежность фирмы (обобщенный внутренний фактор), экономический кризис (обобщенный внешний фактор) (Рис.31). Рассмотрим ситуацию в которой первая вершина «надежность» –«*reliability*» имеет два состояния «высокая» - «*high*», «не высокая» – «*not high*», а вторая «кризис» - «*crisis*» два состояния кризис «повлиял» - «*affected*», «не повлиял» - «*not affected*».

Вершины надежность и экономический кризис не имеют родительских вершин, поэтому они являются маргинальными, то есть ни от чего не зависят (рис.32).

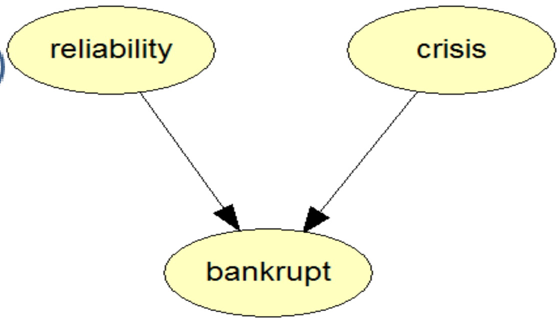


Рис.31 Рис.32

Приведенные таблицы показывают вероятности пребывания вершины «обанкротилась» - *bankrupt* в определенном состоянии, обусловленном состоянием родительских вершин.

априорная вероятность *P* («*reliability*»)

affected	0.1
not affected	0.9

априорная вероятность *P* («*crisis*»)

high	0.1
not high	0.9

условные вероятности *P*(*bankrupt/crisis, reliability*)

crisis	affected		not affected	
reliability	high	not high	high	not high
bankrupot	0.95	0.85	0.9	0.02
not bankrot	0.05	0.15	0.1	0.98

Современные программные средства (Netica, Hugin и др.) позволяют строить более сложные байесовские сети доверия, вводить новые свидетельства и получать решения (выводы) за счет пересчета новых вероятностей в вершинах, соответствующих новым свидетельствам рис.33). *P*(фирма= «*bankrupt*»)=0,1832.

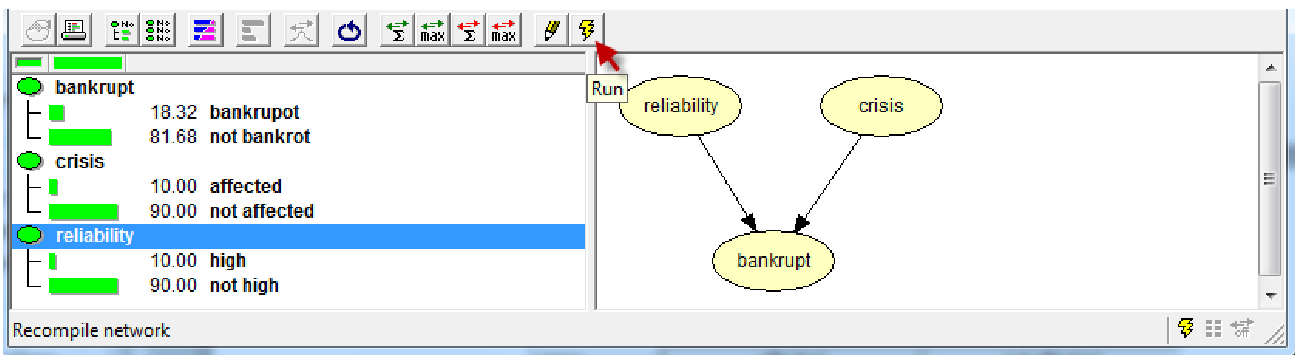


Рис.33

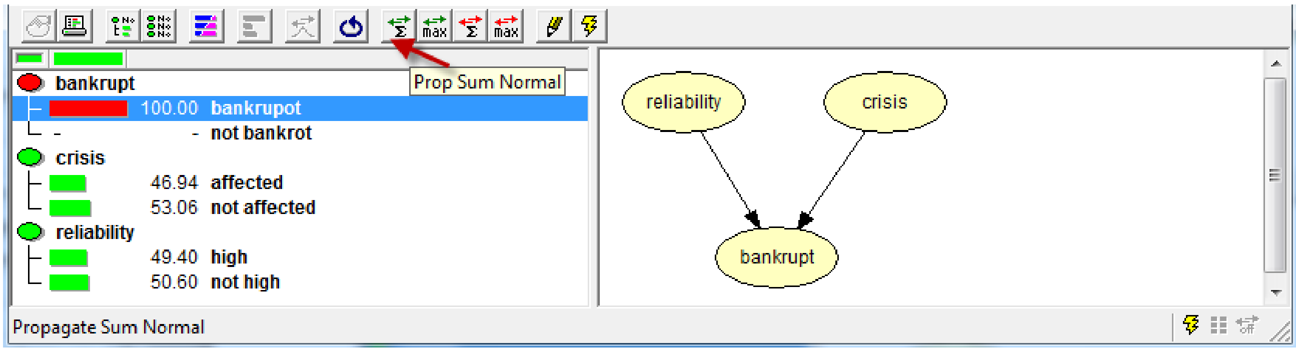


Рис.34.

В приведенном выше примере – *пусть известно, что фирма обанкротилась* (*вероятностью 1 или на 100%*). После этого можно узнать вероятность того, что повлиял экономический кризис. Для приведенных выше данных, опираясь на результаты вывода путем распространения сумм по БСД, можно показать, что кризис повлиял с вероятностью 0,469, а надежность фирмы с вероятностью 0,494.

Можно предположить, что на банкротство фирмы с вероятностью 0,531 не повлияет кризис и с вероятностью 0,506 не повлияет надежность фирмы. Однако этот вывод является преждевременным. Для нахождения наиболее вероятной комбинации состояния вершин в программе *Netica* вместо распространения сумм нужно использовать распространение максимумов. Каждое из состояний вершин, имеющее значение 100.00 будет принадлежать к наиболее вероятной комбинации состояний. В рассматриваемом примере наиболее вероятно, что кризис не повлияет, а надежность фирмы будет высокая.

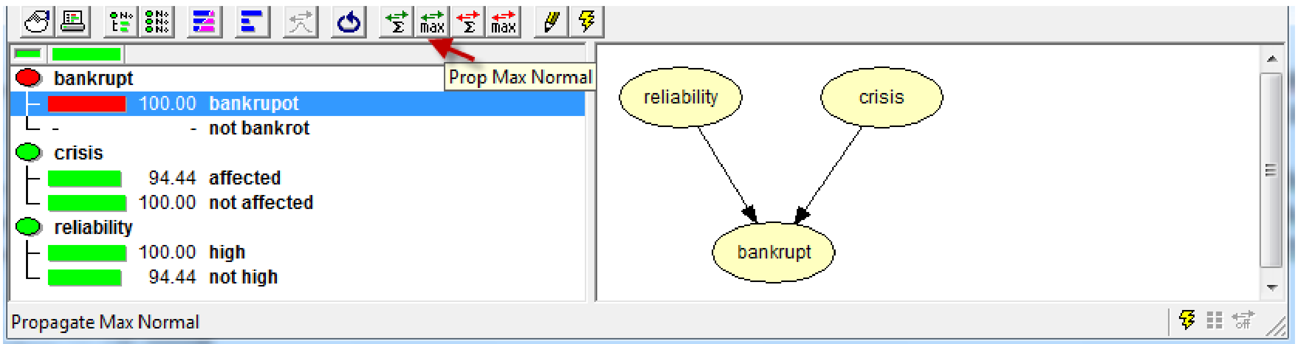


Рис. 35

Рассмотрим наиболее вероятную комбинацию состояний, полученных при наличии факта, что фирма обанкротилась.

P(*reliability = «high», crisis = «not affected» / bankrupt= «bankrupt»*)

Использем формулу $P(A\,|B) = P(AB)/P(B)$. В наших обозначениях

P(*reliability = «high», crisis = «not affected» / bankrupt*)=

=P(*reliability = «high», crisis = «not affected», bankrupt= «bankrupt»)/* P(*bankrupt= «bankrupt»)=0,081/0,0,1832=0,442.*

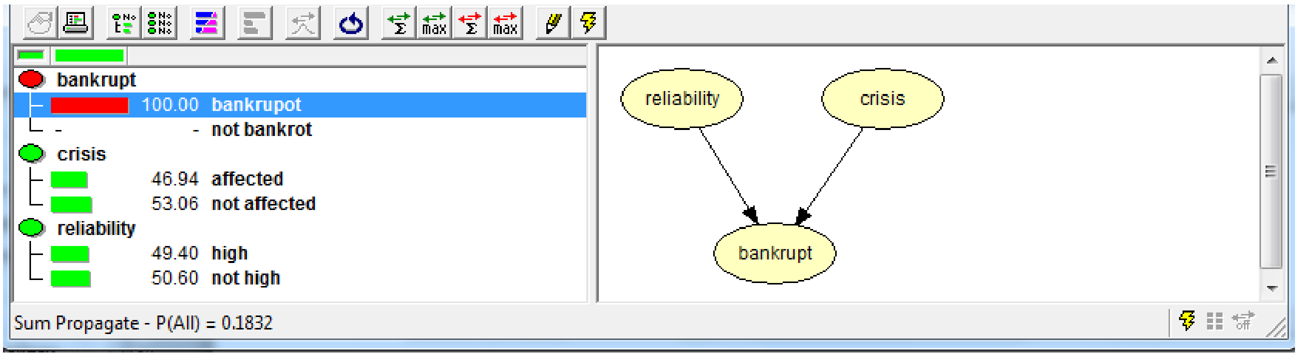


Рис. 36

Итак, наиболее вероятная ситуация, что кризис не повлиял и надежность фирмы была высокой имеет значение вероятности 0,442 (рис. 36-37). При разработке систем принятия решений и экспертных систем используют диаграммы влияния, к БСД добавляют вершины решения - *decisions* (прямоугольники) и вершины пользы – *utility* (ромбы) (рис.38).

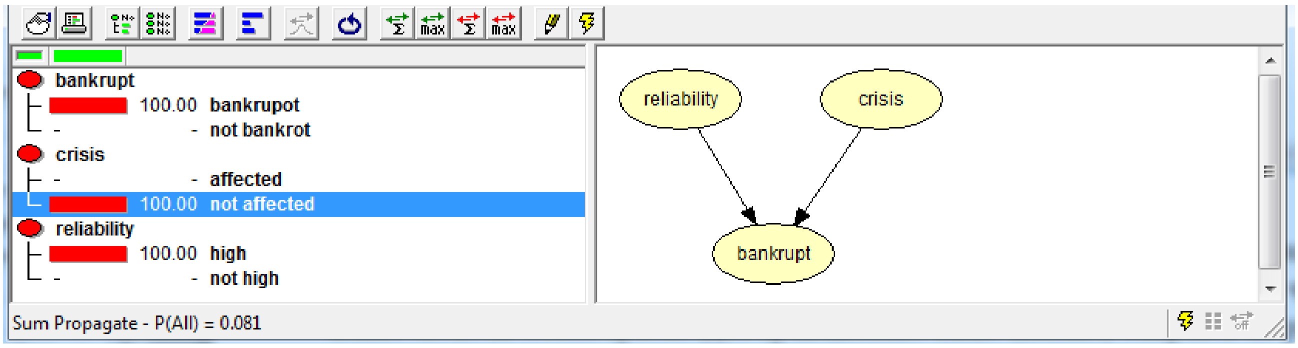


Рис. 37

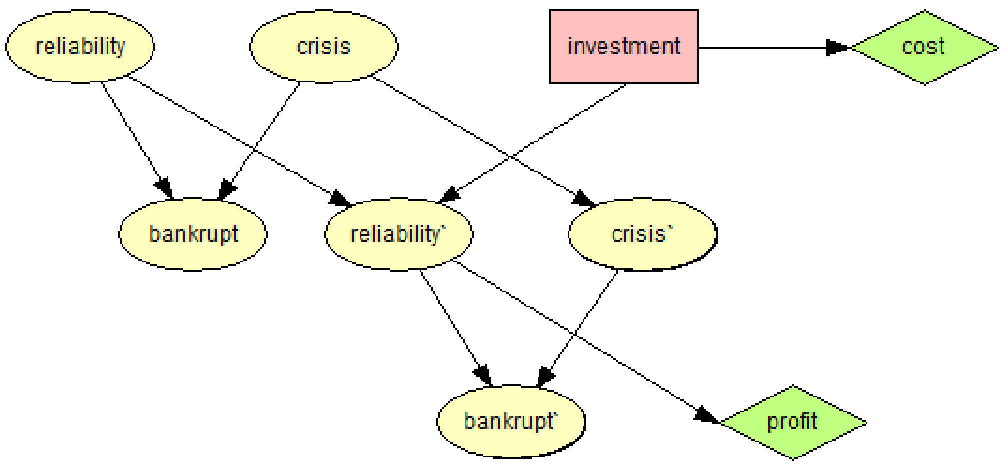


Рис. 38

Предположим, что мы для повышения надежности фирмы инвестируем 8000 д.е. или 0 д.е. и таблицы условных вероятностей для дополнительных переменных *reliability*` *и crisis*`(рис.38), полученные на основе обработки знаний экспертов имеют вид, представленный на рис.39.

investment	investment		not	
reliability	high	not high	high	not high
high	0.2	0.99	0.99	0.02
not high	0.8	0.01	0.01	0.98

crisis	affected	not affected
affected	0.6	0.05
not affected	0.4	0.95

Рис. 39

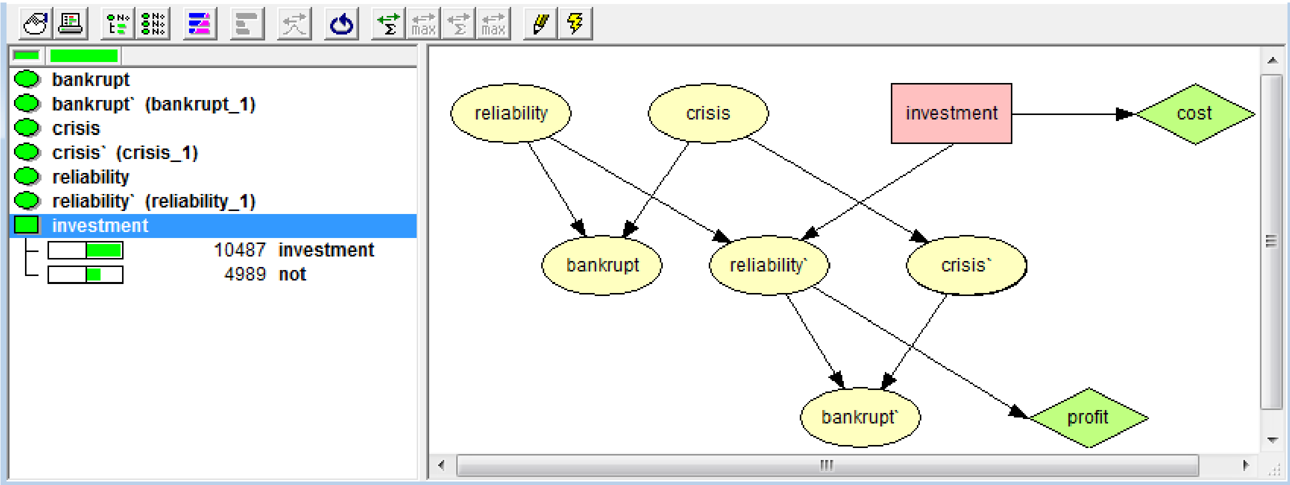


Рис. 40

Тогда полезность от инвестиций составляет 10487 д.е. и 4989 д.е. в противном случае (рис.40). То есть инвестиции более предпочтительны.

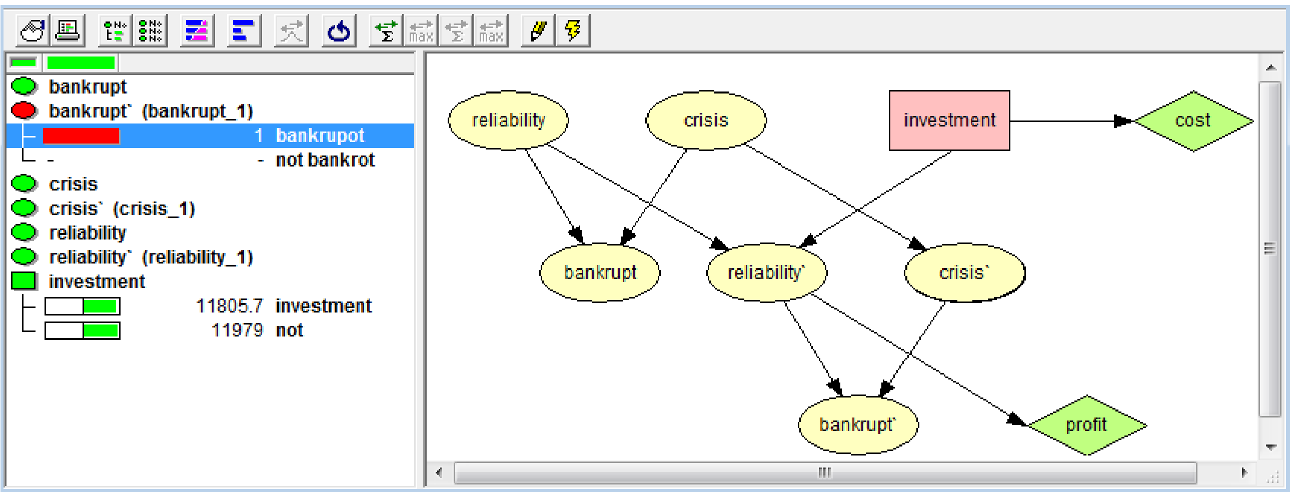


Рис. 41

Если получено свидетельство о том, что фирма банкрот, то результативность инвестиций 11805,7 д.е., а их отсутствия 11979 д.е., то есть инвестировать нет смысла (рис.41).

Сегдня байесовские сети доверия имеют широкое применение при разработке экспертных систем и систем принятия решений для оценки рисков в различных областях деятельности: медицине, финансах, коммерции и т.д.

Литература к лекции

1. Вентцель Е.С., Овчаров Л.А. Теория вероятностей и её инженерные приложения. – М.: Высшая школа, 2000. – 480 с.
2. Горелова Г.В., Кацко И.А. Теория вероятностей и математическая статистика в примерах и задачах с применением EXCEL. 4-е изд. испр.и доп. Ростов н/Д, «Феникс», 2006г. – 475с.
3. Кнут Д. Искусство программирования, том 1. Основные алгоритмы, 3-е изд.:Пер. с англ. – М.: Вильямс, 2001. – 720 с.
4. Грэхэм Р., Кнут Д., Паташник О. Конкретная математика. Основания информатики: Пер с англ. – М.: Мир, 1998. – 703 с., ил.
5. Пугачев В.С. Теории вероятностей и математическая статистика: Учеб. Пособие. – 2-е изд., исправл. и дополн. – М.: ФИЗМАТЛИТ, 2002. – 496 с.
6. Реньи А. Трилогия о математике. (Диалоги о математике. Письма о вероятности. Дневник. – Записки студента по теории информации.); Пер. с вен гер. – М.: Мир, 1980. – 375 с.
7. Стивенс Р. Visual Basic. Готовые алгоритмы: Пер. с англ. – М.: ДМК Пресс, 2000. – 384 с. :ил. (Серия «Для программистов»).
8. Хабаров С. Экспертные системы.: Уч. пос.- Кафедра Информатики и Информационных Систем, ФЭУ, С-Пб. ЛТА

Статистика – это математическая теория,
позволяющая познать мир через опыт.
В. Томпсон

Математическая и прикладная статистика – одно из
наиболее значимых интеллектуальных достижений человечества.
NN

Окружающий нас мир насыщен информацией – разнообразные потоки данных окружают нас, захватывая в поле своего действия, лишая правильного восприятия действительности. Не будет преувеличением сказать, что информация становится частью действительности и нашего сознания.
Без адекватных технологий анализа данных человек оказывается беспомощным в жестокой информационной среде и скорее напоминает броуновскую частицу, испытывающую жестокие удары со стороны и не имеющую возможности рационально принять решение.
Статистика позволяет компактно оценить данные, понять их структуру, провести классификации, увидеть закономерности в хаосе случайных явлений. Удивительно, что даже простейшие методы визуального и разведочного анализа данных позволяют существенно прояснить информацию, первоначально порождающую нагромождением цифр.
В. Боровиков – научный директор StatSoft Russia.
Statistica.Искусство анализа данных на компьютере, 2003

Основная идеология методов многомерного статистического анализа сводится к использованию теории алгебраических инвариантов, не изменяющихся при линейных преобразованиях (например, собственные значения, собственные вектора, определители, декомпозиция матриц, корреляция между переменными и т.д.).

Справочная система Statistica 6.1
Аксиомы одной эпохи – нерешение задачи следующей.
Р. Тони

Часть II. Математическая статистика

Математическая статистика позволяет обосновать ответ на вопросы: случайно или закономерно изучаемое явление (влияет ли доза внесения удобрений на урожайность?); как зависит результативный признак от факторного (как зависит урожайность от дозы внесения удобрений, при прочих равных условиях?); сколько необходимо провести наблюдений для объективного суждения об изучаемом явлении; какой фактор сильнее влияет на результат (влияние вида удобрений на урожайность) и т.д.

Методы математической статистики можно разделить на описательные (дескриптивные) и аналитические. Описательные методы позволяют описать реальные наблюдения с помощью таблиц, графиков, характеристик положения (среднее арифметическое, мода, медиана), характеристик рассеяния (среднее линейное отклонение, среднее квадратическое отклонение, дисперсия, коэффициент вариации) и т. д.

Аналитические методы позволяют на основании выборочных наблюдений сделать статистически значимые выводы о наличии закономерностей для всей совокупности. Аналитические методы обычно основываются на соответствующих вероятностных моделях, предполагающих нормальное (или другое известное) распределение совокупности изучаемого признака - методы *параметрической статистики*.

Другим направлением аналитических методов, являются методы *непараметрической статистики*, которые не опираются на нормальное распределение (или любое другое) и не используют его свойства.

Основная цель математической статистики - это получение и обработка данных для статистически значимой поддержки процесса принятия решения, например, при решении задач планирования, управления, прогнозирования.

Глава 3. Анализ вариационных рядов

3.1. Вариационные ряды распределения

В реальных социально - экономических системах нельзя проводить активные эксперименты, поэтому данные обычно представляют собой наблюдения за происходящим процессом, например: курс валюты на бирже в течение месяца, урожайность пшеницы в хозяйстве за 30 лет, производительность труда рабочих за смену и т.д. Результаты наблюдений − это, в общем случае, ряд чисел, расположенных в беспорядке, который для изучения необходимо упорядочить (проранжировать).

Операция, заключенная в расположении значений признака по возрастанию, называется *ранжированием* опытных данных.

После операции ранжирования опытные данные можно сгруппировать так, чтобы в каждой группе признак принимал одно и то же значение, которое называется *вариантом* (x_i). Число элементов в каждой группе называется *частотой* варианта (n_i).

Размахом вариации называется число $W=x_{max}-x_{min}$, где x_{max}- наибольший вариант, x_{min}- наименьший вариант.

Сумма всех частот равна определенному числу n, которое называется объемом совокупности:

$$\sum_{i=1}^{k} n_i = n_1 + n_2 + \ldots + n_к = n. \tag{3.1.1}$$

Отношение частоты данного варианта к объему совокупности называется *относительной частотой* ($\widehat{p}_i$) или *частостью* этого варианта:

$$\widehat{p} = \frac{n_i}{n}, \tag{3.1.2}$$

$$\sum_{i=1}^{k} \widehat{p}_i = \sum_{i-1}^{k} \frac{n_i}{n} = \frac{\sum_{i=1}^{k} n_i}{n} = \frac{n}{n} = 1. \tag{3.1.3}$$

Последовательность вариант, расположенных в возрастающем порядке, называется *вариационным рядом* (вариация - изменение).

Вариационные ряды бывают дискретными и непрерывными. *Дискретным вариационным рядом* называется ранжированная последовательность вариант с соответствующими частотами и (или) частостями.

Пример 3.1.1. В результате тестирования группа из 24 человек набрала баллы: 4, 0, 3, 4, 1, 0, 3, 1, 0, 4, 0, 0, 3, 1, 0, 1, 1, 3, 2, 3, 1, 2, 1, 2. Построить дискретный вариационный ряд.

Решение. Проранжируем исходный ряд, подсчитаем частоту и частость вариант:

0, 0, 0, 0, 0, 0, 1, 1, 1, 1, 1, 1, 1, 2, 2, 2, 3, 3, 3, 3, 3, 4, 4, 4.

В результате получим дискретный вариационный ряд (табл.5).

Таблица 5

Балл, x_i	Число студентов, n_i	Относительная частота, $\widehat{p}_i$
0	6	6/24
1	7	7/24
2	3	3/24
3	5	5/24
4	3	3/24
$\sum$	24	1

Построение дискретного вариационного ряда нецелесообразно, если число значений признака велико. В этом случае следует построить *интервальный вариационный ряд.* Для построения такого ряда промежуток изменения признака разбивается на ряд отдельных интервалов и подсчитывается количество значений величины в каждом из них.

Будем считать, что отдельные (частичные) интервалы имеют одну и ту же длину. Число интервалов (k), в случае нормально распределённой совокупности, можно определить по формуле Стерджесса

$$k = 1 + 3{,}322\lg(n)$$

или приближённо: $k \in [6;12]$. (3.1.4)

Длина частичного интервала определяется по формуле

$$h = \frac{W}{k} = \frac{x_{\max} - x_{\min}}{k} = \frac{15-4}{7} \approx 1{,}6. \quad (3.1.5)$$

Пример 3.1.2. Пусть дан ряд распределения хозяйств по количеству рабочих на 100 га с/х угодий (n=60):

12	6	8	6	10	11	7	10	12	8	7	7	6	7	8	6	11	9	11
9	10	11	9	10	7	8	8	8	11	9	8	7	5	9	7	7	14	11
9	8	7	4	7	5	5	10	7	7	5	8	10	10	15	10	10	13	12
11	15	6																

Построить интервальный вариационный ряд.

Решение. Для определения числа групп подставим значение n=60 в формулу Стерджесса: $k = 1 + 3{,}322\lg 60 \approx 6{,}907; \quad k = 7.$

Найдем длину частичного интервала:

Построим интервальный вариационный ряд, для этого в качестве начального значения используем x_{min}. Разобьем интервал вариации признака X на k=7 частичных интервалов с шагом h=1,6 и подсчитаем количество рабочих на 100 га сельскохозяйственных угодий в каждом интервале (табл.6).

Таблица 6

Группы хозяйств по численности работников на 100га с/х угодий	Число хозяйств в группе (n_i)	Накопленное число хозяйств (S_i)	Относительная частота ($\hat{P}_i$)
4 - 5,6	5	5	5/60
5,61 - 7,2	17	22	17/60
7,21 - 8,8	9	31	9/60
8,81 - 10,4	15	46	15/60
10,41 - 12,0	10	56	10/60
12,01 - 13,6	1	57	1/60
13,61 -15,2	3	60	3/60
Итого:	60	-	1

Графическое изображение вариационных рядов.

Вариационные ряды изображают графически с помощью полигона и гистограммы.

Полигон частот - это ломаная, отрезки которой соединяют точки

$$(x_1;n_1),(x_2;n_2),...(x_k;n_k).$$

Полигон относительных частот - это ломаная, отрезки которой соединяют точки: $(x_1;\frac{n_1}{n}),(x_2;\frac{n_2}{n}),...(x_k;\frac{n_k}{n}).$

Для примера 3.1.1 построим полигон частот:

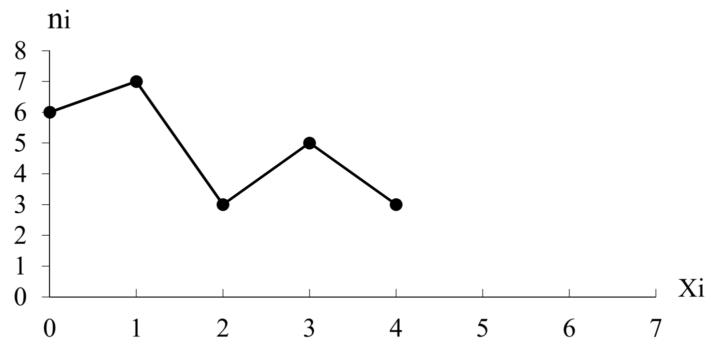


Рис.42.1 – Полигон частот

Полигон относительных частот будет иметь следующий вид (рис.42.2).

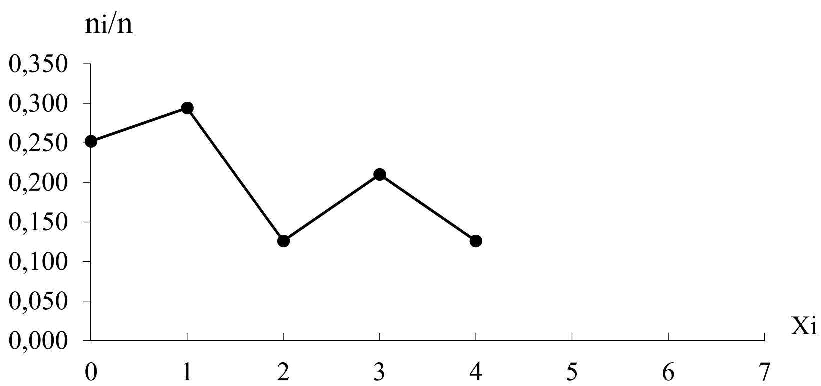


Рис.42.2 – Полигон относительных частот

Гистограммой частот называется фигура, состоящая из прямоугольников с основанием *h* и высотами n_i. Для *гистограммы относительных частот* в качестве высоты рассматривают n_i/n. Гистограмма относительных частот является аналогом дифференциальной функции случайной величины.

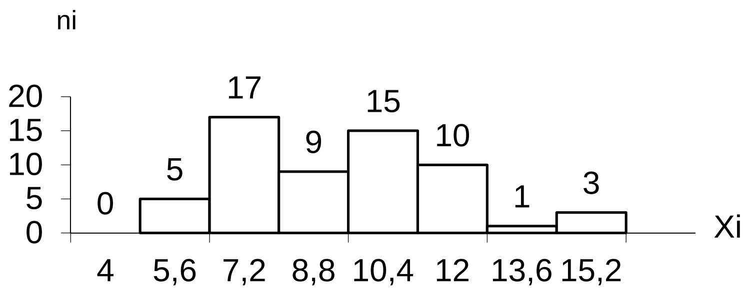


Рис. 42.3 – Гистограмма частот

Построим гистограмму частот для примера 3.1.2.(рис.42.3). Гистограмма построена в Excel (*Пакет анализа* - инструмент **Гистограмма**) подпись под прямоугольником означает верхнюю границу интервала, над прямоугольником – соответствующую частоту.

График гистограммы относительных частот можно получить из графика рис. 44 сжатием в 60 раз вдоль оси ординат.

3.2. Числовые характеристики вариационных рядов

Вариационные ряды позволяют получить первое представление об изучаемом распределении. Далее необходимо исследовать числовые характеристики распределения (аналогичные характеристикам распределения теории вероятностей): характеристики положения (средняя арифметическая, мода, медиана); характеристики рассеяния (дисперсия, среднее квадратическое отклонение, коэффициент вариации); характеристики меры скошенности (коэффициент асимметрии) и островершинности (эксцесс) распределения.

Характеристики положения вариационного ряда.

Средней арифметической ($\overline{X}$) дискретного вариационного ряда называется отношение суммы произведений вариант на соответствующие частоты к объему совокупности:

$$\overline{X} = \frac{\sum x_i n_i}{\sum n_i} = \frac{\sum x_i n_i}{n}. \qquad (3.2.1)$$

Средняя арифметическая имеет те же единицы измерения, что и варианты.

Свойства средней арифметической.

1. Средняя арифметическая суммы соответствующих друг другу значений, принадлежащих двум группам наблюдений, равна алгебраической сумме средних арифметических этих групп:

$$\overline{X \pm Y} = \overline{X} \pm \overline{Y}.$$

2. Если ряд наблюдений состоит из двух непересекающихся групп наблюдений, то средняя арифметическая $\overline{Z}$ всего ряда наблюдений равна взвешенной средней арифметической групповых средних $\overline{X}$ и $\overline{Y}$, причём весами являются объёмы групп $n_1 = \sum n_i, n_2 = \sum m_j$ соответственно

$$\overline{X \pm Y} = \frac{\sum x_i n_i + \sum y_j m_j}{n_1 + n_2}.$$

3. Средняя арифметическая постоянной равна самой постоянной

$$\overline{C} = C.$$

4. Если все результаты наблюдений умножить на одно и то же число, то имеет место равенство:

$$\overline{Z} = \overline{CX} = C\overline{X}.$$

5. Сумма отклонений результатов наблюдений от их средней, взвешенная с соответствующими частотами равна нулю

$$\sum (x_i - \overline{X}) n_i = 0.$$

6. Если все результаты наблюдений увеличить (уменьшить) на одно и то же число, то средняя арифметическая увеличится (уменьшится) на то же число, т.е.:

$$\overline{Z} = \overline{X \pm C} = \overline{X} \pm C.$$

7. Если все частоты вариантов умножить на одно и то же число, то средняя арифметическая не изменится.

Модой $(\mathrm{M}_{\mathrm{o}}^*(\mathrm{X}))$ дискретного вариационного ряда называется вариант, имеющий наибольшую частоту.

Медианой $(\mathrm{M}_{\mathrm{e}}^*(\mathrm{X}))$ дискретного вариационного ряда называется вариант, делящий ряд на две равные части.

Если дискретный вариационный ряд имеет 2n членов в ранжированной совокупности: $x_1, x_2, \ldots, x_n, x_{n+1}, \ldots, x_{2n}$, то

$$M_e^*(X) = \frac{x_n + x_{n+1}}{2}. \quad (3.2.2)$$

Если дискретный вариационный ряд в ранжированной совокупности имеет 2n+1 членов: $x_1, x_2, \ldots, x_{n-1}, x_n, x_{n+1}, \ldots, x_{2n+1}$, то

$$\mathrm{M}_{\mathrm{e}}^*(\mathrm{X}) = \mathrm{x}_{\mathrm{n}+1}. \quad (3.2.3)$$

В примере 3.1.1:

$$\overline{X} = \frac{0 \cdot 6 + 1 \cdot 7 + 2 \cdot 3 + 3 \cdot 5 + 4 \cdot 3}{24} = \frac{40}{24} = \frac{5}{3} = 1{,}67;$$

$$M_{\mathrm{o}}^*(\mathrm{X}) = 1;$$

$$M_e^*(X) = \frac{1+1}{2} = 1.$$

Для интервальных вариационных рядов имеют место формулы:

а) медианы:

$$M_e^*(X) = x_{Me} + h \cdot \frac{0{,}5n - S_{Me-1}}{n_{Me}}, \quad (3.2.4)$$

где x_{Me} – начало медианного интервала,

h – длина частичного интервала,

n – объем совокупности,

$S_{Me\text{-}1}$ – накопленная частота интервала, предшествующего медианному,

n_{Me} – частота медианного интервала;

б) моды:

$$M^*_0(X) = x_{M_0} + h\frac{(n_{M_0} - n_{M_0-1})}{(n_{M_0} - n_{M_0-1}) + (n_{M_0} - n_{M_0+1})}, \tag{3.2.5}$$

где x_{Mo} – начало модального интервала,

h – длина частичного интервала,

n_{Mo} – частота модального интервала,

$n_{Mo\text{-}1}$ - частота предмодального интервала,

n_{Mo+1} - частота послемодального интервала;

в) средней арифметической, совпадающей с формулой (3.2.1) для дискретного вариационного ряда, причем в качестве вариант x_i принимаются середины соответствующих интервалов.

Для примера 3.1.2:

$$M^*_e(x) = 7{,}2 + 1{,}6 \cdot \frac{0{,}5 \cdot 60 - 22}{9} = 8{,}62\ ;$$

$$M^*_0(x) = 5{,}6 + 1{,}6\frac{(17-5)}{(17-5)+(17-9)} = 6{,}56\ .$$

Мода и медиана используются в качестве характеристики среднего положения в случае, если границы ряда нечеткие или если ряд не симметричен.

Замечание. Известно, что множество результатов наблюдений, расположенных на числовой прямой имеет свойство группироваться относительно некоторого центра. Этот центр часто рассматривают, как «свёртку» всех наблюдений, которая используется в анализе. Центр может быть описан различными статистиками или иначе мерами центральной тенденции: модой, медианой, средней арифметической. При выборе меры центральной тенденции следует иметь в виду следующее.

1. В малой группе мода не стабильна − при небольших изменениях она может сильно измениться: M_0(1,1,1,3,5,7,7,8)=1, M_0(0,2,1,3,5,7,7,8)=7.
2. На медиану не влияет величина «больших» и «малых» значений, например, если последнее значение утроить, то медиана не изменится.
3. На величину среднего влияет каждое значение. Если одно из значений изменится на C единиц, то $\overline{X}$ изменится в том же направлении на $\frac{C}{n}$ единиц.
4. Медиана – лучшая характеристика центральной тенденции, когда гистограмма исходных данных унимодальна (имеет одну моду). Поэтому, например, для адекватного описания дохода или заработной платы часто используют медиану.
5. Некоторые множества не имеют центральной тенденции. Д.Ж.Гласс, Д.Стенли в книге «Статистические методы в педагогике и психологии» (Москва.: Прогресс, 1976. – 495 с.) приводят следующий анекдот, обобщающий множество проблем, возникающих в процессе применения разных мер центральной тенденции.

   «Однажды пятеро мужчин сидели рядом на скамейке парка. Двое были бродягами, имущество которых выражалось в 25 центах. Третий был рабочим, чей счёт в банке и имущество составляли 2000 долларов. Четвёртый владел 15 000 долларов в различных формах. Пятый же был мультимиллионером с чистым доходом 5 000 000 долларов. Поэтому модальный актив группы составил 25 центов. Эта цифра точно характеризует двоих, но является чрезвычайно некорректной для трёх других. Медиана, составляющая 2000 долларов, несколько меняет дело для всех, кроме рабочего. Среднее, 1 003 400,10 долларов, не является вполне удовлетворительным даже для мультимиллионера. Если мы должны выбрать одну меру тенденции, возможно, это была бы мода, которая точно описывает 40 процентов группы. Однако, если сказать, что «модальный актив пяти лиц, сидящих на скамье парка, равен 25 центам», то нам пришлось бы сделать вывод о том, что общий актив группы приблизительно составляет 1,25 доллара, что меньше фактического более чем на пять миллионов долларов. Очевидно, нет меры, адекватной этим «странным соседям по скамейке», которые просто не имеют «центральной тенденции».

Показатели вариации.

Показатели центральной тенденции ($M_0, M_e, \overline{X}$) не исчерпывают всех свойств распределения. В одних случаях значения признака концентрируются тесно около среднего значения, в других наблюдается значительное рассеяние.

Для изучения степени изменчивости признака вводят показатели вариации:

1. $W = x_{max} - x_{min}$ – размах вариации.
2. Значения x_i имеют свойство концентрироваться около $\overline{X}$, поэтому вводят следующие характеристики (т.к. $\sum (x_i - \overline{X}) n_i = 0$):

Дисперсия дискретного ряда распределения:

$$D^* = \frac{\sum (x_i - \overline{X})^2 n_i}{n}, \tag{3.2.6}$$

характеризует средний квадрат отклонения x_i от $\bar{x}$.

Среднее квадратическое отклонение дискретного ряда распределения:

$$\sigma^* = \sqrt{\frac{\sum (x_i - \overline{X})^2 n_i}{n}}, \tag{3.2.7}$$

выражается в тех же единицах, что и x_i.

Среднее линейное отклонение:

$$L(X) = \frac{\sum |x_i - \overline{X}| n_i}{n}. \tag{3.2.8}$$

Коэффициент вариации:

$$V^* = \frac{\sigma^*}{\overline{X}} \cdot 100\%, \tag{3.2.9}$$

характеризует относительное значение среднего квадратического отклонения и обычно служит для сравнения колеблемости несоизмеримых показателей.

Свойства дисперсии:

1. Дисперсия постоянной величины равна 0

$$D^*(C) = 0.$$

2. Если все результаты наблюдений увеличить (уменьшить) на одно и то же число C, то дисперсия и среднее квадратическое отклонение не изменятся, т.е.

$$D^*(X \pm C) = D^*(X),\ \sigma^*(X \pm C) = \sigma^*(X)$$

3. Если все результаты наблюдений умножить на одно и то же число, то имеет место равенство:

$$D^*(CX) = C^2 D(X),\ \sigma^*(CX) = |C|\sigma^*(X)$$

4. Если все частоты вариантов умножить на одно и то же число, то дисперсия и среднее квадратическое отклонение не изменятся.
5. Свойство минимальности дисперсии.

$$\frac{\sum (x_i - C)^2 n_i}{n} \to min \text{ при } C = \overline{X}$$

Следствие 1. Средний квадрат отклонений значений x_i от их средней арифметической равен среднему квадрату отклонений x_i от произвольной постоянной а минус квадрат разности между средней арифметической ($\overline{X}$) и этой произвольной постоянной.

Пусть $\sigma^{*2} = \dfrac{\sum (x_i - \overline{X})^2 n_i}{n}, \sigma_a^{*2} = \dfrac{\sum (x_i - a)^2 n_i}{n}$, тогда

$$\sigma_x^{*2} = \sigma_a^{*2} - (\overline{X} - a)^2.$$

Следствие 2. Дисперсия равна средней арифметической из квадратов значений признака минус квадрат средней арифметической

$$\sigma_x^{*2} = \overline{X^2} - (\overline{X})^2.$$

6. *Правило сложения дисперсий*. Если объединяются несколько распределений в одно, то общая дисперсия σ_0^{*2} нового распределения равна средней арифметической из дисперсий объединяемых распределений, сложенной с дисперсией частных средних относительно общей средней нового распределения. Или, иначе говоря, общая дисперсия равна сумме внутригрупповой и межгрупповой дисперсий :

$i \backslash j$	x_1	x_2	…	x_m	Σ
1	n_{11}	n_{12}		n_{1m}	n_1
2	n_{21}	n_{22}		n_{2m}	n_2
3	n_{31}	n_{32}		n_{3m}	n_3
. . .	… … …	… … …	… … …	… … …	. . .
k	n_{k1}	n_{k2}		n_{km}	n_m
Σ	N_1	N_2		N_m	N

$$\sigma_0^{*2} = \overline{\sigma^{*2}} + \delta^{*2} \tag{3.2.10}$$

или

$$\sigma_0^{*2} = \frac{\sum_{i,j} (x_{ij} - \overline{X_0})^2 \cdot n_{ij}}{N} = \frac{\sum x_j^2 N_j}{N} - (\overline{X_0})^2, \tag{3.2.11}$$

где n_{ij} – частота j–го варианта i-го частного распределения (j=1,…, m; i=1,2,…, k)

x_{ij} – j–й вариант i-го частного распределения (j=1,…, m; i=1,2,…, $к$),

n_i – объем i-го частного распределения,

$N_j = \sum\limits_i n_{ij}$ – частота j-го варианта нового распределения,

$N = \sum\limits_{i=1}^{к} n_i$ - объем нового распределения,

$\overline{X}_i = \dfrac{\sum_j n_{ij} x_j}{n_i}$ – средняя арифметическая i-го частного распределения, (i=1,...,k),

$\overline{X}_0 = \dfrac{\sum x_j N_j}{N}$ – средняя арифметическая нового распределения,

$\sigma_i^2 = \dfrac{\sum_j x_{ij}^2 n_{ij}}{n_i} - (\overline{X}_i)^2$ – дисперсия i-го частного распределения,

$\overline{\sigma^{*2}} = \dfrac{\sum \sigma_i^2 n_i}{N}$ – внутригрупповая дисперсия,

$\delta^{*2} = \dfrac{\sum (\overline{X}_i - x_0)^2 n_i}{N}$ - межгрупповая дисперсия.

Моменты для вариационных рядов в математической статистике находятся по формулам, аналогичным формулам (2.7.6), (2.7.7),(2.7.11), (2.10.3):

$\alpha_s^* = \dfrac{\sum x_i^s n_i}{n}$ - начальный момент s – го порядка,

$\mu_s^* = \dfrac{\sum (x_i - \overline{x})^s n_i}{n}$ - центральный момент s – го порядка,

$r_s^* = \dfrac{\sum (x_i - \overline{x})^s n_i}{n\sigma_x^{*s}}$ - основной момент s – го порядка,

$r_{s,h}^* = \dfrac{\sum (x_i - \overline{x})^s \cdot (y_i - \overline{y})^h n_i}{n\sigma_x^{*s} \cdot \sigma_y^{*h}}$ - основной момент порядка s,h.

Соотношения между начальными и центральными моментами в математической статистике соответствуют формулам (2.7.8).

Коэффициент асимметрии: $Sk* = \dfrac{\sum (x_i - \overline{X})^3 n_i}{n\sigma^{*3}}$.

Эксцесс: $Ex* = \dfrac{\sum (x_i - \overline{X})^4 n_i}{n\sigma^{*4}} - 3$.

Рассчитаем среднюю арифметическую, дисперсию, коэффициенты асимметрии и эксцесса для примера 3.1.2. Построим вспомогательную таблицу (табл. 7).

Замечание.

1. Для упрощения расчётов лучше сразу после нахождения второго центрального момента (дисперсии) перейти к расчёту основных моментов (r_3, r_4).
2. Получив основные числовые характеристики (положения, рассеяния, асимметрии, островершинности) распределения, можно сделать в первом приближении суждение о нормальности распределения, для которого, как известно, Sk=0, Ex=0. Далее следует проверить гипотезу о нормальности распределения, принятие которой является довольно сильным утверждением, позволяющим применять классические методы математической статистики (метод доверительных интервалов, дисперсионный анализ, корреляционно – регрессионный анализ и т.д.). Хотя ряд исследований показал, в частности для дисперсионного анализа, что отклонение от нормального закона допустимо.

Таблица 7 – Вспомогательная таблица для расчета числовых характеристик ряда распределения

Группы предприятий по численности работников на 100 га сельхозугодий, чел	Среднее значение интервала (x_i)	Число хозяйств в группе (n_i)	$x_i n_i$	$x_i - \overline{X}$	$(x_i - \overline{X})^2 n_i$	$\frac{x_i - \overline{X}}{\sigma^*}$	$(\frac{x_i - \overline{X}}{\sigma^*})^3 n_i$	$(\frac{x_i - \overline{X}}{\sigma^*})^4 n_i$
4 - 5,6	4,8	5	24	-3,813	72,708	-1,559	-18,954	29,554
5,61 - 7,2	6,4	17	108,8	-2,213	83,280	-0,905	-12,601	11,404
7,21 - 8,8	8	9	72	-0,613	3,386	-0,251	-0,142	0,036
8,81 - 10,4	9,6	15	144	0,987	14,603	0,403	0,985	0,397
10,41 - 12	11,2	10	112	2,587	66,908	1,058	11,832	12,514
12,01 - 13,6	12,8	1	12,8	4,187	17,528	1,712	5,017	8,588
13,61 - 15,2	14,4	3	43,2	5,787	100,457	2,366	39,740	94,030
итого	-	60	516,8	-	358,869	-	25,876	156,523

Среднее значение признака составит: $\overline{X} = \frac{\sum x_i n_i}{\sum n_i} = \frac{516,8}{60} = 8,613$.

Дисперсия и среднее квадратическое отклонение:

$$D^* = \frac{\sum (x_i - \overline{X})^2 n_i}{n} = \frac{358,869}{60} = 5,981,$$

$$\sigma^* = \sqrt{\frac{\sum (x_i - \overline{X})^2 n_i}{n}} = \sqrt{5,981} = 2,446.$$

Коэффициент вариации: $V^* = \frac{\sigma^*}{\overline{X}} \cdot 100\% = \frac{2,446}{8,613} \cdot 100\% = 28,4\%$.

Таким образом, средняя численность работников на 100 га сельскохозяйственных угодий по исследуемой совокупности хозяйств составила 8,61 чел. Плотность работников в среднем колебалась в промежутке $\overline{X} \pm \sigma^* = 8,61 \pm 2,45$, то есть от 6,16 до 11,06 чел. на 100 га сельскохозяйственных угодий. Этот интервал, а так же коэффициент вариации показывают, что имеются большие различия в обеспеченности хозяйств рабочей силой.

Коэффициент асимметрии: $Sk^* = \frac{\sum (x_i - \overline{X})^3 n_i}{n\sigma^{*3}} = \frac{25,876}{60} = 0,43.$

Эксцесс: $Ex^* = \frac{\sum (x_i - \overline{X})^4 n_i}{n\sigma^{*4}} - 3 = \frac{156,523}{60} - 3 = \text{-}0,39.$

Найденное значение коэффициента асимметрии (не достаточно близкое к нулю) указывает, что распределение не симметрично. Эксцесс также отличен от нуля, что говорит о возможном отличии распределения от нормального.

Контрольные задания 3.2.

1. По списку на предприятии числится 100 рабочих, которые имеют следующие разряды:
   1,5,2,4,3,4,6,4,5,1,2,2,3,4,5,3,4,5,2,1,4,5,5,4,3,4,6,1,2,4,4,3,5,6,4,3,3,1,3,4,3,1,2, 4,4,5,6,1,3,4,5,3,4,4,3,2,6,1,2,4,5,3,3,2,3,6,4,3,4,5,4,3,3,2,6,3,3,4,5,4,4,3,3,2,1,2, 1,6,5,4,3,2,3,4,4,3,5,6,1,5.
   Составить ряд распределения рабочих по разрядам. Найти накопленные частоты и частости. Вариационный ряд изобразить графически.
   Определить средний разряд рабочего, модальный и медианный разряд, дисперсию и среднее квадратическое отклонение.
2. Имеются следующие данные о числе производственных подразделений на каждом из 100 сельскохозяйственных предприятий:
   2,4,5,3,4,6,7,4,5,3,3,4,2,6,5,4,7,2,3,4,4,5,4,3,4,6,6,5,2,3,4,3,5,6,7,2,4,3,4,5,4,6,7, 2,5,3,5,4,3,7,2,4,3,4,5,4,3,2,6,7,6,4,3,2,3,4,5,4,3,5,4,3,2,6,4,5,7,5,4,3,4,5,7,4,3,4, 5,6,5,3,4,2,2,4,3,7,5,6,4,5.
   Составить ряд распределения сельскохозяйственных предприятий по числу производственных подразделений на одно хозяйство. Найти накопленные частоты и частости. Вариационный ряд изобразить графически.
   Определить среднее число производственных подразделений на одно хозяйство, модальное и медианное значения числа подразделений, дисперсию и среднее квадратическое отклонение.
3. В задачах 1)-20)по данным приложения 8, составить статистический ряд распределения по одному признаку. Найти накопленные частоты и частости. Ряд распределения изобразить графически. Определить моду, медиану, среднее значение, дисперсию, среднее квадратическое отклонение, показатели асимметрии и эксцесса. Сделать выводы по результатам расчетов.

   1) Площадь сельскохозяйственных угодий на условный эталонный трактор, га.
   2) Валовая продукция на 100 га сельскохозяйственных угодий, тыс. руб.
   3) Валовая продукция на среднегодового работника, тыс. руб.
   4) Валовая продукция на 100 руб. основных фондов, руб.
   5) Валовая продукция на 100 руб. производственных затрат, руб.
   6) Реализованная продукция на 100 руб. основных производственных фондов, руб.

7) Реализованная продукция на 100га сельскохозяйственных угодий, тыс. руб.
8) Реализованная продукция на среднегодового работника, тыс. руб.
9) Реализованная продукция на 100 руб. затрат, руб.
10) Производственные затраты на 100 га сельскохозяйственных угодий, тыс. руб.
11) Производственные затраты на среднегодового работника, тыс. руб.
12) Затраты на реализованную продукцию на 100 га сельскохозяйственных угодий, тыс. руб.
13) Основные производственные фонды на 100 га сельскохозяйственных угодий, тыс. руб.
14) Основные производственные фонды на среднегодового работника, тыс. руб.
15) Энергетические мощности на 100 га сельскохозяйственных угодий, л.с.
16) Энергетические мощности на среднегодового работника, л.с.
17) Площадь сельскохозяйственных угодий на среднегодового работника, га.
18) Численность тракторов на одно хозяйство.
19) Среднегодовая численность работников на одно хозяйство, чел.
20) Площадь сельскохозяйственных угодий на одно хозяйство, га.

4. Определить абсолютную и относительную плотность распределения работников предприятия по стажу их работы на данном предприятии.

Распределение работников по стажу работы

Стаж работы, лет	До 1	1-5	5-10	10-20	20-40	Всего
Число работников	8	12	16	14	10	60

Найти средний стаж работы, среднее квадратическое отклонение и коэффициент вариации.

5. Имеются следующие данные о площади посева овощей в хозяйствах совокупности районов.

Площадь посева овощей на хозяйство

Район	Номер хозяйства						
	1	2	3	4	5	6	7
1	8	10	12	6	15	30	21
2	32	16	26	41	44	38	-
3	101	165	230	144	84	76	260
4	22	30	44	18	16	31	-
5	10	7	4	3	12	7	6
6	255	366	384	273	450	510	-
7	121	84	96	110	161	143	-
8	34	16	84	71	36	8	17
9	46	41	48	52	50	58	-
10	15	24	86	144	34	14	24

Дать сравнительную оценку колеблемости площади посева овощей в хозяйствах двух районов.

6. Путем устного опроса изучалось качество продукции, выпускаемой фирмой и реализуемой в магазине этой фирмы. Посетители давали оценку качества по десятибалльной шкале. Были получены сводные данные.

Бальная оценка продукции предприятия

Оценка качества продукции, балл	1-2	3-4	5-6	7-8	9-10
Число случаев	3	8	36	89	45

Определить средний балл качества продукции, среднее квадратическое отклонение, коэффициент вариации, показатели асимметрии и эксцесса.

7. По данным распределения студентов по результатам сдачи экзаменов определить: средний бал успеваемости студентов по каждому предмету и по всем предметам; дисперсии балла успеваемости по предмету и в целом по всем предметам; межгрупповую дисперсию. Найти общую дисперсию успеваемости, используя правило сложения дисперсий.

Распределение студентов группы по результатам сдачи экзаменов

Оценка на экзамене	Число студентов, получивших оценку по предметам			
	1	2	3	4
2	2	1	4	3
3	6	10	8	8
4	10	8	9	9
5	7	6	4	5

8. Работники предприятия сгруппированы по возрасту.

Распределение работников предприятия по возрасту

Категории работников	Возраст работников, лет					Всего работников
	До 30	30-40	40-50	50-60	свыше 60	
Рабочие	43	141	216	127	118	645
Руководители	2	4	6	8	4	24
Специалисты	3	18	30	34	22	107
Всего работников	48	163	252	169	144	776

Определить: средний возраст работников, предприятия в целом и по отмеченным категориям; модальное и медианное значения возраста работников по категориям и предприятию; дисперсию и среднее квадратическое отклонение возраста по категориям работников и предприятию; межгрупповую дисперсию возраста работников. Найти общую дисперсию возраста работников, используя правило сложения дисперсий.

9. Проведите анализ данных годовых уровней прибыли трёх компаний:

год	*«Cherry Computers»*	*«Lemon Motors»*	*«Orange Electronics»*
1983	14,2	-6,2	37,5
1984	12,3	13,3	-10,6
1985	-16,2	-8,4	40,3
1986	15,4	27,3	5,4
1987	17,2	28,2	6,2
1988	10,3	14,5	10,2
1989	-6,3	-2,4	13,8
1990	-7,8	-3,1	11,5
1991	3,4	15,6	-6,2
1992	12,2	18,2	27,5

Найдите среднее значение и стандартное отклонение прибыли для каждой из компаний. Сравните полученные результаты их деятельности за 10 лет. Деятельность какой из компаний, по вашему мнению, более успешна?

10. Администрацию универсама интересует оптимальный уровень запасов продуктов в торговом зале, а также среднемесячный объём покупок товаров, которые не являются предметом ежедневного потребления в семье (например, таких, кА к сода). Для выяснения этого вопроса менеджер универсама в течение января регистрировал частоту покупок 100-граммовых пакетов с содой и собрал следующие данные x_i:

$$4,4,9,3,3,1,2,0,4,2,3,5,7,10,6,5,7,3,2,9,8,1,4,6,5,4,2,1,0,8.$$

Постройте вариационный ряд, определите его числовые характеристики. Какие рекомендации Вы бы дали администрации универсама?

3.3. Выборочный метод

В реальных условиях обычно бывает трудно или экономически нецелесообразно, а иногда и невозможно, исследовать всю совокупность, характеризующую изучаемый признак *(генеральную совокупность)*. Поэтому на практике широко применяется выборочное наблюдение, когда обрабатывается часть генеральной совокупности *(выборочная совокупность)*. Свойства (закон распределения и его параметры) генеральной совокупности неизвестны, поэтому возникает задача их оценки по выборке. Для получения хороших оценок характеристик генеральной совокупности необходимо, чтобы выборка была *репрезентативной* (представительной). Репрезентативность, в силу закона больших чисел, достигается случайностью отбора.

Различают 5 основных типов выборок.

1) *Собственно-случайная:*

а) *повторная* (элементы после выбора возвращаются обратно);

б) *бесповторная* (выбранные элементы не возвращаются).

2) *Типическая* – генеральная совокупность предварительно разбивается на группы типических элементов, и выборка осуществляется из каждой. Следует различать:

а) *равномерные* выборки (при равенстве объемов исходных групп в генеральной совокупности выбирается одинаковое количество элементов из каждой);

б) *пропорциональные* (численность выборок формируют пропорционально численностям или средним квадратическим отклонениям групп генеральной совокупности);

в) *комбинированные* (численность выборок пропорциональна и средним квадратическим отклонениям, и численностям групп генеральной совокупности).

3) *Механическая* – отбор элементов проводится через определенный интервал.

4) *Серийная* – отбор проводится не по одному элементу, а сериями для проведения сплошного обследования.

5) *Комбинированная* – используются различные комбинации вышеуказанных методов, например, типическая выборка сочетается с механической и собственно случайной.

После осуществления выборки возникает задача оценки числовых характеристик генеральной совокупности по элементам выборочной совокупности. Различают точечные и интервальные оценки.

Точечная оценка характеристики генеральной совокупности - это число, определяемое по выборке.

Пусть $\widehat{\theta} = \widehat{\theta}_n$ - выборочная характеристика, вычисленная по результатам n наблюдений величины X, используемая в качестве оценки θ - характеристики генеральной совокупности (в качестве θ может быть $M(X)$, $D(X)$ и т.д.).

Качество оценки $\widehat{\theta}$ устанавливается по трем свойствам: состоятельность, несмещенность, эффективность.

1. *Состоятельность.* Оценка $\widehat{\theta}_n$ является состоятельной оценкой генеральной характеристики θ, если для любого $\varepsilon > 0$ выполняется следующее равенство

$$\lim_{n \to \infty} P\left(\left|\widehat{\theta}_n - \theta\right| < \varepsilon\right) = 1.$$

Это означает, что при увеличении объема выборки n выборочная характеристика $\theta_n \to \theta$.

2. *Несмещенность.* Оценка $\widehat{\theta}$ генеральной характеристики θ называется несмещенной, если для любого фиксированного числа наблюдений n выполняется равенство $M(\theta_n) = \theta$.

3. *Эффективность*. Несмещенная оценка $\widehat{\theta} = \widehat{\theta}_n$ генеральной характеристики θ называется несмещенной эффективной, если среди всех подобных оценок той же характеристики она имеет наименьшую дисперсию:

$$D(\widehat{\theta}_n) \to \min.$$

Можно показать, что статистики $\overline{X}, \widehat{p}$ являются состоятельными, несмещенными и эффективными характеристиками математического ожидания M(X) и вероятности р соответственно.

Выборочная дисперсия $\widehat{D}$ (далее по тексту $\widehat{D} = \sigma^2$) не обладает свойством несмещенности. На практике используют исправленную выборочную дисперсию S^2, которая является несмещенной оценкой дисперсии генеральной совокупности:

$$S^2 = \frac{n}{n-1} D(x) = \frac{n}{n-1} \frac{\sum (x_i - \overline{X})^2 \cdot n_i}{n} = \frac{\sum (x_i - \overline{X})^2 \cdot n_i}{n-1}, \qquad (3.3.1)$$

S - стандартное отклонение.

Кроме того, в расчётах используют стандартную ошибку выборки:

$$S_{\overline{x}} = \frac{S}{\sqrt{n}}. \qquad (3.3.2$$

Замечание. Точечные оценки получают обычно с помощью метода моментов и метода максимального правдоподобия [2,10,12,14].

Интервальной называют оценку, которая определяется двумя числами-границами интервала. Она позволяет ответить на вопрос: внутри какого интервала и с какой вероятностью находится неизвестное значение оцениваемого параметра θ генеральной совокупности.

Пусть $\widehat{\theta}$ точечная оценка параметра θ Чем меньше разность $\widehat{\theta}$ и θ тем точнее и лучше оценка. Обычно говорят о *доверительной вероятности* (надежности оценки) p=1-α, с которой θ будет находиться в интервале $\widehat{\theta} - \Delta < \theta < \widehat{\theta} + \Delta$, где: Δ(Δ >0) –предельная ошибка выборки, которая может быть либо задана наперёд, либо вычислена; α - риск или уровень значимости (вероятность того, что неравенство будет неверным). Оценка указанного доверительного интервала может быть получена (с наименьшей вероятностью) с помощью неравенства Чебышева (при $\varepsilon = \Delta$). В качестве $1-\alpha$ принимают значения 0,90; 0,95; 0,99; 0,999. Доверительная вероятность показывает, что в

(1-α)100% случаев оценка θ будет накрываться указанным интервалом.

Точечная оценка математического ожидания *M*(*X*)=*a* определяется как средняя арифметическая:

$$\overline{X} = \frac{1}{n} \sum x_i n_i. \qquad (3.3.3)$$

Точечная оценка вероятности p_i определяется как относительная частота:

$$\hat{p}_i = \frac{n_i}{n}. \qquad (3.3.4)$$

Для построения доверительного интервала параметра а – математического ожидания нормального распределения составляют выборочную характеристику (*статистику*), функционально зависимую от наблюдений и связанную с а, например, для повторного отбора:

$$u = \frac{\overline{X} - a}{\sigma(\overline{X})} = \frac{\overline{X} - a}{\frac{\sigma}{\sqrt{n}}}. \qquad (3.3.5)$$

Статистика *u* распределена по нормальному закону распределения с математическим ожиданием $a = 0$ и средним квадратическим отклонением σ=1. Отсюда,

$$P(|u| < u_{\alpha/2}) = 1 - \alpha \text{ или } 2\Phi(u_{\alpha/2}) = 1 - \alpha ,$$

где Ф − функция Лапласа, $u_{\alpha/2}$ − квантиль нормального закона распределения, соответствующая уровню значимости α. Доверительный интервал для параметра а:

$$\overline{X} - u_{\alpha/2}\frac{\sigma}{\sqrt{n}} < a < \overline{X} + u_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \qquad (3.3.6)$$

где $\Delta = u_{\alpha/2}\sigma(\overline{X})$- предельная ошибка выборочной средней.

В более общем случае выбор статистики зависит от вида и объема выборки (табл. 8).

Таблица 8 – Формулы предельной ошибки и необходимого объема выборки для различных случаев отбора

Выборка		Собственно-случайная		Типическая		Серийная	
		повторная	бесповторная	повтор втор-ная	бесповторная	повторная	Бесповторная
Предельная ошибка, Δ	средней, $\overline{X}$	$t\sqrt{\frac{\sigma^2}{n}}$	$t\sqrt{\frac{\sigma^2}{n}(1-\frac{n}{N})}$	$t\sqrt{\frac{\overline{\sigma^2}}{n}}$	$t\sqrt{\frac{\overline{\sigma^2}}{n}(1-\frac{n}{N})}$	$t\sqrt{\frac{\delta^2_{м.с.}}{n_c}}$	$t\sqrt{\frac{\delta^2_{м.с.}}{n_c}(1-\frac{n_c}{N_c})}$
	доли, $\hat{P}$	$t\sqrt{\frac{pq}{n}}$	$t\sqrt{\frac{pq}{n}(1-\frac{n}{N})}$	$t\sqrt{\frac{\overline{pq}}{n}}$	$t\sqrt{\frac{\overline{pq}}{n}(1-\frac{n}{N})}$	$t\sqrt{\frac{pq_{м.с.}}{n_c}}$	$t\sqrt{\frac{pq_{м.с.}}{n_c}(1-\frac{n_c}{N_c})}$
Необходимая численность, n	средней, $\overline{X}$	$\frac{t^2\sigma^2}{\Delta^2}$	$\frac{t^2\sigma^2 N}{t^2\sigma^2 + \Delta^2 N}$	$\frac{t^2\overline{\sigma^2}}{\Delta^2}$	$\frac{t^2\overline{\sigma^2}N}{t^2\overline{\sigma^2} + \Delta^2 N}$	$\frac{t^2\delta^2_{м.с.}}{\Delta^2}$	$\frac{t^2\delta^2_{м.с.}N_c}{t^2\delta^2_{м.с.} + \Delta^2 N_c}$
	доли, $\hat{P}$	$\frac{t^2 pq}{\Delta^2}$	$\frac{t^2 Npq}{t^2 pq + \Delta^2 N}$	$\frac{t^2\overline{pq}}{\Delta^2}$	$\frac{t^2 N\overline{pq}}{t^2\overline{pq} + \Delta^2 N}$	$\frac{t^2 pq_{м.с.}}{\Delta^2}$	$\frac{t^2 pq_{м.с.}N_c}{t^2 pq_{м.с.} + \Delta^2 N_c}$

В табл.8:

1) t – квантиль распределения, соответствующая уровню значимости α,

   а) при $n \geq 30$ $t = u_{\alpha/2}$ - квантиль нормального закона распределения (прил.1),

   б) при $n<30$ t - квантиль распределения Стьюдента с $v=n-1$ степенями свободы для двусторонней области (прил.3);

2) σ^2 – выборочная дисперсия,

   а) при $n \geq 30$ $\sigma^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{n}$,

   б) при $n<30$ вместо σ^2 берут $S^2 = \frac{\sum (x_i - \bar{x})^2 n_i}{n-1}$;

3) pq - дисперсия относительной частоты в схеме повторных независимых испытаний;

4) N - объем генеральной совокупности;

5) n - объем выборки;

6) $\overline{\sigma^2}$ - средняя арифметическая групповых дисперсий (внутригрупповая дисперсия);

7) $\overline{pq}$ - средняя арифметическая дисперсий групповых долей;

8) $\delta^2_{\text{м.с.}}$ - межсерийная дисперсия;

9) $pq_{\text{м.с.}}$- межсерийная дисперсия доли;

10) N_c - число серий в генеральной совокупности;

11) n_c - число отобранных серий (объем выборки);

12) Δ - предельная ошибка выборки ($\Delta = u_{\alpha/2}\sigma(\overline{X})$ или $\Delta = u_{\alpha/2}\sigma(\hat{p})$).

Замечание.

1. Собственно-случайная выборка применяется, например, в сельском хозяйстве, другие типы выборок чаще применяются в социально-экономических исследованиях.
2. «Средняя ошибка типической выборки» ≤ «средняя ошибка механической выборки» ≤ «средняя ошибка собственно случайной выборки», поэтому на практике при расчете оценок механической выборки используют формулы собственно-случайной повторной выборки.
3. Ошибка серийной выборки меньше ошибки собственно-случайной выборки, однако, зависит от числа серий n_c и поэтому может завышаться.

Пример 3.3.1. При уровне значимости α=0,05 определить доверительный интервал для средней численности работников на 100 га сельскохозяйственных угодий в примере 3.1.2, учитывая, что проводилась 10% случайная бесповторная выборка. При тех же условиях найти необходимый объём выборки для уменьшения предельной ошибки в два раза.

Решение. Средняя численность работников на 100 га сельскохозяйственных угодий составляет 8,61. Доверительный интервал для средней определяется по формуле:

$$\overline{X} - \Delta_{\bar{x}} < \overline{X}_0 < \overline{X} + \Delta_{\bar{x}},$$

где $\overline{X}$ - выборочная средняя,

$\overline{X}_0$ - средняя генеральной совокупности,

$\Delta_{\bar{x}}$ - предельная ошибка выборки для средней.

Предельная ошибка выборки при случайном бесповторном отборе определяется по формуле (табл. 6):

$$\Delta_{\bar{x}} = t\sigma(\bar{X}) = t\sqrt{\frac{\sigma^2}{n}(1 - \frac{n}{N})},$$

где t - квантиль нормального закона распределения, при α=0,05, t=1,96 (Приложение 1);

N - объем генеральной совокупности, где $N = 10 \cdot n = 600$;

σ^2 - выборочная оценка дисперсии генеральной совокупности; так как объем выборочной совокупности n=60>30, то в качестве выборочной дисперсии используем σ^2=5,98.

Имеем $\Delta_{\bar{x}} = 1{,}96\sqrt{\frac{5{,}98}{60}\left(1 - \frac{60}{600}\right)} = 0{,}59.$

Значит, с доверительной вероятностью 1-α=0,95, можно утверждать, что средняя численность работников на 100 га сельскохозяйственных угодий во всей совокупности хозяйств находится в границах:

$\bar{X} \pm \Delta_{\bar{x}} = 8{,}61 \pm 0{,}59$, то есть от 8,02 до 9,2.

Необходимый объем выборки, для того, чтобы предельная ошибка не превышала $0{,}5 \cdot \Delta_{\bar{x}}$, при заданном уровне значимости α, в случае случайного бесповторного отбора определяется по формуле (табл. 5):

$$n = \frac{t^2\sigma^2 N}{t^2\sigma^2 + \Delta^2 N}.$$

Следовательно,

$$n = \frac{1{,}96^2 \cdot 600 \cdot 2{,}45^2}{1{,}96^2 \cdot 2{,}45^2 + 0{,}295^2 \cdot 600} = \frac{13835{,}5}{23{,}059 + 52{,}215} = 184.$$

Значит, для уменьшения предельной ошибки в два раза, объем совокупности необходимо увеличить в три раза.

Контрольные задания 3.3.

1. Для определения потерь зерна при уборке случайным способом проведено 100 измерений. Средняя величина потерь составила 1,8 ц с одного гектара посевов, при среднем квадратическом отклонении 0,5 ц с га. С доверительной вероятностью 0,95 определить границы, в которых будет находиться средняя величина потерь зерна с 1 га и возможная величина потерь, если площадь уборки зерновых составила 640 га.
2. С помощью случайной выборки изучалось время выполнения производственной операции рабочими бригады. На основании 60 наблюдений установлено, что в среднем на выполнение производственной операции затрачивалось 0,5 часа, при среднем квадратическом отклонении 0,12 часа. Считая время выполнения производственной операции нормально распределенной

случайной величиной, определить границы, в которых находится среднее время выполнения производственной операции всех рабочих с доверительной вероятностью: а) 0,9; б) 0,95.

3.Случайным бесповторным способом изучались остатки горюче-смазочных материалов на складе предприятий. Обследовано 110 предприятий из 750. Средние остатки составили 150 т, при среднем квадратическом отклонении 42 т. С доверительной вероятностью 0,95 определить границы, в которых будут находиться средние остатки горюче-смазочных материалов на одно предприятие и общие остатки горюче-смазочных материалов.

4. Считая данные задачи 1 контрольных заданий 3.2 результатом 20% выборки, определить: а) несмещенные оценки математического ожидания, дисперсии и среднего квадратического отклонения разряда рабочих; б) доверительный интервал для математического ожидания с доверительной вероятностью 0,95; в) вероятность того, что интервал $(0{,}95\,\overline{X}_в;\ 1{,}05\,\overline{X}_в)$ покроет математическое ожидание разряда рабочего; г) объем выборки, при котором с доверительной вероятностью 0,95 предельная ошибка выборки уменьшится в 1,5 раза, при сохранении уровня остальных характеристик.

5. Считая данные задачи 2 контрольных заданий 3.2 результатом 20% выборки, определить: а) несмещенные оценки математического ожидания, дисперсии и среднего квадратического отклонения числа производственных подразделений в расчете на одно сельскохозяйственное предприятие; б) доверительный интервал для математического ожидания с доверительной вероятностью 0,9; в) объем выборки, при котором с доверительной вероятностью 0,9 предельная ошибка выборки уменьшится в 2 раза, при сохранении уровня остальных характеристик.

6. Считая данные задачи 3 контрольных заданий 3.2 результатом 20% случайной бесповторной выборки, определить: а) несмещенные оценки математического ожидания, дисперсии и среднего квадратического отклонения изучаемого параметра; б) доверительный интервал для математического ожидания с доверительной вероятностью 0,95; в) вероятность того, что интервал $(0{,}95\,\overline{x}_в;\ 1{,}05\,\overline{x}_в)$ покроет математическое ожидание изучаемого параметра; г) объем выборки, при котором с доверительной вероятностью 0,95 предельная ошибка выборки уменьшится в 2 раза, при сохранении уровня остальных характеристик.

7. В районе имеется 10000 дачных участков населения. В результате выборочного обследования 300 дачных участков оказалось, что средняя выборочная урожайность овощей составила 250 ц с гектара при среднем квадратическом отклонении 60 ц с га. Известно, что 40% общей площади посевов овощей занимали помидоры. С доверительной вероятностью 0,95 определить границы, в которых будет находиться средняя урожайность овощей на всех дачных участках и удельный вес посевов помидор. Сколько необходимо обследовать дачных участков, чтобы предельная ошибка выборки по признакам уменьшилась в 1,5 раза?

8. Для определения влажности зерна случайным способом было взято 25 проб. Средний процент влажности зерна составил 16%, а выборочное среднее квадратическое отклонение 2,5%. Определить: а) несмещенные оценки математического ожидания, дисперсии и среднего квадратического отклонения; б) интервал, который покроет математическое ожидание с доверительной вероятностью, 0,95.
9. Вероятность изготовления продукции высшего качества фирмой составляет 0,9. Сколько необходимо обследовать единиц продукции, чтобы с доверительной вероятностью 0,95 можно было утверждать, что доля продукции высшего качества по выборке будет отклоняться от постоянной вероятности по модулю не более чем на 0,03?
10. Случайным бесповторным способом проведено выборочное обследование семей района. Из 1300 семей обследовано 130, по которым определен душевой доход на одного члена семьи, представленный в виде интервального вариационного ряда. Распределение семей по величине месячного дохода на одного члена семьи

Группы семей по месячному доходу на одного члена семьи, руб.	До 500	500-1000	1000-1500	1500-2000	Свыше 2000
Число семей	23	36	44	17	10

С доверительной вероятностью 0,95 определить границы, в которых будет находиться средний месячный доход на одного члена семьи по району, а также доля семей с доходами, менее 1000 руб. на одного члена семьи.
11. На фирме проведен выборочный опрос 10% работников по вопросам изменения условий труда. Из 90 работников основного производства за изменение условий труда высказалось 65 человек, из 30 работников вспомогательного производства – 20, а из 25 работников, занятых управлением фирмой - 21. С доверительной вероятностью 0,95 определить границы, в которых будет находиться доля работников фирмы, поддерживающих изменение условий труда.
12. Для определения влияния микроэлементов на результаты откорма свиней проведен опыт на 8 группах животных. Рационы отличаются набором и дозами микроэлементов:

Результаты откорма свиней в опыте

Рацион	Поголовье свиней, гол.	Среднесуточный прирост живой массы, г	Среднее квадратическое отклонение, г
1	90	500	40
2	75	575	45
3	100	610	54
4	50	450	52
5	70	590	65
6	60	650	70
7	110	490	48
8	80	540	62

С доверительной вероятностью 0,95 определить границы, в которых будет находиться среднесуточный прирост свиней по каждому рациону и по опыту в целом.

13. Взято 16 проб молока, поступившего на реализацию из акционерного сельскохозяйственного предприятия. Средняя жирность молока составила 3,7%, при среднем квадратическом отклонении 0,5%. Какова вероятность того, что средняя жирность молока всех партий не выйдет за границы от 3,6% до 3,8%?
14. Тридцать восемь студентов университета сдали экзамен по статистике на отличные и хорошие отметки. Чему равна вероятность того, что в случайной выборке из 100 студентов по крайней мере 30 окажутся с хорошими и отличными оценками по статистике?
15. Автотранспортная компания желает оценить среднее время транзита грузов из столицы в северные регионы страны. Случайная выборка 20 партий товаров дала $\overline{X}$=2,6 дней, s=0,4 дня. Постройте 99%-ый доверительный интервал для среднего времени транзита товаров.
16. Предположим, что среднее время пребывания в очереди к кассиру универмага составляет 12 мин со средним квадратическим отклонением 3 мин. Если вы отобрали случайным образом 5 покупателей, то чему равна вероятность того, что их время пребывания в очереди составит по крайней мере 10 мин? Чему равна средняя выборочная времени ожидания в очереди? Чему равно среднее квадратическое отклонение выборочной средней?
17. Из 500 выпускников средних школ города 72% собираются поступать в институт. Чему равна вероятность того, что среди случайно отобранных выпускников доля желающих поступить в вуз окажется выше 80%?

3.4. Проверка статистических гипотез

Статистической гипотезой называется всякое высказывание о генеральной совокупности, проверяемое по выборке. Статистические гипотезы делятся на:

1. Параметрические - это гипотезы, сформулированные относительно параметров (среднего значения, дисперсии и т.д.) распределения известного вида;
2. Непараметрические - это гипотезы, сформулированные относительно вида распределения (например, определение по выборке степени нормальности генеральной совокупности).

Процесс использования выборки для проверки гипотезы называется *статистическим доказательством*. Основную выдвигаемую гипотезу называют *нулевой* H_0. Наряду с нулевой гипотезой рассматривают ей *альтернативную* H_1. Например, H_0: $M(x)=1$, математическое ожидание генеральной совокупности равно 1; H_1: $M(x)>1$, или $M(x)<1$, или $M(x)\neq 1$ (математическое ожидание больше 1, или меньше 1, или не равно 1).

Выбор между гипотезами H_0 и H_1 может сопровождаться ошибками двух родов. Ошибка *первого рода* α означает вероятность принятия H_1, если верна гипотеза H_0: $\alpha = p(H_1/H_0)$. Ошибка *второго рода* означает вероятность принятия H_0, если верна гипотеза H_1: $\beta = p(H_0/H_1)$. Существует правильное решение двух видов: $p(H_0/H_0) = 1-\alpha$ и $p(H_1/H_1) = 1-\beta$ (табл. 9).

Таблица 9 – Ошибки первого и второго рода

Принятая гипотеза	H_0	H_1
H_0 - верна	$P(H_0/H_0)$=1-α	$P(H_1/H_0)$=α
H_0 – неверна	$P(H_0/H_1)$=β	$P(H_1/H_1)$=1-β

Правило, по которому принимается решение о том, что верна или не верна гипотеза H_0, называется *критерием,* где:

$\alpha=P(H_1/H_0)$ - *уровень значимости критерия*;

$M=1-\beta=P(H_1/H_1)$ - *мощность критерия*.

Статистическим критерием K называют случайную величину, с помощью которой принимают решение о принятии или отклонении H_0.

Замечание. Для проверки параметрических гипотез используют *критерии значимости*, основанные на статистиках: u, χ^2, t, F (приложения 5-7). Непараметрические гипотезы проверяют с помощью *критериев согласия*, использующих статистики распределений: χ^2, Колмогорова-Смирнова [1, 2, 6, 10] и т.д.

Например, H_0: $M(x)$=10. В зависимости от альтернативной гипотезы рассматривают три случая.

1. Если $H_1 : M(x) \neq 10$.

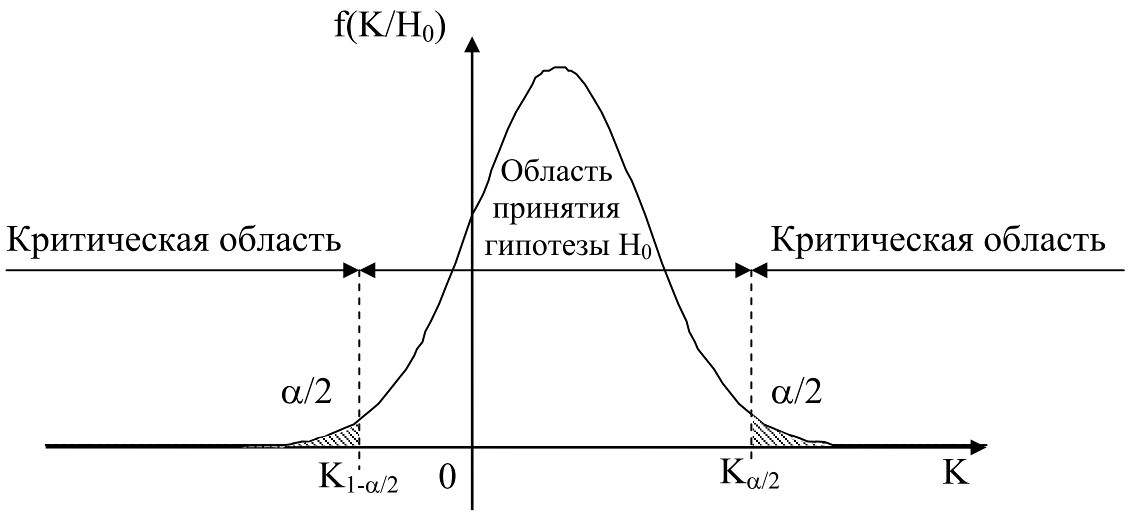


Рис 43.1 - Двусторонняя критическая область

В этом случае рассматривают двустороннюю критическую область и используют дифференциальную функцию $f\,(K/H_0)$, для определения соответствующих квантилей (границ области принятия гипотезы - левой ($K_{1-\alpha/2}$) и правой ($K_{\alpha/2}$)). Площадь под криволинейной трапецией дифференциальной функции слева от $K_{1-\alpha/2}$ и справа от $K_{\alpha/2}$ равна α/2.

Общая площадь ограниченная криволинейной трапецией дифференциальной функции, квантилями и осью абсцисс равна (1 - α) (рис. 43.1):

$$p\left(K_{1-\alpha/2} < K < K_{\alpha/2}\right) = 1 - \alpha. \qquad (3.5.1)$$

2. Если $H_1 : M(x) > 10,$ то рассматривается правосторонняя критическая область (площадь под криволинейной трапецией справа от K_α равна α) (рис. 43.2):

$$P\left(K > K_\alpha\right) = \int_{K_\alpha}^{+\infty} f\left(K/H_0\right) dK = \alpha. \qquad (3.5.2)$$

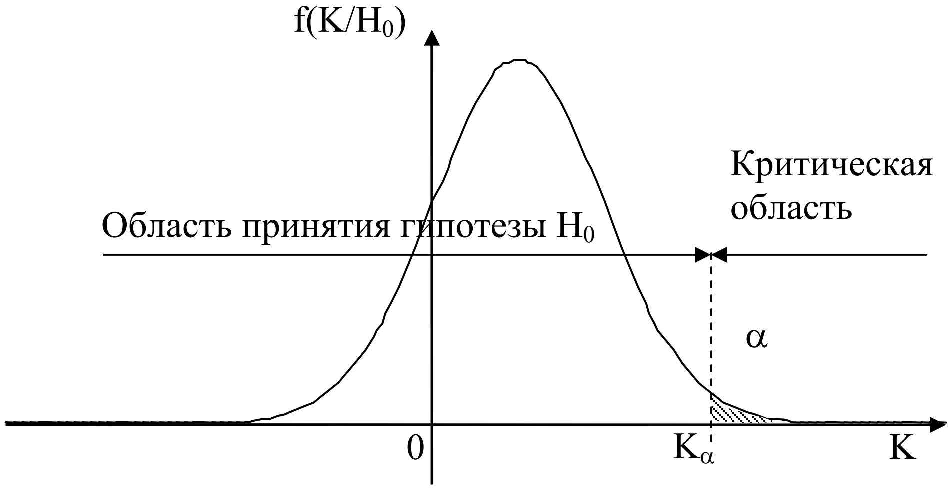


Рис.43.2 - Правосторонняя критическая область

Рис. 43.3- Левосторонняя критическая область

3. Если $H_0 : M(x) < 10$, то рассматривается левосторонняя критическая область (площадь под криволинейной трапецией слева от $K_{1-\alpha}$ равна α) (рис. 43.3):

$$p\left(K < K_{1-\alpha}\right) = \int_{-\infty}^{K_{1-\alpha}} f\left(K/H_0\right) dK = \alpha . \qquad (3.5.3)$$

3.5. Алгоритм проверки статистических гипотез, примеры и контрольные задания

1. Располагая выборочными данными ($x_1, x_2, \dots, x_n$), формируют нулевую гипотезу H_0 и конкурирующую гипотезу H_1.
2. Задают уровень значимости α (обычно принимают α =0,1; 0,01; 0,05; 0,001).
3. Рассматривается выборочная статистика наблюдений (критерий) К, обычно одна из перечисленных ниже:
   u - нормальное распределение;
   χ^2- распределение Пирсона (хи – квадрат);
   t - распределение Стьюдента;
   F - распределение Фишера - Снедекора.
4. На основании выборки $x_1, x_2, \dots x_n$ - определяют значение критерия (статистики) *К* (приложения 5-7). В зависимости от вида альтернативной гипотезы выбирают по соответствующей таблице квантили критерия для двусторонней $\left(K_{1-\frac{\alpha}{2}} \text{ и } K_{\frac{\alpha}{2}}\right)$ или односторонней области ($K_{1-\alpha}$ или K_{α}) (приложения 1-4). Если значения критерия попадают в критическую область, то H_0 отвергается; в противном случае принимается гипотеза H_0 и считается, что H_0 не противоречит выборочным данным (при этом существует возможность ошибки с вероятностью равной α).

Замечание. Следует отметить, что возможность принятия гипотезы происходит из принципа невозможности наступления маловероятных событий. Те же события, вероятность которых близка к 1, принимаются за достоверные. Возникает проблема выбора уровня риска (уровня значимости α).

В одних случаях возможно пренебрегать событиями p<0,05, в других нельзя пренебрегать событиями, которые могут появиться с p=0,001 (разрушение сооружений, транспортных средств и т.д.).

Рассмотрим примеры применения методов проверки статистических гипотез.

Пример 3.5.1. [1] Результаты исследований прочности на сжатие (СВ *X*) 200 образцов бетона представлены в виде сгруппированного статистического ряда.

---

[1] Пример заимствован из [2].

Интервалы прочности, кг/см^2	Среднее значение интервала, x_i	Частота, n_i
190-200	195	10
200-210	205	26
210-220	215	56
220-230	225	64
230-240	235	30
240-250	245	14

$\sum n_i = n = 200$

Требуется проверить нулевую гипотезу о нормальном законе распределения прочности образцов бетона на сжатие. Уровень значимости принять α=0,001 .

Решение. Из условия следует, что точные параметры гипотетического нормального закона нам неизвестны, поэтому нулевую гипотезу (H_0) словесно можно сформулировать следующим образом: $F(x)$ является функцией нормального распределения с параметрами $M(X) = a = \overline{X}$ и $D(X) = \sigma^2$.

Для проверки этой нулевой гипотезы найдем точечные оценки математического ожидания и среднего квадратического отклонения нормально распределенной случайной величины:

$$\overline{X} = \frac{\sum x_i n_i}{n} = 221\,кг/см^2;\ \sigma = \sqrt{\frac{1}{n}\sum (x_i - \overline{X})^2 n_i} = 12{,}33\,кг/см^2.$$

При проверке гипотезы о нормальном распределении генеральной совокупности сравниваются эмпирические (наблюдаемые) и теоретические (вычисленные в предположении нормальности распределения) частоты. Для этого используются статистика χ^2 – Пирсона с $\nu=k-r-1$ степенями свободы (k – число групп, r – число оцениваемых параметров, в настоящем примере оценивались математическое ожидание и среднее квадратическое отклонение, следовательно, r = 2). Если $\chi^2_{расч.} \geq \chi^2_{кр.}$, то нулевая гипотеза отвергается и считается, что предположение о нормальности распределения не согласуется с опытными данными. В противном случае ($\chi^2_{расч.} < \chi^2_{кр.}$) нулевая гипотеза принимается.

Вычислим теоретические вероятности p_i, попадания СВ $X \to N$ (221; 12,33) в частичные интервалы $[x_{i-1}; x_i)$ по формуле:

$$p_i = P(x_i \leq X \leq x_{i+1}) = \Phi(u_{i+1}) - \Phi(u_i),\ (i=1,2,...,k),$$

где

$$u_i = \frac{x_i - \overline{X}}{\sigma};\ \Phi(\mathrm{u_i}) = \frac{1}{\sqrt{2\pi}}\int_0^{u_i} e^{-\frac{t^2}{2}}dt\ .$$

Дальнейшие вычисления, необходимые для определения расчетного значения выборочной статистики χ^2, проведем в следующей таблице.

Интервалы изменения наблюдаемых значений СВ X	Частоты n_i	Нормированные интервалы $[u_i; u_{i+1}]$	$p_i = [\Phi(u_{i+1}) - \Phi(u_i)]$	np_i	$(n_i-np_i)^2$	$\frac{(n_i - np_i)^2}{np_i}$
190—200	10	$(-\infty; -1,70)$	0,045	9	1	0,11
200—210	26	$[-1,70; -0,89)$	0,142	28,4	5,76	0,20
210—220	56	$[-0,89; -0,08)$	0,281	56,2	0,04	0,00
220—230	64	$[-0,08; +0,73)$	0,299	59,8	17,61	0,29
230—240	30	$[+0,73; +1,54)$	0,171	34,2	17,64	0,52
240—250	14	$[+1,54; +\infty)$	0,062	12,4	2,56	0,23
$\sum$	200	-	1,000	200,0		$\chi^2_{\text{расч.}}$=1,35

Замечание.

1. Так как СВ *X*, распределенная по нормальному закону, определена на $(-\infty; \infty)$, то наименьшее значение стандартизованной переменной $\frac{190-221}{12,33} = -2,51$ заменено на $-\infty$, а наибольшее значение $\frac{250-221}{12,33} = 2,35$ заменено на $+\infty$.
2. Применение критерия χ^2, для проверки гипотезы о нормальности распределения предполагает наличие в каждом частичном интервале не менее пяти элементов, в противном случае желательно объединять эти интервалы с соседними.
3. Проверка гипотезы о принадлежности СВ показательному, биномиальному, пуассоновскому или другому распределению основывается на применении в описанном алгоритме соответствующих интегральных функций.

В результате вычислений получили $\chi^2_{\text{расч.}}$=1,35. Найдем по таблице квантилей χ^2-распределения (прил. 2), при заданном уровне значимости α=0,001 и числе степеней свободы $\nu=k-r-1=6-2-1=3$, критическое значение:

$$\chi^2_{\text{кр.}} = \chi^2_{0,001;\,3} = 16,266.$$

Так как $\chi^2_{\text{расч.}} < \chi^2_{\text{кр.}}$ (1,35<16,266), то нет оснований для отклонения нулевой гипотезы о нормальном законе распределения прочности на сжатие с параметрами a=221 и σ^2=152.

Пример 3.5.2. Из нормальной генеральной совокупности сельскохозяйственных предприятий, рассматриваемых по показателю урожайности пшеницы, с известным средним квадратическим отклонением σ=9,4 и генеральной средней $\overline{X}_0$=38,1, извлечена выборка объема n=50. По ней найдена выборочная средняя $\overline{X}$=42. Требуется при уровне значимости α=0,05 проверить нулевую гипотезу H_0:

а) $\overline{X} = \overline{X}_0 = 38,1$, при конкурирующей гипотезе H_1: $\overline{X} \neq 38,1$;

б) $\overline{X} = \overline{X}_0 = 38,1$, при конкурирующей гипотезе H_1: $\overline{X} < 38,1$;

в) $\overline{X} = \overline{X}_0 = 38,1$, при конкурирующей гипотезе H_1: $\overline{X} > 38,1$.

Решение. Необходимо рассмотреть критерий $K=u$, где $u_{расч.}=\dfrac{(\overline{X}-\overline{X}_0)\cdot\sqrt{n}}{\sigma}=\dfrac{(42-38{,}1)\cdot\sqrt{50}}{9{,}4}=2{,}93.$

а) По условию конкурирующая гипотеза имеет вид $\overline{X}\neq38{,}1$, поэтому критическая область двусторонняя. Найдем критическую точку из равенства $\Phi(u_{кр.,\alpha/2})=(1-\alpha)/2=(1-0{,}05)/2=0{,}475$. Согласно приложения 1: $u_{кр.}=1{,}96$.

$\left|u_{расч.}\right|>u_{кр.}$, поэтому следует отклонить нулевую гипотезу, то есть выборочная и гипотетическая генеральная средняя статистически различаются значимо.

б) По условию конкурирующая гипотеза имеет вид $\overline{X}<38{,}1$, поэтому критическая область левосторонняя. Найдем критическую точку из равенства $\Phi(u_{кр.,\alpha})=(1-2\alpha)/2=(1-0{,}1)/2=0{,}45$. Согласно приложения 1: $u_{кр.}=-1{,}65$. $u_{расч.}>u_{кр.}$, поэтому следует принять нулевую гипотезу H_0, то есть выборочная и гипотетическая генеральная средняя статистически различаются не значимо.

в) По условию конкурирующая гипотеза имеет вид $\overline{X}>38{,}1$, поэтому критическая область правосторонняя. Найдем критическую точку из равенства $\Phi(u_{кр.,\alpha})=(1-2\alpha)/2=(1-0{,}1)/2=0{,}45$. Согласно приложения 1: $u_{кр.}=+1{,}65$. $u_{расч.}>u_{кр.}.$ поэтому следует отклонить нулевую гипотезу H_0, то есть выборочная и гипотетическая генеральная средняя статистически различаются значимо.

Замечание. 1) Вид альтернативной гипотезы существенно влияет на статистический вывод, поэтому ее нужно выдвигать, исходя из реального смысла рассматриваемой задачи. Практически в качестве альтернативной гипотезы следует выбрать такую, которая побуждает лицо принимающее решение к действиям. Например, при проверке гипотезы о том, что средний вес взятого продукта соответствует объявленному номиналу (H_0), в качестве альтернативной гипотезы (H_1) следует принять гипотезу – средний вес меньше номинала (покупателя обвешивают). 2) Отклонение нулевой гипотезы не есть доказательство верности альтернативной. Если необходимо проверить H_1, то для этого нужно взять уже другую выборку. 3)Принимая H_0 мы отнюдь не доказываем ее абсолютную истинность. Можно лишь утверждать при принятии H_0 то, что она не противоречит опытным данным с вероятностью (1- α) (например, при α =0,05 нет противоречия в 95 случаях из 100).

Пример 3.5.3. Оценить существенность различий в средней урожайности двух сортов озимой пшеницы. Если для первого сорта средняя урожайность $\overline{X}_1=35{,}6$ ц/га и выборочная дисперсия $S_1^2=8{,}05$, а для второго сорта средняя урожайность $\overline{X}_2=48{,}4$ и выборочная дисперсия $S_2^2=14{,}31$. Объемы выборок n_1=5 и n_2=5 соответственно.

Решение. Выдвигаем нулевую гипотезу о том, что средние урожайности двух сортов пшеницы не отличаются друг от друга, т.е. $H_0:\overline{X}_1=\overline{X}_2$, при альтернативной гипотезе $H_1:\overline{X}_1\neq\overline{X}_2$ -урожайности существенно различны.

Примем уровень значимости $\alpha=0{,}05$.Так как выборки независимы, причем $n_1=n_2$,то применим критерий t – Стьюдента с $k=n_1+n_2-2$ степенями свободы.

$$t_{расч.}=\frac{\left|\overline{X}_1-\overline{X}_2\right|}{\sqrt{\dfrac{S_1^2}{n_1}+\dfrac{S_2^2}{n_2}}}=\frac{48{,}4-35{,}6}{\sqrt{\dfrac{8{,}05}{5}+\dfrac{14{,}31}{5}}}=\frac{12{,}8}{2{,}11}=6{,}07\,.$$

По приложению 3 определим критическое значение t-распределения:

$$t_{кр}=t_{0,5;10}=2{,}31\,,$$

при числе степеней свободы $k = 5 + 5 - 2 = 10$.

Так как $t_{расч.} > t_{кр}$, то нулевую гипотезу следует отклонить. Следовательно, два сорта пшеницы отличаются статистически значимо по величине урожайности.

Пример 3.5.4. Проверить нулевую гипотезу о равенстве двух выборочных площадей посева овощей на одно хозяйство на основании решения задания 3.2.4.($H_0 : \overline{X}_1 = \overline{X}_2$), при конкурирующей гипотезе $H_1 : \overline{X}_1 \neq \overline{X}_2$ (уровень значимости принять равным 0,05). Использовать критерий t – Стьюдента с $k = (n_1 + n_2 - 2)$ степенями свободы:

$$t_{расч} = \frac{\left|\overline{X}_1 - \overline{X}_2\right|}{\sqrt{\frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \cdot \frac{n_1 + n_2}{n_1 \cdot n_2}}}.$$

Пример 3.5.5. Два сорта озимой пшеницы испытывались на одинаковом числе участков на протяжении семи лет (табл.10).

При уровне значимости $\alpha = 0{,}05$ проверить нулевую гипотезу о существенности различий в урожайности двух сортов озимой пшеницы.

Решение. Так как имеются две зависимости выборки, т.е. существует определенная корреляция между урожайностью сортов по годам, то необходимо оценить значимость не разности двух выборочных средних, а средней разности.

Выдвигаем нулевую гипотезу: средняя величина различий в урожайности пшеницы равна нулю, $H_0 : \overline{X_1 - X_2} = 0$ при $H_1 : \overline{X_1 - X_2} \neq 0$.

Таблица 10 – Вспомогательная таблица для расчета ошибки средней разности

Годы	Урожайность, ц/га		Разность $d_i = x_{1i} - x_{2i}$	$(d_i - \overline{d})$	$(d_i - \overline{d})^2$
	x_{2i}	x_{1i}			
1995	46	53	6	1	1
1996	43	48	5	0	0
1997	46	45	-1	-6	36
1998	51	56	5	0	0
1999	52	58	6	1	1
2000	48	55	7	2	4
2001	52	59	7	2	4
Сумма			35	0	40

3. По данным таблицы 10 найдем среднюю разность $\overline{d}$ и ошибку средней разности $S_{\overline{d}}$:

$$\overline{d} = \frac{\Sigma d_i}{n} = \frac{35}{7} = 5\,; \quad S_{\overline{d}} = \sqrt{\frac{\Sigma\left(d_i - \overline{d}\right)^2}{n(n-1)}} = \sqrt{\frac{40}{7(7-1)}} = 0{,}98\,; \quad t_{расч.} = \frac{\overline{d}}{S_{\overline{d}}} = \frac{5}{0{,}98} = 5{,}10\,,$$

где $d_i = x_{1i} - x_{2i}$, n – число пар наблюдений.

При α =0,05; $k=n$-1=7-1=6, $t_{кр.}$=2,45

Сопоставив расчётное значение t с критическим, можно сделать вывод, что два сорта существенно различаются по уровню урожайности.

Контрольные задания 3.5.

1. По данным приложения 8 по одному показателю случайным способом провести 30% выборку. По выборочной совокупности определить среднее значение. При уровне значимости 0,05 проверить нулевую гипотезу о равенстве выборочного параметра среднему значению по всей совокупности.
2. Проверить гипотезу о равенстве средних урожайностей в двух хозяйствах, если в результате случайной выборки получены следующие результаты:

1 хозяйство

Урожайность, ц с 1 га	Число участков
x_i	n_i
25-35	30
35-45	30
45-55	40
	n=100

2 хозяйство

Урожайность, ц с 1 га	Число участков
y_i	m_i
15-25	30
25-35	30
35-45	40
45-55	50
	m=150

3. По двум независимым выборкам объема n_1 и n_2, извлеченным из нормальных генеральных совокупностей, проверить гипотезу о равенстве средних, при уровне значимости α=0,01, если:

   а) $\overline{x} = 50; \quad \overline{y} = 45; \quad D(X) = 1200; \quad D(Y) = 2025; \quad n_1 = 35; \quad n_2 = 45;$

   б) $\overline{x} = 70; \quad \overline{y} = 60; \quad D(X) = 1470; \quad D(Y) = 1320; \quad n_1 = 60; \quad n_2 = 40.$
4. Провести две случайные выборки по одному из показателей приложения 8, объемами n_1 и n_2. Проверить нулевую гипотезу о равенстве выборочных средних, при уровне значимости 0,05 (предполагается, что дисперсии неизвестны и одинаковы): а) $n_1=n_2=20$; б) $n_1=20$; $n_2=10$.
5. Проводилось испытание 8 сортов озимой пшеницы.

Повторения	Сорт							
	1	2	3	4	5	6	7	8
I	45	51	60	49	63	44	55	60
II	44	50	62	52	61	40	53	55
III	46	56	61	45	62	41	51	53
IV	44	52	56	48	56	43	58	57
V	47	54	61	47	62	45	54	54
VI	45	52	59	46	61	41	53	56

Каждый сорт высевался на 6 делянках одинаковой площади. При 5% уровне значимости проверить гипотезу о существенности различий в средней урожайности двух сортов озимой пшеницы (номера сортов даются студенту преподавателем). Урожайность озимой пшеницы, ц/га:

6. При уровне значимости 0,05 проверить гипотезу о равенстве среднего балла по теории вероятностей и математике.

Теория вероятностей	4	5	3	4	5	3	5	2	4	4	3	2	4	4	5	3
Математика	4	5	2	3	4	3	5	2	4	3	4	3	4	3	5	2

7. Произведено выборочное обследование 10% приусадебных участков восьми районов случайным бесповторным способом. Получены следующие результаты об урожайности овощей.

№ п/п	Урожайность, ц/га	Среднее квадратическое отклонение, ц/га	Доля овощей в площади участков, %	Число обследованных участков
1	215	30	30	100
2	246	35	35	80
3	305	32	40	150
4	220	24	50	120
5	164	20	36	60
6	280	23	65	70
7	340	40	45	90
8	316	36	53	100

При уровне значимости 0,05 по двум районам проверить гипотезы о равенстве: дисперсий, средних выборочных урожайностей, долей посевов овощей в площади приусадебных участков.

8. Результаты выступлений спортсменов оценивались двумя судьями по десятибалльной шкале.

Номер спортсмена		1	2	3	4	5	6	7	8	9	10
Оценка судьи	1	8,5	9	7,4	9,4	9,7	6,5	7,1	8,3	9,1	8,0
	2	8,3	9,1	7,7	9,3	9,2	6,0	7,3	8,1	9,1	7,9

При уровне значимости 0,05 проверить гипотезу о значимости различий в оценке выступлений спортсменов двумя судьями.

9. По результатам задачи 1 контрольных заданий 3.2 проверить гипотезу о нормальном распределении рабочих предприятия по разрядам.

10. По результатам задачи 2 контрольных заданий 3.2 проверить гипотезу о том, что число производственных подразделений на предприятиях распределяется по нормальному закону.

11. По результатам задачи 3 контрольных заданий 3.2 проверить гипотезу о нормальном распределении совокупности хозяйств по изучаемому признаку.

Индивидуальные задания к главе 3

1. Цель и задачи работы

Цель: показать знания теории точечных статистических оценок, продемонстрировать умение грамотно проводить обработку статистических рядов и на основании результатов вычислений делать заключение об изучаемом случайном явлении или процессе.

Задачи:
- уметь выбрать схему изучения статистических данных, подобрать необходимые методы и формулы для расчетов;
- сравнить результаты расчетов, полученных по разным выборкам и разными методами;
- закрепить навыки вычислений и анализа;
- правильно сформулировать статистические гипотезы и провести их проверку.

2. Общее описание задания

При выполнении домашнего задания необходимо провести обработку статистических данных с целью получения ответа на те вопросы, которые были поставлены перед статистиком-аналитиком в каждом конкретном случае (в соответствии с вариантом задания).

Исходные данные представлены результатами наблюдений в виде табл.11.

Таблица 11

X_j	X_1	X_2	……..	$X_к$
N_j	n_1	n_2	……..	$n_к$

Исходные данные необходимо представить в виде интервального вариационного ряда.

Каждый вариант задания может выполняться бригадой из четырех человек для того, чтобы можно было бы провести анализ результатов расчетов разными способами.

По данным таблиц наблюдений для каждого ряда распределения необходимо:

1.1. Вычислить статистики (оценки) положения, рассеяния; коэффициенты асимметрии, эксцесса.

1.2. Подобрать аппроксимирующие кривые распределения, используя нормальный закон распределения (дополнительно: используя любой другой закон распределения).

1.3. Проанализировать исходные данные и результаты расчетов, сделать предварительные (интуитивные) выводы, основываясь на практических вопросах задания.

2.1. Провести проверку статистических гипотез для всех статистик (оценок) и кривых распределения.
2.2. Провести сравнение результатов расчетов.
2.3. Ответить на практические вопросы задания (сделать выводы).

3. Варианты задания

Варианты 1.1, 1.2, 1.3, 1.4. Процесс извлечения гелия.

Исследуется работа промышленных агрегатов по процессу извлечения гелия (He) из природного Оренбургского газа. Целью работы установки является более полное (близкое к 100%), извлечение гелия. Технологический процесс считается отлаженным, если в готовом продукте (баллоне с гелием) содержится не менее 99% He (или 0,99 He и только 0,01 примесей); качество продукта считается очень хорошим, если извлечено более 99,8% He. Испытываются два технологических режима N1 и N2 для того, чтобы выбрать лучший по признаку наибольшего процента извлечения гелия. Кроме того, необходимо повысить точность измерений, поэтому испытанию подвергаются два способа измерений: A и B. Еще одной целью исследования является установление стандарта на процент гелия (интервальная оценка математического ожидания), который может быть задан для изучаемого производственного процесса.

Результаты наблюдений в различных ситуациях представлены таблицами 11.1 - 11.4 (данные предварительно упорядочены, в табл. 11.3 и 11.4 данные представлены в абсолютных единицах).

Таблица 11.1 – Технология $N1$, способ A; N_1 = 210

%He; X_j	98,50	98,75	98,80	98,93	98,99	99,10	99,20	99,32	99,40	99,55	99,65	99,90	99,98
n_j	10	10	15	20	25	30	20	20	15	15	5	15	10

Таблица 11.2 – Технология $N2$, способ B; N_2 = 120

%He; Xj	98,30	98,50	98,72	98,91	99,00	99,15	99,2	99,50	99,72	99,85	99,86
nj	2	2	4	10	6	10	24	30	26	4	2

Таблица 11.3 –Технология $N1$, способ B; N_3=42

%He; X_j	98,43	98,50	98,71	98,82	99,22	99,54	99,73	99,92
n_j	1	2	10	6	12	6	4	1

Таблица 11.4 – Технология $N2$, способ A; N_4=25

%He; X_j	98,62	98,83	99,00	99,45	99,62	99,71	99,84	99,90	99,93
n_j	1	3	3	5	6	3	2	1	1

Сформулируйте и проверьте статистические гипотезы, на основании которых можно выяснить:

- отличаются ли технологические режимы *N*1 и *N*2 и если да, то какой из них лучше?
- отличаются ли способы измерений А и В, какой из них точнее?
- какой стандарт можно установить на процентное содержание гелия?
- какое число измерений достаточно для того, чтобы определить, соответствует ли продукция стандарту?

Условия выполнения вариантов 1.1-1.4 (табл.11.5).

Таблица 11.5

Условия	1.1	1.2	1.3	1.4
Уровень значимости α	0,01; 0,05	0,005; 0,05	0,02; 0,05	0,025; 0,05
Точность вычислений (число знаков после запятой)	3	4	3	4

Варианты 2.1, 2.2, 2.3, 2.4. Процесс пропитки стеклоткани.

Стеклоткань после пропитки специальными смолами становится токопроводящей и используется для создания одежды с подогревом, а также для некоторых нагревательных устройств. Для получения ткани с заданным номиналом электрического сопротивления (R, Ом) квадратного сантиметра ткани подбираются соответствующие технологические режимы пропитки. Было проанализировано три режима: *N*1, *N*2, *N*3 производства ткани, обеспечивающих $R = 100$ Ом. Пропитывались ткани двух типов для того, чтобы выбрать один, обеспечивающий меньший разброс значений R около заданного номинала.

Результаты наблюдений представлены таблицами 12.1 - 12.5 (данные предварительно упорядочены).

Таблица 12.1–Режим *N*1, ткань А; $N_1 = 100$

$R; X_j$	86	88	93	98	101	104	109	115	122
n_j	5	12	20	30	10	13	6	3	1

Таблица 12.2 –Режим *N*2, ткань А; $N_2 = 125$

$R; X_j$	84	85	93	97	101	104	107	112	117
n_j	1	8	10	16	35	30	17	6	2

Таблица 12.3–Режим *N*3, ткань А; $N_3 = 90$

$R; X_j$	90	93	98	100	102	104	106
n_j	1	5	12	20	30	10	12

Таблица 12.4–Режим $N1$, ткань В; $N_4 = 25$

$R; X_j$	92	96	101	105	109
n_j	5	12	2	4	2

Таблица 12.5–Режим $N3$, ткань В; $N_5 = 160$

$R; X_j$	90	94	97	100	102	105	108	111	114
n_j	6	15	15	21	32	25	13	30	3

Сформулируйте и проверьте статистические гипотезы, на основании которых можно:

- сравнить различные технологические режимы и выбрать из них тот, который обеспечивает заданный номинал R = 100 Ом с наибольшей точностью;
- выбрать тип ткани А или В, который не влияет на точность процесса и заданный номинал;
- определить, сколько опытов достаточно проводить, чтобы уверенно контролировать качество продукции.

Условия выполнения вариантов 2.1-2.4 (табл.12.6).

Таблица 12.6

Условия	2.1	2.2	2.3	2.4
Уровень значимости α	0,01; 0,05	0,005; 0,05	0,02; 0,05	0,025; 0,05
Точность вычислений (число знаков после запятой)	3	2	3	2

Варианты 3.1, 3.2, 3.3, 3.4. Анализ продуктов питания.

Лаборатория проводит анализ продуктов питания с целью определения в них наличия вредных веществ. С определенным видом продуктов работают два лаборанта, результаты анализов сравниваются. Продукты поступают из двух пунктов. Лаборатория должна дать заключение, где производятся наиболее «чистые» продукты. Кроме того, руководителя лаборатории интересует вопрос: отличаются ли по точности результаты экспериментов у первого и второго лаборанта. Им было предложено независимо проанализировать одни и те же образцы. Для этих образцов необходимо было определить содержание вредного вещества X. В единице объема продукта количество X не должно превышать 0,015. Данные измерений представлены таблицами 13.1 - 13.4.

Таблица 13.1–Лаборант 1, пункт 1; N_1 = 120

X_j	0,0110	0,0120	0,0127	0,0130	0,0138	0,0014	0,0150	0,0156	0,0170	0,0180
n_j	2	2	7	16	30	35	20	5	2	1

Таблица 13.2 –Лаборант 1, пункт 2; N_2 = 25

X_j	0,0120	0,0128	0,0135	0,0140	0,0147	0,0156	0,0160
n_j	1	2	5	10	4	2	1

Таблица 13.3–Лаборант 2, пункт 1; N_3 = 110

X_j	0,0100	0,0120	0,0135	0,0142	0,0149	0,0152	0,0160	0,0175	0,0190
n_j	2	10	17	30	25	17	5	3	1

Таблица 13.4–Лаборант 2, пункт 2; N_4= 20

X_j	0,0115	0,0127	0,0136	0,0142	0,0150	0,0152	0,0165
n_j	1	1	3	10	3	1	1

Сформулируйте и проверьте статистические гипотезы, на основании которых можно выяснить:
- можно или нет двум пунктам поставки продуктов предъявить сертификат качества?
- одинакова ли квалификация обоих лаборантов (то есть, отличаются ли у них значимо результаты анализов)?
- сколько образцов достаточно брать для испытаний на первом и втором пунктах?

Условия выполнения вариантов 3.1-3.4 (табл. 13.5).

Таблица 13.5

Условия	3.1	3.2	3.3	3.4
Уровень значимости α	0,05; 0,02	0,05; 0,01	0,05; 0,005	0,05; 0,025
Точность вычислений (число знаков после запятой)	3	4	4	3

Варианты 4.1, 4.2, 4.3, 4.4. Курс ценных бумаг.

Изучаются колебания X_j (денежные единицы) курсов ценных бумаг (тип *N*1, *N*2, *N*3, *N*4), принадлежащих разным группам риска (риск оценивается величиной дисперсии) и в различные периоды времени. Исследования ведутся двумя независимыми аналитическими центрами А и В. Банк, заинтересованный в результатах анализа, в целях формирования «портфеля ценных бумаг», желает знать результаты классификации по группам. Сделав случайную выборку информации о колебании курсов, аналитики получили следующие данные (табл. 14.1 - 14.6). X_j - цена одного пакета ценных бумаг.

Таблица 14.1 – Бумаги *N*1, центр А; N_1 = 190

$X_j \cdot 10^2$	2	3	6	8	9	11	13	14	16	17	19	20
n_j	5	5	5	10	25	30	40	30	20	10	5	5

Таблица 14.2 – Бумаги *N*2, центр А; N_2 =132

$X_j \cdot 10^2$	7	8	9	10	11	12	13	14	15	16	17
n_j	1	5	5	10	25	20	25	20	15	5	1

Таблица 14.3 – Бумаги *N*2, центр В; N_3 = 93

$X_j \cdot 10^2$	8	9	10	11	12	13	14	15	16
n_j	2	3	15	20	30	15	5	2	1

Таблица 14.4 – Бумаги *N*3, центр А; N_4 = 175

$X_j \cdot 10^2$	3	5	7	8	9	11	13	14	16	17	19	21
n_j	1	5	10	20	30	40	35	15	10	5	3	1

Таблица 14.5 – Бумаги N4, центр В; N_5 = 90

$X_j \cdot 10^2$	9	10	11	12	13	14	15	16
n_j	1	2	10	25	30	15	5	2

Таблица 14.6 –Бумаги *N*4, центр А; N_6 = 22

$X_j \cdot 10^2$	11	12	13	14	15	16
n_j	1	5	10	3	2	1

Сформулируйте и проверьте статистические гипотезы, необходимые для ответа на вопросы:

- какие бумаги можно отнести к бумагам одинаковой группы риска?
- отличаются ли арифметические средние колебания курса?
- различны ли выводы аналитических центров А и В?
- какой тип бумаг Вы предпочтете купить, если Ваши средства ограничены суммой не более С денежных единиц за один пакет ценных бумаг?

Условия выполнения вариантов 4.1-4.4 (табл.14.7).

Таблица 14.7

Условия	4.1	4.2	4.3	4.4
Уровень значимости α	0,05; 0,01	0,05; 0,025	0,05; 0,005	0,05; 0,005
Допустимая цена (С) одного пакета ЦБ	1100	1000	1200	1300

Замечание. Анализу можно подвергать не все типы бумаг (по личному выбору).

Варианты 5.1, 5.2, 5.3, 5.4. Процесс обогащения руды.

На обогатительных фабриках происходит отделение частиц металла от пустой породы (после раздробления руды и последующей ее обработки). Одним из показателей качества готовой продукции - концентрата - являются классы крупности X_j (d, мк) частиц металла, входящих в него. В результате анализов, проведенных на одной из обогатительных фабрик Зангезурского медно-молибденового рудника, были получены данные по распределениям классов крупности при различных технологических режимах. При этом проходили испытания нового автоматического прибора (гранулометра), по измерению классов крупности. Точность анализов гранулометра сравнивалась с точностью при традиционных лабораторных способах измерений. Результаты анализов представлены таблицами 15.1 - 15.6.

Таблица 15.1 – Технология $N1$, лаб. анализ; N_1 =100

X_j	0,63	0,64	0,65	0,67	0,68	0,70	0,73	0,75	0,77	0,79	0,82	0,85
n_j	1	3	1	2	8	5	45	15	5	7	3	5

Таблица 15.2 – Технология $N1$, гранулометр; N_2 = 95

X_j	0,59	0,63	0,65	0,67	0,70	0,72	0,73	0,75	0,78	0,79	0,85
n_j	1	3	1	2	8	5	45	15	5	7	3

Таблица 15.3 – Технология $N2$, лаб. анализ; N_3 = 105

X_j	0,62	0,67	0,69	0,72	0,74	0,75	0,79	0,80	0,81	0,85
n_j	5	5	10	15	5	45	8	5	2	5

Таблица 15.4 – Технология $N2$, гранулометр; N_4 = 100

X_j	0,58	0,64	0,67	0,70	0,72	0,73	0,76	0,79	0,80	0,83	0,89
n_j	5	5	3	7	5	10	40	10	5	5	5

Таблица 15.5 – Технология $N3$, гранулометр; N_5 = 26

X_j	0,66	0,68	0,70	0,72	0,73	0,74	0,76	0,78
n_j	1	2	1	2	5	10	4	1

Таблица 15.6 – Технология N3, лаб. анализ; N_6 = 28

X_j	0,67	0,68	0,71	0,73	0,74	0,75	0,77
n_j	1	1	2	10	8	5	1

Сформулируйте и проверьте статистические гипотезы, необходимые для ответа на вопросы:
- существенно ли различаются между собой три технологии?
- какая из технологий наилучшим образом удовлетворяет ГОСТу на классы крупности: $d \in [0,64; 0,84]$, $\bar{d}$ = 0,74?
- можно ли считать успешными испытания автоматического гранулометра или же лабораторные анализы более точны?

Условия выполнения вариантов 5.1-5.4 (табл.15.7).

Таблица 15.7

Условия	5.1	5.2	5.3	5.4
Уровень значимости α	0,05; 0,025	0,05; 0,02	0,05; 0,01	0,05; 0,005
Точность вычислений (число знаков после запятой)	2	3	2	3

Варианты 6.1, 6.2, 6.3, 6.4.

Процесс листопроката.

В одном из цехов анализируется работа листопрокатного стана по результатам контроля качества продукции. Основным показателем качества является толщина (X_j, мм) готового стального листа. Целью исследования является выяснение вопроса: достаточно ли проводить только настройку технологического процесса или необходимо уже проводить ремонт и замену оборудования для обеспечения заданной точности по толщине металла. Кроме того, попутно решается вопрос, какая доля брака может быть пущена как годная продукция другого сорта (другой толщины листа).

Исследуемая номинальная толщина листа 1,9 мм (допуск ±0,04 мм); 2,0 мм (допуск ±0,04 мм); 2,1 мм (допуск ±0,05 мм). Результаты измерений в разных условиях представлены в таблицах 16.1-16.6.

Таблица 16.1 – Настройка сразу после ремонта; N_1, = 145 (номинал 2 мм)

X_j, мм	1,93	1,94	1,95	1,96	1,97	1,98	1,99	2,01	2,03	2,05	2,06	2,08	2,10
n_j	2	3	3	7	10	20	23	30	28	13	4	1	1

Таблица 16.2 – Настройка без проведения ремонта; N_2 = 115 (номинал 2 мм)

X_j, мм	1,90	1,92	1,96	1,97	1,98	2,00	2,02	2,04	2,05	2,06	2,07	2,09	2,10
n_j	2	1	3	7	12	20	25	17	7	10	7	3	1

Таблица 16.3 – Настройка сразу после ремонта; N_3 = 95 (номинал 1,9 мм)

X_j, мм	1,85	1,86	1,87	1,89	1,90	1,91	1,92	1,93	1,94	1,95	1,96
n_j	1	6	5	20	25	15	20	5	4	3	1

Таблица 16.4 – Настройка без проведения ремонта; N_4 = 76 (номинал 1,9 мм)

X_j, мм	1,85	1,87	1,89	1,91	1,92	1,93	1,94	1,96	1,97	1,98
n_j	1	2	10	10	27	10	6	5	4	1

Таблица 16.5 – Настройка сразу после ремонта; N_5 = 30 (номинал 2,1 мм)

X_j, мм	2,04	2,06	2,08	2,09	2,10	2,11	2,12	2,14
n_j	1	1	2	4	10	8	3	1

Таблица 16.6 –Настройка без проведения ремонта; N_6 = 29 (номинал 2,1 мм)

X_j, мм	2,04	2,06	2,08	2,09	2,10	2,11	2,12	2,13	2,14	2,16
n_j	1	1	2	1	3	4	8	1	7	1

Сформулируйте и проверьте статистические гипотезы, необходимые для ответа на вопросы:

- существенно ли разнится точность настройки процесса до ремонта и после ремонта?
- существенно ли разнится точность настройки в зависимости от того номинала, на который ведется настройка?
- какая доля брака при различных настройках может быть использована как годная продукция другого сорта (номинала)?
- сколько замеров толщины стенки листа стали необходимо произвести, чтобы быть уверенными в статистических выводах?

Условия выполнения вариантов 6.1-6.4 (табл.16.7).

Таблица 16.7

Условия	6.1	6.2	6.3	6.4
Уровень значимости α	0,05; 0,001	0,05; 0,025	0,05; 0,01	0,05; 0,005
Точность вычислений (число знаков после запятой)	4	3	4	3

Варианты 7.1, 7.2, 7.3, 7.4. Задача инвестирования.

Инвестиционная компания *N*1 объявила средний годовой доход по акциям от определенного производства равным 11,2%. Инвестор желает проверить, действительно ли это так, и делает случайные выборки объемом N акций интересующей его отрасли индустрии. Попутно инвестор проводит анализ годового дохода по акциям инвестиционной компании *N*2, по которым объявлен средний годовой доход 12,0%. Имеет ли инвестор достаточно оснований, чтобы отказаться от инвестирования в компанию *N*1? Необходимые статистические данные представлены таблицами 17.1 - 17.4.

Таблица 17.1 – Компания *N*1, 1-ый год; $N_1 = 115$

X_j, %	9,7	10,1	10,4	10,6	10,9	11,0	11,2	11,3	11,4	11,6	12,1
n_j	1	3	10	18	29	25	15	8	3	2	1

Таблица 17.2 – Компания *N*2, 1-ый год; $N_2 = 70$

X_j, %	10,3	10,6	10,8	10,9	11,1	11,4	11,7	12,2	12,4
n_j	1	3	5	10	20	18	8	3	2

Таблица 17.3 – Компания *N*1, 2-ой год; $N_3 = 25$

X_j, %	10,7	10,9	11,0	11,1	11,4	11,7	12,0
n_j	1	1	5	10	5	2	1

Таблица 17.4 – Компания N2, 2-ой год; $\mathrm{N}_4 = 25$

X_j, %	10,6	11,0	11,2	11,5	11,6	11,8	11,9
n_j	1	3	5	8	5	2	1

Сформулируйте и проверьте статистические гипотезы, необходимые для ответа на вопросы:

- отличаются ли результаты анализа за 1-ый и 2-ой года?
- существенно ли различие средних годовых доходов компаний *N*1, *N*2?
- отличаются ли наблюдаемые средние годовые доходы от объявленных инвестиционными компаниями?
- акции какой компании Вы предпочтете, учитывая риск (измеряемый дисперсией) от приобретения этих акций? Если уровень риска: ограничен, неограничен?
- можно ли пользоваться результатами анализа по малым выборкам?

Условия выполнения вариантов 7.1 - 7.4 (табл.17.5).

Таблица 17.5

Условия	7.1	7.2	7.3	7.4
Уровень значимости α	0,05; 0,005	0,05; 0,01	0,05; 0,02	0,05; 0,025
Допустимая доля риска	0,35	0,30	0,40	0,28
Точность вычислений (число знаков после запятой)	3	4	3	2

Варианты 8.1, 8.2, 8.3, 8.4. Процесс трубосварки.

Трубосварочный цех металлургического завода выпускает сварные стальные трубы разного диаметра X_j (D, мм). По данным отдела технического контроля для исследуемого трубоэлектросварочного агрегата желательно установить оптимальное время планового ремонта (17 или 20 суток). Ремонт необходимо проводить тогда, когда возникает опасность того, что доля брака превысит допустимый уровень. Замеры диаметров производятся двумя контролерами: А и В. Результаты измерений приведены в таблицах 18.1-18.6. Продукция считается годной, если отклонение от номинала не превышает ±1%.

Таблица 18.1 – Контролер А, 5 суток после ремонта; N_1 = 36 (номинал 18 мм)

X_j	16, 0	16,6	17,0	17,4	17,8	18,0	18,2	18,6	19,2	19,8
n_j	1	2	1	3	8	10	5	3	2	1

Таблица 18.2 – Контролер А, 20 суток после ремонта; N_2=105 (номинал 18 мм)

X_j	16,8	17,2	17,6	18,2	18,4	18,8	19,2	19,4	19,8	20,2	20,8
n_j	1	1	10	15	18	25	10	15	7	2	1

Таблица 18.3 – Контролер В, 5 суток после ремонта; N_3 = 30 (номинал 20 мм)

X_j	17,6	18,0	18,6	19,0	19,4	19,8	20,2	20,6	21,4	22,0
n_j	1	1	2	3	2	10	6	3	1	1

Таблица 18.4 – Контролер В, 17 суток после ремонта; N_4 = 95 (номинал 20 мм)

X_j	18,2	18,4	18,8	19,2	19,8	20,4	20,6	21,0	21,2	21,4	22,0	22,6	22,8
n_j	1	2	3	5	10	20	10	15	12	8	5	3	1

Таблица 18.5 – Контролер А, 20 суток после ремонта; N_5=140 (номинал 22 мм)

X_j	19,6	20,2	20,8	21,4	21,6	22,0	22,2	22,6	22,8	23,0	23,6	23,8
n_j	1	2	5	5	12	15	30	18	25	15	8	4

Таблица 18.6 – Контролер В, 20 суток после ремонта; N_6 = 140 (номинал 22 мм)

X_j	19,8	20,6	21,0	21,6	21,8	22,4	22,6	23,0	23,4	23,8	24,0
n_j	1	2	7	10	18	24	28	20	19	10	1

Сформулируйте и проверьте статистические гипотезы, необходимые для ответа на вопросы:

- существенно ли отличаются результаты работы агрегата сразу после ремонта, или через 17 и 20 суток?

- когда лучше проводить плановый ремонт: через 17, 20 суток или раньше? (Допустимая доля брака представлена в таблице 18.7.)
- разнятся ли по точности результаты двух контролеров?
- какая доля брака одного номинала может быть использована как годная продукция другого номинала?
- сколько измерений диаметра труб желательно проводить для того, чтобы быть уверенными в статистических выводах? Какова степень уверенности в различных случаях?
-

Условия выполнения вариантов 8.1-8.4 (табл.18.7).

Таблица 18.7

Условия	8.1	8.2	8.3	8.4
Уровень значимости α	0,05 0,01	0,05 0,001	0,05 0,025	0,05; 0,005
Доля брака	12%	10%	8%	11%
Точность вычислений (число знаков после запятой)	3	4	3	4

4. Контрольные вопросы

1) Определите разницу в понятиях: «статистика (оценка)» и «параметр» распределения.
2) Каким требованиям должны удовлетворять выборки из генеральной совокупности?
3) Какие требования предъявляются к оценкам параметров?
4) Зависят ли результаты определения статистик от методов вычислений?
5) Определите понятия «точечная», «интервальная» оценка.
6) Влияет ли (как, и почему, если влияет) выбор числа опытов n, числа разрядов К на расчетные значения статистик?
7) Определите понятия «гипотеза», «критерий проверки гипотезы», «ошибка первого и второго рода».
8) Влияет ли выбор уровня значимости α на надежность оценки?
9) Взаимосвязаны ли заданная точность, статистическая надежность, доверительная вероятность, объем выборки?
10) Как изменение уровня значимости влияет на вероятность совершения ошибки первого или второго рода?
11) Какие критерии проверки гипотез Вам известны?
12) Каким образом могут быть определены критические точки в разных ситуациях?

13) Опишите наиболее употребимые распределения статистик, используемых для проверки гипотез.
14) Каким образом можно быстро сделать приблизительное заключение, подчиняется ли случайная величина нормальному закону распределения?
15) Для проверки каких гипотез применяются критерии значимости?
16) Для проверки каких гипотез применяются критерии согласия?
17) Какие проблемы возникают при статистическом оценивании?
18) К каким ошибкам в статистических выводах может привести отклонение от нормального распределения?
19) Какие критерии наиболее чувствительны к отклонениям случайной величины от нормального распределения?
20) В каких случаях применяются односторонние и двусторонние критерии?
21) Как могут быть проверены гипотезы об однородности, репрезентативности выборки?
22) Определите понятие «мощность критерия».
23) Каким образом формируются гипотезы о вероятностях, о средних, о дисперсиях, о законах распределения?
24) Какой из методов измерений на практике предпочтительней: метод с малой систематической погрешностью и более высокой точностью или метод, который дает несмещенное среднее значение с невысокой точностью?
25) Какие оценки положения и формы ряда распределения Вам известны и какие из них лучше в разных ситуациях?
26) Основываясь на разложении бинома $(\alpha-(1-\alpha))^n$, где α - уровень значимости (или вероятность ошибки), n - число опытов, определить вероятность получения при ограниченном числе n (n=2, n=3, . . .) случайного статистически значимого результата.
27) Определить понятие «стандартное отклонение». В каких формулах для вычисления значений критериев присутствует эта величина?
28) Как определить наименьшее число опытов для оценки стандартного отклонения и среднего значения?
29) Каким образом неравенство Чебышева используется при статистическом оценивании?
30) Как определить наименьший уровень значимости α, при котором можно отбросить нуль-гипотезу в каждом конкретном примере?

5.Требования к оформлению пояснительной записки

Пояснительная записка может быть представлена в тетради, объемом около 12 листов, либо на листах формата А4.

Содержание пояснительной записки.

1. Введение.
    1.1. Цель и задачи работы.
    1.2. Исходные данные, соответствующие конкретному варианту.
    1.3. Описание задачи статистического анализа (формулировка проблем, необходимые формулы, соответствующие варианту задания).
2. Расчетная часть.
    2.1.
        2.1.1. Расчеты 1 части.
        2.1.2. Анализ результатов.
        2.1.3. Предварительные выводы.
    2.2.
        2.2.1. Расчеты 2 части.
        2.2.2. Анализ результатов.
        2.2.3. Выводы.
3. Заключение (перечень сделанного, рекомендации, ответы на контрольные вопросы).

Глава 4. Анализ и построение зависимостей

4.1. Дисперсионный анализ

4.1.1. Постановка задачи и сущность дисперсионного анализа

Методы сравнительного измерения дисперсии были предложены в 70-х годах XIX века Лексисом, В. Борткевичем и А.А. Чупровым в целях определения стабильности статистических рядов.

Дисперсионный анализ, как метод исследования, появился в работах Р. Фишера (1918-1935 гг.) в связи с исследованиями в сельском хозяйстве для выявления условий, при которых испытываемый сорт с/х культуры даёт максимальный урожай. (В агрономических исследованиях первый фактор − сорт почвы, второй фактор − способ обработки.) Дальнейшее развитие дисперсионный анализ получил в работах Йетса [16,18].

Дисперсионный анализ позволяет ответить на вопрос о наличии существенного влияния некоторых факторов на изменчивость фактора, значения которого могут быть получены в результате опыта. При проверке статистических гипотез предполагается случайность вариации изучаемых факторов. В дисперсионном анализе один или несколько факторов изменяются заданным образом, причём, эти изменения могут влиять на результаты наблюдений. Исследование такого влияния и является целью дисперсионного анализа.

В настоящее время наблюдается все более широкое использование дисперсионного анализа в экономике, социологии, биологии и др., особенно, после появления программных средств, снявших проблемы громоздкости статистических вычислений.

В практической деятельности, в различных областях науки мы часто сталкиваемся с необходимостью оценить влияние различных факторов на те или иные показатели. Часто эти факторы имеют качественный характер (например, качественным фактором, влияющим на экономический эффект, может быть введение новой системы управления производством) и тогда дисперсионный анализ приобретает особую ценность, так как становится единственным статистическим способом исследования, дающим такую оценку.

Сейчас теорию дисперсионного анализа можно считать в достаточной мере сформировавшейся, но способы организации эксперимента и вычислительные схемы продолжаются совершенствоваться. В пособии используется одна из нетрадиционных схем изложения материала и алгоритмы вычислений (основой послужила работа В.А.Юденкова [Юденков В.А. Дисперсионный анализ. Уч. пос., 1982. – 95 с.]), позволяющие облегчить понимание сути дисперсионного анализа и вычислительных операций.

Постановка задачи. В любом ряде испытаний имеется несколько факторов, вызывающих изменчивость средних значений наблюдаемых случайных величин – результативных признаков. Эти факторы могут принадлежать к одному или нескольким источникам изменчивости (например, расположение торговых заведений в центре и на окраине города, изменения в законодательстве, разные климатические условия, разные уровни образования и т.п.). Очевидно, что даже при самом тщательном исследовании не удается выявить все источники изменчивости, а иногда в этом нет необходимости или смысла. Но при наличии опыта эксперта и в зависимости от цели исследования всегда можно выдвинуть гипотезу о существовании влияния тех или иных факторов на результативный признак.

Дисперсионный анализ дает возможность установить, существенное ли влияние оказывает тот или иной из рассматриваемых факторов на изменчивость признака, а также определить количественно «удельный вес» каждого из источников изменчивости в их общей совокупности. Но дисперсионный анализ позволяет дать положительный ответ лишь о наличии существенного влияния, в противном случае вопрос остается открытым и требует дополнительных исследований (чаще всего – увеличения числа опытов).

В дисперсионном анализе используются следующие термины.

Фактор (X) – то, что как мы считаем, должно оказывать влияние на результат (результативный признак) Y.

Уровень фактора (или способ обработки, иногда буквально, например – способ обработки почвы) – значения (X_i, $i = 1, 2, \ldots, I$), которые может принимать фактор.

Отклик – значение измеряемого признака (величина результата Y_i).

Техника дисперсионного анализа меняется в зависимости от числа изучаемых независимых факторов. Если факторы, вызывающие изменчивость среднего значения признака, принадлежат одному источнику, то мы имеем простую группировку, или однофакторный дисперсионный анализ и далее, соответственно, двойная группировка – двухфакторный дисперсионный анализ, трехфакторный дисперсионный анализ,…, m- факторный. Факторы в многофакторном анализе принято обозначать латинскими буквами: A, B, C и т.д.

Задача дисперсионного анализа - исследование влияния тех или иных факторов (или уровней факторов) на изменчивость средних значений наблюдаемых случайных величин.

Сущность дисперсионного анализа. Дисперсионный анализ состоит в выделении и оценке отдельных факторов, вызывающих изменчивость. С этой целью производят разложение общей дисперсии σ^2 наблюдаемой частичной совокупности (общей дисперсии признака), вызванной всеми источниками изменчивости, на составляющие дисперсии, порожденные независимыми факторами. Каждая из этих составляющих дает оценку дисперсии σ_A^2, $\sigma_B^2, \ldots$, вызванную конкретным источником изменчивости, в общей совокупности. Для проверки значимости этих составляющих оценок дисперсии их сравнивают с общей дисперсией в общей совокупности (по критерию Фишера).

Например, в двухфакторном анализе мы получим разложение вида:

$$\sigma_C^2 = \sigma_A^2 + \sigma_B^2 + \sigma_{AB}^2 + \sigma_\varepsilon^2,$$

где

σ_C^2 - общая дисперсия изучаемого признака *C*;

σ_A^2 - доля дисперсии, вызванная влиянием фактора *A*;

σ_B^2 - доля дисперсии, вызванная влиянием фактора *B*;

σ_{AB}^2 - доля дисперсии, вызванная взаимодействием факторов *A* и *B*;

σ_ε^2 - доля дисперсии, вызванная неучтёнными случайными причинами (случайная дисперсия);

В дисперсионном анализе рассматривается гипотеза: H_0 – ни один из рассматриваемых факторов не оказывает влияния на изменчивость признака. Значимость каждой из оценок дисперсии проверяется по величине её отношения к оценке случайной дисперсии и сравнивается с соответствующим критическим значением, при уровне значимости α, с помощью таблиц критических значений *F*-распределения Фишера-Снедекора (прил.4). Гипотеза H_0 относительно того или иного источника изменчивости отвергается, если $F_{расч.} > F_{кр.}$ (например, для фактора *B*: $S_B^2/S_\varepsilon^2 > F_{кр.}$).

В дисперсионном анализе рассматриваются эксперименты 3-х видов:

а) эксперименты, в которых все факторы имеют систематические (фиксированные) уровни;

б) эксперименты, в которых все факторы имеют случайные уровни;

в) эксперименты, в которых есть факторы, имеющие случайные уровни, а так же факторы, имеющие фиксированные уровни.

Случаи а), б), в) соответствуют трем моделям, которые рассматриваются в дисперсионном анализе.

Таблица 19

Номер наблюдения *j*	Уровни фактора			
	A_1	A_2	...	A_p
1	X_{11}	X_{21}	...	X_{p1}
2	X_{12}	X_{22}	...	X_{p2}
3	X_{13}	X_{23}	...	X_{p3}
.	.	.	...	...
.	.	.	...	...
.	.	.	...	...
n	X_{1n}	X_{2n}	...	X_{pn}
ИТОГИ				

Рассмотрим единичный фактор, который принимает p различных уровней, и предположим, что на каждом уровне сделано n наблюдений, что дает $N=np$ наблюдений. (Ограничимся рассмотрением первой модели дисперсионного анализа – все факторы имеют фиксированные уровни.)

Пусть результаты представлены в виде X_{ij} (i=1,2…,p; j=1,2,…,n).

Данные обычно располагают в виде таблицы (табл.19).

Предполагается, что для каждого уровня n наблюдений имеется средняя, которая равна сумме общей средней и ее вариации обусловленной выбранным уровнем:

$$X_{ij} = \mu + A_i + \varepsilon_{ij},$$

где μ - общая средняя;

A_i - эффект, обусловленный i – м уровнем фактора;

ε_{ij} – вариация результатов внутри отдельного уровня фактора. С помощью члена ε_{ij} принимаются в расчет все неконтролируемые факторы [6, 16].

Пусть наблюдения на фиксированном уровне фактора нормально распределены относительно среднего значения $\mu + A_i$ с общей дисперсией σ^2.

Таблица 20 – Таблица дисперсионного анализа

Источник изменчивости	Суммы квадратов (*SS*)	Степени свободы (*df*)	Средние квадраты (*MS*)
Различия между уровнями	$S_1 = n\sum_i (X_{i.} - X_{..})^2$	*p-1*	$M_1 = \frac{S_1}{p-1}$
Различия внутри уровней	$S_2 = \sum_{i,j} (X_{i.j} - X_{i.})^2$	*N-p*	$M_2 = \frac{S_2}{N-p}$
Сумма	$S_2 = \sum_{i,j} (X_{ij} - X_{..})^2$	*N-1*	

Тогда (точка вместо индекса обозначает усреднения соответствующих наблюдений по этому индексу):

$$X_{ij} - X.. = (X_{i\cdot} - X..) + (X_{ij} - X_i.).$$

После возведения (4.1.2) в квадрат и суммирования по i и j получим:

$$\sum_{i,j} (X_{ij} - X..)^2 = \sum_{i,j} (X_i. - X..)^2 + \sum_{i,j} (X_{ij} - X_i.)^2,$$

так как $\sum_{i,j} (X_i. - X..)(X_{ij} - X_i.) = \sum_i (X_i. - X..) \sum_j (X_{ij} - X_i.)$, но $\sum_j (X_{ij} - X_i.) = 0.$

Иначе сумму квадратов можно записать: $S = S_1 + S_2$. Величина S_1 вычисляется по отклонениям p средних от общей средней $X..$, поэтому S_1 имеет (p-1) степеней свободы. Величина S_2 вычисляется по отклонениям N наблюдений от p выборочных средних и, следовательно, имеет

$N\text{-}p = np - p = p(n\text{-}1)$ степеней свободы. S имеет (N-1) степеней свободы. По результатам вычислений строится таблица дисперсионного анализа (табл.20).

Если гипотеза о том, что влияние всех уровней одинаково, справедлива, то обе величины M_1 и M_2 (средние квадраты) будут несмещенными оценками

σ^2. Значит, гипотезу можно проверить, вычислив отношение (M_1/M_2) и сравнив его с $F_{кр.}$ с $v_1 = (p-1)$ и $v_2 = (N-p)$ степенями свободы [15,16,].

Если $F_{расч.} > F_{кр.}$, то гипотеза о незначимом влиянии фактора A на результат наблюдений не принимается. (Примеры рассматриваются ниже и в части III.)

Для оценки существенности различий при $F_{расч.} \rangle\ F_{табл.}$ вычисляют:

а) *ошибку опыта*

$$S_{\overline{X}} = \sqrt{\frac{S^2{}_z}{n}}$$

б) *ошибку разности средних*

$$S_d = S_{\overline{X}} \cdot \sqrt{2}$$

в) *наименьшую существенную разность*

$$HCP_{\alpha,k} = t_{\alpha,k} \cdot S_d$$

Сравнивая разность средних значений $\overline{X}_i.$ по вариантам с НСР, делают вывод о существенности различий в уровне средних.

Замечание. Применение дисперсионного анализа предполагает, что:

1) $M(\varepsilon_{ij})=0$,
2) $D(\varepsilon_{ij})=\sigma^2=$ const,
3) $\varepsilon_{ij} \rightarrow N(0, \sigma^2)$ или $x_{ij} \rightarrow N(a, \sigma^2)$.

4.1.2. Модели однофакторного и многофакторного дисперсионного анализа.

4.1.2.1. Варианты эксперимента дисперсионного анализа.

Рассмотрим несколько наиболее распространенных вариантов эксперимента, организуемого для проведения дисперсионного анализа: однофакторный, двухфакторный и трехфакторный анализ с разным числом уровней факторов и разным числом опытов на каждом уровне.

А. Однофакторный эксперимент (один фактор А)

Значения измеряемого признака – X_{im}.

1.Эксперимент на двух уровнях, i =1,2 (рис.44 а):

- без повторных опытов, m = 1;
- с повторными опытами, одинаковое число опытов на каждом уровне, m = 1, 2, …, n.
- с повторными опытами, разное число опытов на каждом уровне m = 1, 2, …, n_i.

2.Эксперимент на нескольких уровнях, i =1, 2, …, a (рис.44 б):

- без повторных опытов, m = 1;
- с повторными опытами, одинаковое число опытов на каждом уровне

$m = 1, 2, \ldots, n$;

- с повторными опытами, разное число опытов на каждом уровне $m = 1, 2, \ldots, n_i$.

Рис. 44 – Точки эксперимента в однофакторном анализе: а) два уровня A_i, $i=1,2$; б) несколько уровней A_i, $i=1,2,\ldots,a$

Таблица 21 представляет исходные данные однофакторного эксперимента на двух уровнях с одинаковым числом повторных опытов. Число групп (H) равно числу уровней: A_1, A_2; i=1,2.

Таблица 21– Данные для однофакторного анализа, равное число опытов

Уровни (группы)	Результаты опытов: X_{im}, $m = 1,2,\ldots, n$				
	X_{i1}	…	X_{im}	…	X_{in}
A_1	$\boldsymbol{X}_{11}$	**…**	$\boldsymbol{X}_{1m}$	**…**	$\boldsymbol{X}_{1n}$
A_2	$\boldsymbol{X}_{21}$	**…**	$\boldsymbol{X}_{2m}$	**…**	$\boldsymbol{X}_{2n}$

А.В. Двухфакторный эксперимент (факторы А и В)

Значения измеряемого признака - X_{ijm}.

1.Эксперимент на двух уровнях, $i=1,2$; $j = 1,2$ (рис.45):

- без повторных опытов, $m = 1$;

- с повторными опытами, одинаковое число опытов на каждом уровне; $m = 1, 2, \ldots, n$.

- с повторными опытами, разное число опытов на каждом ij-уровне, $m_{ij} = 1, 2, \ldots, n_{ij}$.

2.Эксперимент на нескольких уровнях, $i =1, 2,\ldots, a$; $j = 1, 2, \ldots, b$:

- без повторных опытов, $m = 1$;

- с повторными опытами, одинаковое число опытов на каждом ij-уровне, $m_{ij} = 1, 2, \ldots, n$;

- с повторными опытами, разное число опытов на каждом ij-уровне, $m_{ij} = 1, 2, \ldots, n_{ij}$.

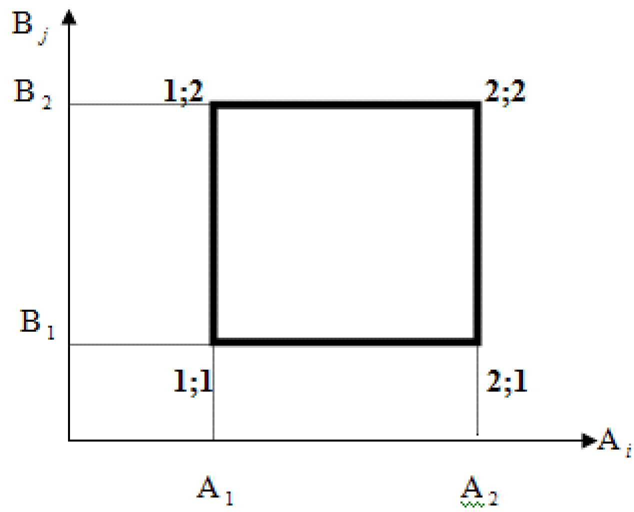


Рис. 45 – Точки эксперимента двухуровневого двухфакторного анализа

Приведем различные формы представления исходных данных в зависимости от вида эксперимента (таблицы 22 и 23).

Таблица 22 – Данные для двухфакторного анализа на двух уровнях, разное число опытов

№ строки (группы)	Сочетания уровней A B	Результаты опытов: X_{ijm}; $m = 1, 2, \ldots, n_{ij}$					
		X_{ij1}	…	X_{ijm}	…	X_{ijn1-1}	X_{ijn1}
1	1; 1	$\boldsymbol{X}_{111}$	…	$\boldsymbol{X}_{11m}$	…	$\boldsymbol{X}_{11n1-1}$	$\boldsymbol{X}_{11n1}$
2	1; 2	$\boldsymbol{X}_{121}$	…	$\boldsymbol{X}_{12n2}$	–	–	–
3	2; 1	$\boldsymbol{X}_{211}$	…	$\boldsymbol{X}_{21m}$	…	$\boldsymbol{X}_{21n3-1}$	$\boldsymbol{X}_{21n3}$
4	2; 2	$\boldsymbol{X}_{221}$	…	$\boldsymbol{X}_{22m}$	…	$\boldsymbol{X}_{22n4}$	–

Таблица 23 – Данные для двухфакторного анализа на нескольких уровнях, равное число опытов

№ строки	Сочетания уровней A B	**Наблюденные значения признака в группах, X_{ijm}**				
		1-й опыт	**…**	***m*- опыт**	**…**	***n*-опыт**
1	**1; 1**	X_{111}	…	X_{11m}	…	X_{11n}
2	**1; 2**	X_{121}	…	X_{12m}	…	X_{12n}
…	**…**	…	…	…	…	…
ij	***i*; *j***	X_{ij1}	…	X_{ijm}	…	X_{ijn}
…	**…**	…	…	…	…	…
H	***a*; *b***	X_{ab1}	…	X_{abm}	…	X_{abn}

Число групп (H) равно числу перестановок уровней: $ij = 1,2,\ldots,H$

А.В.С. Трехфакторный эксперимент (факторы А, В и С)

Значения измеряемого признака - X_{ijkm}.

1.Эксперимент на двух уровнях, i =1,2; j = 1,2; k =1,2 (рис.46):

- без повторных опытов, m = 1;

- с повторными опытами, одинаковое число опытов на каждом ijk-уровне, m_{ijk} = 1, 2, …, n;

- с повторными опытами, разное число опытов на каждом ijk -уровне, m_{ijk} = 1, 2, …, n_{ijk}.

2.Эксперимент на нескольких уровнях, i =1, 2, …, a; j = 1, 2, …, b; k =1, 2, …, c:

- без повторных опытов, m = 1;

- с повторными опытами, одинаковое число опытов на каждом ijk - уровне, m_{ijk} = 1, 2, …, n.- с повторными опытами, разное число опытов на каждом ijk -уровне, m_{ijk} =1, 2, …, n_{ijk}.

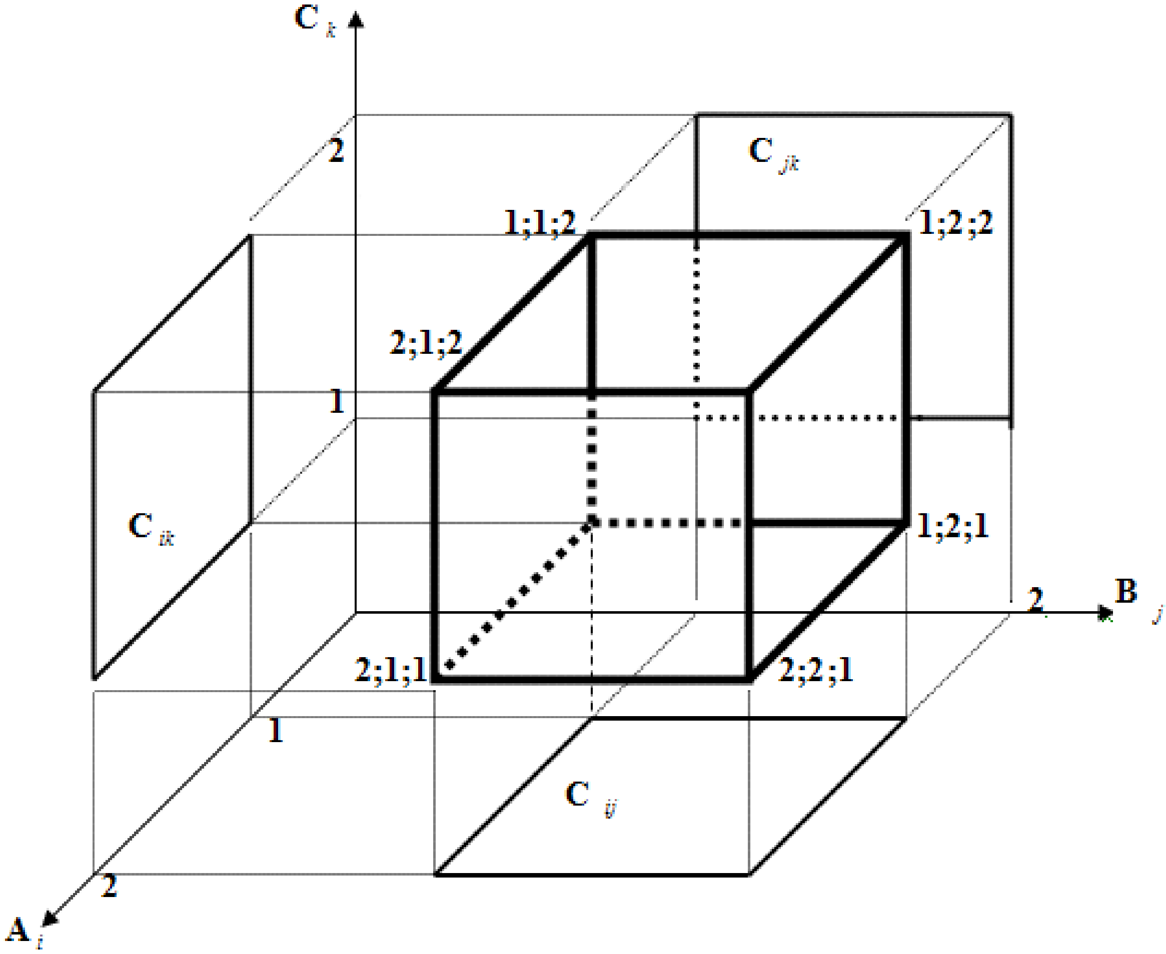


Рис. 46 – Точки эксперимента двухуровневого трехфакторного анализа

Таблица 24 – Данные для трехфакторного анализа на двух уровнях, разное число опытов

№ строки (группы)	Сочетания уровней A B C	Результаты опытов: X_{ijkm}; m = 1, 2, …, n_{ijk}					
		X_{ijk1}	…	X_{ijkm}	…	X_{ijkn-1}	X_{ijkn}
1	1; 1; 1	$\boldsymbol{X}_{1111}$	...	$\boldsymbol{X}_{111m}$	...	$\boldsymbol{X}_{111n1}$	$\boldsymbol{X}_{111n1}$
2	1; 2; 1	$\boldsymbol{X}_{1211}$	...	$\boldsymbol{X}_{121m}$	...	$\boldsymbol{X}_{121n2}$	–
3	2; 1; 1	$\boldsymbol{X}_{2111}$	...	$\boldsymbol{X}_{211n3}$	–	–	–
4	2; 2; 1	$\boldsymbol{X}_{2211}$	...	$\boldsymbol{X}_{221m}$	...	$\boldsymbol{X}_{221n4}$	–
5	1; 1; 2	$\boldsymbol{X}_{1121}$	...	$\boldsymbol{X}_{112n5}$	–	–	–
6	1; 2; 2	$\boldsymbol{X}_{1221}$	...	$\boldsymbol{X}_{122n6}$	–	–	–
7	2; 1; 2	$\boldsymbol{X}_{2121}$	...	$\boldsymbol{X}_{212m}$	...	$\boldsymbol{X}_{212n7-1}$	$\boldsymbol{X}_{212n7}$
8	2; 2; 2	$\boldsymbol{X}_{2221}$	...	$\boldsymbol{X}_{222m}$	...	$\boldsymbol{X}_{222n8-1}$	$\boldsymbol{X}_{222n8}$

Число групп (H) равно числу перестановок уровней: ijk= 1,2,…,H; $H = 8$.

4.1.2.2. Модель однофакторного дисперсионного анализа. Фактор *А*.

Основное уравнение дисперсионного анализа:

$$\boldsymbol{SS} = \boldsymbol{SS}_a + \boldsymbol{SS}_\varepsilon \tag{4.1.1}$$

Одинаковое число повторных опытов (m = 1,2,…,n):

$$\boldsymbol{SS} = \sum_{i-1}^{a}\sum_{m=1}^{n}\left(X_{ij} - \overline{X}\right)^2 \tag{4.1.2}$$

где $\boldsymbol{SS}$ - *общая сумма квадратов* разностей наблюдений и их среднего значения;

$$\boldsymbol{SS}_a = \mathrm{n}\sum_{i=1}^{a}\left(\overline{X}_i - \overline{X}\right)^2 \tag{4.1.3}$$

где $\boldsymbol{SS}_a$ **-** *сумма квадратов между группами* (вклад в общую сумму квадратов, обусловленный различиями в уровнях фактора А);

$$\boldsymbol{SS}_e = \sum_{i=1}^{a}\sum_{m=1}^{n}\left(X_{im} - \overline{X}_i\right)^2, \tag{4.1.4}$$

где $\boldsymbol{SS}_e$ **-** *сумма квадратов внутри групп* – остаток, вклад в общую сумму квадратов, вызванный случайной изменчивостью данных внутри групп (или сумма квадратов случайных эффектов - ошибка опыта).

$$\overline{X} = \frac{\sum_{i=1}^{a}\sum_{m=1}^{n} X_{im}}{N}, \tag{4.1.5}$$

где $\overline{X}$ **-** *общее среднее*, $\boldsymbol{N = an}$ – *общее число опытов*;

$$\overline{X}_i = \frac{\sum_{m=1}^{n} X_{im}}{n}, \tag{4.1.6}$$

где $\overline{X}_i$ - *среднее значение на i уровне фактора А.*

Разное число повторных опытов (m =1, 2, …, n_i):

$$\mathbf{SS} = \sum_{i=1}^{a}\sum_{m=1}^{n_i}\left(X_{im}-\overline{X}\right)^2; \quad \mathbf{SS}_a = \sum_{i=1}^{a} n_i\left(\overline{X}_i-\overline{X}\right)^2; \quad \mathbf{SS}_e = \sum_{i=a}^{a}\sum_{m=1}^{n_i}\left(X_{im}-\overline{X}_i\right)^2; \tag{4.1.7}$$

$$\overline{X} = \frac{\sum_{i=1}^{a}\sum_{m=1}^{n_i} X_{im}}{N}; \quad \boldsymbol{N} = \sum_{i=1}^{a} n_i; \quad \overline{X}_i = \frac{\sum_{m=1}^{n_i} X_{im}}{n_i} \tag{4.1.8}$$

Оценки дисперсий и определение числа степеней свободы

$\boldsymbol{S}^2 = \frac{SS}{\nu}$ - *оценка общей дисперсии;* $\boldsymbol{\nu} = \boldsymbol{N} - \mathbf{1}$ - *число степеней свободы при определении общей дисперсии;*

$\boldsymbol{S}_a^2 = \frac{SS_a}{\nu_a}$ - *оценка дисперсии по уровням фактора А;* $\boldsymbol{\nu}_a = \boldsymbol{a} - \mathbf{1}$ - *число степеней свободы фактора А;*

$\boldsymbol{S}_e^2 = \frac{SS_e}{\nu_e}$ - *остаточная оценка дисперсии (дисперсия ошибки);*

$\boldsymbol{\nu}_e = \boldsymbol{N} - \boldsymbol{a}$ - *число степеней свободы при определении ошибки.*

$$\boldsymbol{\nu} = \boldsymbol{\nu}_a + \boldsymbol{\nu}_e = \boldsymbol{N} - \mathbf{1} = (\boldsymbol{a} - \mathbf{1}) + (\boldsymbol{N} - \boldsymbol{a}) \tag{4.1.9}$$

Проверка H_0- гипотезы

Расчетное значение критерия:

$$\boldsymbol{F}_{расч.} = \frac{S_a^2}{S_e^2}. \tag{4.1.10}$$

Критическое значение $\boldsymbol{F}_{табл}$ определяется по прил.4 при $\boldsymbol{\alpha}$, $\boldsymbol{\nu}_1 = \boldsymbol{\nu}_a$ и $\boldsymbol{\nu}_2 = \boldsymbol{\nu}_e$. Если

$$\boldsymbol{F}_{расч.} \le \boldsymbol{F}_{табл.} \text{ при } \boldsymbol{\alpha},\ \boldsymbol{\nu}_1,\ \boldsymbol{\nu}_2, \tag{4.1.11}$$

то гипотеза H_0- принимается. В противном случае – отклоняется.

4.1.2.3. Двухфакторный дисперсионный анализ. Факторы А и В

Основное уравнение двухфакторного дисперсионного анализа

$$\boldsymbol{SS} = \boldsymbol{SS}_a + \boldsymbol{SS}_b + \boldsymbol{SS}_{ab} + \boldsymbol{SS}_e \tag{4.1.12}$$

Одинаковое число повторных опытов (m = 1, 2, …, n):

$$\boldsymbol{SS} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n}\left(X_{ijm}-\overline{X}\right)^2, \tag{4.1.13}$$

где $\boldsymbol{SS}$ - *общая сумма квадратов* разностей наблюдений и их среднего значения (сумма квадратов общих эффектов);

$$\boldsymbol{SS}_a = bn\sum_{i=1}^{a}\left(\overline{X}_i-\overline{X}\right)^2, \tag{4.1.14}$$

где $\boldsymbol{SS}$ **-** вклад в общую сумму квадратов, обусловленный различиями в уровнях фактора *А*, или *взвешенная сумма квадратов эффектов фактора А* **(**сумма квадратов между группами);

$$\boldsymbol{SS}_b = an\sum_{j=1}^{b}\left(\overline{X}_j - \overline{X}\right)^2, \qquad (4.1.15)$$

где $\boldsymbol{SS}_b$ - *взвешенная сумма квадратов эффектов фактора B* (сумма квадратов между группами);

$$\boldsymbol{SS}_{ab} = \mathrm{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{X} + \overline{X}_{ij} - \overline{X}_i - \overline{X}_j\right)^2, \qquad (4.1.16)$$

где $\boldsymbol{SS}_{ab}$ - *взвешенная сумма квадратов взаимодействия уровней факторов A и B* или *смешанный эффект факторов A и B (*сумма квадратов между группами);

$$\boldsymbol{SS}_e = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij} - \overline{X}_i\right)^2, \qquad (4.1.17)$$

где $\boldsymbol{SS}_e$ - *сумма квадратов внутри групп* – остаток, вклад в общую сумму квадратов, вызванный случайной изменчивостью данных внутри групп;

$$\overline{X} = \frac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n} X_{ijm}}{N}, \qquad (4.1.18)$$

где $\overline{X}$ - *общее среднее*, *N = abn – общее число опытов*;

$$\overline{X}_i = \frac{\sum_{i=1}^{a}\sum_{m=1}^{n} X_{ijm}}{bn}, \qquad \overline{X}_j = \frac{\sum_{j=1}^{b}\sum_{m=1}^{n} X_{ijm}}{an}, \qquad (4.1.19)$$

где $\overline{X}_i$, $\overline{X}_j$ - *средние значения на i уровне фактора A, j уровне фактора B соответственно.*

$$\overline{X}_{ij} = \frac{\sum_{m=1}^{n} X_{ijm}}{n}, \qquad (4.1.20)$$

где $\overline{X}_{ij}$ - *среднее значение при различных сочетаниях уровней ij.*

При разном числе повторных опытов (m =1, 2, …, n_i)

суммирование ведется не до n, а до n_i, т.е. - $\sum_{m=1}^{n_i} X_{ijm}$.

Оценки дисперсий и определение числа степеней свободы

$$\boldsymbol{S}^2 = \frac{SS}{\nu}, \qquad (4.1.21)$$

где S^2 - *оценка общей дисперсии;* $\boldsymbol{\nu} = \boldsymbol{N} - \mathbf{1}$ *- число степеней свободы при определении общей дисперсии;*

$$\boldsymbol{S}_a^2 = \frac{SS_a}{\nu_a}, \qquad \mathrm{S}_b^2 = \frac{SS_b}{\nu_b}, \qquad (4.1.22)$$

где S_a^2 *- оценка дисперсии по уровням фактора A;* $\boldsymbol{\nu}_a = \boldsymbol{a} - \mathbf{1}$ *- число степеней свободы фактора A;* S_b^2 *- оценка дисперсии по уровням фактора B ;*
$\boldsymbol{\nu}_b = \boldsymbol{b} - \mathbf{1}$ *- число степеней свободы фактора B;*

$$\mathbf{S}_{ab}^2 = \frac{SS_{ab}}{\nu_{ab}},\qquad(4.1.23)$$

где S_{ab}^2 *- оценка дисперсии по уровням факторов А и В;*

$\boldsymbol{\nu_{ab}} = (\boldsymbol{a} - 1)(\boldsymbol{b} - 1)$ *-число степеней свободы взаимодействия факторов А и В;*

$$\boldsymbol{S}_e^2 = \frac{SS_e}{\nu_e},\qquad(4.1.24)$$

где S_e^2 **-** *остаточная оценка дисперсии (дисперсия ошибки);*

$\boldsymbol{\nu}_e = \boldsymbol{N} - \boldsymbol{ab} = \boldsymbol{ab}(\boldsymbol{n} - 1)$ **-** *число степеней свободы при определении ошибки.*

Общее число степеней свободы:

$$\nu = \boldsymbol{\nu}_a + \nu_b + \nu_{ab} + \boldsymbol{\nu}_e = \boldsymbol{N} - 1 = (\boldsymbol{a} - 1)(\boldsymbol{N} - \boldsymbol{a})\qquad(4.1.25)$$

Проверка H_0- гипотезы

Определение расчетного значения критерия:

$$\boldsymbol{F} = \frac{S^2}{S_e^2};\quad \boldsymbol{F}_{a\ расч.} = \frac{S_a^2}{S_e^2};\quad \boldsymbol{F}_{b\ расч.} = \frac{S_b^2}{S_e^2};\quad \boldsymbol{F}_{ab\ расч.} = \frac{S_{ab}^2}{S_e^2}.\qquad(4.1.26)$$

Критическое значение $\boldsymbol{F}_{табл}$ определяется при $\boldsymbol{\nu}_1 = \boldsymbol{\nu}_a$ *и* $\boldsymbol{\nu}_2 = \boldsymbol{\nu}_e$.

Если $F_{расч} \le F_{табл}$ при α, ν_1, ν_2, то гипотеза H_0- принимается.

В противном случае – отклоняется и продолжается анализ гипотез о влиянии уровней факторов.

4.1.2.4.Трехфакторный дисперсионный анализ. Факторы *А*, *В* и *С*.

Основное уравнение трехфакторного дисперсионного анализа

$$\boldsymbol{SS} = \boldsymbol{SS}_a + \boldsymbol{SS}_b + \boldsymbol{SS}_c + \boldsymbol{SS}_{ab} + \boldsymbol{SS}_{ac} + \boldsymbol{SS}_{bc} + \boldsymbol{SS}_{abc} + \boldsymbol{SS}_e\qquad(4.1.27)$$

Одинаковое число повторных опытов ($\boldsymbol{m}$ = 1, 2, …, $\boldsymbol{n}$):

$\boldsymbol{SS} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n}\left(X_{ijkm} - \overline{X}\right)^2$ - *общая сумма квадратов* разностей наблюдений и их среднего значения (сумма квадратов общих эффектов);

$\boldsymbol{SS}_a = bn\sum_{i=1}^{a}\left(\overline{X}_i - \overline{X}\right)^2$ **-** вклад в общую сумму квадратов, обусловленный различиями в уровнях фактора *А*, или *взвешенная сумма квадратов эффектов фактора А* **(**сумма квадратов между группами);

$\boldsymbol{SS}_b = an\sum_{j=1}^{b}\left(\overline{X}_j - \overline{X}\right)^2$ - *взвешенная сумма квадратов эффектов фактора В* **(**сумма квадратов между группами);

$\boldsymbol{SS}_c = cn\sum_{k=1}^{c}\left(\overline{X}_k - \overline{X}\right)^2$ - *взвешенная сумма квадратов эффектов фактора С* **(**сумма квадратов между группами);

$SS_{ab} = \mathrm{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{X} + \bar{X}_{ij} - \bar{X}_i - \bar{X}_j\right)^2$ **-** *взвешенная сумма квадратов взаимодействия уровней факторов А и В* или *смешанный эффект факторов А и В (*сумма квадратов между группами);

$SS_{ac} = n\sum_{i=1}^{a}\sum_{k=1}^{c}\left(\bar{X} + \bar{X}_{ik} - \bar{X}_i - \bar{X}_k\right)^2$ **-** *смешанный эффект факторов А и С;*

$SS_{bc} = n\sum_{j=1}^{b}\sum_{k=1}^{c}\left(\bar{X} + \bar{X}_{jk} - \bar{X}_j - \bar{X}_k\right)^2$ **-** *смешанный эффект факторов В и С;*

$SS_{abc} = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(\bar{X}_{ijk} + \bar{X}_i + \bar{X}_j + \bar{X}_k - \bar{X}_{ij} - \bar{X}_{ik} - \bar{X}_{jk} - \bar{X}\right)^2$ **-** *смешанный эффект факторов А , В и С;*

$SS_e = \sum_{i=1}^{a}\sum_{j=1}^{b}\left(X_{ij} - \bar{X}_i\right)^2$ **-** *сумма квадратов внутри групп –* остаток, вклад в общую сумму квадратов, вызванный случайной изменчивостью данных внутри групп;

$\bar{X} = \dfrac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n} X_{ijm}}{N}$ **-** *общее среднее***,** N = abn – *общее число опытов*;

$\bar{X}_i = \dfrac{\sum_{i=1}^{a}\sum_{m=1}^{n} X_{ijm}}{bn}$**;** $\bar{X}_j = \dfrac{\sum_{j=1}^{b}\sum_{m=1}^{n} X_{ijm}}{an}$ **-** *средние значения на i уровне фактора А, j уровне фактора В соответственно.*

$\bar{X}_{ij} = \dfrac{\sum_{m=1}^{n} X_{ijm}}{n}$ **-** *среднее значение при различных сочетаниях уровней ij.*

При разном числе повторных опытов (m =1,2,…,n_i) суммирование ведется не до *n*, а до n_i,т.е. - $\sum_{m=1}^{n_i} X_{ijm}$ **.**

Оценки дисперсий и определение числа степеней свободы

$S^2 = \dfrac{SS}{\nu}$ **-** *оценка общей дисперсии;* **ν =** N **- 1** *- число степеней свободы при определении общей дисперсии;*

$S_a^2 = \dfrac{SS_a}{\nu_a}$**;** $\mathbf{S}_b^2 = \dfrac{SS_b}{\nu_b}$**;** $\mathbf{S}_c^2 = \dfrac{SS_c}{\nu_c}$ *- оценки дисперсий по уровням соответственно факторов А, В и С;*

$\nu_a = a - 1$; $\nu_b = b - 1$; $\nu_c = c - 1$ **-** *число степеней свободы факторов А,В и С соответственно;*

$S_{ab}^2 = \dfrac{SS_{ab}}{\nu_{ab}}$; $SS_{ac}^2 = \dfrac{SS_{ac}}{\nu_{ac}}$; $SS_{bc}^2 = \dfrac{SS_{bc}}{\nu_{bc}}$; $S_{abc}^2 = \dfrac{SS_{abc}}{\nu_{abc}}$ *- оценки дисперсий по взаимосвязям уровней факторов;*

$\nu_{ab} = (a-1)(b-1)$; $\nu_{ac} = (a-1)(c-1)$; $\nu_{bc} = (a-1)(c-1)$;

ν_{abc} = (a –1)(b –1)(с-1) - *числа степеней свободы взаимодействий факторов;*

$\boldsymbol{S}_e^2 = \dfrac{SS_e}{\nu_e}$ **-** *остаточная оценка дисперсии (дисперсия ошибки);*

$\boldsymbol{\nu}_e$ **= *N* – *abc* = *abc*(*n* - 1) -** *число степеней свободы при определении ошибки.*

$$\nu = \boldsymbol{\nu}_a + \boldsymbol{\nu}_b + \nu_c + \nu_{ab} + \nu_{ac} + \nu_{bc} + \nu_{abc} + \boldsymbol{\nu}_e .$$

Проверка H_0- гипотезы

Определение расчетных значений критерия:

$$\boldsymbol{F}_{a\ расч.} = \frac{S_a^2}{S_e^2};\ \boldsymbol{F}_{b\ расч.} = \frac{S_b^2}{S_e^2};\ \boldsymbol{F}_{c\ расч} = \frac{S_c^2}{S_e^2};\ \boldsymbol{F}_{ab\ расч.} = \frac{S_{ab}^2}{S_e^2};\ \boldsymbol{F}_{ac\ расч.} = \frac{S_{ac}^2}{S_e^2};$$

$$\boldsymbol{F}_{bc\ расч.} = \frac{S_{bc}^2}{S_e^2};\quad \boldsymbol{F}_{abc\ расч.} = \frac{S_{abc}^2}{S_e^2}.$$

Критическое значение $F_{табл.}$ определяется по прил.4 при **α, $\boldsymbol{\nu}_1$** и **$\boldsymbol{\nu}_2$**

Если $F_{расч.} \le F_{табл.}$ при α, ν_1, ν_2, то гипотеза H_0- принимается. В противном случае – отклоняется (устанавливается существенное влияние факторов на изменчивость признака) и исследуется значимость средних признака на отдельных уровнях.

4.1.3. Схемы дисперсионного анализа

4.1.3.1. Преобразования вычислительных формул

Расчет непосредственно по вышеприведенным формулам удобен только в случае малого числа уровней и опытов. В противном случае используются преобразованные формулы сумм квадратов.

В основе вычислительных формул лежат преобразования, которые проиллюстрируем на общей сумме квадратов однофакторного эксперимента:

$$\boldsymbol{SS} = \sum_{i=1}^{a}\sum_{m=1}^{n}\left(X_{im} - \overline{X}\right)^2 = \sum_{i=1}^{a}\sum_{m=1}^{n}\left(X_{im}^2 - 2X_{im}\overline{X} - \overline{X}^2\right) =$$

$$= \sum_{i=1}^{a}\sum_{m=1}^{n}X_{im}^2 - \boldsymbol{2}\sum_{i=1}^{a}\sum_{m=1}^{n} X_{im}\overline{X} + \sum_{i=1}^{a}\sum_{m=1}^{n} \overline{X}^2 = \sum_{i=1}^{a}\sum_{m=1}^{n}X_{im}^2 -$$

$$-2\mathrm{N}\frac{\sum_{i=1}^{a}\sum_{m=1}^{n}X_{im}}{N}\frac{\sum_{i=1}^{a}\sum_{m=1}^{n}X_{im}}{N} + +\mathbf{N}\left(\frac{\sum_{i=1}^{a}\sum_{m=1}^{n}X_{im}}{N}\right)^2 = \sum_{i=1}^{a}\sum_{m=1}^{n}X_{im}^2 - \mathbf{N}\,\overline{X}^2 =$$

$$= \sum_{i=1}^{a}\sum_{m=1}^{n}X_{im}^2 - \frac{\left(\sum_{i=1}^{a}\sum_{m=1}^{n}X_{im}\right)^2}{N} = C_0 - \frac{C^2}{N} \qquad (4.1.28)$$

В результате аналогичных преобразований имеем нижеследующие формулы.

Однофакторный эксперимент, разное число параллельных опытов; общее число опытов $N = a\sum_{i=1}^{a} n_i$.

$$SS_a = \sum_{i=1}^{a} n_i\left(\bar{X}_i - \bar{X}\right)^2 = \sum_{i=1}^{a} n_i \bar{X}_i^{\,2} - \frac{1}{N}\left(\sum_{i=1}^{a}\sum_{m=1}^{n} X_{im}\right)^2 = \frac{a}{N}\sum_{i=1}^{a} C_i^{\,2} - \frac{C^2}{N}; \qquad (4.1.29)$$

$$SS_e = \sum_{i=1}^{a}\sum_{m=1}^{n}\left(X_{im} - \bar{X}_i\right)^2 = \sum_{i=1}^{a}\sum_{m=1}^{n} X_{ij}^2 - \sum_{i=1}^{a} n_i \bar{X}_i^2 = \mathbf{C}_0 - \frac{1}{n}\sum_{i=1}^{a} C_i^2, \qquad (4.1.30)$$

где $$C = \sum_{i=1}^{a}\sum_{m=1}^{n} X_{im} = \sum_{i=1}^{a} C_i; \quad C_i = \sum_{m=1}^{n} X_{im}; \ \mathbf{C}_0 = \sum_{i=1}^{a}\sum_{m=1}^{n} X_{ij}^2 \qquad (4.1.31)$$

Двухфакторный эксперимент, равное число параллельных опытов; общее число опытов $N = abn$

$$SS = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n}\left(X_{ijm} - \bar{X}\right)^2 = \sum\left(X_{ijm}^2 - 2X_{ijm}\frac{\sum X_{ijm}}{N} + \frac{\left(\sum X_{ijm}\right)^2}{N^2}\right) =$$

$$= \sum X_{ijm}^2 - \frac{2}{N}\left(\sum X_{ijm}\right)^2 + \frac{abn}{N^2}\left(\sum X_{ijm}\right)^2 = \sum X_{ijm}^2 - \frac{1}{N}\left(\sum X_{ijm}\right)^2 = \mathbf{C}_0 - \frac{C^2}{N}; \qquad (4.1.32)$$

$$SS_a = \mathrm{bn}\sum_{i=1}^{a}\left(\bar{X}_i - \bar{X}\right)^2 = \frac{a}{N}\sum_{i=1}^{a}\left(\sum_{j=1}^{b}\sum_{m=1}^{n} X_{ijm}\right)^2 - \frac{1}{N}\left(\sum X_{ijm}\right)^2 = \frac{a}{N}C_i^2 - \frac{C^2}{N}; \qquad (4.1.33)$$

$$SS_b = \mathrm{an}\sum_{j=1}^{b}\left(\bar{X}_j - \bar{X}\right)^2 = \frac{b}{N}\sum_{j=1}^{b}\left(\sum_{i=1}^{a}\sum_{m=1}^{n} X_{ijm}\right)^2 - \frac{1}{N}\left(\sum X_{ijm}\right)^2 = \frac{b}{N}C_j^2 - \frac{C^2}{N}; \qquad (4.1.34)$$

$$SS_{ab} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\bar{X} + \bar{X}_{ij} - \bar{X}_i - \bar{X}_j\right)^2 = \frac{ab}{N}\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\sum_{m=1}^{n} X_{ijm}\right)^2 - \mathrm{SS}_a - \mathrm{SS}_b - \frac{C^2}{N} =$$

$$= \frac{ab}{N}\sum_{i=1}^{a}\sum_{j=1}^{b} C_{ij}^2 - \mathrm{SS}_a - \mathrm{SS}_b - \frac{C^2}{N}; \qquad (4.1.35)$$

$$SS_e = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n}\left(X_{ijm} - \bar{X}_{ij}\right)^2 = \mathrm{C}_0 - \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b} C_{ij}^2, \qquad (4.1.36)$$

где: $\sum = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n}$;

$$\mathbf{C} = \sum X_{ijm}; \quad \mathbf{C}_0 = \sum X_{ijm}^2; \quad \mathbf{C}_i = \sum_{j=1}^{b}\sum_{m=1}^{n} X_{ijm}; \quad \mathbf{C}_j = \sum_{i=1}^{a}\sum_{m=1}^{n} X_{ijm}; \quad \mathbf{C}_{ij} = \sum_{m=1}^{n} X_{ijm} \qquad (4.1.37)$$

Трехфакторный эксперимент, равное число параллельных опытов; общее число опытов N = abcn

$$SS = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=}^{n}\left(X_{ijkm} - \bar{X}\right)^2 = \sum\left(X_{ijkm} - \bar{X}\right)^2 = \mathrm{C}_0 - \frac{C^2}{N}; \qquad (4.1.38)$$

$$SS_a = bcn\sum_{i=1}^{a}\left(\bar{X}_i - \bar{X}\right)^2 = \frac{a}{N}\sum_{i=1}^{a} C_i^2 - \frac{C^2}{N}; \ \mathbf{SS}_b = \mathrm{acn}\sum_{j=1}^{b}\left(\bar{X}_j - \bar{X}\right)^2 = \frac{b}{N}\sum_{j=1}^{b} C_j^2 - \frac{C^2}{N}; \qquad (4.1.40)$$

$$SS_c = abn\sum_{k=1}^{c}\left(\overline{X}_k - \overline{X}\right)^2 = \frac{c}{N}\sum_{k=1}^{c}C_k^2 - \frac{C^2}{N}; \tag{4.1.41}$$

$$SS_{ab} = cn\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{X} + \overline{X}_{ij} - \overline{X}_i - \overline{X}_j\right)^2 = \frac{ab}{N}\sum_{i=1}^{a}\sum_{j=1}^{b}C_{ij}^2 - \mathrm{SS}_a - \mathrm{SS}_b - \frac{C^2}{N}; \tag{4.1.42}$$

$$SS_{ac} = bn\sum_{i=1}^{a}\sum_{k=1}^{c}\left(\overline{X} + \overline{X}_{ik} - \overline{X}_i - \overline{X}_k\right)^2 = \frac{ac}{N}\sum_{i=1}^{a}\sum_{k=1}^{c}C_{ik}^2 - \mathrm{SS}_a - \mathrm{SS}_c - \frac{C^2}{N}; \tag{4.1.43}$$

$$SS_{bc} = an\sum_{j=1}^{b}\sum_{k=1}^{c}\left(\overline{X} + \overline{X}_{jk} - \overline{X}_j - \overline{X}_k\right)^2 = \frac{bc}{N}\sum_{j=1}^{b}\sum_{k=1}^{c}C_{jk}^2 - \mathrm{SS}_b - \mathrm{SS}_c - \frac{C^2}{N}; \tag{4.1.44}$$

$$SS_{abc} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\left(\overline{X}_{ijk} + \overline{X}_i + \overline{X}_j + \overline{X}_k - \overline{X}_{ij} - \overline{X}_{ik} - \overline{X}_{jk} - \overline{X}\right)^2 =$$

$$= \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}C_{ijk}^2 - \mathrm{SS}_a - \mathrm{SS}_b - \mathrm{SS}_c - \mathrm{SS}_{ab} - \mathrm{SS}_{ac} - \mathrm{SS}_{bc} - \frac{C^2}{N}; \tag{4.1.45}$$

$$\boldsymbol{SS}_e = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n}\left(X_{ijkm} - \overline{X}_{ijk}\right)^2 = \mathbf{C}_0 - \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}C_{ijk}^2 . \tag{4.1.46}$$

Где $\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n} = \sum$;

$$\boldsymbol{C} = \sum X_{ijm};\quad \boldsymbol{C}_0 = \sum X_{ijm}^2;\quad \boldsymbol{C}_i = \sum_{j=1}^{b}\sum_{k=1}^{c}\sum_{m=1}^{n}X_{ijkm};\quad \boldsymbol{C}_j = \sum_{i=1}^{a}\sum_{k=1}^{c}\sum_{m=1}^{n}X_{ijkm};$$

$$\boldsymbol{C}_k = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n}X_{ijkm};\quad \boldsymbol{C}_{ij} = \sum_{k=1}^{c}\sum_{m=1}^{n}X_{ijkm};\quad \boldsymbol{C}_{ik} = \sum_{j=1}^{b}\sum_{m=1}^{n}X_{ijkm};\quad \boldsymbol{C}_{jk} = \sum_{i=1}^{a}\sum_{m=1}^{n}X_{ijkm};$$

$$\boldsymbol{C}_{ijk} = \sum_{m=1}^{n}X_{ijkm} . \tag{4.1. 47}$$

Результаты вычислений принято представлять в виде таблицы дисперсионного анализа, например, схема двухфакторного дисперсионного анализа (таблица 25).

4.1.3.2.Алгоритмы расчетов

Алгоритм 1.

1. Построение вспомогательной таблицы

2. Вычисление средних: (например, формулы 4.1.5, 4.1.6, 4.1.8 - однофакторный эксперимент; формулы 4.1.18, 4.1.19, 4.1.20 – двухфакторный эксперимент)

3. Вычисление сумм квадратов: (например, формулы 4.1.2, 4.1.3, 4.1.4, 4.1.7- однофакторный эксперимент; формулы 4.1.13, 4.1.14, 4.1.15, 4.1.16 – двухфакторный эксперимент)

4. Вычисление оценок дисперсий (например, формулы 4.1.21, 4.1.22, 4.1.23, 4.1.24–двухфакторный эксперимент)

5. Проверка гипотезы H_0 (например, формула 4.1.26)

Алгоритм 2.

1. Построение вспомогательной таблицы.

2. Вычисление коэффициентов (вспомогательных сумм), например, для трехфакторного эксперимента: $\boldsymbol{C}$, $\boldsymbol{C}_0$, $\boldsymbol{C}_i$, $\boldsymbol{C}_j$, $\boldsymbol{C}_k$, $\boldsymbol{C}_{ij}$, $\boldsymbol{C}_{ik}$, $\boldsymbol{C}_{jk}$, $\boldsymbol{C}_{ijk}$ (формулы 4.1.47)

3. Вычисление сумм квадратов: (например, формулы 4.1.38 - 46 для трехфакторного эксперимента)

4. Вычисление оценок дисперсий

5. Проверка гипотезы H_0

6. Если H_0 не отклоняется, то - проверка значимости уровней факторов

Таблица 25 – Схема двухфакторного дисперсионного анализа

Источник изменчивости	**Сумма квадратов эффектов**	ν	**Оценка дисперсии**	***F***	η *100,%
Общая	$SS = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n}\left(X_{ijm} - \overline{X}\right)^2$	$\nu =$ $= N\text{-}1$	$S = \dfrac{SS}{\nu}$		
Фактор *A*	$SS_a = bn\sum_{i=1}^{a}\left(\overline{X}_i - \overline{X}\right)^2$	$\nu_a =$ $= a\text{-}1$	$S_a = \dfrac{SS_a}{\nu_a}$	$F_a = \dfrac{S_a}{S_e}$	$\eta_a = \dfrac{S_a}{S}$
Фактор *B*	$SS_b = an\sum_{j=1}^{b}\left(\overline{X}_j - \overline{X}\right)^2$	$\nu_b =$ $= b\text{-}1$	$S_b = \dfrac{SS_b}{\nu_b}$	$F_b = \dfrac{S_b}{S_e}$	$\eta_b = \dfrac{S_b}{S}$
Взаимодействие *AB*	$SS_{ab} = n\sum_{i=1}^{a}\sum_{j=1}^{b}\left(\overline{X} + \overline{X}_{ij} - \overline{X}_i - \overline{X}_j\right)^2$	$\nu_{ab} =$ $= (a\text{-}1)\times(b\text{-}1)$	$S_{ab} = \dfrac{SS_{ab}}{\nu_{ab}}$	$F_{ab} = \dfrac{S}{S_e}$	$\eta_{ab} = \dfrac{S_{ab}}{S}$
Остаточная (внутри групп)	$SS_e = \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n}\left(X_{ijm} - \overline{X}_{ij}\right)^2$	$\nu_e =$ $= ab\times(n\text{-}1)$	$S_e = \dfrac{SS_e}{\nu_e}$		$\eta_e = \dfrac{S_e}{S}$

4.1.4.ПРИМЕРЫ
4.1.4.1. Однофакторный дисперсионный анализ

Пример 1. Равное число наблюдений. Анализ влияния производственной смены на количество брака в выпускаемой продукции (Юденков В.А. Дисперсионный анализ. Уч. пос., 1982. – с.95).

На некотором предприятии, работающем в три смены, получены данные о проценте брака выпускаемой продукции в каждой из смен за семь последовательных дней. Данные опыта (однофакторный эксперимент, вариант 1.2 - три уровня, равное число опытов), приведены в табл.26.

Таблица 26 – Данные о % брака выпускаемой продукции

Смена (уровни), i = 1,2,3	**Результаты опытов: X_{im}, % брака; m = 1, 2, …, 7, день наблюдения**						
	Пн. 1	**Вт. 2**	**Ср. 3**	**Чт. 4**	**Пт. 5**	**Сб. 6**	**Вс. 7**
1	2,0	1,5	3,0	6,0	0,2	0	1,0
2	1,5	4,0	4,0	0	0	2,5	1,5
3	1,5	1,5	6,0	6,0	0	3,0	1,0

Проверяемая гипотеза H_0: отсутствие влияния фактора А – смена – на % брака выпускаемой продукции.

Решение. 2-й алгоритм

1) Построение вспомогательной таблицы

Построим вспомогательную таблицу (табл.27) для промежуточных вычислений сумм квадратов.

Таблица 27 – Вспомогательные вычисления при однофакторном дисперсионном анализе

i	X_{im}							C_i	C_i^2	X_{im}^2						
	1	2	3	4	5	6	7			1	2	3	4	5	6	7
1	2,	1,	3,	6,	0,	0	1,	13,	187,6	4,0	2,2	9,0	36,	0,0	0	1,0
2	1,	4,	4,	0	0	2,	1,	13,	182,2	2,2	16,	16,	0	0	6,2	2,2
3	1,	1,	6,	6,	0	3,	1,	19,	361,0	2,2	2,2	36,	36,	0	9,0	1,0

2) Вычисления вспомогательных сумм

$$C^2 = \left(\sum_{i=1}^{a}\sum_{m=1}^{n} X_{ij}\right)^2 = \left(\sum_{i=1}^{a} C_i\right)^2 = \mathbf{46{,}2^2} = \mathbf{2134{,}44;}$$

$$\boldsymbol{C_0} = \sum_{i=1}^{a}\sum_{m=1}^{n} X_{ij}^2 = \mathbf{181{,}54;} \quad \sum_{i=1}^{a} C_i^{\,2} = \mathbf{730{,}94.}$$

3)Вычисления сумм квадратов

Общая сумма квадратов: $\boldsymbol{SS} = \sum_{i-1}^{a}\sum_{m=1}^{n}\left(X_{ij}-\overline{X}\right)^2 = \sum_{i=1}^{a}\sum_{m=1}^{n}X_{ij}^2 - \frac{1}{N}\left(\sum_{i=1}^{a}\sum_{m=1}^{n}X_{ij}\right)^2 =$

$= \mathbf{C}_0 - \frac{C^2}{N} = 181{,}54 - \frac{46{,}2^2}{21} = \mathbf{79{,}9}.$

Сумма квадратов между группами: $\boldsymbol{SS}_a = \sum_{i=1}^{a} n_i\left(\overline{X}_i-\overline{X}\right)^2 = n\sum_{i=1}^{a}\left(\overline{X}_i-\overline{X}\right)^2$

$= \sum_{i=1}^{a} n_i\overline{X}_i^{\,2} - \frac{1}{N}\left(\sum_{i=1}^{a}\sum_{m=1}^{n}X_{im}\right)^2 = \frac{a}{N}\sum_{i=1}^{a}C_i^{\,2} - \frac{C^2}{N} = \frac{3}{21}730{,}94 - 101{,}64 = \mathbf{2{,}78}.$

Сумма квадратов внутри групп: $\boldsymbol{SS}_e = \sum_{i=1}^{a}\sum_{m=1}^{n}\left(X_{im}-\overline{X}_i\right)^2 = \sum_{i=1}^{a}\sum_{m=1}^{n}X_{ij}^2 -$

$\sum_{i=1}^{a} n_i\overline{X}_i^2 = \boldsymbol{C}_0 - \frac{1}{n}\sum_{i=1}^{a}C_i^2 = 181{,}54 - \frac{1}{7}730{,}94 = \mathbf{77{,}12};$

Проверка: $\boldsymbol{SS}_e = \boldsymbol{SS} - \boldsymbol{SS}_a = 79{,}90 - 2{,}78 = \mathbf{77{,}12}$.

4) Проверка гипотезы:

$\boldsymbol{S}_a^2 = \frac{SS_a}{\nu_a} = \frac{2{,}78}{2} = \mathbf{1{,}39}$; $\boldsymbol{S}_e^2 = \frac{SS_e}{\nu_e} = \frac{77{,}12}{18} = \mathbf{4{,}28}$;

$\boldsymbol{\nu}_a = \boldsymbol{a} - \mathbf{1} = 3-1 = \mathbf{2}$; $\boldsymbol{\nu}_e = \boldsymbol{N} - \boldsymbol{a} = 21 - 3 = \mathbf{18}$;

$\boldsymbol{F}_{a\ расч.} = \frac{S_a^2}{S_e^2} = \frac{1{,}39}{4{,}28} = \mathbf{0{,}32}$; $\mathbf{F}_{табл.} = \mathbf{3{,}55}$ при $\boldsymbol{\alpha} = \mathbf{0{,}05}$; $\boldsymbol{\nu}_1 = \mathbf{2}$;

$\boldsymbol{\nu}_2 = \mathbf{8}.$

Так как $\boldsymbol{F}_{расч.} \le \boldsymbol{F}_{табл.}$,

то гипотеза H_0 об отсутствии влияния фактора смены на % брака выпускаемой продукции, принимается. Т.е. производственная смена не влияет на появление брака.

Пример 2. Неравное число наблюдений по уровням. Срок службы электрических ламп. [с. 370-372, Митропольский А.К. Техника статистических вычислений.- М.: Наука, 1971.-576с.]

Для изготовления каждой партии ламп была взята проволока разных заводов изготовителей. Все же прочие условия производства были одинаковы для каждой партии. Требуется выяснить, отличаются ли партии ламп между собой по сроку службы. Если ответ будет положительным, то можно думать, что качество проволоки варьирует реально, и следовательно, для достижения стандартизации производства электрических ламп необходимо достигнуть большей однородности проволоки во всех партиях и пересмотреть контракты с поставщиками. Данные наблюдений представлены табл.28.

Таблица 28 – Данные о сроке службы электрических ламп

Партия ламп (группа)	**Результаты наблюдений: X_{im} - срок службы, тыс. час, m = 1, 2, …, n_i**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
№1	1,60	1,61	1,65	1,68	1,70	1,72	1,80	-
№2	1,58	1,64	1,64	1,70	1,75	-	-	-
№3	1,46	1,55	1,60	1,62	1,64	1,66	1,74	1,82
№4	1,51	1,52	1,53	1,60	1,67	1,68	-	-

Проверяемая гипотеза H_0: «партии ламп не отличаются между собой по сроку службы».

Решение. 1-й алгоритм

1) Построение вспомогательной таблицы

Построим вспомогательную таблицу (табл.28) для вычисления сумм квадратов.

Пояснения к вычислениям в табл.29:

$$C_i = \sum_{m=1}^{n_i} X_{im};$$

$$\overline{X} = \frac{\sum_{i=1}^{a} C_i}{N}; \quad \overline{X} = \frac{42,57}{26} = 1,64 \quad \text{или } \overline{X} = \frac{\sum_{i=1}^{a} \overline{X}_i}{a}; \quad \overline{X} = \frac{1,68 + 1,66 + 1,64 + 1,57}{4} = 1,64.$$

Таблица 29 – Вычисления средних значений и сумм квадратов

Партия ламп (группа)	n_i	C_i	$\overline{X}_i$	$\mathbf{x}_{im} = X_{im} - \overline{X}$	$\sum_{m=1}^{n_i} x_{im}$	$\overline{x}_i$	$\sum_{m=1}^{n_i} x_{im}^2$	$\mathbf{n}_i\ \overline{x}_i^2$
№1	7	11,76	1,68	-0,04; -0,03; 0,01; 0,04; 0,06; 0,08; 0,16	0,28	0,04	0,0398	0,0112
№2	5	8,31	1,66	-0,06; 0; 0; 0,06; 0,11	0,11	0,02	0,0193	0,0020
№3	8	13,09	1,64	-0,18; -0,09; -0,04; -0,02; 0; 0,02; 0,10; 0,18	- 0,03	0	0,0853	0
№4	6	9,41	1,57	-0,13; -0,12; -0,11; -0,07; -0,04; 0,04	- 0,43	- 0,07	0,0515	0,0294
Σ	26	42,57	1,64		- 0,07	0,01	0,1959	0,0426

$\overline{X}_i = \dfrac{C_i}{n_i}$; $\quad i$=1; $\quad \overline{X}_{i=1} = \dfrac{11,76}{7} = 1,68$; $\quad \overline{X}_2 = 1,66$ и т.д.

$x_{im} = X_{im} - \bar{X}$: i=1; $x_{i=1\ m=1} = x_{11}$= 1,60 -1,64 = - 0,04; x_{12}= 1,61-1,64 = - 0,03 и т.д.

$$\bar{x}_i = \frac{\sum_{m=1}^{n_i} x_{im}}{n_i}; \quad i=1; \quad \bar{x}_{i=1} = \frac{0,28}{7} = 0,04; \quad \bar{x}_2 = 0,02 \text{ и т.д.}$$

2) Вычисления сумм квадратов

Общая сумма квадратов:

$$\boldsymbol{SS} = \sum_{i=1}^{a}\sum_{m=1}^{n_i}\left(X_{im} - \bar{X}\right)^2 = 0,0398 + 0,0193 + 0,0853 + 0,0515 = \mathbf{0,1959;}$$

$$\boldsymbol{SS}_a = \sum_{i=1}^{a} n_i\left(\bar{X}_i - \bar{X}\right)^2 = \mathbf{0,0426;} \quad \boldsymbol{SS}_e = \sum_{i=a}^{a}\sum_{m=1}^{n_i}\left(X_{im} - \bar{X}_i\right)^2 = \mathbf{0,1533;}$$

3)Оценки дисперсий

S^2 **= 0,0078;** S_a^2 **= 0,0142;** S_e^2 **= 0,0070**

4)Проверка гипотезы

$$F_{расч.} = \frac{0,0142}{0,0070} = 2,03;$$

при $\nu_1 = 3, \nu_2 = 22$ и $\alpha = 0,05$ $F_{табл.} = 3,05$.

Так как $F_{расч.} \le F_{табл.}$, то $\boldsymbol{H}_0$**:** «партии ламп не отличаются между собой по сроку службы» – принимаем.

4.1.4.2. Двухфакторный дисперсионный анализ

Пример 1. Исследование склонности к алкоголизму и наркомании в колледжах.

Были проведены опросы учащихся в 6 группах голландских колледжей с целью выяснить: имеется ли у учащихся склонность к вредным привычкам? Для этого учащемуся задавались вопросы: употребляют они А - алкоголь (1) или нет (2), принимают ли они В - наркотики (1) или нет (2), или же употребляют и то , и другое В зависимости от ответов в каждом коллективе подсчитывалось число человек, соответствующих тем или иным сочетаниям уровней факторов А и В. Данные опросов систематизированы в таблице 30.

Таблица 30 – Результаты исследований склонности к алкоголизму (A) и наркомании (B)

№ группы	Сочетания уровней A и B	Число наблюдений в группах X_{ijm}, m = 1,2,…,n					
		1-й коллектив	**2-й коллектив**	**3-й коллектив**	**4-й коллектив**	**5-й коллектив**	**6-й коллектив**
1	1 1	3	5	2	4	3	1
2	1 2	4	6	8	7	3	2
3	2 1	6	5	5	7	5	5
4	2 2	6	8	10	8	11	8

Решение. Алгоритм 2

1) Промежуточные вычисления, построение вспомогательной таблицы

Таблица 31– Вспомогательная таблица для вычисления сумм квадратов

№	*A B*	C_{ij}	C_i	C_j	C_{ij}^2	X_{ijm}^2						$\sum_{m=1}^{n} X_{ijm}^2$
						1	**2**	**3**	**4**	**5**	**6**	
1	**1 1**	18	**48**		324	9	25	4	16	9	1	64
2	**1 2**	30		**51**	900	16	36	64	49	9	4	178
3	**2 1**	33	**84**	**81**	1089	36	25	25	49	25	25	185
4	**2 2**	51			2601	36	64	100	64	121	64	449
$\sum$		**С= =132**			**4914**	проверка						**C_0= = 876**
						97	180	193	178	164	94	

Представим результаты промежуточных вычислений в пространстве эксперимента двухфакторного анализа (рис.47)

2) Вычисления сумм квадратов

Общая сумма квадратов:

$$\boldsymbol{SS} = \mathbf{C}_0 - \frac{C^2}{N} = 876 - \frac{132^2}{24} = 876 - 726 = \mathbf{150}$$

Взвешенная сумма квадратов эффектов факторов А и В (сумма квадратов между группами):

$$\boldsymbol{SS}_a = \frac{a}{N}\sum_{i=1}^{a} C_i^2 - \frac{C^2}{N} = \frac{2}{24}(2304+7056) - 726 = 780 - 726 = \mathbf{54}$$

$$\boldsymbol{SS_b} = \frac{b}{N}\sum_{j=1}^{b} C_j^2 - \frac{C^2}{N} = 12\,(2601+6561) - 726 = 763{,}5 - 726 = \mathbf{37{,}5}$$

Взвешенная сумма квадратов взаимодействия уровней факторов A и B или смешанный эффект факторов A и B (сумма квадратов между группами):

$$\boldsymbol{SS_{ab}} = \frac{ab}{N}\sum_{i=1}^{a}\sum_{j=1}^{b} C_{ij}^2 - SS_a - SS_b - \frac{C^2}{N} = \frac{1}{6}4914 - (54+37{,}5+726) = 879 - 817{,}5 = \mathbf{1{,}5}$$

Ошибка эксперимента (сумма квадратов внутри групп):

$$\boldsymbol{SS_e} = \boldsymbol{C_0} - \frac{ab}{N}\sum_{i=1}^{a}\sum_{j=1}^{b} C_{ij}^2 = 876 - 819 = \mathbf{57}$$

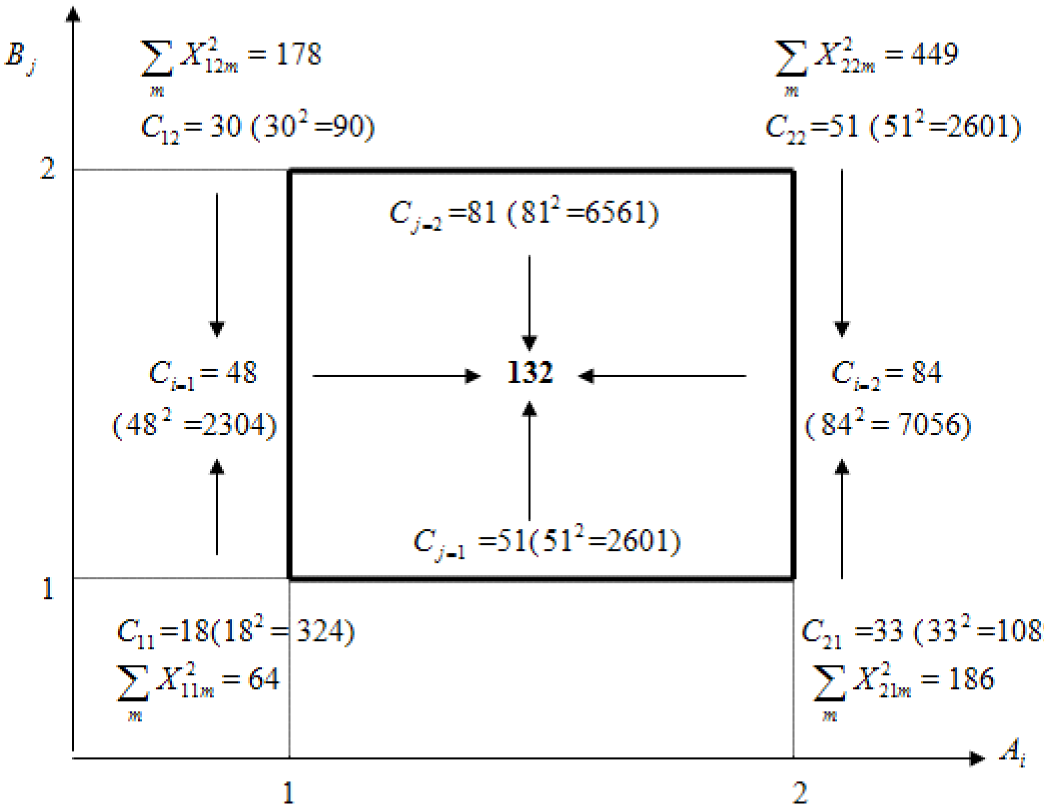


Рис. 47 – Результаты промежуточных вычислений в пространстве эксперимента двухфакторного анализа

Проверка:

$$\boldsymbol{SS_e} = \boldsymbol{SS} - \boldsymbol{SS_a} - \boldsymbol{SS_b} - \boldsymbol{SS_{ab}} = 150 - (54+37{,}5+1{,}5) = 150 - 93 = \mathbf{57.}$$

3) Оценки дисперсий

$$S^2 = \frac{SS}{\nu} = \frac{150}{23} = \mathbf{6{,}5;}\quad \nu = N - 1 = 24 - 1 = \mathbf{23;}$$

$$S_a^2 = \frac{SS_a^2}{\nu_a} = \frac{54}{1} = \mathbf{54;}\quad \nu_a = a - 1 = 3\text{-}1 = \mathbf{2;}\quad S_b^2 = \frac{SS_b^2}{\nu_b} = \frac{37{,}5}{1} = \mathbf{37{,}5};$$

$$\nu_b = \boldsymbol{b} - \mathbf{1} = \mathbf{3\text{-}1=2}; \quad S_{ab}^2 = \frac{SS_{ab}^2}{\nu_{ab}} = \frac{1{,}5}{4} = \mathbf{0{,}375};$$

$$\nu_{ab} = (a\text{-}1)(b\text{-}1) = \mathbf{4}; \; S_e^2 = \frac{SS_e^2}{\nu_e} = \frac{57}{13} \approx \mathbf{4{,}38}; \; \nu_e = N\text{-}ab = 24\text{-}9 = \mathbf{13}.$$

4) Проверка гипотез

$\boldsymbol{F}_{a\ расч.} = \frac{S_a^2}{S_e^2} = \frac{54}{4{,}38} \approx$ **12,3**; при $\alpha = 0{,}05$, $\nu_1 = 2$, $\nu_2 = 13$ табличное значение $F_{кр} =$ **3,80**;

Так как $\boldsymbol{F}_{a\ расч.}$ **= 12,3 >** $F_{кр.}$ **= 3,80, то гипотеза** $\boldsymbol{H}_0$**: отклоняется**

$\boldsymbol{F}_{a\ расч.} = \frac{S_b^2}{S_e^2} = \frac{37{,}5}{4{,}38} \approx$ **8,33;** при $\alpha = 0{,}05$, $\nu_1 = 2$, $\nu_2 = 13$ табличное значение $F_{кр.} =$ **3,80**;

Так как $\boldsymbol{F}_{b\ расч.}$ **= 8,33 >** $F_{кр.}$ **= 3,80, то гипотеза** $\boldsymbol{H}_0$**: отклоняется**

$\boldsymbol{F}_{ab\ расч.} = \frac{S_{ab}^2}{S_e^2} = \frac{1{,}5}{4{,}38} \approx$ **0,34 ;** при $\alpha = 0{,}05$, $\nu_1 = 4$, $\nu_2 = 13$ табличное значение $F_{кр.} =$ **3,18**

Так как $\boldsymbol{F}_{ab\ расч.}$ **= 0,34 <** $F_{кр.}$ **= 3,18, то гипотеза** $\boldsymbol{H}_0$ **принимается**

Наблюденные значения $F_{a\ расч.}$, $\mathrm{F}_{b\ расч.}$ превышают соответствующие значения $F_{кр.}$. Таким образом, дисперсионный анализ выявил существенное влияние на результативный признак факторов *A* и *B*.

Построим таблицу дисперсионного анализа (табл.32).

Таблица 32 – Результаты двухфакторного дисперсионного анализа

Источник изменчивости	**Сумма квадратов эффектов**	ν	**Оценка диспер-сии**	$\boldsymbol{F}$	η***100,%**
Общая	$\boldsymbol{SS} = 150$	$\nu = 23$	$S^2 = 6{,}5$		
Фактор *A*	$\boldsymbol{SS}_a = 54$	$\nu_a = 1$	$S_a^2 = 54$	$F_a = 18{,}9$	$\eta_a = 36$
Фактор *B*	$\boldsymbol{SS}_b = 37{,}5$	$\nu_b = 1$	$S_b^2 = 37{,}5$	$F_b = 13{,}15$	$\eta_b = 25$
Факторы *AB*	$\boldsymbol{SS}_{ab} = 1{,}5$	$\nu_{ab} = 1$	$S_{ab}^2 = 1{,}5$	$F_{ab} = 0{,}52$	$\eta_{ab} = 1$
Остаточная	$\boldsymbol{SS}_e = 57$	$\nu_e = 20$	$S_e^2 = 2{,}85$		$\eta_e = 38$

Из таблицы дисперсионного анализа (табл.32) по величине η видно также, что доля влияния фактора A составляет 36%, доля фактора B – 25%, взаимодействия факторов – 1%, не учтенных в опыте причин – 38%.

Так как установлено существенное влияние факторов на изменчивость признака, исследуем значимость его средних на отдельных уровнях.

Из табл.32 найдем средние значения признаков на каждом из уровней факторов A и B. Так:

общее среднее $\overline{X} = \dfrac{\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{m=1}^{n} X_{ijm}}{N} = \dfrac{132}{24} = \mathbf{5{,}5}$;

$$\overline{X}_i = \frac{\sum_{i=1}^{a}\sum_{m=1}^{n} X_{ijm}}{bn}; \quad \overline{X}_{i=1} = \frac{48}{12} = \mathbf{4}; \quad \overline{X}_{i=2} = \frac{84}{12} = \mathbf{7};$$

$$\overline{X}_j = \frac{\sum_{j=1}^{b}\sum_{m=1}^{n} X_{ijm}}{an} \quad \overline{X}_{j=1} = \frac{51}{12} = \mathbf{4{,}25}; \quad \overline{X}_{j=2} = \frac{81}{12} = \mathbf{6{,}75}.$$

Стандартные ошибки разности средних $\overline{X}_i$, $\overline{X}_j$:

$$\overline{S}_i = \sqrt{\frac{2S_e^2}{n}} = \sqrt{\frac{2*2{,}85}{6}} \approx \mathbf{0{,}97} = \overline{S}_j = \overline{S}.$$

Из таблицы распределения Стьюдента определяем при $\alpha = 0{,}05$ и ν_e =20 значение $t_{табл} = 2{,}09$.

Тогда предельная ошибка разности средних: $\overline{S}$ *t = 0,97*2,09 = 2,02.

Попарное сравнение между собой значений $\overline{X}_i$,а также значений $\overline{X}_j$ показывает, что разности между ними превышают величину предельной ошибки разности средних. Следовательно, можно сделать вывод о существенности влияния на результативный признак каждого уровня факторов А и В.

Так как гипотеза об отсутствии влияния взаимодействия АВ на результативный признак принята, то исследовать стандартную ошибку средних $\overline{X}_{ij}$ не будем.

4.1.4.3.Трехфакторный дисперсионный анализ

Пример. Трехфакторный эксперимент на двух уровнях с равным числом повторных опытов, n = 3. Менеджер фирмы, занимающейся торговлей промышленными товарами, решил провести анализ причин, влияющих на выручку сети магазинов, принадлежащих фирме. В первую очередь он решил выяснить, зависит ли средняя по магазинам еженедельная выручка (результативный признак X_{ijkm}, $1*10^2$ тыс.у.е.): от уровня образования персонала (фактор A), от поставщиков (фактор B), от дня недели (фактор C). Каждый из факторов варьировался первоначально только на двух уровнях: для A – среднее (уровень 1) и высшее (уровень 2) профессиональное образование, для B – фирма-поставщик №1 (уровень 1) и фирма-поставщик №2 (уровень 2), для C – начало недели (уровень 1) и конец недели (уровень 2). Анализ проводился по данным трех недель наблюдений: n = 3 (табл.33).

Проверяемая гипотеза H_0: ни уровень образования торговцев, ни поставщики, ни начало и конец недели не влияют на выручку магазинов.

Решение. Алгоритм 2

1) Промежуточные вычисления, построение вспомогательной таблицы

В табл. 33 помещены данные эксперимента и часть промежуточных вычислений, необходимых для определения сумм квадратов.

Для облегчения вычислений, изобразим пространство трехфакторного эксперимента (рис.48).

Вычисления

С =72; $\dfrac{C^2}{N} = \dfrac{5184}{24} = 216$; $C_0 = 266$

Вычисления C_{ij}:

$C_{i=1\,j=1} = 12 + 7 = \mathbf{19}$; $C_{i=1\,j=2} = 8 + 14 = \mathbf{22}$; $C_{i=2\,j=1} = 9 + 9 = \mathbf{18}$; $C_{i=2\,j=2} = 8 + 5 = \mathbf{13}$.

Вычисления C_{ik}:

$C_{i=1\,k=1} = 12 + 8 = \mathbf{20}$; $C_{i=1\,k=2} = 7 + 14 = \mathbf{21}$ $C_{i=2\,k=} = 9 + 5 = \mathbf{14}$; $C_{i=2\,k=2} = 9 + 8 = 17$.

Вычисления C_{jk}:

$C_{j=1\,k=1} = 9 + 12 = \mathbf{21}$; $C_{j=1\,k=2} = 9 + 7 = \mathbf{16}$; $C_{j=2\,k=1} = 5 + 8 = 13$; $C_{j=2k=2} = 8 + 14 = \mathbf{22}$.

Вычисления C_i:

$C_{i=1} = 17 + 14 = \mathbf{31} = 18 + 13$; $C_{i=2} = 19 + 22 = \mathbf{41} = 21 + 20$.

Таблица 33 – Данные о выручке сети промтоварных магазинов; часть промежуточных вычислений

№	A B C	X_{ijkm}, $1{*}10^2$ у.е.			C_{ijk}	C^2_{ijk}	C_{ij}	C_i	X^2_{ijkm}			$\sum_{m=1}^{n} X^2_{ijkm}$
1	**1 1 1**	1	3	5	9	81	18	31	1	9	25	35
2	**1 1 2**	2	4	3	9	81			4	16	9	29
3	**1 2 1**	2	2	1	5	25	13		4	4	1	9
4	**1 2 2**	1	5	2	8	64			1	25	4	30
5	**2 1 1**	5	3	4	12	144	19	41	25	9	16	50
6	**2 1 2**	4	2	1	7	49			16	4	1	21
7	**2 2 1**	3	1	4	8	64	22		9	1	16	26
8	**2 2 2**	5	4	5	14	196			25	16	25	66
$\sum$		23	24	25	**С =72**	**704**	72	72	85	84	97	C_0 **=266**

Вычисления C_j:

$C_{j=1} = 16 + 21 = \mathbf{37} = 18 + 19$; $C_{j=2} = 22 + 13 = \mathbf{35} = 13 + 22$.

Вычисления C_k:

$C_{k=1}$ = 21 + 13 = **34** = 20 + 14 ; $C_{k=2}$ = 21 + 17 = **38** = 16 + 22.

1) Вычисления сумм квадратов

$$\boldsymbol{SS} = C_0 - \frac{C^2}{N} = 266 - \frac{5184}{24} = 266 - 216 = \mathbf{50};$$

$$\boldsymbol{SS}_a = \frac{a}{N}\sum_{i=1}^{a} C_i^2 - \frac{C^2}{N} = \frac{2}{24}\left(31^2 + 41^2\right) - 216 = 220{,}17 - 216 = \mathbf{4{,}17}\ ;$$

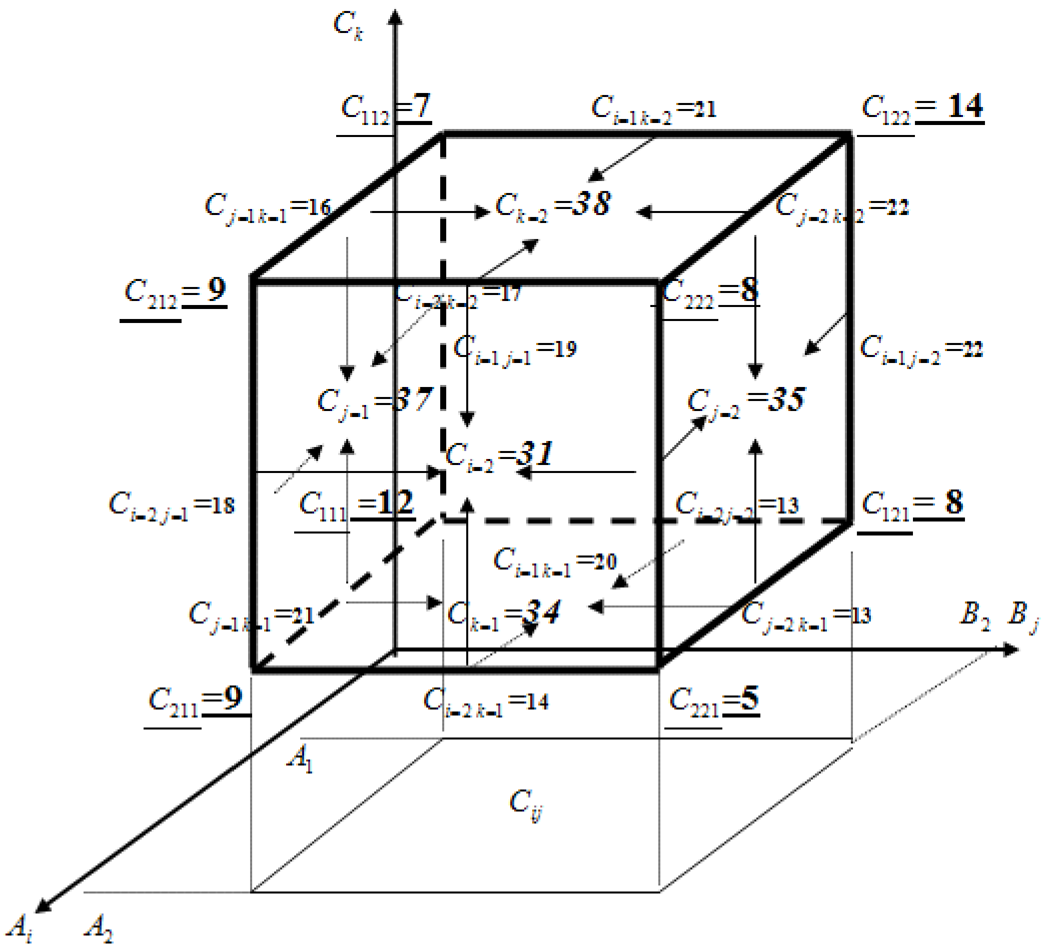


Рис. 48 – Вычисления в пространстве трехфакторного дисперсионного анализа

$$\boldsymbol{SS}_b = \frac{b}{N}\sum_{j=1}^{b} C_j^2 - \frac{C^2}{N} = \frac{2}{24}\left(37^2 + 35^2\right) - 216 = 216{,}17 - 216 = \mathbf{0{,}17};$$

$$\boldsymbol{SS}_c = \frac{c}{N}\sum_{k=1}^{c} C_k^2 - \frac{C^2}{N} = \frac{2}{24}\left(34^2 + 38^2\right) - 216 = 216{,}67 - 216 = \mathbf{0{,}67}\ ;$$

$$\boldsymbol{SS}_{ab} = \frac{ab}{N}\sum_{i=1}^{a}\sum_{j=1}^{b} C_{ij}^2 - \mathrm{SS}_a - \mathrm{SS}_b - \frac{C^2}{N} = \frac{4}{24}\left(18^2 + 13^2 + 19^2 + 22^2\right) - 4{,}17 - 0{,}17 - 216 =$$
$$= 223 - 220{,}33 = \mathbf{2{,}66};$$

$$\boldsymbol{SS}_{ac} = \frac{ac}{N}\sum_{i=1}^{a}\sum_{k=1}^{c} C_{ik}^2 - \mathrm{SS}_a - \mathrm{SS}_c - \frac{C^2}{N} = \frac{4}{24}\left(20^2 + 14^2 + 21^2 + 17^2\right) - 4{,}17 - 0{,}67 - 216 =$$
$$= 221 - 220{,}83 = \mathbf{0{,}16};$$

$$\boldsymbol{SS}_{bc} = \frac{bc}{N}\sum_{j=1}^{b}\sum_{k=1}^{c} C_{jk}^2 - \mathrm{SS}_b - \mathrm{SS}_c - \frac{C^2}{N} = \frac{4}{24}\left(21^2 + 13^2 + 16^2 + 22^2\right) - 0{,}17 - 0{,}67 - 216 =$$

$= 225 – 216{,}84 = \mathbf{8{,}16}$;

$$\boldsymbol{SS}_{abc} = \frac{1}{n}\sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}C_{ijk}^2 - SS_a - SS_b - SS_c - SS_{ab} - SS_{ac} - SS_{bc} - \frac{C^2}{N} =$$

$$= \frac{1}{3}704 – 15{,}99 – 216 = 234{,}67 - 231{,}99 = \mathbf{2{,}68;}$$

$$\boldsymbol{SS}_e^2 = C_0 - \sum_{i=1}^{a}\sum_{j=1}^{b}\sum_{k=1}^{c}C_{ijk}^2 = 266 – 234{,}67 = \mathbf{31{,}33.}$$

Проверка:

$$\boldsymbol{SS}_e^2 = SS - SS_a - SS_b - SS_c - SS_{ab} - SS_{ac} - SS_{bc} - SS_{abc} = 50 - 18{,}67 = \mathbf{31{,}33.}$$

4) Оценки дисперсий

$\boldsymbol{S}^2 = \frac{SS}{\nu} = \frac{50}{23} = \mathbf{2{,}17;}$ $\boldsymbol{\nu} =$ N - 1 $= 24 – 1 = \mathbf{23;}$ $\mathbf{S}_a^2 = \frac{SS_a}{\nu_a} = \mathbf{4{,}17;}$

$\boldsymbol{\nu}_a = a – 1 = 2 – 1 = \mathbf{1;}$

$\boldsymbol{S}_b^2 = \frac{SS_b}{\nu_b} = \mathbf{0{,}17;}$ $\boldsymbol{\nu}_b = b – 1 = 2 – 1 = \mathbf{1;}$ $\boldsymbol{S}_c^2 = \frac{SS_c}{\nu_c} = \mathbf{0{,}67;}$

$\nu_c =$ c -1 $= 2 – 1 = \mathbf{1;}$

$\boldsymbol{S}_{ab}^2 = \frac{SS_{ab}}{\nu_{ab}} = \mathbf{2{,}66;}$ $\nu_{ab} = (a – 1)(b – 1) = \mathbf{1;}$ $\boldsymbol{S}_{ac}^2 = \frac{SS_{ac}}{\nu_{ac}} = \mathbf{0{,}16;}$

$\nu_{ac} = (a - 1)(c - 1) = \mathbf{1;}$

$\boldsymbol{S}_{bc}^2 = \frac{SS_{bc}}{\nu_{bc}} = \mathbf{8{,}16;}$ $\nu_{bc} = (a – 1)(c – 1) = \mathbf{1;}$ $\boldsymbol{S}_{abc}^2 = \frac{SS_{abc}}{\nu_{abc}} = \mathbf{2{,}68;}$

$\nu_{abc} =$ (a – 1)(b – 1)(c-1) = **1;**

$\boldsymbol{S}_e^2 = \frac{SS_e}{\nu_e} = \frac{31{,}33}{16} = \mathbf{1{,}96;}$ $\boldsymbol{\nu}_e = N – abc = abc(n - 1) = 24 – 8 = \mathbf{16}.$

5) Проверка $\mathbf{H}_0$- гипотезы

Определение расчетных значений критерия:

$\boldsymbol{F}_{a\ расч.} = \frac{S_a^2}{S_e^2} = \frac{4{,}17}{1{,}96} = \mathbf{2{,}13;}$ $\boldsymbol{F}_{b\ расч.} = \frac{S_b^2}{S_e^2} = \frac{0{,}17}{1{,}96} = \mathbf{0{,}09;}$ $\boldsymbol{F}_{c\ расч.} = \frac{S_c^2}{S_e^2} = \frac{0{,}67}{1{,}96} =$

$\mathbf{=0{,}34;}$ $\boldsymbol{F}_{ab\ расч.} = \frac{S_{ab}^2}{S_e^2} = \frac{2{,}66}{1{,}96} = \mathbf{1{,}36;}$ $\boldsymbol{F}_{ac\ расч.} = \frac{S_{ac}^2}{S_e^2} = \frac{0{,}16}{1{,}96} = \mathbf{0{,}08;}$

$\boldsymbol{F}_{bc\ расч.} = \frac{S_{bc}^2}{S_e^2} = \frac{8{,}16}{1{,}96} = \mathbf{4{,}16;}$

$\boldsymbol{F}_{abc\ расч.} = \frac{S_{abc}^2}{S_e^2} = \frac{2{,}68}{1{,}96} = \mathbf{1{,}37.}$

Построим таблицу дисперсионного анализа (табл. 34).

Таблица 34 – Результаты трехфакторного дисперсионного анализа

Источник изменчиво-сти	**Сумма квадра-тов эффектов**	ν	**Оценка диспер-сии**	F	η *100, %
Общая	$SS = 50$	$\nu = 23$	$S^2 = 2,17$		
Фактор *A*	$SS_a = 4,17$	$\nu_a = 1$	$S_a^2 = 4,17$	$F_a = 18,9$	$\eta_a = 36$
Фактор *B*	$SS_b = 0,17$	$\nu_b = 1$	$S_b^2 = 0,17$	$F_b = 13,15$	$\eta_b = 25$
Фактор *C*	$SS_c = 0,67$	$\nu_{ab} = 1$	$S_{ab}^2 = 0,67$	$F_{ab} = 0,52$	$\eta_{ab} = 1$
Факторы *AB*	$SS_{ab} = 2,66$		2,66		
Факторы *AC*	$SS_{ac} = 0,16$		0,16		
Факторы *BC*	$SS_{bc} = 8,16$		8,16		
Факторы *ABC*	$SS_{abc} = 2,68$		2,68		
	31,33				
Остаточ-ная	$SS_e = 57$	$\nu_e = 20$	$S_e^2 = 2,85$		$\eta_e = 38$

Критическое значение $F_{табл.}$ определено при $\alpha = 0,05$, $\nu_1 = 1$ и $\nu_2 = 16$; $F_{табл.} = 4,49$. **Так как** $F_{расч.} \le F_{табл.}$ при α, ν_1, ν_2, **то гипотеза H_0- принимается.** Нет необходимости проводить исследование значимости средних признака на отдельных уровнях. Таким образом, ни один из факторов: уровень образования торговых работников (*A*), поставщики товара (*B*), день недели (*C*) - не влияли на выручку исследуемых промтоварных магазинов.

Контрольные задания 4.1.

1. Оценить существенность различий в успеваемости студентов по четырем предметам и группам. Численность студентов в каждой группе составляет 25 человек.

Уровень успеваемости студентов, балл

Пред-мет	Группы							
	1	2	3	4	5	6	7	8
1	4,3	4,1	4,1	4,2	4,4	4,5	4,0	4,3
2	4,2	4,0	3,9	4,0	4,3	4,3	3,7	3,9
3	4,4	4,5	4,2	4,2	4,3	4,3	4,4	4,4
4	3,9	3,9	4,0	4,1	4,2	4,4	4,1	4,2
5	4,1	4,3	4,1	4,3	4,1	4,4	4,1	3,8
6	4,3	4,4	4,2	4,4	4,4	4,1	4,2	4,0
7	3,9	3,7	3,6	3,8	4,1	3,7	3,9	3,7
8	4,1	4,2	4,0	4,3	4,3	4,1	4,2	4,1

2.Доказывает ли опыт влияние различных доз удобрений на урожайность озимой пшеницы?

Урожайность озимой пшеницы с 1 га, ц

Дозы удобрений	Повторения							
	1	2	3	4	5	6	7	8
1	39	39	37	41	36	38	43	40
2	41	40	39	40	38	38	41	42
3	42	40	39	42	40	39	43	45
4	47	45	43	42	40	41	45	50
5	55	50	50	48	44	43	51	53
6	41	42	44	44	43	40	42	47
7	43	50	47	49	50	46	44	49
8	39	40	37	43	39	41	40	41

3. Оценить различия в среднемесячной заработной плате механизаторов различной квалификации.

Среднемесячная заработная плата механизаторов, руб.

Класс Механизаторов	Бригады									
	1	2	3	4	5	6	7	8	9	10
I	1590	1610	1800	1380	1290	1570	1640	1720	1590	1400
II	1460	1270	1360	1460	1240	1350	1230	1310	1455	1450
III	1175	1180	1210	1200	1190	1085	1195	1170	1005	1230

4. Оценить существенность влияния различных сортов и доз удобрений на урожайность риса.

Урожайность риса с 1га, ц

Сорт	Доза удобрений	Повторения							
		1	2	3	4	5	6	7	8
1	1	42	39	44	41	38	39	37	42
	2	44	47	46	45	43	42	41	44
	3	45	55	53	50	45	46	54	53
2	1	45	42	44	40	44	43	46	45
	2	49	47	49	47	45	47	48	47
	3	57	56	55	50	47	45	47	47
3	1	39	42	44	41	42	40	42	40
	2	49	51	55	53	51	54	53	49
	3	47	49	45	53	52	50	54	50
4	1	41	44	39	40	43	1	43	45
	2	48	49	46	51	52	49	46	51
	3	52	49	47	50	50	48	47	50
5	1	38	40	39	42	44	43	40	41
	2	49	52	50	52	48	49	50	54
	3	53	58	49	50	50	53	49	50
6	1	42	41	43	41	39	40	44	42
	2	49	52	53	53	50	54	53	53
	3	49	51	50	49	50	52	53	51

5. Оценить существенность различий уровня производительности при культивации в различных хозяйствах по пропашным культурам и стажу работы механизаторов.

Объем выполненных работ механизаторами за 1 час работы, эт. га

Культура	Стаж работы, лет	Хозяйства			
		1	2	3	4
Кукуруза на зерно	до 5 от 5 до 10 от 10 до 15	0,75 1,40 1,25	0,9 1,55 1,35	0,95 1,35 1,35	1,00 1,50 1,40
Кукуруза на силос	до 5 от 5 до 10 от 10 до 15	0,85 1,50 1,35	0,95 1,40 1,40	0,85 1,55 1,55	1,10 1,45 1,50
Подсолнечник	до 5 от 5 до 10 от 10 до 15	0,80 1,35 1,45	0,90 1,45 1,40	0,75 1,35 1,30	0,85 1,40 1,30

4.2. Корреляционно-регрессионный анализ и многомерные статистические методы

Социально-экономические процессы и явления зависят от большого числа параметров, их характеризующих. При решении задач определения структуры статистических взаимосвязей параметров, степени влияния параметров друг на друга, характера этих взаимосвязей возникают трудности, обусловленные неполнотой информации, сложностью получения статистических данных, многообразием моделей и методов, разработанных к настоящему времени в теории статистики. «Многомерные статистические методы среди множества возможных вероятностно-статистических моделей позволяют обоснованно выбирать ту, которая наилучшим образом соответствует исходным статистическим данным, характеризующим реальное поведение исследуемой совокупности объектов, оценить надежность и точность выводов, сделанных на основании ограниченного статистического материала.» [Дубров А. М., Мхитарян В. С., Трошин Л. И. Многомерные статистические методы: Уч. – М.: финансы и статистика, 1988. – 352 с., с.3].

Многомерный экономико-статистический анализ опирается на широкий спектр методов. В настоящем разделе представлены методы множественного корреляционно-регрессионного анализа, которые к настоящему времени получили наибольшее распространение в практической работе и чаще всего применяются при пользовании стандартными вычислительными программами по статистике. При проведении системного анализа до изучения взаимосвязей в многомерной совокупности эти методы позволяют оценить тесноту связи между системами показателей (факторов) и дать представление о стохастических связях между отдельной зависимой переменной и группой влияющих на нее факторов.

О методах многомерной классификации, такие как кластерный, дискриминантный, главных компонент, факторный и др. можно получить представление в [Л IV,10]. Методы кластерного и дискриминантного анализа предназначены для разделения рассматриваемых совокупностей объектов, субъектов или явлений на группы в определенном смысле однородные. Наличие множества исходных признаков, характеризующих процесс функционирования объектов заставляет отбирать из них наиболее существенные и изучать затем меньший набор показателей. Метод главных компонент и факторный анализ позволяют с минимальной потерей «сжать» информацию за счет некоторого преобразования исходных признаков, приводящего к уменьшению их числа. Они вскрывают объективно существующие, непосредственно не наблюдаемые закономерности при помощи полученных факторов или главных компонент, которые лаконично или более просто объясняют многомерные структуры.

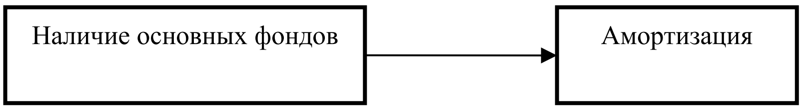
Причинная связь

В экономике большую роль играет исследование зависимостей и взаимосвязей между объективно существующими явлениями и процессами. Оно дет возможность глубже понять сложный механизм причинно-следственных отношений между явлениями. Для исследования *интенсивности, вида и формы* причинных влияний в эконометрике применяется корреляционно - регрессионный анализ. Он находит широкое применение при прогнозировании, при решении задач народнохозяйственного и внутрихозяйственного планирования, при выявлении факторов, воздействуя на которые можно вмешиваться в экономический процесс с целью получения нужных результатов

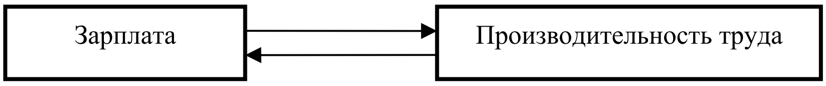
Под причинной связью понимают такое соединение явлений и процессов реальной действительности, когда изменение одного из них - это следствие изменения другого.

Часто одно и тоже явление может выступать как результат одной или нескольких причин и в то же время само служит причиной наступления других явлений и процессов. На рисунках 49 - 51 изображен ряд примеров, иллюстрирующих основные типы причинно-следственных связей между явлениями.

а) $X \rightarrow Y$, X – причина, Y – следствие,



б) $Y \leftrightarrow X$, причинные связи между двумя явлениями, между которыми существует взаимодействие,



в) $\begin{array}{ll} X & \rightarrow Y_1 \\ \downarrow & \\ Y_2 & \end{array}$, явление X влечет за собою несколько других явлений.

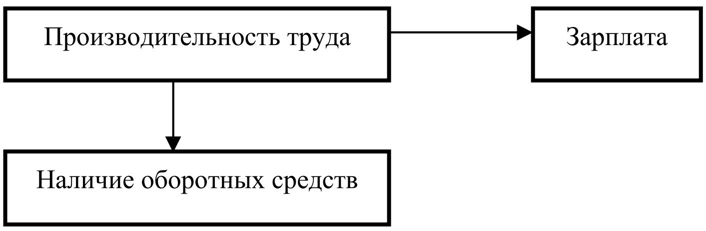


Рис. 49 – Основные типы причинно-следственных связей между явлениями

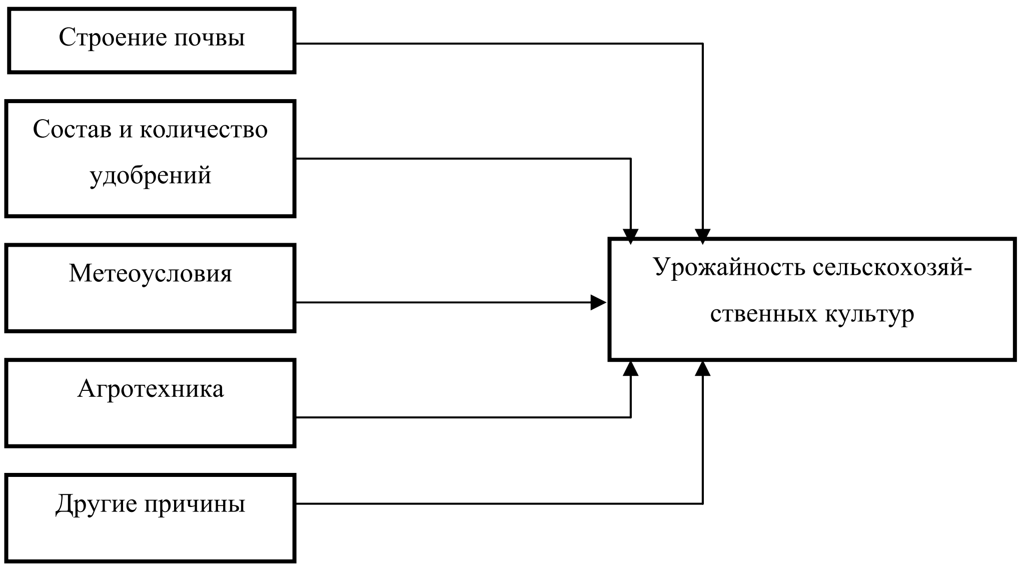
Признание факта множественности причин и множественности следствий нашло отражение при исследовании закономерностей окружающего мира методами многомерного статистического анализа.

Приведенные основные схемы причинно – следственных отношений лежат в основе различных видов корреляций и регрессий.

г) $X_1 \rightarrow$

$X_2 \rightarrow\ Y,$ несколько явлений *X* являются причинами *Y,*

$X_3 \rightarrow$



д) Причинно-следственные комплексы Вариант №1.

$X_3 \rightarrow X_2 \rightarrow\ X_1 \rightarrow Y$,

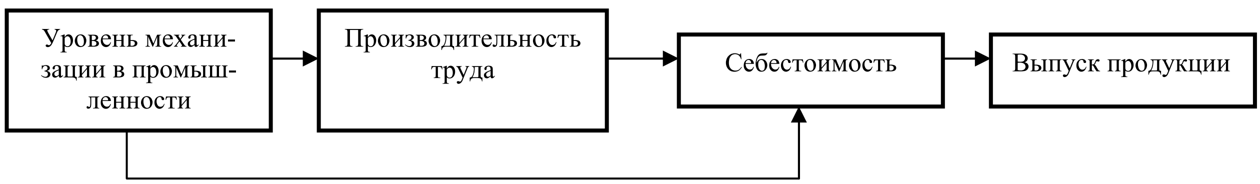


Рис. 50 – Основные типы причинно-следственных связей между явлениями

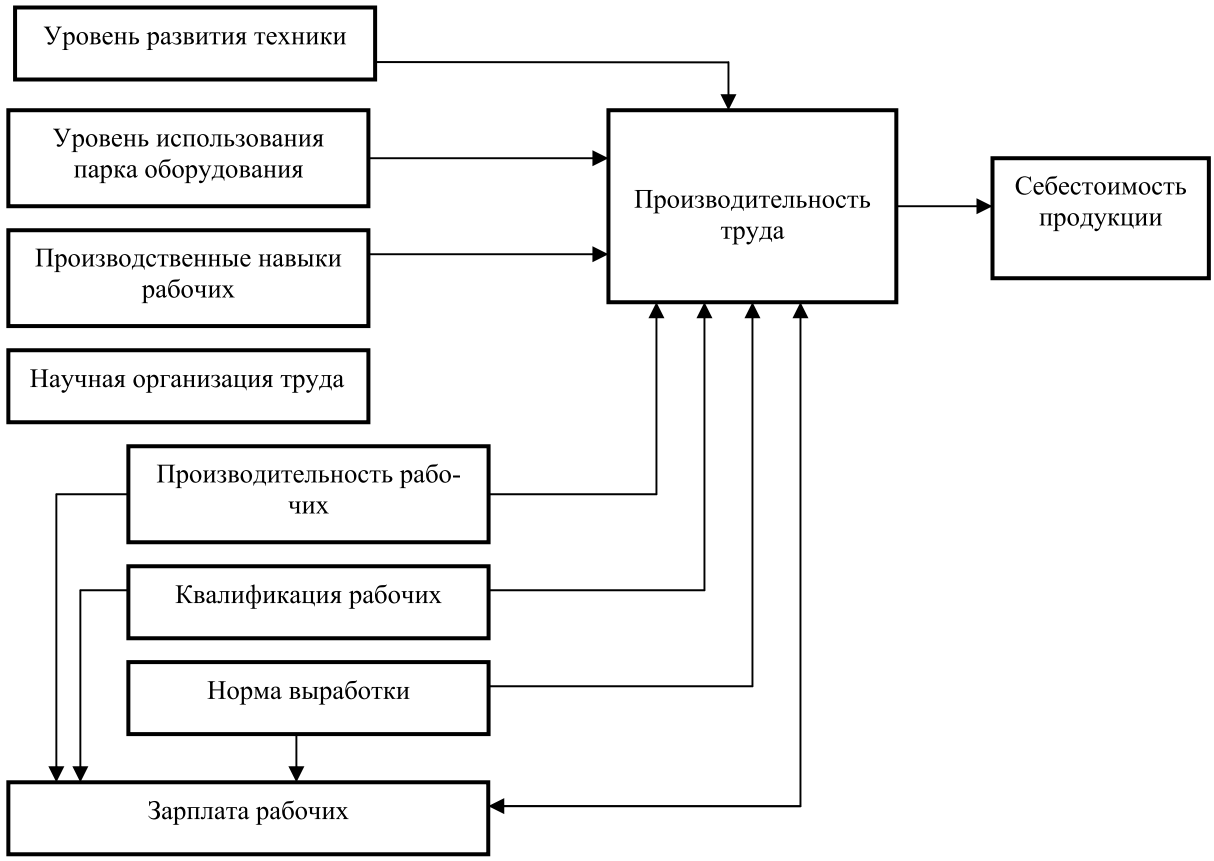


Рис. 51 – Основные типы причинно-следственных связей между явлениями, причинно-следственные комплексы

Понятие о корреляционной и регрессионной связи.

Различают два типа зависимостей между экономическими явлениями: функциональную и стохастическую.

При изучении случайных величин в общем случае необходимо рассматривать *стохастическую зависимость*, когда каждому значению СВ X может соответствовать одно и более значение СВ Y, причем до опыта нельзя предсказать возможное соответствие (рис.52).

Рис. 52 – Виды зависимостей

В случае стохастической связи изменение СВ Y, вследствие изменения СВ X, можно разбить на две компоненты: а) функциональную, связанную с зависимостью Y от X, б) случайную, связанную со случайным характером самих СВ X и Y. Соотношение между функциональной и случайной компонентой определяет силу связи. Отсутствие первой компоненты указывает на независимость СВ *X* и *Y*, отсутствие второй компоненты показывает, что между СВ *X* и *Y* существует функциональная связь (таким образом, функциональную связь можно рассматривать как частный случай стохастической).

Функциональная связь *y = f(x): для каждой независимой переменной X существует вполне определенное значение зависимой переменной Y.*

Статистическая или стохастическая (вероятностная) связь отражает закономерности только в массовых явлениях.

Статистической зависимостью *называют зависимость, при которой изменение одной из величин влечет изменение распределения другой.*

Статистическую зависимость называют ***корреляционной****, если при изменении значений одной величины меняется среднее значение другой.*

Корреляция означает соответствие, соотношение.

При изучении конкретных зависимостей вводят понятия:

- *факторные признаки (факторы)* - объясняющие, независимые переменные, причины *X*. Могут быть случайными и неслучайными;

- ***результативные признаки (показатели)*** – объясняемые зависимые переменные *Y*, случайные.

Иногда *X* и *Y* можно менять местами.

При сравнении функциональных и корреляционных зависимостей следует иметь в виду, что при функциональной зависимости, зная X, можно вычислить величину Y. При корреляционной зависимости устанавливается лишь тенденция изменения Y при изменении X.

Содержание теории корреляции составляет изучение зависимости вариации признака от окружающих условий.

С корреляционным анализом тесно связан регрессионный. Их объединяют методы обработки данных, отличают цели и формы установления связи.

В корреляционном анализе оценивается сила стохастической связи, в регрессионном – форма.

Кроме того, если не известно, какой их признаков - зависимый, какой - независимый, или же это безразлично, то *X* и *Y* равноправны в этом смысле. В такой ситуации говорят о взаимосвязи корреляционного типа в широком смысле. Если переменные не равноправны, т.е. четко ясно, какая из них причина, какая – следствие, то говорят о регрессионной зависимости.

Регрессия – это односторонняя стохастическая зависимость, когда одна из переменных служит причиной для изменения другой.

Например:

− при изучении потребления электроэнергии (*Y*) в зависимости от объема производства (*X*) речь идет об односторонней связи, следовательно, о регрессии;

− рост доходов ведет к увеличению потребления;

− снижение процентной ставки увеличивает инвестиции;

− увеличение валютного курса сокращает чистый экспорт.

Известно, что при функциональной зависимости можно построить обратную функцию. Если регрессионная зависимость отражает реальную функциональную, то она может быть обратимой тоже. Если же связь только стохастическая, то регрессия не обратима.

Необратимость может быть обусловлена:

− структурой явления, определяющей направление связи;

− постановкой цели исследования (когда решается задача, как по значениям аргумента предсказать значение функции);

− способом измерения эмпирических точек.

Рассмотрим особенности, связанные с постановкой задачи. Если изучают стохастическую зависимость Y от X, то устанавливают регрессию Y на X, например, исследование потребления энергии Y от объема производства X: $Y = f(X)$.

Если X от Y - то регрессия X на Y:

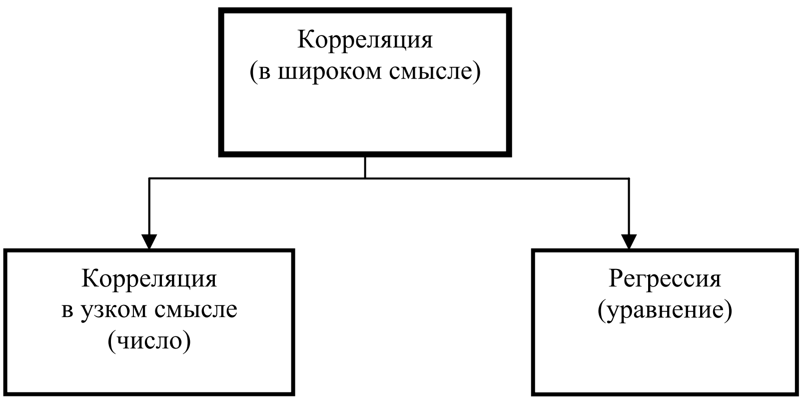
− изучается механизм влияния объема производства X на величину потребления энергии Y: $X = g(Y)$ (это представляет интерес для планирования народного хозяйства);

− стоимость товара влияет на спрос и спрос влияет на стоимость.

Но могут быть ситуации, когда обратная регрессия не имеет физического смысла, например, урожайность зависит от количества осадков, обратная зависимость бессмысленна.

Функция регрессии может формально устанавливать зависимость, хотя переменные и не состоят в причинно- следственной связи.

Для понимания соотношения между корреляционной и регрессионной связью вводят понятие корреляции «в широком смысле» (как связь, соотношение между явлениями), корреляции «в узком смысле» (степень тесноты, интенсивности, прямолинейности, четкости, строгости – измеряемых числом). Тогда соотношение между корреляцией и регрессией условно можно представить в виде схемы:



Ход рассуждений, постановка задач, результаты различны в корреляционном и регрессионном анализе. Но очень часто эти два вида анализа проводят параллельно на одном и том же массиве исходных данных.

Виды регрессий и корреляций, задачи корреляционного и регрессионного анализа

При изучении взаимосвязи факторных и результативных признаков могут быть ситуации: а) X и Y – случайные величины; б) X - не случайные, Y – случайная.

Виды корреляций.

а) Относительно характера корреляции:

− положительная (равнонаправленная, прямая);

− отрицательная (обратная).

б) Относительно числа переменных:

− простая или парная;

− множественная, с ее помощью можно охватить весь причинно-следственный комплекс;

− частная корреляция, корреляция между двумя переменными при фиксированном влиянии других (вскрывается внутренняя структура соотношений, т.е. элиминируется влияние других факторов).

в) Относительно формы связи:

− линейная;

− нелинейная.

г) Относительно типа связи явлений:

− непосредственная корреляция;

− косвенная корреляция;

− ложная корреляция.

Виды регрессии.

а) Относительно числа явлений (переменных), учитываемых в регрессии:

− простая регрессия (парная) регрессия;

− множественная или частная регрессия.

б) Относительно формы зависимости:

− линейная регрессия;

− нелинейная регрессия.

в) Относительно характера регрессии (имеет смысл только для простой линейной регрессии):

− положительная регрессия;

− отрицательная регрессия.

г) Относительно типа связи явлений:

− непосредственная регрессия – причина прямо воздействует на следствие;

− косвенная регрессия, Y и X не состоят в прямой зависимости, а детерминируются общей для них причиной, через третью переменную;

− нонсенс- регрессия (абсурдную).

Задачи корреляционного и регрессионного анализа

Корреляционный анализ

а) Измерение степени связности (тесноты, силы, строгости, интенсивности) двух и более явлений:

− уточняет (верифицирует) известные связи;

− обнаруживает неизвестные.

б) Отбор факторов, оказывающих наибольшее влияние на Y, на основании степени связности – эти факторы потом используют в регрессионном анализе.

в) Обнаружение неизвестных причинных связей – непосредственно не определяет их, но устанавливает степень необходимости этих связей и достоверность суждения о них, чтобы не было «ложной» корреляции.

Регрессионный анализ

а) Установление формы зависимости (для случая парной регрессии):

− убывающая;

− возрастающая.

б) Определение функции регрессии.

в) Оценка неизвестных значений зависимой переменной – можно воспроизвести значение Y при заданных значениях X внутри интервала (интерполяция) и вне интервала (экстраполяция заданных изменений X).

Имеется различие при исследовании корреляционных связей между экономическими явлениями и между явлениями в естественных науках и технике. В последних можно планировать эксперимент, когда добиваются элиминирования побочных факторов и поддержания условий эксперимента на неизменном уровне. В экономическом эксперименте эти действия практически невозможны – одно и то же следствие может быть порождено слишком многими причинами.

В экономике массовое исследование носит апостериорный характер, в естественных – априорный.

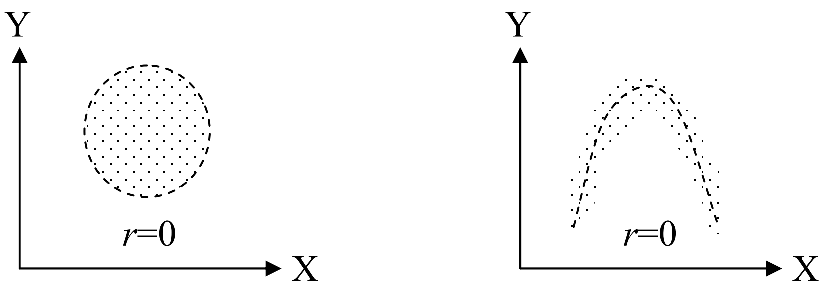
4.3. Корреляция

Изучение реальных процессов обычно предполагает наблюдение над целым рядом случайных величин (например, количество внесённых удобрений и урожайность, количество решённых задач и оценка на контрольной и т.д.). Возникает задача изучения взаимосвязи между случайными величинами. На первом уровне эта задача (оценка существенности влияния одного фактора на другой) может быть решена средствами дисперсионного анализа. Для численной оценки связи используется корреляционный анализ.

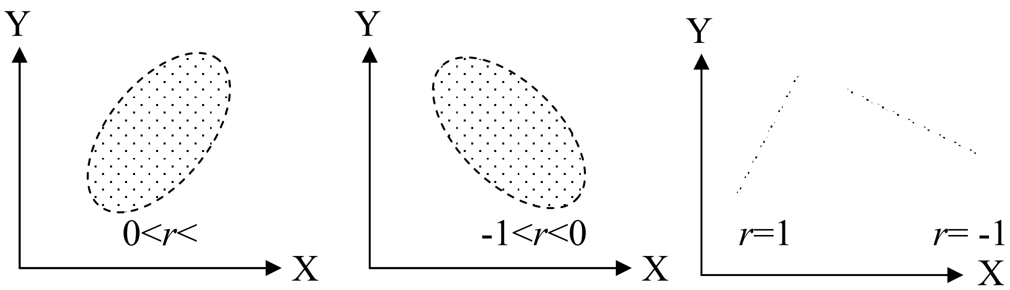
В математике зависимость между двумя величинами X и Y выражается с помощью функции $y = f(x)$, где каждому возможному значению X ставится в соответствие не более одного значения Y (подобная зависимость называется функциональной).

В общем виде задача выявления и оценки силы стохастической связи не решена до сих пор. Важным частным случаем стохастической зависимости является корреляционная. *Корреляционная зависимость* между переменными величинами – это та функциональная зависимость, которая существует между значениями одной из них и групповыми средними другой. (Корреляционные зависимости Y на X и X на Y обычно не совпадают). Корреляционная связь чаще всего характеризуется выборочным коэффициентом корреляции r, который характеризует степень линейной функциональной зависимости между СВ X и Y. Для двух СВ X и Y коэффициент корреляции имеет следующие свойства (рис.53):

1. $-1 \le r \le 1$;
2. Если $r = \pm 1$, то между СВ X и Y существует функциональная линейная зависимость;
3. Если $r=0$, то СВ X и Y некоррелированны, что не означает независимости вообще (на втором рисунке сверху рис.39 СВ X и Y некоррелированны, однако, зависимость между ними можно описать параболой);
4. Если X и Y образуют систему нормально зависимых СВ, то из их некоррелированности следует их независимость;
5. Коэффициенты корреляции Y на X и X на Y совпадают.



Отсутствие корреляции



Положительная корреляция | Отрицательная корреляция | Линейная зависимость

Рис. 53 – Оценка взаимозависимости между выборочными значениями СВ X и Y с помощью выборочного коэффициента корреляции r

Обычно исследователь располагает выборочными данными, по которым требуется определить (с помощью выборочного коэффициента корреляции r) меру зависимости случайных величин. Раздел математической статистики, изучающий методы оценки выборочного коэффициента корреляции r, проверку значимости коэффициента корреляции (насколько существенной является гипотеза, что $r=0$ - СВ X и Y некоррелированны), построение для него (в случае значимости) доверительных интервалов, называется *корреляционным анализом*.

Выборочные данные обычно можно представить в виде корреляционного поля и (или) корреляционной таблицы.

Существует достаточно много методов оценки r, основные из них:

а) в случае парной зависимости - коэффициент корреляции Пирсона:

$$r_{xy} = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\sum (x-\bar{x})^2 \sum (y-\bar{y})^2}} = \frac{\overline{xy} - \bar{x}\cdot\bar{y}}{\sqrt{\overline{x^2}-(\bar{x})^2}\cdot\sqrt{\overline{y^2}-(\bar{y})^2}}, \qquad (4.3.1)$$

где $\overline{xy} = \frac{\sum xy}{n}, \quad \bar{x} = \frac{\sum x}{n}, \quad \bar{y} = \frac{\sum y}{n}, \quad \overline{x^2} = \frac{\sum x^2}{n}, \quad \overline{y^2} = \frac{\sum y^2}{n}.$

Рассмотрим, в силу большой применимости, идеологию построения коэффициента Пирсона r_{xy}. Положение объекта относительно других в выборке (для парной зависимости) зависит от $\bar{x}$, $\bar{y}$ и проявляется в величине и знаках $(x_i - \bar{x})$ и $(y_i - \bar{y})$. Если объект имеет высокий уровень по обеим переменным, то произведение $(x_i - \bar{x})\cdot(y_i - \bar{y})$ будет большим и положительным. Аналогично, если он относительно низок как по х, так и по у, то $(x_i - \bar{x})\cdot(y_i - \bar{y})$ для него также будет большим и положительным (поскольку произведение двух отрицательных чисел положительно). Если x и y в основном связаны прямо (большие значения с большими, а малые – с малыми), то большинство произведений $(x_i - \bar{x})\cdot(y_i - \bar{y})$будет положительно, следовательно,

$$\sum (x_i - \bar{x})\cdot(y_i - \bar{y})$$

будет большой и положительной.

Если x и y имеют обратную связь – большое x встречается с малым y и наоборот, то большинство произведений $(x_i - \bar{x})\cdot(y_i - \bar{y})$ будет отрицательно, следовательно, $\sum (x_i - \bar{x})\cdot(y_i - \bar{y})$будет отрицательной (когда x и y связаны обратной зависимостью).

Если между переменными нет систематической связи, (большие х сочетаются с малыми у так же часто, как и с большими и тоже самое справедливо для малых х), то $\sum (x_i - \bar{x})\cdot(y_i - \bar{y})$ будет близка к нулю.

Обычно $\sum (x_i - \bar{x})\cdot(y_i - \bar{y})$ усредняют (для устранения зависимости от числа пар наблюдений) и получают ковариацию

$$\operatorname{cov}(x, y) = \frac{\sum (x_i - \bar{x})\cdot(y_i - \bar{y})}{n} -$$

величину, которая велика и положительна, когда x и y сильно связаны прямой связью; близка к нулю, в случае отсутствия связи; велика и отрицательна когда переменные сильно связаны обратной связью.

Для устранения влияния на ковариацию разброса случайных величин ее нормируют − делят на σ_X и σ_Y. В результате получается коэффициент корреляции Пирсона (4.3.1).

б) в случае изучения зависимости СВ X_1 от СВ X_2 и СВ X_3 − *множественный коэффициент корреляции*:

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2r_{12}r_{13}r_{23}}{1 - r_{23}^2}}, \qquad (4.3.2)$$

где r_{ij} - парные коэффициенты корреляции (например, r_{12} − коэффициент корреляции между X_1 и X_2).

$$0 \le R_{1.23} \le 1.$$

Можно рассматривать так же $R_{2.13}$ и $R_{3.12}$.

$D=R^2$ - *множественный коэффициент детерминации*, характеризующий долю зависимости независимой переменной от зависимых.

Если $R_{1.23} = 1$, то X_1 однозначно определяется функциональной линейной зависимостью от X_2 и X_3: $X_1=a+bX_2+cX_3$.

Замечание. 1. Использование коэффициентов парной и множественной корреляции (r и R) неявно предполагает нормальное распределение генеральных совокупностей, из которых производится выборка.

2. Как отмечено выше, коэффициент корреляции характеризует степень линейной зависимости между переменными. Нелинейная связь не обнаруживается. В этом случае для оценки используется корреляционное отношение. *Корреляционным отношением* Y и X называется отношение межгруппового среднего квадратического отклонения δ_y переменной Y к её общему среднему квадратическому отклонению σ_y

$$\eta_{yx} = \frac{\delta_y}{\sigma_y},$$

где межгрупповая дисперсия определяется по формуле

$$\delta_y^2 = \frac{\sum_i (\overline{y_i} - \overline{y})^2 \cdot n_i}{n}.$$

Аналогично определяется корреляционное отношение X и Y η_{xy}.Основные свойства корреляционных отношений:

1) $0 \le \eta_{YX} \le 1, \quad 0 \le \eta_{XY} \le 1$;
2) если $\eta=0$,то корреляционная связь отсутствует;
3) если $\eta=1$,то переменные связаны функционально;
4) для линейной зависимости между переменными X иY необходимо и достаточно, чтобы выполнялось равенство $|r| = \eta_{YX}$;
5) $\eta_{XY} \ne \eta_{YX}$;
6) $0 \le |r| \le \eta \le 1$.

В других случаях (когда вид распределения неизвестен) используют меры связи, не регламентирующие нормальность выборок (методы *непараметрической статистики*), например, *коэффициент ранговой корреляции* Спирмена - r_s:

$$r_s = 1 - \frac{6\sum d_i^2}{n(n^2-1)}, \qquad (4.3.3)$$

где d_i^2 - квадраты разности рангов,

n - число наблюдений (число пар рангов).

Ранг - это порядковый номер значений признака, расположенных в порядке возрастания или убывания их величин.

Пример 4.3.1. Преподавателю и студенту было предложено расположить 10 профессий в порядке их общественной значимости. Ответы перечислены ниже.

Оценка преподавателя, x_i	Профессии	Оценка студента, y_i
3	профессор	2
1	врач	1
4	учитель школы	7
2	директор магазина	4
8	бухгалтер	5
6	банкир	3
9	водитель	9
5	журналист	8
10	ди-джей	10
7	программист	6

Какова корреляция рангов между двумя рядами оценок? Одинаково ли мнение преподавателя и студента по этому вопросу?

Решение. Определим разности рангов, их квадраты и суммы:

d_i=x_i-y_i	1	0	-3	-2	3	3	0	-3	0	1	$\sum$=0
d_i^2	1	0	9	4	9	9	0	9	0	1	$\sum$=42

Имеем: $r_s = 1 - \dfrac{6\sum d_i^2}{n(n^2-1)} = 1 - \dfrac{6\cdot 42}{10(100-1)} = 0{,}7456.$

Проверим, существует ли положительная корреляционная связь между мнениями преподавателей и студентов, для этого (при $n \geq 10$, в других случаях см.[6]) используем t - статистику Стьюдента с ν=(n-2) степенями свободы:

$$t = |r_s|\sqrt{\frac{n-2}{1-r_s^2}}. \qquad (4.3.4)$$

Нулевая гипотеза - коэффициент корреляции не является статистически значимым (H_0: r_s=0). Альтернативная гипотеза - существует положительная корреляционная зависимость (H_1: r_s>0).

При уровне значимости α=0,05 для односторонней (правосторонней) критической области (прил. 3):

$$t_{\text{кр.}} = t_{0{,}05{,}8} = 1{,}86,$$

$$t_{\text{расч.}} = 0{,}7456\sqrt{\frac{10-2}{1-0{,}7456^2}} = 2{,}8284.$$

2,8284>1,86 ($t_{\text{расч.}} > t_{\text{кр.}}$). Следовательно, связь между мнениями преподавателя и студента является статистически значимой, при 5% уровне значимости.

Непосредственное вычисление и оценка коэффициента корреляции еще не говорит о наличии причинной зависимости между случайными величинами. Изучение корреляционной связи должно сопровождаться пониманием внутренних особенностей изучаемых процессов и возможных причинно-следственных связей.

Так Езикил М. и Фокс К. пишут по этому поводу «Исследователи думают, что они правильно поступают когда изучают зависимость данной переменной от нескольких факторов, пренебрегая теми из них, корреляция с которыми не проявляется, и отбирая для анализа множественной корреляции лишь те факторы простая корреляция с которыми высока… Эти действия могут привести к пренебрежению такими факторами…, действие которых может обнаружиться лишь после устранения влияния других, находящихся во взаимодействии факторов» (Езикил М., Фокс К. Методы анализа корреляций и регрессий: линейных и криволинейных. М.: 1966.-557с.(с211).

Теснота связи результативного признака с изучаемым факторным окажется равной парному коэффициенту корреляции только в случае, если одновременно с изменением факторного соответственно изменяются и все другие признаки имеющие связи между собой, что вряд ли возможно. В этом случае целесообразно использовать частные коэффициенты корреляции, имеющие для случая двух факторных признаков следующий вид:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_2}) \cdot (1 - r^2_{x_1x_2})}}; \quad r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{yx_1}) \cdot (1 - r^2_{x_1x_2})}};$$

$$r_{x_1x_2 \cdot y} = \frac{r_{x_1x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{(1 - r^2_{yx_1}) \cdot (1 - r^2_{yx_2})}},$$

которые позволяют измерить связь между факторным и результативным признаком (x_j и y) при исключении влияния всех остальных переменных, что тоже не совсем соответствует истине, ибо и в этом случае на результативный признак отражается влияние других факторов кроме изучаемого.

Рис. 54 – Схема определения вида корреляции

Различают корреляцию обусловленную [6]:

- *причинной зависимостью* *X* от *Y*;
- *зависимостью* *X* и *Y* *от третьей величины*;
- *неоднородностью выборки* (например, при изучении зависимости размеров кровяных шариков от содержания гемоглобина в крови у ново-

рождённых, женщин и мужчин коэффициент корреляции порядка нескольких сотых. Однако, при объединении статистического материала коэффициент корреляции равен 0,75.);

- *формально* (числовыми данными).

Для определения вида корреляции можно использовать схему, приведённую на рисунке 54. Обычно теснота связи между результативным и факторным признаком определяется по величине коэффициента корреляции (шкала Шеддока):

- ✓ $r < 0{,}2$ - связи нет;
- ✓ $0{,}2 \le r < 0{,}5$ - связь слабая;
- ✓ $0{,}5 \le r < 0{,}75$ - связь средняя;
- ✓ $0{,}75 \le r < 0{,}95$ - связь тесная;
- ✓ $0{,}95 \le r < 1$ - связь очень тесная.

4.4. Регрессионный анализ

4.4.1.Основные понятия и определения, модель регрессии

Регрессионный анализ - один из основных методов современной математической статистики, позволяющий аналитически представить связь между переменными объекта. Если корреляционный анализ позволяет установить существует или не существует факт зависимости между парами наблюдений, то регрессионный анализ дает целый арсенал методов построения соответствующих зависимостей.

В предисловии к русскому изданию книги [5, с.5] ее научные редакторы Ю.П.Адлер и В.Г.Горский написали: "Среди известных методов математической статистики регрессионный анализ занимает исключительное положение. Это связано не столько с тем, что регрессионный анализ является одним из самых распространеннейших методов обработки результатов наблюдений, но и с тем, что он, по существу, служит основой целого ряда других методов математической статистики и прежде всего – планирования экспериментов, дисперсионного анализа, многомерного статистического анализа… В научной литературе этому методу посвящены тысячи журнальных статей и десятки монографий…"

В настоящее время серьезные исследовательские работы с применением методов регрессионного анализа выполняются исключительно с помощью ЭВМ. В связи с этим в регрессионном анализе возникли и возникают алгоритмы и их модификации в необозримом количестве. Разобраться в их достоинствах и недостатках очень не просто, особенно, если известно только назначение регрессионного анализа, но не его суть, не его система. Изучение регрессионного анализа можно провести на первых стадиях и без применения ЭВМ, главное – отработать навыки "правильного" принятия решений после каждого шага вычислений и умение делать "правильные" выводы.

Объект исследования в регрессионном анализе – экономические, социальные, политические, экологические, технические и другие системы и процессы.

Предмет исследования – математические модели регрессионного анализа.

Цель исследования – установление по результатам статистических наблюдений (пассивных или активных) адекватной аналитической зависимости (уравнения регрессии) между показателями и факторами, которые характеризуют изучаемые системы. Это соответствует одной из наиболее общих задач статистики - оценивания степени и формы связи между величинами.

Регрессия – "возвращение", "движение назад", "перестановка слов в обратном порядке"; *регрессионное доказательство* – ход рассуждения идет от следствий к основанию или - доказательство восходит от фактов, как следствий, к доказываемому положению, как к основанию.

Термин «регрессия» появился впервые более 100 лет назад в работе английского физиолога Ф. Гальтона, исследовавшего 928 пар родителей и их детей и пришедшего к выводу о том, что имеет место «регресс» − чем выше родители, тем ниже дети, поэтому проведенный анализ назвал регрессионным. Хотя анализ который он проводил, скорее, можно назвать корреляционным, но термин исторически прижился. Кстати термин «корреляция» так же придумал он (обозначение r происходит от слова регрессия).

Рассмотрим первоначально более простую задачу. Пусть изучается связь между двумя величинами: X и Y; в результате эксперимента определены их совместные значения (табл.35).

Таблица 35 – Результаты статистических наблюдений

X	x_1	x_2	…	x_i	…	x_n
Y	y_1	y_2	…	y_i	…	y_n

Данные табл.35 можно изобразить на графике – диаграмме рассеяния. На рис.53 представлены точки эксперимента, по которым можно предположить наличие или отсутствие стохастической зависимости между переменными X и Y. Задача исследователя – найти аналитическую функцию, которая наилучшим образом описывает экспериментальные данные.

Переменные X и Y могут быть:

- обе случайными;
- одна случайная, другая – нет; обычно X – не случайна, фиксирована и управляема, Y – случайная.

Если обе X и Y – случайны, то имеет место некоторое совместное распределение $f(x, y)$, причем степень зависимости между X и Y, как известно, может быть охарактеризована коэффициентом корреляции r_{xy} (в случае линейной зависимости), корреляционным отношением (в случае нелинейной зависимости) или двумя функциями регрессии X на Y и Y на X, т.е. зависимостями:

$$x = M\,[Y/X], \qquad y = M[X/Y],$$

где M – соответствующие математические ожидания. Все это - задачи *корреляционного анализа*.

Если переменная X – не случайна, то каждому ее значению x соответствует некоторое распределение $f(y)$ случайной величины.

Может быть не одна переменная X, а система k величин -
$\boldsymbol{X}=\{X_1,X_2,\ldots,X_j,\ldots,X_k\}$.

Мы в основном будем знакомиться с методом регрессионного анализа для изучения объектов, которые можно представить моделью типа «черный ящик» (рис.41.).

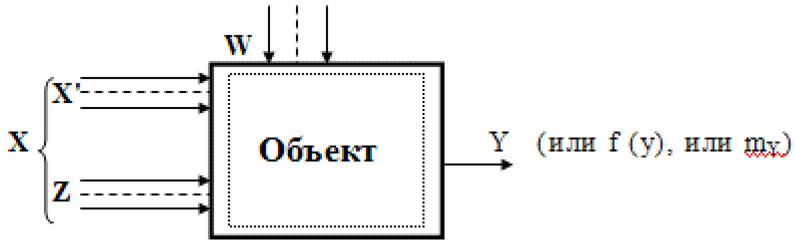


Рис. 55 – Модель объекта

На рис.55:

$\boldsymbol{X}$ = (X', Z) – факторы (вектор входных переменных);

$\boldsymbol{X'}$ – управляемые, независимые переменные;

$\boldsymbol{Z}$ – контролируемые, но неуправляемые факторы;

$\boldsymbol{Y}$ – "отклик" ("показатель качества управления", "выход"), $f(y)$ – закон распределения , m_Y - математическое ожидание случайной величины Y;

$\boldsymbol{W}$ – помехи. Заметим, что вектор мы обозначаем «жирно».

Исследование объекта может быть пассивным, когда фиксируются «естественные» значения $\boldsymbol{X}$ и соответствующие им значения $\boldsymbol{Y}$ (пассивный эксперимент), и активным, когда осуществляются целенаправленные изменения $\boldsymbol{X}$ (активный эксперимент). В первом случае вопросы организации сбора данных не являются первостепенными и чем их больше, тем лучше. Обработка результатов пассивного эксперимента ведется методами «классического» регрессионного анализа. Во втором случае имеем дело с планированием эксперимента и соответствующими специальными формами регрессионного анализа. Планирование эксперимента, если то позволяет объект исследования, существенно эффективнее пассивного эксперимента в смысле минимизации числа опытов и точности получаемых выводов. Но для таких объектов, как, например, социально-экономические системы, активный эксперимент применим лишь в редких случаях. В основном он используется в точных науках, при изучении технологических объектов, в сельскохозяйственном эксперименте и др., т.е. только там, где изменения параметров $\boldsymbol{X}$ на границах допуска не приводят к фатальным последствиям и где физически возможно и экономически не дорого «управлять» факторами $\boldsymbol{X}$.

Регрессионная модель представляет собою математическое выражение, связывающее входные переменные $\boldsymbol{X}$ *с одним «выходом»* Y. Поэтому в реальных условиях, когда результат функционирования объекта должен характеризоваться несколькими показателями Y, возникает отдельная проблема выбора единственного показателя, наиболее полно характеризующего особенности изучаемого объекта. Хотя эта задача практически очень важна, она выходит за рамки математической теории регрессионного анализа и в данном случае мы не будем ее рассматривать.

По результатам наблюдений определяется уравнение регрессии. Итак:

Уравнение регрессии - это зависимость случайной величины Y *от неслучайных факторов* $\boldsymbol{X}$, *т.е. зависимость «следствия»* Y *от «причин»* $\boldsymbol{X}$:

$$Y=\eta(\boldsymbol{X},\boldsymbol{\beta})+\boldsymbol{\varepsilon}, \qquad (4.4.1)$$

где $\boldsymbol{X}=\{X_1, X_2, \ldots, X_j, \ldots, X_k\}$ – вектор факторов, $j=1, 2, \ldots, k$;

$\boldsymbol{\beta}=\{\beta_1, \beta_2, \ldots, \beta_j, \ldots, \beta_d\}$ - вектор параметров модели; $\eta(\boldsymbol{X},\boldsymbol{\beta})$- функция регрессии (или функция отклика) случайной величины $\boldsymbol{Y}$ на неслучайные $\boldsymbol{X}$; $\boldsymbol{\varepsilon}=\{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_i, \ldots, \varepsilon_n\}$- вектор ошибок наблюдений. При многократном однотипном воздействии $\boldsymbol{X}$ на входе получаем на выходе объекта различные значения Y. Т.е. определенному X_i $= (x_{i1}, x_{i2}, \ldots, x_{ig}, \ldots, x_{im})$, где $i=1, 2, \ldots, n$, n – число всех опытов (объем выборки), $l= 1, 2, \ldots, m_i$, m_i - число повторных (или «параллельных») опытов, встречающихся в выборке, соответствует некоторое распределение $f(y_i)$, $i=1, 2, \ldots, n$ случайной величины Y (рис.42),

Уравнение регрессии вида (4.4.1) описывает только статику объекта, т.е. предполагается, что взаимосвязь показателя $\boldsymbol{Y}$ и факторов $\boldsymbol{X}$, установленная в определенный момент (интервал) времени от времени не зависит (т.е. параметры модели (4.4.1) не зависят от времени).

Регрессионные модели в зависимости от рассматриваемых факторов могут быть использованы в целях: объяснения сути явления (предсказательная модель), прогнозирования (прогнозная модель), управления. Если установлена зависимость Y только от управляемых факторов $\boldsymbol{X'}$, то это уравнение теоретически может быть использовано в целях управления объектом (заметим, что при построении уравнения по результатам пассивного эксперимента ошибка в управлении может быть неприемлемой). Функциональная модель и модель для прогнозирования содержит все группы факторов $\boldsymbol{X'}$, $\boldsymbol{Z}$ (рис.41):

$$Y=\eta(\boldsymbol{X'}, \boldsymbol{Z}, \boldsymbol{\beta})+\boldsymbol{\varepsilon} .$$

Обычно, функциональная модель более сложная, чем предсказательная.

Чаще всего исследователь не знает вида функции отклика η и его цель – найти ее, в простейшем случае – найти параметры $\boldsymbol{\beta}$.

В целях построения $\eta(\boldsymbol{X},\boldsymbol{\beta})$ обычно предполагают, что это – гладкая функция в области допустимых значений: $\boldsymbol{X} \in \boldsymbol{X}_{доп}$. В этом случае возможно ее разложение в ряд Тейлора в окрестности некоторой точки, например, точки, соответствующие «центру» эксперимента - среднему значению $\overline{X}$. В результате получаем полином степени р вида:

$$Y= \beta_0+ \Sigma_j\ \beta_j\, x_j + \Sigma_{uj}\ \beta_{uj}\, x_u\, x_j+ \Sigma_j\ \beta_{jj}\, x_j^{\,2} +\ldots +\varepsilon, \qquad (4.4.2)$$

где Σ_j – сумма по j = 1, 2,…, k, Σ_{uj} – сумма парных взаимодействий $x_u\, x_j$, u,j =1, 2,…, k, $u \neq j$, k – число факторов; β_{uj} – коэффициент парного взаимодействия, β_{jj} – коэффициент при квадрате переменной и т. д.; в формуле (2) степень полинома p = 2.

Обычно разложение ограничивают конечным числом членов ряда. Например:

$$Y=\beta_0+\Sigma_j\beta_j x_j \qquad (4.4.3)$$

Выражение (4.4.3) – это линейная регрессия (p = 1).

Наибольшее возможное число d коэффициентов регрессии для заданного числа факторов k и степени полинома р определяется формулой: $d\ = k+p$.

По результатам эксперимента могут быть определены не «истинные» коэффициенты регрессии $\boldsymbol{\beta}$, соответствующие генеральной совокупности, а лишь их оценки $\boldsymbol{B}$ = (b_0,b_1, … ,b_j,…, b_d), вычисленные по выборке объемом $\boldsymbol{n}$. В этом случае уравнение регрессии в векторной форме имеет вид:

$$\hat{Y}=\eta(\boldsymbol{X},\boldsymbol{B}), \qquad (4.4.4)$$

где $\hat{Y}$ – предсказанные (прогнозируемые) значения выходной величины.

При выводе и использовании формул регрессионного анализа удобнее пользоваться векторной формой представления уравнений регрессии:

$$\boldsymbol{Y}=\boldsymbol{X\beta}+\boldsymbol{\varepsilon}\ ; \qquad \hat{\boldsymbol{Y}}=\boldsymbol{XB}\ , \qquad (4.4.5)$$

где $\boldsymbol{Y}$ - вектор наблюдений; $\boldsymbol{X}$ – матрица значений независимых переменных; $\boldsymbol{\beta}$, $\boldsymbol{B}$ – векторы коэффициентов и их оценок соответственно; $\boldsymbol{\varepsilon}$ - вектор ошибок:

$$Y=\begin{bmatrix} y_1 \\ y_2 \\ . \\ y_i \\ . \\ y_n \end{bmatrix}, X=\begin{bmatrix} 1x_{11}\ x_{12}\ldots x_{1j}..x_{1k} \\ 1x_{21}\ x_{22}\ldots x_{2j}..x_{2k} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ 1x_{i1}\ x_{i2}\ldots x_{ij}..x_{ik} \\ \ldots\ldots\ldots\ldots\ldots\ldots \\ 1x_{n1}\ x_{n2}\ldots x_{nj}..x_{nk} \end{bmatrix}, B=\begin{bmatrix} \beta_0 \\ \beta_1 \\ . \\ \beta_i \\ . \\ \beta_d \end{bmatrix}, \varepsilon=\begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ \varepsilon_i \\ . \\ \varepsilon_n \end{bmatrix}.$$

Первый столбец матрицы $\boldsymbol{X}$ содержит фиктивную переменную $x_{i0}=1$, $i=1, 2, \ldots, n$.

Замечание [5,10,11,18]. Анализ и построение зависимостей - одна из основных проблем прикладной статистики и науки вообще. В классической постановке проблема сводится к оценке по результатам наблюдений некоторой линейной модели

$$y_i = \beta_0 + \sum_{j=1}^{k} \beta_j x_{j\mathrm{i}} + \varepsilon_i\,, \qquad (i=1, 2,\ldots, n),$$

где y_i – n случайных величин (наблюдаемые входные переменные), являющихся линейными комбинациями x_{ji} с m неизвестными постоянными $\beta_1, \beta_2,\ldots,\beta_m$ плюс ошибки $\varepsilon_1, \varepsilon_2,\ldots,\varepsilon_m$; $\{x_{ji}\}$ – известные значения наблюдений (постоянные коэффициенты).

Указанная модель имеет общий характер и в зависимости от значений $\{x_{ji}\}$ может описывать три различные схемы [18]:

а) если $x_{ji} = \{0;1\}$, то это модель дисперсионного анализа ;

б) если x_{ji} – пробегает непрерывное множество значений (например, время t,
температура T), то это модель регрессионного анализа;

в) если x_{ji} – совмещает переменные а) и б), то это модель ковариационного анализа.

Входные переменные x_j можно разделить на ненаблюдаемые (латентные) и наблюдаемые (явные). Среди последних выделяют контролируемые и управляемые переменные. Посредством управляемых переменных проводятся активные (планируемые) эксперименты. Контролируемые переменные позволяют провести только пассивный эксперимент.

В общем случае, когда $d> k$ (число коэффициентов регрессии больше числа анализируемых факторов k), можно записать уравнение регрессии в следующей векторной форме:

$$\hat{\boldsymbol{Y}}=\boldsymbol{FB}, \qquad (4.4.6)$$

где $\boldsymbol{F}[f_{iq}(X)\ ]_{nd}$ – матрица известных функций f_{iq} от независимых переменных.

Например, пусть k=2, а d=5, т.е. необходимо вычислить пять коэффициентов: b_0, b_1, b_2, b_3, b_4. Тогда:

$\boldsymbol{F}=$ $(f_{i1}, f_{i2}, f_{i3}, f_{i4}, f_{i5})$, где $f_{i0}=1= x_{i0}$, $f_{i1}=x_{i1}$, $f_{i2}= x_{i2}, f_{i3}= x_{i1}x_{i2}, f_{i5}= x_{i1}^2, f_{i5}= x_{i2}^2$, т.е. введены переобозначения
и вместо уравнения

$\hat{Y} = b_0+ b_1x_1 + b_2x_2\ b_{12}x_1\ x_2+ b_{11}x_1^2 + b_{22}x_2^2$

имеем

$$\hat{Y} = b_0+ b_1 f_1 + b_2 f_{2+}\ b_3 f_3 + b_4 f_4 + b_5 f_5 \qquad (4.4.7)$$

После вычисления коэффициентов регрессии нужно вернуться к первоначальным обозначениям для того, чтобы облегчить интерпретацию результатов.

Задачи регрессионного анализа:

- вычисление коэффициентов регрессии;
- проверка значимости коэффициентов регрессии;
- проверка адекватности модели;
- выбор “лучшей” регрессии;
- вычисление стандартных ошибок.

Вычисление коэффициентов регрессии осуществляется методом наименьших квадратов (МНК-метод).

Проверка значимости коэффициентов регрессии основана на методах проверки “гипотез о средних”.

Проверка адекватности модели основана на методах дисперсионного анализа.

Для выбора “лучшей” регрессии разработано много правил [4,5], так что возникает дополнительная проблема: каким правилом воспользоваться?

Вычисление стандартных ошибок, по которым можно судить о точности предсказаний, осуществляется по обычным формулам расчета средних квадратичных отклонений.

Используя процедуры регрессионного анализа, следует иметь в виду, что полученным результатам можно доверять, если выполняются постулаты регрессионного анализа

Постулаты регрессионного анализа

Для того, чтобы правильно построить модель типа (4.4.6), а также, чтобы выводы, полученные при ее анализе были верными, необходимо, чтобы эксперимент удовлетворял ряду условий.

Первое условие. Результаты эксперимента должны быть свободны от систематических ошибок, т.е. математическое ожидание $M\{Y\}$ величины Y должно быть равно действительному значению $\tilde{Y}$, т.е. :

$$M\left\{Y\right\} = \tilde{Y},$$

$$\tilde{Y} = (\tilde{Y}_1,.., \tilde{Y}_i,.., \tilde{Y}_N)$$

Следовательно, математическое ожидание ошибки ε будет равно нулю

$$M\ \left\{e\right\} = \ E\left\{Y\ - \tilde{Y}\ \right\} = \ 0$$

или если действительным значением считать предсказанное по уравнению регрессии значение $\hat{Y}$, то:

$$M\ \left\{e\right\} = \ E\left\{Y\ - \hat{Y}\ \right\} = \ 0$$

Рассмотрим отдельный опыт в точке x_i. Пусть для «истинной модели» (4.4.1) величина среднего $M\ \{\ \tilde{Y}_i\ \}$ значения $\tilde{y}_i$ равна $M\{\ \tilde{Y}_i\ \} = \tilde{y}_i$.

Определим ошибку ε_i i-го опыта:

$$\varepsilon_i = Y_i - \hat{Y} = \left[(Y_i - \hat{Y}_i) - (\tilde{Y}_i - M\{\tilde{Y}_i\})\right] + \left[\tilde{Y}_i - M\{\tilde{Y}_i\}\right] = A_i + B_i,$$

где $A_i = (Y_i - \hat{Y}_i) - (\tilde{Y} - M\ \{\ \hat{Y}_i\ \})$ – случайная переменная с нулевым средним;

$B_i = \left[\tilde{Y}_I - M\{\tilde{Y}_i\}\right]$ -ошибка смещения.

Если построенная модель верна (корректна), то ошибка смещения равна нулю и первое условие соблюдено.

Второе условие – дисперсия результатов наблюдения во всех x_i точках одинакова, т.е.:

$D\{Y_i\} = \sigma^2,\ D\{\varepsilon_i\} = \sigma^2$ для $\forall i$.

Третье условие – результаты наблюдений в точке x_i не зависят от результатов наблюдений в предыдущей точке x_{i-1}, т.е. Y_{i-1} и Y_i - не коррелированны, так что ковариации равны нулю:

$Cov\{Y_{i-1}, Y_i\} = M\{(Y_{i-1} - \hat{Y}_{i-1})(Y_i - \hat{Y}_i)\} = 0;$

$Cov\{\varepsilon_{i-1}, \varepsilon_i\} = M\{\varepsilon_{i-1} - \varepsilon_i\} = 0;$

Поэтому для уравнения регрессии имеем, например:

$$M\{Y_i\} = \beta_0 + \Sigma_j \beta_j x_{ij} + \Sigma_{uj} \beta_{uj} x_{iu} x_{ij} + \Sigma_j \beta_{jj} x_{ij}^2 + \ldots + \varepsilon_i,$$

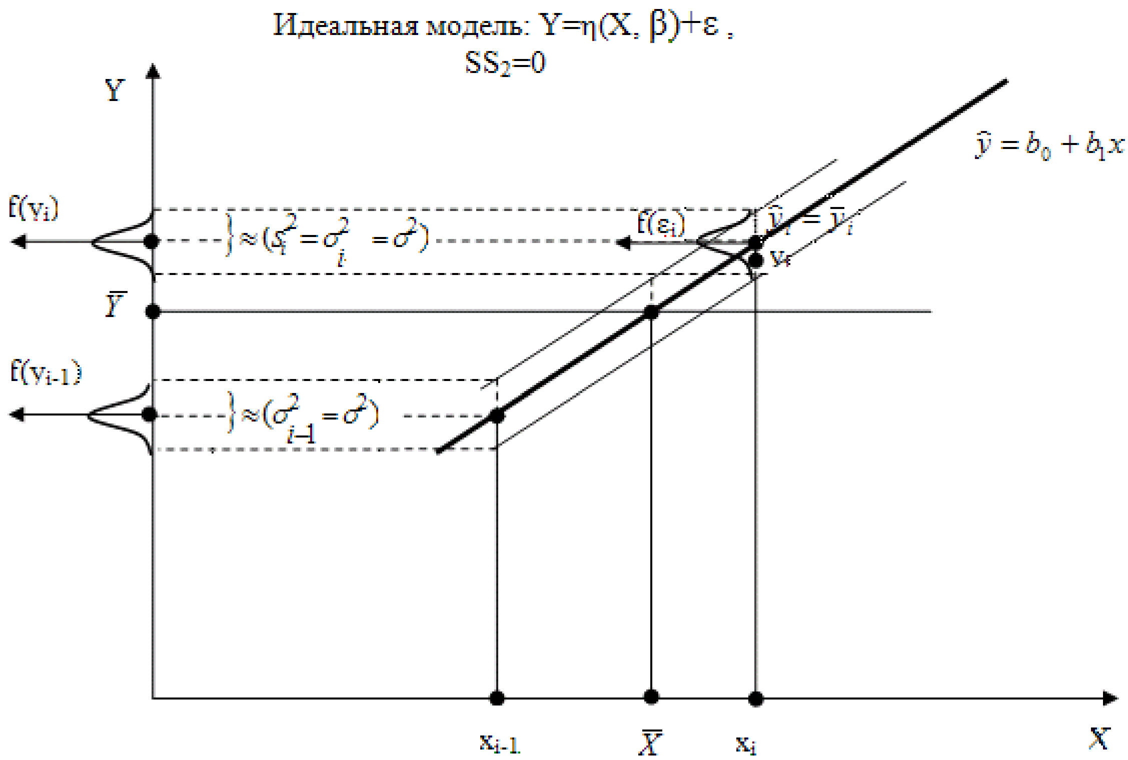


Рис. 56 – Графическая иллюстрация постулатов регрессионного анализа

Четвертое условие: Y_i, ε_i – случайные величины, подчиненные нормальному закону распределения со средними

$M\{\tilde{Y}_i\}$ и дисперсиями $D\{\tilde{Y}_i\} = \sigma^2$, т.е

$Y_i \approx N(\tilde{Y}_i, \sigma^2);$

$\varepsilon_i \approx N(0, \sigma^2)$

где N – обозначение нормальных распределений наблюдаемой величины Y и ее ошибки ε.

Представленные условия формулируются в виде следующих постулатов.

1. Случайная величина Y и ее ошибка ε подчинены нормальному закону распределения.
2. Дисперсия выходной величины Y постоянна и не зависит от величины Y_i, $i = 1, 2,\ldots, n$.
3. Результаты наблюдений Y_i в разных точках эксперимента независимы и не коррелированны.

К этим постулатам добавляют еще один, который практически в большой степени обеспечивает выполнение первых трех.

4. Входные переменные X_j – независимы, неслучайны, измеряются без ошибок. Проиллюстрируем постулаты рис.56.

4.4.2.Оценка коэффициентов регрессии, метод наименьших квадратов

Классическим методом оценивания коэффициентов уравнения регрессии является метод наименьших квадратов (МНК).

Замечание. Первое изложение элементов МНК дано А.М. Лежандром в 1806 г. в связи с вопросами вычисления космических орбит. К.Ф. Гаусс дал вероятностное обоснование МНК (1809 г.), разработал вычислительную сторону вопроса (1810 г., 1821 г.). П.С. Лаплас получил ряд важных результатов и применил их к МНК (1812 г.). П.Л. Чебышев разработал теорию интерполяции по МНК с помощью ортогональных полиномов, носящих его имя. Дальнейшее развитие МНК связано с именами: А.А. Маркова, Ф. Гельмерта, Ю. Неймана, Ф. Дэвид, А. Эйткина, С. Рао. А.Н. Колмогоров в 1946 г. нашел изящное геометрическое изложение МНК [11]. В настоящее время классической книгой является труд Ю.В. Линника "Метод наименьших квадратов и основы теории обработки наблюдений" [11], в котором с использованием матричного подхода изложена современная теория МНК.

Коэффициенты регрессии b, вычисленные по результатам наблюдений, являются только оценками «истинных» коэффициентов регрессии β. Полагают, что будучи случайными величинами, оценки b подчиняются нормальному закону распределения:

$$b \approx N(\beta, D\{b\});,$$

где
$$M\{b\} = \beta$$
$$D\{b\} = D\{(b-\beta)^2\} = E\{(b-\beta)^2\} = \min$$

в котором β – математическое ожидание, D{b} – дисперсия.

Метод наименьших квадратов (МНК) – это метод вычисления коэффициентов b при соблюдении условий минимизации суммы квадратов ошибок (потерь):

$$\sum_{i=1}^{n} g(\varepsilon_i) \to \min.$$

Если функция потерь $g(\varepsilon_i)= \varepsilon_i{}^2$ - квадратичная, то способ получения оценок β_j называется Методом наименьших квадратов (МНК).

$\varepsilon_i = y_i - \hat{y}_i$

$$SS_2 = \sum_{i=1}^{N} \varepsilon_i^2 = \sum_{i=1}^{N} (Y_i - \widetilde{Y}_i)^2 \to \min$$

или в векторной форме

$$SS_2 = \sum_{i=1}^{N} \varepsilon_i^2 = Y^T Y - 2B(F^T Y) + B^T (F^T F) B \to \min.$$

Условия минимума квадрата ошибок:

$$\frac{dSS_2}{db_j} = 0,\ j = \overline{0,d} \quad \text{или} \quad (F^T F)B = F^T Y \text{ -} \qquad (4.4.8)$$

нормальные уравнения.

Решение системы нормальных уравнений (МНК– оценка вектора B):

$$B = (F^T F)^{-1} F^T Y \qquad (4.4.8')$$

Условие разрешимости: детерминант не равен нулю

$$\mathrm{Det}\ (F^T F)^{-1} \neq 0.$$

Линейная регрессия.

Парная регрессия. Рассмотрим решение системы нормальных уравнений в простейшем случае – линейная регрессия от одного фактора x, число определяемых коэффициентов: d=2; $\hat{y} = b_0 + b_1 x$
Система нормальных уравнений

$$\begin{cases} b_0 N + b_1 \sum x_i = \sum y_i; \\ b_0 \sum x_i + b_1 \sum x_i^2 = \sum x_i y_i \end{cases} \qquad (4.4.8'')$$

$$\sum = \sum_{i=1}^{N}$$

Решение

$$b_1 = \frac{\sum x_i y_i - \left[\left(\sum x_i\right)\left(\sum y_i\right)\right]/N}{\sum x_i^{\ 2} - \left(\sum x_i\right)^2 / N} = \frac{\sum \left(x_i - \overline{x}\right)\left(y_i - \overline{y}\right)}{\sum \left(x_i - \overline{x}\right)^2};$$

$b_0 = \hat{y} - b_1$; $\bar{x} = \sum x_i / N$; $\bar{y} = \sum y_i / N$;

тогда $\hat{y}_i = \overline{y} + b_1 (x_i - \overline{X})$ или $\hat{y}_i - \overline{y} = (x_i - \overline{X})$.

Если k=1, d=2, q=1, 2

$F = [f_{iq}(x)]_{n,d} = [f_{i,q=1}(x_0) = 1; f_{i,q=2}(x) = x_i] = [1, x_i]$, $F^T F = X^T X$.

$$X^T X = \begin{bmatrix} N & \sum x_i \\ \sum x_i & \sum x_i^{\ 2} \end{bmatrix} \qquad X^T Y = \begin{bmatrix} \sum y_i \\ \sum y_i x_i \end{bmatrix} = F^T Y.$$

Обращение матрицы 2x2:

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} = \begin{bmatrix} c/D & -b/D \\ -b/D & a/D \end{bmatrix}; \qquad D = ac - b^2 = N\sum x_i^2 - (\sum x_i)^2.$$

$$(X^TX)^2 = \begin{bmatrix} \dfrac{\sum x_i^2}{N\sum\left(x_i - \overline{X}\right)^2} & \dfrac{-\overline{X}}{\sum\left(x_i - \overline{X}\right)^2} \\ \dfrac{-x_i}{\sum\left(x_i - \overline{X}\right)^2} & \dfrac{1}{\sum\left(x_i - \overline{X}\right)^2} \end{bmatrix} = \frac{1}{N\sum\left(x_i - \overline{X}\right)^2}\begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & N \end{bmatrix};$$

$$B = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \frac{1}{N\sum\left(x_i - \overline{X}\right)^2}\begin{bmatrix} \sum x_i^{\,2} & -\sum x_i^{\,2} \\ -\sum x_i^{\,2} & N \end{bmatrix}\begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}.$$

Итак, решив систему (4.4.8″) относительно b_0 и b_1, получим искомое уравнение, которое носит название регрессии Y на X.

Аналогично, предположив, что искомая зависимость имеет вид:

$$x = b_0' + b_1' y, \tag{4.4.9}$$

где значения y_i - фиксированы, мы получим (решив соответствующую систему нормальных уравнений) уравнение регрессии X на Y.

Если X и Y - система двух нормально распределенных случайных величин, то, преобразуя (4.4.4), уравнение регрессии Y на X можно записать:

$$y - \overline{Y} = r\frac{\sigma_y}{\sigma_x}(x - \overline{X}). \tag{4.4.10}$$

Соответственно уравнение регрессии X на Y:

$$x - \overline{X} = r\frac{\sigma_x}{\sigma_y}(y - \overline{Y}). \tag{4.4.11}$$

Множественная линейная регрессия.

Если k=2, d=3, q=1,2,3 и F=(1, x_i, x_{2i})=X, то

$$X^TX = \begin{bmatrix} N\sum x_{1i}\sum x_{2i} \\ \sum x_{1i}\sum x_{1i}^2\sum x_{1i}x_{2i} \\ \sum x_{21i}\sum x_{1i}x_{2i}\sum x_{2i}^2 \end{bmatrix}; \qquad X^TY = \begin{bmatrix} \sum y_i \\ \sum x_{1i}y_i \\ \sum x_{2i}y_i \end{bmatrix}.$$

Обращение матрицы 3x3

$$\begin{bmatrix} a & b & c \\ b & e & f \\ c & f & k \end{bmatrix}^{-1} = \begin{bmatrix} A & B & C \\ B & E & F \\ C & F & K \end{bmatrix};$$

$$D = aA' - bB' + cC' = aek + 2bcf - af^2 - b^2k - c^2e,$$

$$A = A' / D = (ek - f)^2 / D,\ E = (ak - c^2) / D,$$
$$B = B' / D = -(bk - cf) / D,\ F = (af - bc) / D,$$
$$C = C' / D = (bf - ce) / D,\ k = (ae - b^2) / D.$$

4.4.3. Проверка адекватности модели регрессии

После построения уравнения регрессии возникает вопрос о качестве решения. Пусть при исследовании n пар наблюдений (x_i, y_i) получено уравнение регрессии Y на X.

Рассмотрим тождество:

$$y_i - \hat{y}_i = y_i - \overline{Y}_i - (\hat{y}_i - \overline{Y}_i). \qquad (4.4.3.1)$$

Геометрически это тождество можно проиллюстрировать рисунком 43.

Если переписать (4.4.3.1) в виде:

$$(y_i - \bar{y}) = (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i), \qquad (4.4.3.2)$$

возвести обе части в квадрат и просуммировать по i, то получим:

$$\sum (y_i - \bar{y})^2 = \sum (y_i - \hat{y}_i)^2 + \sum (\hat{y}_i - \bar{y})^2, \qquad (4.4.3.3)$$

(можно показать, что $2\sum (\hat{y}_i - \bar{y})(y_i - \hat{y}_i) = 0$).

Уравнение (4.5.3.1) является основополагающим в дисперсионном анализе.

Выражение (4.4.3.3) можно переписать в виде

$$SS_1 = SS_2 + SS_3$$

Для сумм квадратов обычно вводятся названия:

$SS_1 = \sum (y_i - \bar{y})^2 -$ сумма квадратов отклонений i-го наблюдения от общего среднего, или сумма квадратов отклонений относительно среднего наблюдений (скорректированная сумма квадратов Y-ов, в Excel[1]: SS - итог);

$SS_2 = \sum (y_i - \hat{y}_i)^2 -$ сумма квадратов отклонений i-го наблюдения от предсказанного, или сумма квадратов, обусловленная регрессией (в Excel: SS - остаток);

$SS_3 = \sum (\hat{y}_i - \overline{Y})^2 -$ сумма квадратов отклонений предсказанного значения от общего среднего, или сумма квадратов относительно регрессии (в Excel: SS - регрессия);

$(\sum y_i)^2$ - не скорректированная сумма квадратов Y-ов,

$\dfrac{(\sum y_i)^2}{n} = SS_3(b_0)$ - коррекция на среднее суммы квадратов Y-ов.

---

[1] Здесь и далее "в Excel" понимается инструмент **Регрессия** в *Пакете анализа*

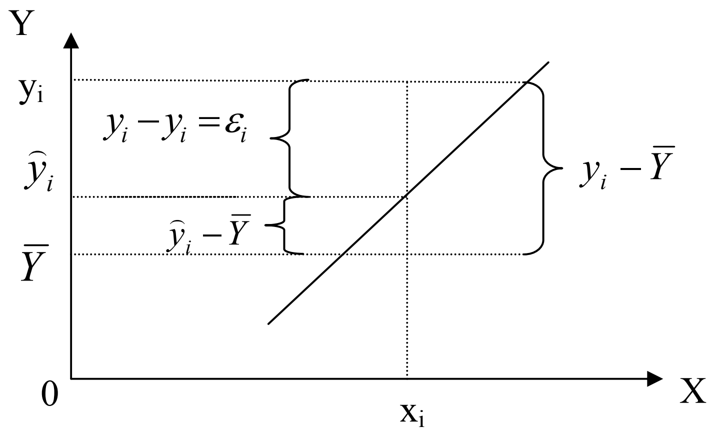


Рис. 57 – Геометрическая интерпретация отклонений

Каждой сумме квадратов соответствует собственное число степеней свободы, аналогично основному уравнению дисперсионного анализа можно записать

$$\nu_{SS1} = \nu_{SS2} + \nu_{SS3}\,,$$

где $\nu_{SS1} = n-1,\ \nu_{SS2} = n-d, \nu_{SS3} = d-1.$

Определим расчетные значения сумм квадратов

$SS_1 = \sum (y_i - \overline{Y})^2 = \sum y_i^2 - n\overline{Y}^2 = SS_1' - SS_3(b_0);$

$SS_3 = \sum (\hat{y}_i - \overline{Y})^2 = SS_3(B/b_0);$

$SS_2 = \sum (y_i - \hat{y}_i)^2 = SS_1 - SS_3\,.$

Для d=2: $SS_3(b_1 | b_0) = b_1[\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{N}];$

$$SS_3(B) = b_1[\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{N}] + N\overline{y^2}.$$

$$SS_2 = \sum y_i^2 - N\overline{y^2} - b_1[\sum x_i y_i - \frac{(\sum x_i)(\sum y_i)}{N}].$$

Для множественной регрессии более удобна матричная форма записи.

$$SS_1 = SS_1^{`} - SS_3 = Y^T - N\bar{y}^2;$$

$$SS_3 = SS_3\left(B/b_0\right) = B^T X^T Y - N\bar{y}^2 = B^T F^T Y - N\bar{y}^2;$$

$$SS_3(B) = B^T X^T Y = B^T F^T Y;$$

$$SS_2 = SS_1' - SS_3(B) = Y^T Y - B^T X^T Y = Y^T Y - B^T F^T Y.$$

Адекватность линии регрессии зависит от того, какая часть суммы квадратов относительно среднего обусловлена суммой квадратов относительно регрессии, а какая суммой квадратов обусловленной регрессией. Суммы квадратов связаны с некоторым числом - числом их степеней свободы $\nu=df$. Это число показывает, сколько независимых элементов информации (из n чисел $y_1, y_2,...,y_n$) необходимо для образования данной суммы квадратов. Например, для $\sum(y_i-\bar{Y})^2$, $df=(n-1)$. Действительно, из n разностей $(y_1-\bar{Y}),(y_2-\bar{Y}),...,(y_n-\bar{Y})$ только (n-1) независимы (или иначе, для образования рассматриваемой суммы из $y_1, y_2,...,y_n$ достаточно (n - 1) значение, так как оставшееся можно определить, зная $\bar{Y}$). Аналогично для $\sum(\hat{y}_i-\bar{Y})^2$, $df=1$; для $\sum(y_i-\hat{y}_i)^2$, $df=(n-2)$ (число степеней свободы определяется как n минус число оцениваемых параметров).

Для построения таблицы дисперсионного анализа необходимо получить средние квадраты (MS), для этого каждая сумма SS делится на соответствующие число степеней свободы df ($MS_R=\frac{SS_{рег.}}{1}, S^2=\frac{SS}{n-2}$).

Если в уравнении регрессии ($y=b_0+b_1x$) $b_1=0$, то величина $F=\frac{MS_R}{S^2}$ распределена по распределению Фишера с (1, n-2) степенями свободы. Этот факт используется для проверки гипотезы H_0: $b_1=0$ с уровнем значимости α (риском ошибиться не более чем в α100% случаев), против альтернативы H_1: $b_1\neq0$.

Обобщим все в таблице дисперсионного анализа (табл.35).

Если $F_{расч.}>F_{кр.}$ при заданном уровне значимости α и соответствующих числах степеней свободы, то гипотеза H_0: $b_1=0$ отбрасывается с риском ошибиться не более чем в α100% случаев (и уравнение регрессии считается статистически значимым).

Доля суммы квадратов, объясняемая регрессией называется множественным коэффициентом детерминации (квадратом множественного коэффициента корреляции R):

$$R^2=\frac{\sum(\hat{y}_i-\bar{y})^2}{\sum(y_i-\bar{y})^2}, 0\le R\le 1, \qquad (4.4.3.4)$$

или в матричной форме

$$R^2=\frac{B^TF^TY-N\overline{y^2}}{Y^TY-N\overline{y^2}}=\frac{SS_3}{SS_1}\approx 1.$$

Значимость R^2 для уравнения (4.4.3.4) определяется по F-критерию. $F_{кр.}=F_\alpha(n,n-d-1)$, $F_{расч.}=\frac{R^2(n-d-1)}{(1-R^2)d}$. Если $F_{расч.}>F_{кр.}$, то гипотезу H_0: $R=0$ отвергают и связь между x и y считают статистически значимой.

Если Y зависит только от одной переменной X, то $R=r$ - парному коэффициенту корреляции.

Таблица 35 – Таблица дисперсионного анализа (основное разложение)

Источник вариации	Число степеней свободы, *df*	Суммы квадратов, *SS*	Средние квадраты, *MS*	$F_{\text{расч.}}$	$F_{\text{кр.}}$
Обусловленный регрессией	1	$\sum_{i=1}^{n}(\hat{y}_i-\bar{y})^2$	MS_R	$F=\frac{MS_R}{S^2}$	$F_\alpha(1, n\text{-}2)$
Относительно регрессии (остаток)	*n*-2	$\sum_{i=1}^{n}(y_i-\hat{y}_i)^2$	$S^2=\frac{SS}{n-2}$		
Общий, скорректированный на среднее Y	*n*-1	$\sum_{i=1}^{n}(y_i-\bar{y})^2$			

Выбор структуры уравнения наилучшей регрессии (наиболее точно описывающей исследуемый процесс) можно осуществить, используя R^2 - квадрат множественного коэффициента корреляции или дисперсионный анализ. Структура уравнения регрессии усложняется (например, в полиномиальном случае повышается степень многочлена) до тех пор, пока увеличение соответствующего критерия не станет пренебрежительно малым.

Однако, вывод о корректности модели по условию $R^2 \approx 1$ не всегда верен. И вот почему. Результата $R^2 \approx 1$ можно добиться, увеличивая число оцениваемых параметров β_j и в случае «насыщенности» d=N R^2 будет равен 1, но модель при этом не обязательно корректна. Особенно вывод опасен при неверном определении N, когда проводятся параллельные опыты. С другой стороны, если $\hat{y}_i=\bar{y}_i$, т.е. $b_0=b_1=...=b_d=0$ или адекватна модель $y=\beta_0+\varepsilon$, то $R^2=0$.

Итак, величина R^2 и суммы квадратов не всегда дают однозначный ответ на вопрос − адекватна ли модель. Для более точной проверки адекватности модели применяют оценки чистой ошибки, определяемые по повторным опытам. Чистая ошибка является оценкой истинной дисперсии σ^2.

Пусть число опытов в каждой точке x_i равно l=1, 2, …, m_i, наблюдаемые значения − y_{il}, тогда вклад в SS_2, связанный с чистой ошибкой, при каждом x_i будет равен

$$s_i^2=\sum_{l=1}^{m_i}(y_{li}-\bar{Y}_i)^2,\ \bar{y}_i=\frac{\sum_{l=1}^{m_i}y_{li}}{m_i},\ \nu_{si}=m_i-1.$$

Общая сумма квадратов, то есть чистая ошибка будет равна

$$SS_2'=\sum_{i=1}^{N}\sum_{l=1}^{m_i}(y_{li}-\bar{Y}_i)^2.$$

Средний квадрат для чистой ошибки

$$s_e^2=\frac{SS_2'}{\nu_e},\ \nu_e=\sum_{i=1}^{N}(m_i-1).$$

N − это число серий повторных опытов.

Для проверки адекватности модели необходимо «расщепить» SS_2 на две составляющие:

$$SS_2 = SS_2' + S_{ад}.$$

Откуда дисперсия адекватности $S_{ад} = SS_2 - SS_2', \nu_{ад} = \nu_{ss2} - \nu_e$.

Средний квадрат $MS_{ад} = \dfrac{S_{ад}}{\nu_{ад}}$.

Откуда $F_{расч.} = \dfrac{MS_{ад}}{s_e^2}$

Если $F_{расч.} > F_{кр.}$, то модель не адекватна. Таким образом мера адекватности F– есть оценка среднего разброса относительно линии регрессии, обусловленная случайными причинами.

4.4.4. Проверка значимости коэффициентов регрессии и стандартных ошибок регрессии

Значимость коэффициентов регрессии. Гипотеза H_0: b_j=0 о значимости коэффициентов регрессии $y = \beta_0 + \sum_{j=1}^{k} b_j x_j$ определяется с использованием t-критерия Стьюдента для двусторонней области при заданном уровне значимости α и $\nu = n$-k-1 степенями свободы:

$$t_{расч} = \frac{|b_j|}{\sqrt{S_\varepsilon^2 \cdot c_{jj}}},$$

где c_{jj} – элемент главной диагонали матрицы $(X'X)^{-1}$, $t_{кр}=t_{дв.\alpha}(n\text{-}k\text{-}1)$, n – число наблюдений, k – число факторов, $S_\varepsilon^2 = \dfrac{\sum (y_i - \hat{y}_i)^2}{n - k - 1}$ - дисперсия остатков. Если гипотеза принимается для всех j, то на этом регрессионный анализ заканчивается (в противном случае незначимые факторы отбрасываются, однако при коррелированности факторов такое исключение оказывается не надежным и приводит к исключению слишком большого числа факторов). Для значимых факторов уравнения регрессии рассматривают интервальные оценки коэффициентов и самого уравнения регрессии.

Доверительный интервал для β_j:

$$b_j - t_{кр.}\sqrt{S_\varepsilon^2 \cdot c_{jj}} \le \beta_j \le b_j + t_{кр.}\sqrt{S_\varepsilon^2 \cdot c_{jj}}. \qquad (4.4.4.1)$$

2. Оценка значимости коэффициента корреляции для уравнения регрессии $y=b_0+b_1x$, осуществляется по t-критерию Стьюдента для двусторонней области при заданном уровне значимости α и $\nu = n$-2 степенями свободы

$$t_{расч.} = |r|\sqrt{\frac{n-2}{1-r^2}}. \qquad (4.4.4.2)$$

3. Для парной регрессии. Значимость коэффициентов регрессии $y=b_0+b_1x$ при заданном уровне значимости α:

а) H_0: $b_1= 0$, H_1: $b_1 \neq 0$,

$$t_{расч.} = \frac{|b|}{S_\varepsilon}\sqrt{\sum_{i=1}^{n}(x_i - \overline{X})^2}\ ; \qquad (4.4.4.3)$$

б) H_0: $b_0= 0$, H_1: $b_0 \neq 0$,

$$t_{расч.} = \frac{b_0}{S_\varepsilon \cdot \sqrt{\frac{1}{n} + \frac{\overline{X}^2}{\sum_{i=1}^{n}(x_i - \overline{X})^2}}}. \qquad (4.4.4.4)$$

4. Доверительные интервалы для b_0, b_1, s^2, r (оценок «истинных» коэффициентов регрессии ($y=b_0+b_1x$) b_0, b_1, дисперсии остатков σ^2 и коэффициента корреляции r_0) при заданном уровне значимости α для двусторонней области:

$$\beta_0 \in \left[b_0 \pm t_{дв.\alpha,(n-2)} \cdot S_\varepsilon \sqrt{\frac{\sum x_i^2}{n\sum (x - \overline{X})^2}} \right], \qquad \beta_1 \in \left[b_1 \pm t_{дв.\alpha,(n-2)} \cdot \frac{S_\varepsilon}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{X})^2}} \right],$$

$$\sigma^2 \in \left[\frac{nS_\varepsilon^2}{\chi^2_{1-\alpha/2,(n-2)}}; \frac{nS_\varepsilon^2}{\chi^2_{\alpha/2,(n-2)}} \right],$$ при больших выборках ($n \geq 30$) r_0

$$\in \left[r \pm u_{\alpha/2} \cdot \frac{\sqrt{1-r^2}}{\sqrt{n}} \right].$$

При $n>10$ проверку гипотезы H_0: $r = r_0$, а также нахождение доверительного интервала для r осуществляют с использованием статистики Фишера[2]

$$Z_r = \frac{1}{2}\ln\frac{1+r}{1-r}.$$

Стандартные ошибки. Доверительный интервал для уравнения регрессии ($y=b_0+b_1x$), при заданном $X=x_0$, $\varphi(x_0) = M(Y/x_0)$:

$$\tilde{y} \in \left[(b_0 + b_1x_0) \pm t_{дв.\alpha,(n-2)} S_\varepsilon \cdot \sqrt{\frac{1}{n} + \frac{(x_0 - \overline{X})^2}{\sum_{i=1}^{n}(x_i - \overline{X})^2}} \right].$$

$$D(\hat{Y}) = D(\tilde{Y}) = D(\overline{Y})(x_i - \overline{X})D(b_1) = \frac{\sigma^2}{N} + \frac{(x_i - \overline{X})^2\sigma^2}{\sum (x_i - \overline{X})^2}$$ – дисперсия истинного

[2] Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: Учебник. – М.: Финансы и статистика, 1998. С. 49-50.

среднего.

$S(\widetilde{Y}_i) = s\sqrt{\dfrac{1}{N} + \dfrac{(x_i - \overline{X})^2}{\sum (x_i - \overline{X})^2}}$ – оценка стандартного отклонения для истинного среднего.

$\Delta_1 = \Delta_{\widetilde{Y}_i} = \widehat{y}_i \pm t_{[(N-d),1-0.5\alpha]} \cdot s(\widetilde{Y}_i)$ – доверительный интервал истинного среднего.

$D(Y_i) = \sigma^2 + D(\widetilde{Y}_i) = \sigma^2\left[1 + N + \dfrac{(x_i - \overline{X})^2}{\sum (x_i - \overline{X})^2}\right]$ – дисперсия индивидуального наблюдения.

$S(Y_i) = \sqrt{s^2 + s^2(\widetilde{Y}_i)} = s\sqrt{1 + N + \dfrac{(x_i - \overline{X})^2}{\sum (x_i - \overline{X})^2}}$ – оценка стандартной ошибки индивидуального наблюдения.

$\Delta_2 = \Delta_{y_i} = \widehat{y}_i \pm t_{[(N-d),1-0.5\alpha]} \cdot s(y_i)$ – доверительный интервал индивидуального наблюдения.

$s(\overline{Y}_i) = s\sqrt{\dfrac{1}{m_i} + \dfrac{1}{N} + \dfrac{(x_i - \overline{X})^2}{\sum (x_i - \overline{X})^2}}$ – оценка средней по повторным опытам.

$\Delta_3 = \Delta_{\overline{y}_i} = \widehat{y}_i \pm t_{[(N-d),1-0.5\alpha]} \cdot s(\overline{y}_i)$ – доверительный интервал средней по повторным опытам.

$$\Delta_1 < \Delta_2 < \Delta_3.$$

(Для построения однофакторных, многофакторных регрессионных моделей и оценки их параметров пользуются стандартными статистическими программами, например, Пакет анализа в *Excel*, программа *Statistica*).

4.4.5. Модели и методы регрессионного анализа

В регрессионном анализе рассматривают [4,5,10,11]:

1) линейную относительно параметров регрессию:
   а) парную линейную ($y=b_0+b_1x$),
   б) парную криволинейную (например, $y=a+bx+cx^2$, $y=a+b\sin x$),
   в) множественную линейную ($y = b_0 + \sum_{j=1}^{k} b_j x_j$),
   г) множественную нелинейную (например, $y=a_0+a_1x_1+a_2x_2+a_3x_1^2+a_4x_2^2+a_5x_1x_2$),
   д) ортогональную полиномиальную регрессию ($y = b_0 + \sum_{j=1}^{k} b_j \varphi_j(x)$, где φ_j - некоторые функции, например, ортогональные полиномы Чебышева);
2) нелинейную относительно параметров (например, $y=a+be^{cx}$).

Практически все перечисленные типы моделей можно построить с помощью той или иной модификации МНК. Адекватность построенных моделей можно проверить с помощью дисперсионного анализа (по схемам аналогичным случаю рассмотренному в параграфе 4.5) и квадрата множественного коэффициента корреляции R^2 (в последнем случае, если независимые переменные детерминируемы, это недопустимо [4]с.34 – 41) .

Вопрос выбора наилучшей регрессии из нескольких построенных моделей регрессии далеко не однозначен. Один из способов, основывающийся на дисперсионном анализе указан выше. Однако он неявно предполагает выполнение классических условий (условий Гаусса – Маркова) применения МНК:

1) вид зависимости $y=f(x)$ известен;
2) входные переменные (x_i) заданы без ошибок;
3) ошибки $\varepsilon_i = y_i - \hat{y}_i$ подчиняются нормальному закону распределения, причем $M(\varepsilon)=0$, $D(\varepsilon)=c= const$;
4) если входных переменных несколько, то они некоррелированны (отсутствие мультиколлинеарности);
5) последовательные значения ошибок ε_i (наблюдений y_i) должны быть некоррелированны - отсутствие автокорреляции ошибок (наблюдений).

На практике не все выше перечисленные условия применения МНК соблюдаются. Обычно, вид зависимости априори неизвестен, ошибки не подчиняются нормальному закону распределения (так как законы больших чисел для конечных выборок не выполняются). Случайные ошибки имеются не только на выходе, но и на входе и т.д.

В настоящее время существует ряд методов, позволяющих получать эффективные параметры регрессии при нарушении условий нормальности ошибок, использующие функции ошибок (функции потерь): Хубера, Пуанкаре, Винзора, Андрюса, Мешалкина, Рамсея, Гуды; джекknайф - оценка и т.д. [4]. При наличии ошибок, как на выходе, так и на входе, вводится понятие ортогональной регрессии (параметры регрессии находят из условия минимизации суммы квадратов расстояний до поверхности регрессии). Если входные переменные коррелированы (мультиколлинеарность), то используются ридж - оценки, метод главных компонент, метод автоматического отсева переменных и т.д.

В случае множественной регрессии выбор "наилучшей регрессии" осуществляется с помощью пошаговой регрессии последовательно включающей (отбрасывающей) входные переменные, факторного анализа, анализа главных компонент.

Если часть наблюдений, используемых в регрессионном анализе, имеет сильно отличающиеся дисперсии, то используется "взвешенный" МНК [4,5,11].

В последние десятилетия развитие регрессионного анализа характеризуется привлечением топологии, теории групп, функционального анализа. Так в последнем случае – обобщение регрессионных моделей на бесконечномерные гильбертовы пространства носит название коллокационных моделей[3] и позволяет решать ряд задач прогнозирования, например, в финансовой сфере.

Замечание.
1. МНК чувствителен к ошибкам округления, поэтому при больших значениях наблюдений x переходят к новой переменной $x' = x - \overline{x}$ (а иногда и $y' = y - \overline{y}$). (При этом лучше иметь нечётное число измерений, так как в случае равных интервалов между последовательными значениями x, сумма нечётных степеней x' становится равной 0 и, следовательно, число членов в системе нормальных уравнений уменьшается, что облегчает её решение.)

Анализ графического изображения наблюдений и знание природы изучаемого процесса
может привести к нелинейной относительно переменной x функции регрессии. В этом случае, для оценки параметров регрессии можно использовать таблицу A[1].

Многие парные нелинейные относительно переменной X уравнения регрессии можно свести к парным линейным, путём соответствующих преобразований переменных и параметров исходных уравнений регрессии (см. табл. B[4]). Хотя следует отметить, что в современных статистических пакетах (например, Statistica) построение большинства нужных уравнений регрессии происходит автоматически.

2. Важным моментом при построении искомой зависимости является отбор факторов x_j, существенно влияющих на результативную переменную y. Известно достаточно много путей отбора, условно их можно разделить на два класса: формальные и содержательные (семантические или смысловые).

Формальные методы основываются на идее перебора различных уравнений (например, различные модификации пошаговой регрессии с последовательным включением или исключением независимых переменных) до момента достижения некоторого критерия, например, F Фишера (дисперсионный анализ), характеризующего (при заданном уровне значимости α) значимость вклада переменной в регрессию.

Содержательные методы предполагают достижение целей моделирования, при этом, как отмечалось выше, различают:

а) Физические модели, описывающие функциональные особенности изучаемых процессов. Построение функциональной модели – редкий случай, так как принципиально невозможно учесть все причинно – следственные связи и их взаимодействия.

б) Модель для управления процессом. Предполагается возможным для любого y_j найти такие x_{ji} (управляющие воздействия), что, задав их в модели, получим требуемое y_j.

в) Модель для предсказания. Даёт возможность по заданным x_{ji} определить прогнозируемое y_i.

Наблюдения за функционированием сложного объекта можно представить в виде множества точек некоторого фазового пространства. Тогда физические модели, модели управления и предсказания – это возможные проекции объекта на различные плоскости, поэтому на практике эти модели не совпадают.

Очевидно, что практически более необходимы модели, описывающие содержательные стороны изучаемого процесса. А попытка совмещения формальных и содержательных критериев – это типичная многокритериальная задача, решение которой обычно не однозначно.

---

[3] Бабешко Л.О. Коллокационные модели прогнозирования в финансовой сфере. -М.: «Экзамен»,2001г. – 288с.
[4] Львовский Е.Н. Статистические методы построения эмпирических формул: Учеб. пособие для втузов.-2-е изд. перераб. и доп.-М.:Высш.шк., 1988. – 239 с.: ил.

Таблица А

№ п/п	Функции	Нормальные уравнения
1	$y = a + bx$	$an + b\sum x = \sum y$ $a\sum x + b\sum x^2 = \sum (xy)$
2	$\lg y = a + bx$ или $y = 10^{a+bx}$	$an + b\sum x = \sum \lg y$ $a\sum x + b\sum x^2 = \sum (x\lg y)$
3	$y = a + b\lg x$	$an + b\sum \lg x = \sum y$ $a\sum \lg x + b\sum (\lg x)^2 = \sum (y\lg x)$
4	$\lg y = a + b\lg x$ или $\mathrm{y} = 10^{\mathrm{a+b\lg x}}$	$\mathrm{an} + \mathrm{b}\sum \lg \mathrm{x} = \sum \lg \mathrm{y}$ $a\sum \lg x \cdot b\sum (\lg x)^2 = \sum (\lg x \lg y)$
5	$y = ab^x$ или $\lg y = \lg a + x\lg b$	$n\lg a + \lg b\sum x = \sum \lg y$ $\lg a\sum x + \lg b\sum x^2 = \sum (\lg x \lg y)$
6	$y = a + bx + cx^2$	$an + b\sum x + c\sum x^2 = \sum y$ $a\sum x + b\sum x^2 + c\sum x^3 = \sum xy$ $a\sum x^2 + b\sum x^3 + c\sum x^4 = \sum (x^2 y)$
7	$y = a + bx + c\sqrt{x}$	$an + b\sum x + c\sum \sqrt{x} = \sum y$ $a\sum x + b\sum x^2 + c\sum \sqrt{x^3} = \sum xy$ $a\Sigma\sqrt{x} + b\Sigma\sqrt{x^3} + c\Sigma x = \Sigma\sqrt{xy}$
8	$y = ab^x c^{x^2}$ или $\lg y = \lg a + x\lg x + x^2 \lg c$	$n\lg a + \lg b\sum x + \lg c\sum x^2 = \sum \lg y$ $\lg a\sum x + \lg b\sum x^2 + \lg c\sum x^3 = \sum (x\lg y)$ $\lg a\sum x^2 + \lg b\sum x^3 + \lg c\sum x^4 = \sum (x^2 \lg y)$

Таблица В

№ п/п	Функция	Линеаризующие преобразования			
		Преобразования переменных		выражения для величин a и b	
		y'	x'	a'	b'
1	$y = a + b/x$	y	$1/\mathrm{x}$	a	b
2	$y = 1/(a + bx)$	$1/\mathrm{y}$	x	a	b
3	$y = x/(a + bx)$	x/y	x	a	b
4	$y = ab^x$	$\lg \mathrm{y}$	x	$\lg a$	$\lg b$
5	$y = ae^{bx}$	$\ln \mathrm{y}$	x	$\ln a$	b
6	$y = 1/(a + be^{-x})$	$1/\mathrm{y}$	$\mathrm{e}^{-\mathrm{x}}$	a	b
7	$y = ax^b$	$\lg \mathrm{y}$	$\lg \mathrm{x}$	$\lg a$	b
8	$y = a + b\lg x$	y	$\lg \mathrm{x}$	a	b
9	$y = a/(b + x)$	$1/\mathrm{y}$	x	b/a	$1/a$
10	$y = ax/(b + x)$	$1/\mathrm{y}$	$1/\mathrm{x}$	b/a	$1/a$
11	$y = ae^{b/x}$	$\ln \mathrm{y}$	$1/\mathrm{x}$	$\ln a$	b
12	$y = a + bx^n$	y	x^{n}	a	b

3. Альтернативным методом "выбора наилучшей регрессии" является метод группового учета аргумента (МГУА), позволяющий определить в данном классе функций с помощью ЭВМ оптимальную структуру искомой зависимости и идентифицировать параметры по внешним критериям, установленным человеком. Человек задает среду решения (список переменных и перечень опорных функций), а так же критерий селекции (отбора). Для проверки адекватности последовательность наблюдений разбивается на части (последовательности). Первая часть (обучающая последовательность) используется для определения оценок коэффициентов по методу наименьших квадратов. Вторая часть (проверочная последовательность) используется для селекции моделей по внешнему критерию, например, построение лучшей прогнозирующей модели.

Автором МГУА является академик АН Украины А.Г. Ивахненко. МГУА хорошо проявил себя при решении многих практических задач (прогнозирования, принятия решений и т.д.), где другие методы оказывались не приемлемы.

4.4.6. Алгоритм корреляционно-регрессионного анализа, примеры и контрольные задания

Корреляционно-регрессионный анализ проводится в следующей последовательности.

1. Исходя из целей и задач исследования зависимости устанавливается результативный (y) и факторные (x_j) признаки.
2. По совокупности объектов определяются значения результативного и факторных признаков.
3. Обосновывается, для случая парной зависимости − обычно графическим методом, модель уравнения регрессии.
4. Методом наименьших квадратов определяются параметры уравнения регрессии.
5. Определяется теснота связи между изучаемыми признаками.
6. Оценивается значимость уравнения связи, его параметров и показателей тесноты связи.

Пример 1. Имеются следующие выборочные данные о стоимости квартир и общей их площади в городе N, июль 2004г.

y	13,8	13,8	14	22,5	24	28	32	20,9	22	21,5	32	35	24	37,9	27,5
x	33	40	36	60	55	80	95	70	48	53	95	75	63	112	70

X – общая площадь квартиры, кв. м;
Y – рыночная стоимость квартиры, тыс. у.е.
Требуется:
1. Построить график зависимости между переменными, по которому необходимо подобрать модель уравнения регрессии.
2. Рассчитать параметры уравнения регрессии методом наименьших квадратов.
3. Оценить качество уравнения с помощью средней ошибки аппроксимации.
4. Найти коэффициент эластичности.
5. Оценить тесноту связи между переменными с помощью показателей корреляции и детерминации.

6. Оценить значимость коэффициентов корреляции и регрессии по критерию t – Стьюдента при уровне значимости $\alpha = 0{,}05$.

7. Охарактеризовать статистическую надежность результатов регрессионного анализа с использованием критерия F – Фишера при уровне значимости $\alpha = 0{,}05$.

8. Определить прогнозное значение результативного признака, если возможное значение факторного признака составит 1.2 от его среднего уровня по совокупности.

Решение. 1. График зависимости переменных *X* и *Y* строится в прямоугольной системе координат. На оси абсцисс откладываются значения факторного признака *X*, а по оси ординат – результативного признака *Y*. Учитывая небольшое число пар значений переменных, по каждой из них выделим пять интервалов, используя формулу:

$$h = \frac{x_{\max} - x_{\min}}{k},$$

где h – длина интервала,

x_{min} - наименьшее значение признака,

x_{max} – наибольшее значение признака,

k – число интервалов.

Для переменной *X*:

$h = \frac{112 - 33}{5} = 15{,}8$. Длина интервала округляется в сторону увеличения до удобного значения, h = 16.

В результате получим следующие границы интервалов:

33, 33+16=49;49+16=65; 65+16=81; 81+16=97; 97+16=113.

Аналогично, для переменной *Y*: $h = \frac{37{,}9 - 13{,}8}{5} = 4{,}82; h = 5$. Границы интервалов составят: 13; 18; 23; 28; 33; 38.

На график наносятся точки, координаты которых соответствуют значениям *X* и *Y*.

Характер расположения точек на графике показывает, что связь между переменными может выражаться линейным уравнением регрессии

$$\hat{y}_x = b_0 + b_1 x.$$

2. Параметры уравнения регрессии находим методом наименьших квадратов, путем составления и решения системы нормальных уравнений (4.4.8″).

Для проведения всех расчетов строится вспомогательная таблица 36.

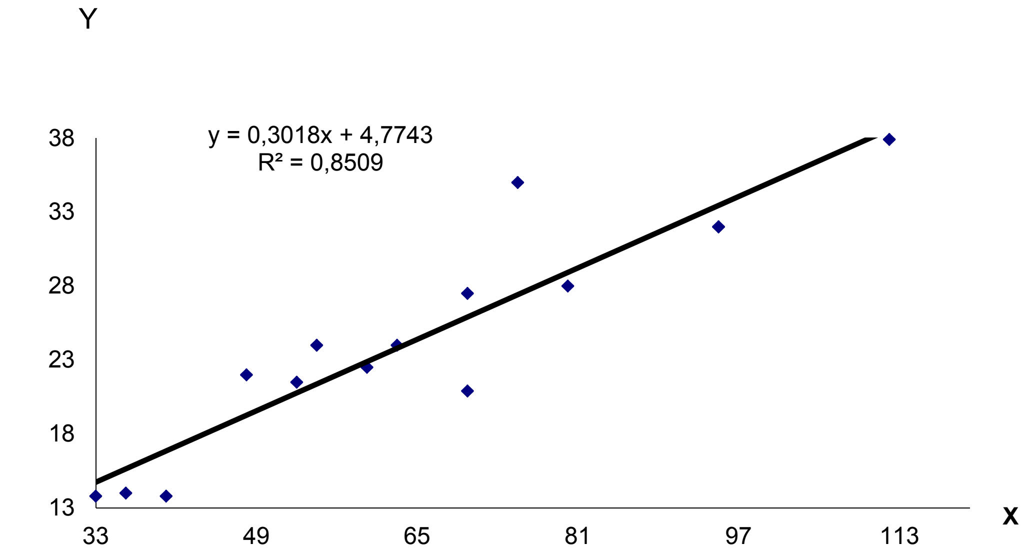


Рис. 58 – Зависимость стоимости квартир от их общей площади

Таблица 36 – Вспомогательная таблица регрессионного анализа

№ п/п	x	y	x^2	y^2	xy	$\hat{y}$	$y-\hat{y}$	$(y-\hat{y})^2$	$A=\left\lvert\frac{y-\hat{y}}{y}\right\rvert\cdot 100\%$
1	33	13,8	1089	190,44	455,4	14,734	-0,934	0,873	6,769
2	40	13,8	1600	190,44	552	16,847	-3,047	9,283	22,078
3	36	14	1296	196	504	15,640	-1,640	2,688	11,711
4	60	22,5	3600	506,25	1350	22,883	-0,383	0,147	1,702
5	55	24	3025	576	1320	21,374	2,626	6,896	10,942
6	80	28	6400	784	2240	28,919	-0,919	0,845	3,283
7	95	32	9025	1024	3040	33,447	-1,447	2,092	4,520
8	70	20,9	4900	436,81	1463	25,901	-5,001	25,012	23,929
9	48	22	2304	484	1056	19,261	2,739	7,500	12,449
10	53	21,5	2809	462,25	1139,5	20,770	0,730	0,532	3,394
11	95	32	9025	1024	3040	33,447	-1,447	2,092	4,520
12	75	35	5625	1225	2625	27,410	7,590	57,604	21,685
13	63	24	3969	576	1512	23,788	0,212	0,045	0,881
14	112	37,9	12544	1436,41	4244,8	38,577	-0,677	0,459	1,787
15	70	27,5	4900	756,25	1925	25,901	1,599	2,556	5,814
Итого	985	368,9	72111	9867,85	26466,7	368,899	-0,001	118,625	135,465
Среднее значение	65,667	24,593	4807,4	657,857	1764,447	24,593	-	7,908	9,031

В таблице все средние находятся по формуле средней арифметической простой: $\bar{X} = \Sigma x : n$.

Подставим полученные суммы в систему уравнений, учитывая, что n=15.

$$\begin{cases} 368{,}9 = 15b_0 + 985b, \\ 26466{,}7 = 985b_0 + 72111b. \end{cases}$$

Решив систему, получим $b_0 = 4{,}7743; b_1 = 0{,}3018$.

Параметры уравнения регрессии также можно найти по формулам, вытекающим из системы нормальных уравнений.

$$b = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\overline{X^2} - \left(\overline{X}\right)^2} = \frac{1764{,}447 - 65{,}667 \cdot 24{,}593}{4807{,}4 - (65{,}667)^2} = 0{,}3018.$$

$$b_0 = \overline{Y} - b_1 \cdot \overline{X} = 24{,}593 - 0{,}3018 \cdot 65{,}667 = 4{,}7743.$$

Небольшие расхождения в результатах расчетов могут происходить за счет округления средних значений во втором случае.

Таким образом, уравнение регрессии имеет вид

$$\hat{y}_x = 4{,}7743 + 0{,}3018x.$$

Коэффициент регрессии показывает, что при увеличении общей площади квартиры на 1 м^2 стоимость квартиры в среднем увеличивается на 0,3018 тыс. у.е. или на 301,8 у.е.

Если в уравнение регрессии подставить фактические значения переменной X, то определяются возможные (теоретические) значения переменной $\hat{y}$, которые наносятся на график в виде уравнения прямой.

3. Качество уравнения регрессии оценивается с помощью средней ошибки аппроксимации

$$\overline{A} = \frac{1}{n} \cdot \sum_{i=1}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right| \cdot 100\%, \quad \overline{A} = \frac{135{,}465}{15} = 9{,}031\%.$$

Значит, фактические значения стоимости квартир от расчетных по уравнению регрессии в среднем различаются на 9,03 %.

Качество уравнения регрессии считается хорошим, если ошибка аппроксимации не превышает 8-10 %. Полученное уравнение регрессии можно оценить как вполне хорошее.

При линейной форме связи средний коэффициент эластичности находится по формуле $Э = b_1 \cdot \frac{\overline{X}}{\overline{Y}}$, где $\overline{X}$ и $\overline{Y}$ - средние значения признаков.

$$Э = 0{,}3018 \cdot \frac{65{,}667}{24{,}593} = 0{,}806.$$

Коэффициент эластичности показывает, что при увеличении общей площади квартиры на 1 % ее стоимость в среднем возрастает на 0,806 %.

5. При линейной зависимости теснота связи между переменными X и Y определяется с помощью коэффициента корреляции:

$r = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{\sigma_x \cdot \sigma_y}$, где σ_x и σ_y – средние квадратические отклонения по X и Y.

$$\sigma_x = \sqrt{\overline{X^2} - \overline{X}^2} = \sqrt{4807{,}4 - 65{,}667^2} = 22{,}254\ ;$$

$$\sigma_y = \sqrt{\overline{Y^2} - \overline{Y}^2} = \sqrt{657{,}857 - 24{,}593^2} = 7{,}283;$$

$$r = \frac{1764{,}447 - 65{,}667 \cdot 24{,}593}{22{,}254 \cdot 7{,}283} = 0{,}922.$$

Так как значение коэффициента корреляции близко к единице, то между признаками связь очень тесная, прямая, близкая к линейной функциональной.

Коэффициент детерминации $r^2 = 0{,}922^2 = 0{,}850$ показывает, что 85 % различий в стоимости квартир объясняется вариацией их общей площади, а 15 % другими, неучтенными факторами (местоположение квартир, благоустроенность территории и другими).

Так как исходные данные являются выборочными, то необходимо оценить существенность или значимость величины коэффициента корреляции. Выдвигаем нулевую гипотезу: коэффициент корреляции в генеральной совокупности равен нулю и изучаемый фактор не оказывает существенного влияния на результативный признак. $H_o : r_г = 0$, при $H_1 : r_г \neq 0$.

Для проверки нулевой гипотезы применим *t*-критерий Стьюдента. Найдем наблюдаемое значение *t*-критерия

$$t_н = \frac{|r|}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{0{,}922}{\sqrt{\frac{1-0{,}922^2}{15-2}}} = 8{,}58 .$$

Критическое значение t находится по таблицам t-распределения Стьюдента при уровне значимости α=0,05 и числе степеней свободы $k = n$-2 = 15-2 =13 для двухсторонней критической области, $t_{кр} = 2{,}16$. Сравниваем $t_{расч.}$ с $t_{кр}$. Так как $t_{расч.} > t_{кр}$, то нулевая гипотеза отвергается, коэффициент корреляции существенно отличен от нуля в генеральной совокупности. Значит, общая площадь квартир оказывает статистически существенное влияние на их стоимость.

Статистическая значимость коэффициента регрессии также проводится с использованием *t*-критерия Стьюдента.

Находится расчетное значение критерия

$$t_{расч.} = \frac{b_1}{m_{b1}}; m_{b1} = \sqrt{\frac{\Sigma(y-\hat{y})^2}{(n-2)\cdot\Sigma(x-\bar{X})^2}} = \sqrt{\frac{\Sigma(y-\hat{y})^2}{(n-2)\cdot\sigma_x^2\cdot n}} =$$

$$= \sqrt{\frac{118{,}625}{(15-2)\cdot 22{,}254^2\cdot 15}} = 0{,}03505;$$

$$t_{расч.} = \frac{0{,}3018}{0{,}035} = 8{,}62 .$$

Критическое значение *t* также равно 2,16. Так как $t_{расч.} > t_{кр.}$, то коэффициент регрессии статистически значим. Подтверждается вывод о значимости влияния общей площади на стоимость квартир.

7. Статистическая надежность уравнения регрессии проверяется с использованием критерия *F* – Фишера.

Расчетное (фактическое) значение *F* – критерия находится по формуле:

$$F_{расч.} = \frac{\Sigma(\hat{y}-\bar{y})^2 / k}{\Sigma(y-\hat{y})^2 / (n-k-1)};$$

где *k* – число параметров при переменных *X*.

Если применяется линейное уравнение регрессии, то расчет $F_{\text{расч.}}$ – упрощается.

$$F_{расч.} = \frac{r^2}{1-r^2} \cdot (n-2) = \frac{0,85}{1-0,85} \cdot 13 = 73,67 .$$

При уровне значимости α=0,05 и числе степеней свободы $k_1 = k = 1, k_2 = n - k - 1 = 15 - 1 - 1 = 15 - 2 = 13$ по таблице находится критическое значение F – критерия. $F_{кр} = F_{05,1,13} = 4,67.$

Так как $F_{\text{расч.}} > F_{\text{кр.}}$, то уравнение регрессии статистически значимое или надежное.

При парной линейной зависимости оценка значимости всего уравнения, коэффициентов корреляции и регрессии дает одинаковые результаты, так как $t_{b_1}^2 = t_r^2 = F_{\alpha,k_1,k_2}$.

8. Прогнозное значение результативного признака определяется путем подстановки в уравнение регрессии прогнозного или возможного значения факторного признака (x_p).

По условию $x_p = \bar{x} \cdot 1,2 = 65,667 \cdot 1,2 = 78,8.$

Тогда прогнозное значение стоимости квартиры составит

$\hat{y}_p = b_0 + b_1 x_p = 4,7743 + 0,3018 \cdot 78,8 = 28,56$.

Значит, при общей площади квартиры в 78,8 м^2 возможная ее стоимость составляет 25,56 тыс. у.е.

Понятие о множественном корреляционно - регрессионном анализе.

В экономических исследованиях результативный признак Y формируется, как правило, под влиянием нескольких факторных признаков $X_1, X_2, \ldots X_k$. Уравнение множественной регрессии имеет вид $y = f(x_1, x_2, \ldots x_k)$.

При построении уравнения множественной регрессии обычно используются следующие функции:

линейная - $y = b_o + b_1 x_1 + b_2 x_2 + \ldots + b_k x_k + \varepsilon$;

степенная - $y = b_o \cdot x_1^{b_1} \cdot x_2^{b_2} \ldots x_k^{b_k} \cdot \varepsilon$;

экспонента - $y = e^{b_o + b_1 x_1 + b_2 x_2 + \ldots + bkx_k + \varepsilon}$.

Часто применяются и другие виды функций, например

$y = b_o + b_1 x_1 + b_2 x_2 + b_3 x_1^2 + b_4 x_2^2 + b_5 x_1 x_2 + \varepsilon$.

Если уравнение нелинейное, то (если это возможно) оно вначале приводится к линейному. Параметры линейного уравнения множественной регрессии находятся методом наименьших квадратов, для чего строится и решается следующая система нормальных уравнений:

$$
\begin{cases}
\Sigma y = nb_o + b_1 \Sigma x_1 + b_2 \Sigma x_2 + ... + b_k \Sigma x_k, \\
\Sigma y x_1 = b_o \Sigma x_1 + b_1 \Sigma x_1^2 + b_2 \Sigma x_1 x_2 + ... + b_k \Sigma x_1 x_k, \\
\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots \\
\Sigma y x_k = b_o \Sigma x_k + b_1 \Sigma x_1 x_k + b_2 \Sigma x_2 x_k + ... + b_k \Sigma x_k^2.
\end{cases}
$$

Множественный коэффициент регрессии b_j показывает, на сколько единиц изменяется в среднем результативный признак Y, если j – ый признак X увеличить на единицу, при условии, что все другие факторы в линейной модели закреплены на постоянном, обычно среднем, уровне.

Уравнение множественной регрессии может быть построено в стандартизованном масштабе, когда единицей измерения признаков принимается их среднее квадратическое отклонение:

$$t_y = \beta_1 t_{x_1} + \beta_2 t_{x_2} + ...\beta_m t_{x_m}\text{, где } t_y = \frac{y - \overline{Y}}{\sigma_y},\ \ t_{x_i} = \frac{x_i - \overline{X}_i}{\sigma_{x_i}},$$

β_i - стандартизованные коэффициенты регрессии (следует отличать их от параметров истинной модели), σ – среднее квадратическое отклонение.

Для оценки тесноты связи между признаками применяются парные, частные и множественные коэффициенты (индексы) корреляции и детерминации. Если изучение зависимости проводится по выборочным данным, то оценивается значимость коэффициентов регрессии, корреляции и всего уравнения множественной регрессии в целом.

Пример 2. Исследовать влияние продуктивности коров и специализации животноводства на трудоемкость производства молока. Результативным признаком (У) являются прямые затраты труда на 1 ц молока, которые характеризуют трудоемкость производства продукции.

Факторные признаки: x_1 – среднегодовой удой молока на корову, характеризующий уровень продуктивности коров; x_2 – удельный вес молока в реализованной продукции животноводства, выражающий уровень специализации животноводства. Исходные данные представлены в таблице 37. Известны так же парные коэффициенты корреляции $r_{yx_1} = -0{,}7930; r_{yx_2} = -0{,}4738; r_{x_1 x_2} = 0{,}5845$.

Требуется определить:

1) параметры множественного уравнения регрессии в натуральной и стандартизованной форме;
2) средние коэффициенты эластичности для каждого фактора;
3) коэффициенты частной и множественной корреляции;
4) общий и частные критерии F – Фишера.

Таблица 37 – Средние значения и колеблемость изучаемых признаков

Признак	Обозначение переменной	Среднее значение	Среднее квадратическое отклонение
Прямые затраты труда на 1 ц молока, чел. ч	Y	5,427	2,983
Среднегодовой удой молока на корову за год, ц	X_1	41,47	12,86
Удельный вес молока в реализованной продукции животноводства, %	X_2	57,48	10,71

Решение. 1. Линейное уравнение множественной регрессии в натуральной форме имеет вид $y = b_o + b_1 x_1 + b_2 x_2$, а в стандартизованной $t_y = \beta_1 \cdot t_{x_1} + \beta_2 t_{x_2}$.

Найдем β – коэффициенты, используя парные коэффициенты корреляции (таблица 37).

$$\beta_1 = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}} = \frac{-0{,}793 - (-0{,}4738) \cdot 0{,}5845}{1 - 0{,}5845^2} = -0{,}7838;$$

$$\beta_2 = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}} = \frac{-0{,}4738 - (-0{,}793) \cdot 0{,}5845}{1 - 0{,}5845^2} = -0{,}0156.$$

Линейное уравнение множественной регрессии в стандартизованном масштабе имеет вид

$$t_y = -0{,}7838 t_{x_1} - 0{,}0156 t_{x_2}.$$

По абсолютной величине β – коэффициентов можно сделать вывод об относительной силе влияния факторов на изменение результативного признака. На трудоемкость производства молока более сильное влияние оказывает удой молока на корову, а влияние удельного веса молока в реализованной продукции животноводства невелико.

Для определения параметров множественного уравнения регрессии в натуральной форме воспользуемся формулами:

$$b_i = \beta_i \cdot \frac{\sigma_y}{\sigma_x}, b_o = \overline{Y} - b_1 \cdot \overline{X}_1 - b_2 \overline{X}_2.$$

$$b_1 = \beta_1 \cdot \frac{\sigma_y}{\sigma_{x_1}} = -0{,}7838 \cdot \frac{2{,}983}{12{,}86} = -0{,}1818;$$

$$b_2 = \beta_2 \cdot \frac{\sigma_y}{\sigma_{x_2}} = -0{,}0156 \cdot \frac{2{,}983}{10{,}71} = -0{,}0043;$$

$$b_o = 5{,}427 - (-0{,}1818) \cdot 41{,}47 - (-0{,}0043) \cdot 57{,}48 = 13{,}2134.$$

Получим линейное уравнение множественной регрессии

$$y = 13{,}2134 - 0{,}1818 x_1 - 0{,}0043 x_2.$$

Коэффициенты множественной регрессии показывают, что при увеличении среднегодового удоя молока на корову на 1ц прямые затраты труда на 1ц молока в среднем снижаются на 0,18 чел-ч, а при увеличении доли молока в реализованной продукции животноводства на 1%, трудоемкость производства молока в среднем снижается на 0,0043 чел-ч.

2. Средние коэффициенты эластичности находятся по формуле:

$$Э_{yx_i} = b_i \cdot \frac{\overline{X}_i}{\overline{Y}};$$

$$Э_{yx_1} = b_1 \cdot \frac{\overline{X}_1}{\overline{Y}} = -0{,}1818 \cdot \frac{41{,}47}{5{,}427} = -1{,}389;$$

$$Э_{yx_2} = b_2 \cdot \frac{\overline{X}_2}{\overline{Y}} = -0{,}0043 \cdot \frac{57{,}48}{5{,}427} = -0{,}046.$$

Значит, при увеличении среднегодового удоя молока на корову на 1%, трудоемкость производства молока снижается в среднем на 1,389 %, при исключении влияния второго фактора. Если увеличить долю молока в реализованной продукции животноводства на 1 %, то трудоемкость молока в среднем уменьшается на 0,046 %, при исключении влияния продуктивности коров.

3. Коэффициенты частной корреляции определяются через парные коэффициенты корреляции по формулам:

$$r_{yx_1 \cdot x_2} = \frac{r_{yx_1} - r_{yx_2} \cdot r_{x_1x_2}}{\sqrt{(1-r^2_{yx_2}) \cdot (1-r^2_{x_1x_2})}} = \frac{-0{,}793 - (-0{,}4738) \cdot 0{,}5845}{\sqrt{(1-0{,}4738^2)(1-0{,}5845^2)}} = -0{,}7222;$$

$$r_{yx_2 \cdot x_1} = \frac{r_{yx_2} - r_{yx_1} \cdot r_{x_1x_2}}{\sqrt{(1-r^2_{yx_1}) \cdot (1-r^2_{x_1x_2})}} = \frac{-0{,}4738 - (-0{,}793) \cdot 0{,}5845}{\sqrt{(1-0{,}193^2)(1-0{,}5845^2)}} = -0{,}0208;$$

$$r_{x_1x_2 \cdot y} = \frac{r_{x_1x_2} - r_{yx_1} \cdot r_{yx_2}}{\sqrt{(1-r^2_{yx_1}) \cdot (1-r^2_{yx_2})}} = \frac{0{,}5845 - (-0{,}793) \cdot (-0{,}4738)}{\sqrt{(1-0{,}793^2)(1-0{,}4738^2)}} = 0{,}3892.$$

Коэффициенты частной корреляции характеризуют тесноту связи между двумя переменными, при исключении влияния третьей переменной. Значит, связь между трудоемкостью производства молока и продуктивностью коров обратная и тесная, между трудоемкостью и долей молока в реализованной продукции животноводства при исключении продуктивности очень слабая. Связь между факторами X_1 и X_2 довольно слабая.

Коэффициент множественной корреляции находится по формуле:

$$R_{yx_1x_2} = \sqrt{\beta_1 \cdot r_{yx_1} + \beta_2 r_{yx_2}} = \sqrt{(-0{,}7838)(-0{,}793) + (-0{,}0156)(-0{,}4738)} =$$
$$= \sqrt{0{,}6216 + 0{,}0074} = \sqrt{0{,}629} = 0{,}7931.$$

Величина коэффициента показывает, что связь между Y, X_1 и X_2 довольно тесная, причем 62,9 % вариации трудоемкости производства молока объясняется влиянием продуктивности коров и удельным весом молока в реализованной продукции животноводства.

4. Оценим значимость уравнения регрессии и коэффициента R^2 с помощью критерия F – Фишера. Наблюдаемое или фактическое значение критерия находится по формуле:

$F_{расч.} = \dfrac{R^2_{yx_1x_2}}{1 - R^2_{yx_1x_2}} : \dfrac{k}{n-k-1}$, где k – число факторов в линейном уравнении регрессии; n – число единиц наблюдения.

$$F_{расч.} = \frac{0{,}629}{1 - 0{,}629} : \frac{2}{15-2-1} = 10{,}17.$$

При уровне значимости α=0,05 и числе степеней свободы $k_1=m=2$, $k_2=n-m-1=15-2-1=12$ по таблице значений критерия F – Фишера критическое значения составляет 3,80, т.е. $F_{кр}=3{,}80$. Сравниваем $F_н$ с $F_{кр}$. Так как $F_н > F_{кр}$, то нулевую гипотезу о незначимости величины R^2 отклоним, т.е. уравнение множественной регрессии и R^2 статистически значимы.

В уравнении множественной регрессии не все факторы могут оказывать статистически существенное влияние на изменение результативного признака. Оценка значимости факторов в уравнении регрессии может быть дана с помощью частного F – критерия или t-критерия Стьюдента.

$$F_{расч.x_1} = \frac{R^2_{yx_1x_2} - r^2_{yx_2}}{1 - R^2_{yx_1x_2}} \cdot \frac{n-k-1}{1} = \frac{0{,}629 - 0{,}4738^2}{1 - 0{,}629} \cdot \frac{15-2-1}{1} = 13{,}08.$$

При α=0,05, $k_1=1$, $k_2=12$, $F_{кр}=4{,}75$.

Так как $F_{расч.x_1} > F_{кр}$, то в уравнение регрессии целесообразно включение фактора x_1 после x_2. Фактор X_1 оказывает статистически значимое влияние на Y.

$$F_{расч.x_2} = \frac{R^2_{yx_1x_2} - r^2_{yx_1}}{1 - R^2_{yx_1x_2}} \cdot \frac{n-k-1}{1} = \frac{0{,}629 - 0{,}793^2}{1 - 0{,}629} \cdot \frac{15-2-1}{1} = 0{,}0049.$$

$F_{кр}=4{,}75$.

Очень низкое значение $F_{нx_2}$ (причем $F_{расч.x_2} < F_{кр}$) свидетельствует о статистической незначимости влияния фактора X_2 и нецелесообразности включения его в уравнение множественной регрессии. В данной задаче на уровень трудоемкости производства молока статистически значимое влияние оказывает продуктивность коров.

Уравнение парной регрессии между X_1 и Y находится по формулам:

$$b_1 = r_{yx_1} \cdot \frac{\sigma_y}{\sigma_{x_1}} = -0,793 \cdot \frac{2,983}{12,86} = -0,1839;$$

$$b_o = \bar{y} - b_1 \cdot \bar{x}_1 = 5,427 - (-0,1839 \cdot 41,47) = 13,053.$$

$$Y = 13,053 - 0,1839 X_1; r = -0,793, r^2 = 0,6288.$$

Значит, при увеличении среднегодового удоя молока на корову на 1 ц прямые затраты труда на 1ц молока в среднем снижаются на 0,1839 чел-ч. Полученное уравнение объясняется 62,88 % различий в трудоемкости производства молока (что на 0,012 % меньше, чем уравнение с двумя факторами).

Контрольные задания 4.4.

1.

№ п/п	Удой молока на фуражную корову, ц	Расход кормов на фуражную корову, ц корм. ед.
1	31,2	33,6
2	44,3	39,7
3	54,5	50,2
4	34,8	36,1
5	46,9	41,2
6	37,2	39,0
7	50,0	45,6
8	34,2	37,4
9	35,0	38,4
10	38,0	40,2
11	53,8	55,7

На основании имеющихся данных определить параметры линейного уравнения регрессии между уровнем кормления и продуктивностью коров, рассчитать коэффициенты корреляции и детерминации. Оценить существенность величины коэффициентов корреляции и регрессии при уровне значимости 0,05.

2. Данные опыта приведены в таблице:

X_i	2	4	6	8	10	12	14
Y_i	4,5	7,0	8,0	7,5	9,0	8,5	9,5

Полагая, что X и Y связаны зависимостью вида y=a+вх, найти коэффициенты a и b способом наименьших квадратов.

3. Дана таблица результатов наблюдений:

X_i	2	4	6	8	10	12	14
Y_i	3,5	6,0	7,0	6,0	7,5	8,5	10

Найти выборочный коэффициент корреляции и определить его значимость.

4. Рейтинг 9 банков был оценен тремя экспертами. С помощью коэффициента ранговой корреляции найти пары экспертов, оценки которых наиболее

близко соответствуют друг другу. Оценить значимость различий в оценке рейтинга банков экспертами.

Рейтинг банков (номер предпочтительности)

Эксперт	Номер банка								
	1	2	3	4	5	6	7	8	9
1	3	2	1	4	5	6	7	8	9
2	2	3	1	4	7	9	8	5	6
3	1	2	5	3	4	6	9	7	8

5. По данным приложения 8 по первым 15 предприятиям требуется:

1. Построить график зависимости между переменными, по которому необходимо подобрать модель уравнения регрессии.
2. Рассчитать параметры уравнения регрессии методом наименьших квадратов.
3. Оценить качество уравнения с помощью средней ошибки аппроксимации.
4. Найти коэффициент эластичности.
5. Оценить тесноту связи между переменными с помощью показателей корреляции и детерминации.
6. Оценить значимость коэффициентов корреляции и регрессии по критерию t – Стьюдента при уровне значимости $\alpha = 0{,}05$.
7. Охарактеризовать статистическую надежность результатов регрессионного анализа с использованием критерия F – Фишера при уровне значимости $\alpha = 0{,}05$.
8. Определить прогнозное значение результативного признака, если возможное значение факторного признака составит 1.2 от его среднего уровня по совокупности.

Изучить зависимость и выявить влияние факторов на изменение результативных признаков по следующим парам признаков:

1) Площадь сельскохозяйственных угодий на среднегодового работника (га) и валовая продукция на 100 га сельскохозяйственных угодий (тыс. руб.);
2) Площадь сельскохозяйственных угодий на эталонный трактор (га) и валовая продукция на 100 га сельскохозяйственных угодий (тыс. руб.);
3) Затраты на 100 га сельскохозяйственных угодий (тыс. руб.) и валовая продукция на 100 га сельскохозяйственных угодий (тыс. руб.);
4) Основные фонды сельскохозяйственного назначения на одного работника (тыс. руб.) и численность работников на 100 га сельскохозяйственных угодий (тыс. руб.);
5) Энергетические мощности на 100 га сельскохозяйственных угодий (л.с.) и численность работников на 100 га сельскохозяйственных угодий (чел.);
6) Энергетические мощности на среднегодового работника (л.с.) и валовая продукция на среднегодового работника (тыс. руб.);

7) Затраты на 100 га сельскохозяйственных угодий (тыс. руб.)и реализованная продукция на 100 га сельскохозяйственных угодий (тыс. руб.);
8) Затраты на реализованную продукцию на 100 га сельскохозяйственных угодий (тыс. руб.) и реализованная продукция на 100 га сельскохозяйственных угодий (тыс. руб.);
9) Энергетические мощности на работника (л.с.) и реализованная продукция на работника (тыс. руб.);
10) Энергетические мощности на работника (л.с.) и валовая продукция на работника (тыс. руб.);
11) Среднегодовая численность работников (чел.) и валовая продукция на одно хозяйство (млн. руб.);
12) Численность тракторов (эт. ед.) и валовая продукция на хозяйство (млн. руб.);
13) Площадь сельскохозяйственных угодий (га) и валовая продукция на хозяйство (млн. руб.);
14) Энергетические мощности (тыс. л.с.) и валовая продукция на хозяйство (млн. руб.);
15) Основные фонды сельскохозяйственного назначения (млн. руб.) и валовая продукция на одно хозяйство (млн. руб.);
16) Затраты на производство валовой продукции (млн. руб.) и валовая продукция на одно хозяйство (млн. руб.);
17) Среднегодовая численность работников (чел.) и реализованная продукция на одно хозяйство (млн. руб.);
18) Численность тракторов (эт. ед.) и реализованная продукция на хозяйство (млн. руб.);
19) Площадь сельскохозяйственных угодий (га) и реализованная продукция на хозяйство (млн. руб.);
20) Энергетические мощности (тыс. л.с.) и реализованная продукция на хозяйство (млн. руб.).

1. Построить график зависимости между переменными, по которому необходимо подобрать модель уравнения регрессии.
2. Рассчитать параметры уравнения регрессии методом наименьших квадратов.
3. Оценить качество уравнения с помощью средней ошибки аппроксимации.
4. Найти коэффициент эластичности.
5. Оценить тесноту связи между переменными с помощью показателей корреляции и детерминации.
6. Оценить значимость коэффициентов корреляции и регрессии по критерию t – Стьюдента при уровне значимости $\alpha = 0{,}05$.
7. Охарактеризовать статистическую надежность результатов регрессионного анализа с использованием критерия F – Фишера при уровне значимости $\alpha = 0{,}05$.

8. Определить прогнозное значение результативного признака, если возможное значение факторного признака составит 1.2 от его среднего уровня по совокупности.

4.5. Анализ временных рядов

Большое количество данных в экономике, коммерции, технике и т.д. могут рассматриваться как в пространстве, так и во времени, путем построения и анализа одного или нескольких временных рядов.

Определим *дискретный временной ряд* как последовательность измерений значений переменной (процесса) за определенный период через одинаковые промежутки времени:

$$Z_1,\ Z_2, Z_3, \ldots Z_t, \ldots, Z_n\ . \qquad (4.5.1)$$

Последовательные наблюдения в (4.7.1) обычно зависимы. С детерминистской точки зрения (4.4.7.1) можно представить как:

$$Z_t = f(t) + \varepsilon_t\,,$$

где t=1, 2,…, n;

f- гладкая (непрерывная и дифференцируемая) функция, характеризующая долгосрочное движение в зависимости от времени - тренд;

ε_t - случайный ряд возмущений, наложенный на систематическую часть.

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционная зависимость между последовательными уровнями временного ряда называют *автокорреляцией уровней ряда*. Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на один или несколько шагов во времени, называемого коэффициентом автокорреляции.

Коэффициент автокорреляции уровней ряда первого порядка, смещенных на одну единицу времени определяется по формуле:

$$r_1 = \frac{\sum_{t=2}^{n}(y_t - \bar{y}_1)\cdot(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^{n}(y_t - \bar{y}_1)^2 \cdot \sum_{t=2}^{n}(y_{t-1} - \bar{y}_2)^2}},$$

Коэффициент автокорреляции уровней ряда второго порядка:

$$r_2 = \frac{\sum_{t=3}^{n}(y_t - \bar{y}_3)\cdot(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^{n}(y_t - \bar{y}_3)^2 \cdot \sum_{t=3}^{n}(y_{t-2} - \bar{y}_4)^2}},$$

где $\bar{y}_3 = \dfrac{\sum_{t=3}^{n} y_t}{n-2}; \qquad \bar{y}_4 = \dfrac{\sum_{t=3}^{n} y_{t-2}}{n-2}.$

Аналогично можно определить коэффициенты автокорреляции более высоких порядков.

Так как коэффициент автокорреляции строится по аналогии с линейным коэффициентом корреляции, то по нему можно судить о наличии линейной или близкой к линейной тенденции. Чем ближе коэффициент автокорреляции первого порядка к единице, тем более выражена линейная тенденция. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию, коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют ***автокорреляционной функцией временного ряда.*** Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка τ, то ряд содержит циклические или сезонные колебания с периодичностью в τ моментов времени. Если ни один коэффициент не является значимым, можно сделать вывод о том, что либо ряд не содержит тенденции и циклических колебаний, либо содержит сильную не линейную тенденцию.

Число периодов или моментов времени, по которым рассчитывается коэффициент автокорреляции, называют ***лагом***.

Построение аналитической функции для моделирования тенденции (тренда) временного ряда называют аналитическим выравниванием временного ряда. Тенденция во времени может принимать разные формы, для ее формализации используют функции, рассмотренные в 4.4.5.

Такой подход, несмотря на заслуженную критику, используется и в настоящее время.

Второй подход (стохастический) заложил Эдни Юл в 1927 г. Он предложил для его объяснения пример, ставший классическим [8]: "Если рассматривать свободное качание маятника, отклоняющегося на малый угол под действием силы тяжести, то хорошо известно, что его движение является гармоническим, т.е. оно может быть представлено синусоидальной и косинусоидальной волной с постоянными амплитудами и периодами колебаний. Но если маленький мальчик обстреливает маятник горохом нерегулярным образом, то его движение будет возмущено. Маятник будет качаться, но с нерегулярными амплитудами и периодами колебаний. Фактически вместо такого поведения, при котором расхождение между теорией и наблюдением можно отнести за счет незначительной ошибки, горох вызывает ряд толчков, *воздействующих на будущее движение системы.* Эта концепция приводит к теории *стохастических процессов*, важнейшим разделом которой является теория стохастических временных рядов".

Третий подход к анализу временных рядов - это спектральный анализ в частотной области. В частном случае можно получить выравнивание по ряду Фурье (при этом обычно рассматривается не более 5 гармоник (j= 1,2,3,4,5)):

$$z_t = a_0 + \sum_{j=1}^{k} (a_j \cos jt + b_j \sin jt)\,. \tag{4.5.2}$$

Параметры a_j и b_j находятся с помощью МНК, в результате применения которого, получим:

$$a_0 = \frac{1}{n}\sum_{t=1}^{n} z_t\,, \quad \mathrm{a_j} = \frac{2}{n}\sum_{t=1}^{n} z_t \cos jt, \quad \mathrm{b_j} = \frac{2}{n}\sum_{t=1}^{n} z_t \sin jt. \tag{4.5.3}$$

Анализ временных рядов преследует целый ряд целей, например [8]:

1) описание поведения ряда;
2) построение модели для объяснения наблюдений;
3) пункты 1) и 2) используют для прогноза, исходя из предположения, о сохранении тенденции развития в будущем.

Для достижения поставленных целей используют модели, основанные на перечисленных выше: детерминистском, стохастическом, спектральном и других подходах. В общем случае можно предположить в модели наличие следующих компонент:

1) тренд или долгосрочное колебание;
2) регулярное движение относительно тренда;
3) сезонная компонента;
4) остаток.

Модель, в которой временной ряд представлен как сумма перечисленных выше компонент, называется *аддитивной моделью* временного ряда. Если временной ряд представлен как произведение компонент, то она называется *мультипликативной моделью* временного ряда.

Отделить тренд и сезонность в общем случае невозможно, так как они взаимно проникают друг в друга. При выделении тренда и сезонности остается колеблющийся ряд. Удаление тренда (сглаживание временного ряда) можно осуществить с помощью скользящей средней (СС). Скользящая средняя, в отличие от простой средней для всей выборки, содержит сведения о тенденциях изменения данных.

Для этого к первым (2m+1) точкам ряда (4.7.1) подбирают полином

$$Q_P(t) = a_p\, t^p + a_{p-1}\, t^{p-1} + a_{p-2}\, t^{p-2} + \ldots + a_1 t + a_0$$

(для определения значения тренда в (m+1) точке) и минимизируют:

$$\sum_{t=-m}^{m} (z_t - a_p\, t^p - a_{p-1}\, t^{p-1} - a_{p-2}\, t^{p-2} - \ldots - a_1 t - a_0)^2\,. \tag{4.5.4}$$

Затем подбирают полином того же порядка для второго, третьего,…, ($2m$+2) наблюдения. Эта процедура продолжается вдоль всего ряда до последней группы из ($2m$+1) точек. На самом деле нет необходимости подбирать полином каждый раз. Покажем, что эта процедура соответствует некоторой линейной комбинации наблюдений с постоянными коэффициентами.

Например, пусть $2m+1=5$ или t: -2,-1,0,1,2; $p=3$

$(Q_3(t)= a_3t^3 + a_2t^2 + a_1t^1+a_0)$,
тогда (4.7.4) примет вид:

$$\sum_{t=-m}^{m}(z_t - a_3t^3 - a_2t^2 - a_1t - a_0)^2 \quad . \tag{4.5.5}$$

После дифференцирования (4.7.5) по a_i и преобразований (так как суммы t с нечётной степенью равны 0) , получим систему уравнений

$$\begin{cases} \sum z_t = 5a_0 + \quad + 10a_2, \\ \sum tz_t = \quad 10a_1 + \quad + 34a_3, \\ \sum t^2 z_t = 10a_0 + \quad + 34a_2, \\ \sum t^3 z_t = \quad 34a_1 + \quad + 130a_3. \end{cases}$$

(4.5.6)

Из (1)и (3) уравнения системы (4.7.6) следует, что:

$$a_0 = \frac{1}{35}(-3z_{-2} + 12z_{-1} + 17z_0 + 12z_1 - 3z_2),$$

но $z_0= a_0$. Итак, значение тренда в какой-либо точке равно средневзвешенному значению пяти точек с данной точкой в качестве центральной и весами $\frac{1}{35}[-3,12,17,12,-3]$. Для пяти точек и $p = 1$ получаем простую скользящую среднюю:

$$a_0 = \frac{1}{5}(z_{-2} + z_{-1} + z_0 + z_1 + z_2).$$

Кроме рассмотренного подхода к выводу формул взвешенных СС существуют другие способы определения СС: использование простых СС, формулы Спенсера и т.д.

При рассмотрении СС, в рамках нашего примера (p=3, m=2), следует отметить проблему крайних двух точек - они не оцениваются, хотя её можно решить, определив из (4.7.6) коэффициенты a_1, a_2, a_3.

Для прогноза в следующей точке следует определить Q_3 (3).

Рассмотренные выше СС (их называют иногда линейными фильтрами), являются симметрическими (т.е. коэффициенты (веса) симметричны относительно среднего).

Для прогнозирования в статистике используют асимметричные фильтры. Так в *Excel* простая *скользящая средняя* заменяет не средний, а последний уровень ряда в промежутке сглаживания, она (СС) используется для расчета значений в прогнозируемом периоде, на основе среднего значения переменной для указанного числа предшествующих периодов, по формуле:

$$F_{t+1} = \frac{1}{N}\sum_{h=0}^{N} Z_{t-h+1}\,, \tag{4.5.7}$$

где N - число предшествующих периодов, входящих в СС;
Z_h - фактическое значение в момент времени h;
F_h- прогнозируемое значение в момент времени h.

Асимметричные СС иногда могут учитывать степень "устаревания данных", т.е. каждое новое наблюдение будет иметь вес больше предыдущих, например:

$$F_{t+1} = (1-\alpha)(A_{t+1} + \alpha A_t + \alpha^2 A_{t-1} + \ldots) = (1-\alpha)\sum_{i=0}^{\infty} \alpha^i A_{t-i+1}, \quad 0 < \alpha < 1 \quad . \qquad (4.5.8)$$

Рассмотренный подход к определению асимметричных СС в Excel носит название *экспоненциальное сглаживание* (ЭС) (или экспоненциальных средних). ЭС предназначается для предсказания значения F_{t+1}на основе прогноза для предыдущего периода F_t, скорректированного с учетом погрешностей в этом прогнозе (A_t - F_t), из (4.5.8) можно получить, что:

$$F_{t+1}=F_t+\alpha\ (A_t-F_t)=\alpha A_t+(1-\alpha)F_t.$$

Существует ещё целый ряд методов сглаживания и экстраполяции: модель Хольта - Уинтерса (содержит три параметра α_1 , α_2, α_3 и позволяет учесть сезонность), модель Харрисона является модификацией предыдущей и выражает сезонность через гармоники . Указанные методы были разработаны для анализа экономических процессов. Широко известны также модель Бокса – Дженкинса, фильтры Калмана и Бюсси [8].

Практически все рассмотренные методы содержат предположения относительно исходных данных, генерирующих временной ряд [8,17]. Критерием адекватности той или иной модели может служить только практическое достижение первоначальных целей анализа временных рядов (описание поведения ряда, объяснение изменения наблюдений, прогноз и т.д.).

Замечание. Тренд даёт возможность прогнозировать основную тенденцию изменения явления во времени. Значения тренда $\hat{z}_t$ обычно сопровождаются ошибками. Хотя при анализе временных рядов обычно нельзя утверждать нормальность распределения отклонений относительно тренда, но, тем не менее, в качестве одного из возможных подходов это предположение используется. Доверительный интервал для тренда определяется как

$$\hat{z}_t \pm t_{дв.\alpha,\nu} \cdot S_{\hat{z}},$$

где z_t - значение тренда в момент времени t, $t_{дв.\alpha,\nu}$ - квантиль распределения Стьюдента для двусторонней критической области при уровне значимости α с ν = n-d степенями свободы (n – число наблюдений, d – число оцениваемых параметров, например, для уравнения прямой d = 2);

$$S_{\hat{z}} = \sqrt{\frac{\sum (z_t - \hat{z}_t)^2}{\nu}}$$ -среднеквадратическое отклонение членов ряда от тренда.

Стационарные временные ряды. Временной ряд, не имеющий тренда (либо с исключённым трендом), называется *стационарным* или иначе - если его свойства не зависят от начала отсчёта времени (механизм, генерирующий ряд не меняется со временем, хотя и носит вероятностный характер). Поэтому перечисленные ниже параметры для данного ряда являются постоянными:

$M(z_t)=M$, $M(z_t-M)^2=\sigma^2=D(z_t)$; $M[(z_t - M)(z_{t+k} - M)]= c_k$ - k- ая автоковариация, $\rho_k = \rho_{-k} = \dfrac{c_k}{\sigma^2}$ - соответствующая автокорреляция.

Иногда совокупность значений ρ_k представляется на графике и называется коррелограммой (но не каждая последовательность констант является коррелограммой [8]).

Для стационарного процесса рассматривают 3 основных типа моделей (соответствующих определённым типам стационарных стохастических процессов).

1. Авторегрессии (АР) порядка p. В этой модели текущее значение t выражается через линейную комбинацию p предыдущих значений процесса плюс случайный импульс ε_t:

$$z_t= a_1z_{t-1}+ a_2z_{t-2} + a_3z_{t-3}+\ldots+ a_{t-p}z_{t-p}+ \varepsilon_t\,. \qquad (4.5.9)$$

Важными частными случаями являются:

а) при p=1, модель процесса Маркова (процесс с отсутствием последействия - каждое следующее значение зависит только от предыдущего):

$$z_t= a_1z_{t-1}+ \varepsilon_t; \qquad (4.5.10)$$

б) при p=2, модель процесса Юла:

$$z_t= a_1z_{t-1}+ a_2z_{t-2} + \varepsilon_t. \qquad (4.5.11)$$

2. Скользящего среднего (СС):

$$z_t= b_1\, \varepsilon_{t-1}+ b_2\, \varepsilon_{t-2} + b_3\, \varepsilon_{t-3}+\ldots+ b_{t-q}\, \varepsilon_{t-q}+ \varepsilon_t \qquad (4.5.12)$$

(термин СС не означает, что сумма весов при ε_i равна 1).

3. Авторегрессии - скользящего среднего (АРСС):

$$z_t= a_1z_{t-1}+ a_2z_{t-2} + a_3z_{t-3}+\ldots+ a_{t-p}z_{t-p}+ b_1\, \varepsilon_{t-1}+ b_2\, \varepsilon_{t-2} + b_3\, \varepsilon_{t-3}+\ldots+ b_{t-q}\, \varepsilon_{t-q}+ \varepsilon_t.$$

Обычно, на практике достаточно рассматривать модели АР, СС, АРСС при p и q не превышающих 2.

Для описания *нестационарных* процессов пользуются экспоненциально взвешенными средними, а в более общем случае моделями авторегрессии-проинтегрированного скользящего среднего (АРПСС) [8,17].

Замечание. 1.Альтернативным методом построения модели прогноза является метод группового учета аргумента (МГУА), позволяющий с помощью ЭВМ определить в данном классе функций оптимальную структуру искомой зависимости и идентифицировать параметры по внешним критериям, установленным человеком. В средине XX века Кенуй и Тьюки предложили метод складного ножа (бут-стреп метод), согласно которому вся совокупность данных разбивается на части и проводится статистический анализ всех частей для достижения несмещённости оценок. Этот метод и был положен в основу МГУА. Человек задает среду решения (список переменных и перечень опорных функций), а так же критерий селекции (отбора). Для проверки адекватности последовательность наблюдений разбивается на части (последовательности). Первая часть (обучающая последовательность) используется для определения оценок коэффициентов по методу наименьших квадратов. Вторая часть (проверочная последовательность) используется для селекции моделей по внешнему критерию, например, построения лучшей прогнозирующей модели.

2. С точки зрения современной науки развитие сложных процессов является детерминированным только между определёнными точками структурных изменений – точками бифуркаций, появление которых случайно. Поэтому можно говорить о существовании пределов предсказуемости, прогнозировать далее которых принципиально невозможно из-за случайного появления бифуркационных точек.

При стандартных подходах (детерминистском, стохастическом и спектральном) наличие неучтенных факторов описывается присутствием случайной составляющей (ε_t), имеющей тот или иной закон распределения (обычно нормальный). Линейные модели окружающего мира, к которым сводятся перечисленные выше методы давно перестали удовлетворять потребностям человека. В настоящее время назрела необходимость описания сложных объектов (процессов) с помощью нелинейных методов, дающих возможность учитывать реальную жизнь (в которой нелинейные связи преобладают), а также учитывать ненаблюдаемые и неконтролируемые факторы. Для этого совокупность данных разбивается на две части: по первой строятся различные модели (обучающая последовательность), по второй отбираются лучшие модели (проверочная последовательность), дающие наилучший прогноз.

Следует отметить, что предлагаемая идеология соответствует принципу Soft Calculation – «мягких вычислений» (терпимость к нечеткости и частичной истинности используемых данных для достижения интерпретируемости, гибкости и низкой стоимости решений), сформулированному известным математиком Лотфи Заде (1994).

В середине 90-х годов оформилось название нового направления в науке – DATA MINING (добыча данных) или иначе – интеллектуальный анализ данных, которое объединило в себе все, что раньше считалось «незаконным ребенком» прикладной науки. Идеология DATA MINING появилась на стыке прикладной статистики, искусственного интеллекта, баз данных и т.д.

Согласно идеологии DATA MINING главным является практическая применимость метода, а не его математическое обоснование, (которое присутствует, но по принципиальным моментам не может быть аксиоматизировано. Например, из теоремы К. Гёделя о неполноте формальной арифметики следует принцип внешних дополнений, согласно которому целью построения математической модели должно быть не достижение формального критерия (минимизация квадратов ошибок, F-Фишера и т.д.), а решение поставленной задачи, например, получение прогнозирующей модели.

Пример 1. Имеются данные о валовом сборе винограда в хозяйствах Краснодарского края.

Год	**1992**	**1993**	**1994**	**1995**	**1996**	**1997**	**1998**	**1999**	**2000**
Валовой сбор, тыс.т	246	229	152	155	190	160	107	155	160

Требуется:

а) построить график временного ряда;

б) рассчитать коэффициент автокорреляции первого порядка;

в) обосновать выбор типа уравнения тренда и рассчитать его параметры;

г) дать интерпретацию параметров тренда и сделать выводы по задаче.

Решение. а) Рассмотрим систему координат $Y0t$, где Y_t – валовой сбор, t - порядковый номер года и нанесем в ней данные примера на график (рис. 59).

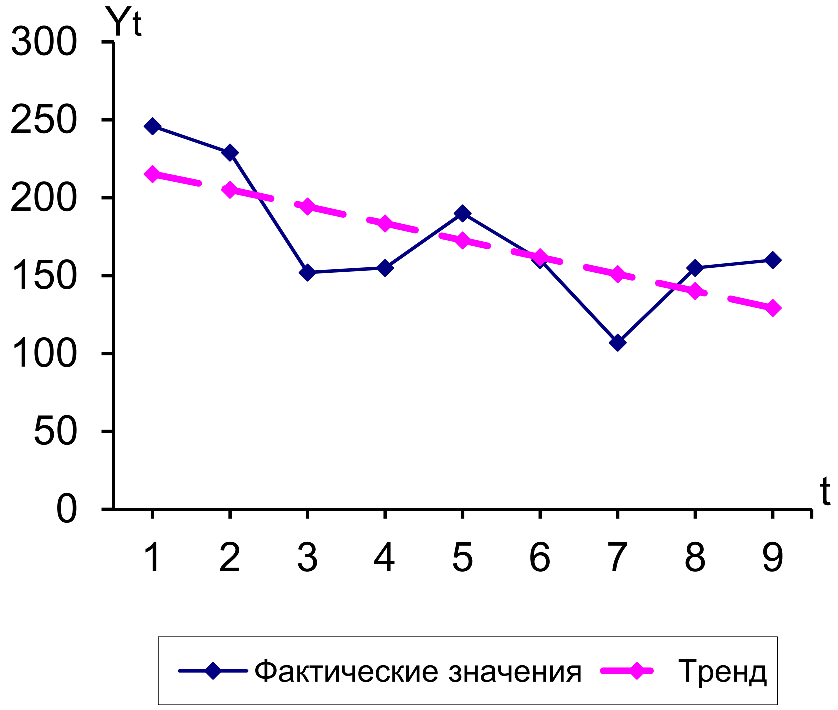


Рис. 59 – Динамика валового сбора винограда во всех категориях хозяйств

б) Определим коэффициент автокорреляции первого порядка, для чего заполним вспомогательную таблицу 38.

Таблица 38 – Вспомогательная таблица для расчета коэффициента автокорреляции

t	y_t	y_{t-1}	$y_t - \bar{y}_1$	$y_{t-1} - \bar{y}_2$	$(y_t - \bar{y}_1) \cdot (y_{t-1} - \bar{y}_2)$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	246	-	-	-	-	-	-
2	229	246	65,5	71,75	4699,625	4290,25	5148,0625
3	152	229	-11,5	54,75	-629,625	132,25	2997,5625
4	155	152	-8,5	-22,25	189,125	72,25	495,0625
5	190	155	26,5	-19,25	-510,125	702,25	370,5625
6	160	190	-3,5	15,75	-55,125	12,25	248,0625
7	107	160	-56,5	-14,25	805,125	3192,25	203,0625
8	155	107	-8,5	-67,25	571,625	72,25	4522,5625
9	160	155	-3,5	-19,25	67,375	12,25	370,5625
Сумма	1554	1394			5138	8486	14355,5

$$r_1 = \frac{\sum\limits_{t=2}^{n}(y_t - \overline{y_1})\cdot(y_{t-2} - \overline{y_2})}{\sqrt{\sum\limits_{t=2}^{n}(y_t - \overline{y_1})^2 \cdot \sum\limits_{t=2}^{n}(y_{t-1} - \overline{y_2})^2}} = \frac{5138}{\sqrt{8486 \cdot 14355{,}5}} = 0{,}4655.$$

$$\overline{y_1} = \frac{\sum\limits_{t=2}^{n} y_t}{n-1} = \frac{1554-246}{8} = 163{,}5;$$

$$\overline{y}_2 = \frac{\sum\limits_{t=2}^{n} y_{t-1}}{n-1} = \frac{1394}{8} = 174{,}25;$$

в) Полученное значение коэффициента автокорреляции и графическое изображение временного ряда позволяют сделать вывод о том, что ряд валового сбора винограда содержит тенденцию, близкую к линейной. Поэтому для моделирования его тенденции используем линейную функцию *y=a+bt.*

Для расчета параметров линейного тренда *a* и *b* используем метод наименьших квадратов, для чего составим и решим следующую систему:

$$\begin{cases} na + b\sum t = \sum y, \\ a\sum t + b\sum t^2 = \sum yt. \end{cases}$$

Заполним вспомогательную таблицу 39.

Таблица 39 – Вспомогательная таблица для расчета параметров тренда

№ п/п	y	t	yt	t^2	y_t
1	246	1	246	1	216,1
2	229	2	458	4	205,2
3	152	3	456	9	194,4
4	155	4	620	16	183,5
5	190	5	950	25	172,7
6	160	6	960	36	161,8
7	107	7	749	49	151,0
8	155	8	1240	64	140,1
9	160	9	1440	81	129,3
Сумма	1554	45	7119	285	1554,1
Среднее значение	172,6667	5	791	31,6667	

Воспользуемся формулами, вытекающими из системы:

$$b = \frac{\overline{yt} - \overline{y} \cdot \overline{t}}{\overline{t^2} - \overline{t}^2} = \frac{791 - 172{,}6667 \cdot 5}{31{,}6667 - 5^2} = \frac{-72{,}3335}{6{,}6667} = -10{,}85;$$

$$a = \overline{y} - b\overline{t} = 172{,}6667 - (-10{,}85) \cdot 5 = 226{,}9167 \Rightarrow \hat{y}_t = 226{,}9167 - 10{,}85t.$$

Таким образом, в среднем ежегодно валовой сбор винограда во всех категориях хозяйств за 1992 – 2000 г. снижался на 10, 85 тыс. тонн

Контрольные задания 4.7

1. На основании данных об урожайности одной сельскохозяйственной культуры: а) построить график динамики урожайности; б) определить параметры тренда урожайности, использовав приемы линейного и нелинейного сглаживания; в) найти выровненные значения урожайности и доверительные интервалы для этих значений; г) определить прогнозные значения урожайности на период до 2015 года.

Урожайность сельскохозяйственных культур

Годы	Пшеница озимая	Кукуруза	Картофель	Сахарная свекла	Подсолнечник	Овощи	Табак
1	2	3	4	5	6	7	8
1994	31,1	26,4	70	237	18,9	109	6,3
1995	32,4	24,9	79	175	18,6	102	6,9
1996	33,1	32,2	83	324	20,4	115	6,1
1997	31,6	33,5	85	264	19,9	117	10,4
1998	37,6	38	68	316	13,4	112	10
1999	28,8	34,8	71	271	22,1	111	9,7
2000	33,2	27,8	81	225	22,3	105	10,6
2001	39,5	30,2	77	289	20,1	129	8,2
2002	37,5	39,4	83	307	15,6	99	7,3
2003	43,2	30,9	76	380	18,2	113	10
2004	36,4	35,3	81	336	23,5	125	12,4
2005	44,1	36,3	86	298	20,4	109	9
2006	39,8	33,3	70	250	17,8	90	6,6
2007	42	35,4	92	278	16,9	123	9,6
2008	36,2	36,4	70	187	16	89	8,4
2009	32,9	31,3	83	259	17,5	82	5,9
2010	38,9	44,6	92	309	12,8	54	7,2
2011	44,5	35,1	95	336	8,4	58	10,8
2012	39,9	42,9	107	442	12,4	65	12,8

2. Имеются следующие данные об объеме подрядных работ строительной организации:

Объем подрядных работ, тыс. руб.

Месяцы	2009г.	2010г.	2011г.	2012г.
Январь	125	158	146	181
Февраль	188	196	232	220
Март	232	264	289	263
Апрель	441	487	466	466
Май	410	405	434	448
Июнь	421	458	411	464
Июль	503	594	603	668
Август	541	605	574	631
Сентябрь	487	511	534	540
Октябрь	317	407	485	391
Ноябрь	246	386	423	378
Декабрь	328	315	398	408

Построить график динамики объема подрядных работ. Определить параметры тренда объема подрядных работ, включающего общую закономерность изменения объема работ и периодическую составляющую, используя периодическую функцию ряда Фурье.

3. На основании данных приложения 9, об урожайности одной сельскохозяйственной культуры: а) построить графики динамики урожайности; б) определить параметры трендов урожайности, использовав приемы: линейного, нелинейного, экспоненциального сглаживания и т.д.; в) найти выровненные значения урожайности и доверительные интервалы для этих значений; г) определить прогнозные значения урожайности на период до 2015 года.

Индивидуальные задания к главе 4

1. Цель и задачи работы

Цель: показать умение моделировать объект или процесс методами регрессионного анализа; продемонстрировать вычислительные навыки и навыки статистических выводов и решений; проявить умение анализировать исходные данные и результаты расчетов, умение практически использовать полученную регрессионную модель.

Задачи:

- уметь построить уравнения регрессии, используя необходимые формулы и методы для расчетов;
- проверить адекватность моделей, выбрать лучшую модель;
- определить значения управляющих воздействий, обеспечивающие заданный номинал с наименьшей ошибкой;

- закрепить навыки вычислений и анализа, приобретенные в процессе изучения раздела «Регрессионный анализ».

Замечание. Важно осознать, что "*анализ и построение зависимостей*" - это следующий этап статистического исследования после "анализа вариационных рядов", поэтому задания главы 4 фактически продолжают задания главы 3.

2.Общее описание задания

Объект исследования может быть представлен схемой (рис.60).

Рисунок 60 – Объект исследования

На рис. 60: $X=\{x_j\}$, $j=1,..., k$ - входы объекта (независимые факторы, заданы в относительных единицах); Y - выход объекта (некоторый технический, производственный, экономический, социальный, экологический и т.п. показатель, задан в абсолютных единицах).

Объект подвергается сериям испытаний или наблюдению в процессе его естественного функционирования. В результате накапливаются данные по $i=1,..., N$ опытам: $X=\{x_{ij}\}$ и соответствующие им Y_i (табл.40).

Таблица 40

№	Факторы	Значение показателя Y
1	$x_{1,1}...x_{1,k}$	Y_1
.	. ... .	.
.	. ... .	.
N	$x_{N,1}...x_{N,k}$	Y_N

Целью наблюдений является построение по данным таблицы 40 регрессионной модели

$$\widehat{Y} = f(X),$$

где $f(X)$ - функция отклика, чаще всего имеющая вид полинома:

$$\widehat{Y} = b_0 + \sum_{j=1}^{k} b_j x_j + \sum_{\substack{j,i=1 \\ j\neq i}}^{k} b_{ij} x_i x_j + \sum_{j=1}^{k} b_{jj} x_j^2 + \ldots,$$

b - оценки коэффициентов β "истинной" регрессии:

$$Y = \beta_0 + \sum_{j=1}^{k} \beta_j x_j + \sum_{\substack{j,i=1 \\ j\neq i}}^{k} \beta_{ij} x_i x_j + \sum_{j=1}^{k} \beta_{jj} x_j^2 + \ldots$$

Регрессионные модели могут быть построены и использованы, если выполнены постулаты регрессионного анализа (условия Гаусса-Маркова – параграф 4.4.1). Для проверки выполнения постулатов используются данные табл.41.

Таблица 41

Фиксированные наборы значений $X=\{x_j\}$	Значение показателя Y при $j=1,..,m$ опытах
$x_1=\{x_{1,1}, x_{1,2}, \ldots, x_{1,k1}\}$ $x_2=\{x_{2,1}, x_{2,2}, \ldots, x_{2,k2}\}$ …	$y_1=\{y_{1,1}, y_{1,j}, \ldots, y_{1,m1}\}$ $y_2=\{y_{2,1}, y_{2,j}, \ldots, y_{2,m2}\}$ …

Регрессионная модель в зависимости от входящих в нее факторов x_j может быть использована:
- для управления объектом;
- для прогнозирования его состояния;
- для объяснения физики явления.

Прежде чем использовать модель, необходимо провести ее анализ: проверить адекватность модели, значимость коэффициентов регрессии, выбрать наилучшее уравнение регрессии.

Каждый вариант задания может выполняться бригадой из четырех человек.

1. Расчетная часть.
   1.1. Записать 8 уравнений регрессии в общем виде, например:
   а) $Y=b_0+b_1x_1$,
   б) $Y=b_0+b_1x_2$,
   в) $Y=b_0+b_1x_1+b_2x_2$,
   г) $Y=b_0+b_1x_1+b_2x_1x_2$,
   д) $Y=b_0+b_1x_1+b_2x_1^2$ и т.д. (по 2 уравнения для каждого члена бригады). Вычислить коэффициенты регрессии.
   1.2. Проверить адекватность модели.
   1.2.1. Дисперсионный анализ.
   1.2.2. Проверка значимости коэффициентов регрессии.
   1.3. Рассчитать стандартные ошибки для Y.
   1.4. Выбрать лучшую регрессию.
   1.5. Для случаев а), б) определить X_1 и X_2, обеспечивающие номинал, и найти соответствующие этому выбору ошибки ε.
   1.6. Для парных линейных регрессий а), б) изобразить результаты расчетов графически.
   1.7. Дать физическую интерпретацию модели, показать возможности ее использования.
2. Выводы.

3. Варианты задания
Варианты 1.1, 1.2, 1.3, 1.4.
Процесс извлечения гелия.

Исследуется работа промышленных агрегатов по процессу извлечения гелия (*He*) из природного Оренбургского газа. Целью исследования является определение зависимости выходного показателя Y (% извлечения гелия, %He), от технологических переменных X_1 (температура, °C) и X_2 (давление, атм.). В реальных условиях процент извлечения гелия зависит от сотен технологических факторов. В настоящей работе исследуется одна из простейших производственных ситуаций. Из многих переменных выбраны только две, используемые на последних этапах процесса извлечения. Данные наблюдений приведены в таблице. Величины Y заданы в абсолютных единицах, X_2 и X_1 - в относительных (абсолютные значения: $X_1 \in$ [-250 °C; -230 °C]; X_2; $\in$ [4,4 атм.; 5,0 атм.]). Определенные наборы значений X_1 и X_2 при фиксированных значениях других параметров соответствуют различным технологиям производства. Предполагается, что измерения достаточно точны и погрешность невелика. Необходимо найти зависимость $Y = f(X_1,X_2)$, оцениваемую результатами наблюдений, и определить значения X_1 и X_2, обеспечивающие заданный номинал $Y_{ном.}$ = 99,8; 99,5; 99,9; 99,85. Определить точность ε, с которой достигается это значение.

Таблица 40.1

№	X_{1i}	X_{2i}	Y_i
1.	3	7	99,25
2.	3	7	99,35
3.	3	7	99,10
4.	6	3	98,50
5.	6	3	98,70
6.	5	6	99,00
7.	7	4	98,50
8.	7	4	98,00
9.	7	4	98,40
10.	4	6	98,75
11.	5	5	98,75
12.	5	5	98,60
13.	5	5	98,55
14.	3	8	99,75
15.	3	8	99,60
16.	10	1	97,50
17.	1	10	99,98
18.	1	10	99,50
19.	8	2	97,80
20.	2	9	99,85

21.	2	9	99,70
22.	9	1	97,75
23.	9	1	98,25
24.	9	1	97,90
25.	2	10	99,95
26.	2	10	99,85
27.	6	8	99,25
28.	6	8	99,35
29.	3	6	99,20
30.	3	6	99,40

Варианты 2.1, 2.2, 2.3, 2.4.

Процесс пропитки стеклоткани.

В результате пропитки стеклоткани специальными смолами она становится токопроводящей и используется в различных нагревательных устройствах. Выходной показатель Y (сопротивление 1 см^2 ткани; Ом) измеряется в случайно выбранных квадратах ткани из полотна (1,2x40)м^2. Электрическое сопротивление квадрата ткани зависит от некоторых технологических переменных. В исследуемых лабораторных условиях наибольшее влияние на величину Y и точность ее поддержания оказывали переменные X_1 (удельное сопротивление пропитывающей смолы; ρ, Ом) и X_2 (зазор между обжимным валиком, снимающим излишки смолы; δ, мм), заданные в относительных единицах (абсолютные значения $\rho \in [60; 110]$; $\delta \in [4{,}2; 4{,}7]$). Данные одной серии испытаний приведены в табл.20.2. Значения Y заданы в абсолютных единицах.

Таблица 40.2

№	X_{1i}	X_{2i}	Y_I
1.	1	1	75
2.	1	1	85
3.	10	10	115
4.	10	10	110
5.	1	2	85
6.	1	2	70
7.	1	2	80
8.	8	9	100
9.	8	9	115
10.	8	9	120
11.	2	3	75
12.	2	3	90
13.	9	8	120
14.	9	8	115

15.	4	3	100
16.	4	3	105
17.	6	7	110
18.	6	7	100
19.	5	5	90
20.	5	5	105
21.	5	5	85
22.	7	6	95
23.	3	4	90
24.	3	4	95
25.	3	3	80
26.	3	3	85
27.	10	8	100
28.	2	1	85
29.	2	1	70
30.	9	7	95

Необходимо определить зависимости $Y = f(X_1,X_2)$, выбрать из них наилучшие и подобрать значения X_1 и X_2, которые обеспечивают номинал $Y_{ном.}$ = 80; 90; 100; 110 Ом. С какой точностью ε могут поддерживаться эти значения?

Варианты 3.1, 3.2, 3.3, 3.4.

Анализ продуктов питания.

Таблица 40.3

№	X_{1i}	X_{2i}	Y_i
1.	3	6	0,016
2.	3	6	0,015
3.	3	6	0,014
4.	6	4	0,014
5.	4	7	0,013
6.	4	7	0,013
7.	9	1	0,011
8.	9	1	0,012
9.	1	10	0,012
10.	1	10	0,017
11.	1	10	0,015
12.	9	2	0,009
13.	9	2	0,010
14.	2	9	0,014
15.	2	9	0,018

16.	2	9	0,016
17.	8	1	0,009
18.	5	5	0,013
19.	5	5	0,011
20.	5	5	0,014
21.	3	7	0,016
22.	4	6	0,012
23.	4	6	0,011
24.	7	3	0,013
25.	7	3	0,012
26.	2	8	0,011
27.	10	2	0,010
28.	10	2	0,009
29.	7	4	0,010
30.	7	4	0,011

Лаборатория производит анализ продуктов, которые обрабатываются при определенной температуре X_1 (t, °С), и в которые добавляются для увеличения срока годности определенные консерванты X_2 (мг). В готовом продукте может содержаться некоторое количество нежелательных веществ Y (в долях к общей массе). X_1 и X_2 даны в относительных единицах (абсолютные значения $t \in [60; 80]$; консервант $X_2 \in [0{,}5; 1]$), Y - в абсолютных.

Необходимо определить зависимость $Y = f(X_1,X_2)$ и установить значения X_1 и X_2, которые обеспечивают номинал $Y_{\text{ном.}}$ =0,009; 0,010; 0,011; 0,01 г. Определить ошибку ε, которая соответствует установленному номиналу $Y_{\text{ном}}$.

Варианты 4.1, 4.2, 4.3, 4.4.

Рынок ценных бумаг.

Исследуются три типа ценных бумаг, доходность которых равна соответственно: Y, X_1 и X_2. Целью исследования является определение зависимости Y от переменных X_1 и X_2, т.е. изучаются взаимосвязанные колебания курсов ценных бумаг. X_1 и X_2 даны в относительных единицах (абсолютные значения: $X_1 \in [500; 900]$, $X_2 \in [800; 1100]$), Y - в абсолютных денежных единицах.

Необходимо определить зависимость $Y=f(X_1,X_2)$ и установить значения X_1 и X_2, которые обеспечивают номинал $Y_{\text{ном.}}$ = 900; 1000; 1100; 1200. Определить ошибку ε, которая соответствует установленному номиналу $Y_{\text{ном}}$.

Таблица 40.4

№	X_{1i}	X_{2i}	Y_i
1.	10	10	1100
2.	10	10	1000
3.	1	2	500

4.	1	2	450
5.	6	7	1100
6.	6	7	700
7.	6	7	900
8.	5	5	800
9.	5	5	1000
10.	5	5	900
11.	9	8	1400
12.	9	8	1250
13.	2	1	600
14.	2	1	800
15.	2	1	700
16.	7	7	1300
17.	7	7	1100
18.	3	3	700
19.	3	3	650
20.	4	2	600
21.	4	2	750
22.	3	4	1000
23.	8	9	800
24.	8	9	1200
25.	1	1	700
26.	1	1	400
27.	6	3	900
28.	8	7	1600
29.	10	1	1550
30.	10	1	1400

Варианты 5.1, 5.2, 5.3, 5.4.

Процесс обогащения руды.

Исследуются различные технологические режимы на фабрике обогащения руды. Режимы отличаются друг от друга по многим параметрам. Для анализа выбраны два основных: X_1 (скорость дробления; v, обор/мин) и X_2 (степень смачивания отдельных частиц измельченной руды; %). X_1 и X_2 даны в относительных единицах (абсолютные единицы: v$\in$ [20; 50]; % $\in$ [80; 98]). Показателем качества дробления Y является класс крупности (d, мм) частиц в пульпе (смесь измельченной в пыль руды, смоченной жидкостью). Значения Y в таблице данных представлены в абсолютных единицах.

Необходимо найти зависимость $Y = f(X_1, X_2)$ и установить значения X_1 и X_2, которые обеспечивают номинал $Y_{ном.}$ = 68; 70; 72; 75. Определить ошибку ε, которая соответствует установленному номиналу $Y_{ном}$.

Таблица 40.5

№	X_{1i}	X_{2i}	Y_I
1.	10	10	80
2.	1	2	90
3.	1	2	80
4.	6	7	80
5.	6	7	75
6.	4	4	75
7.	4	4	85
8.	4	4	80
9.	8	7	85
10.	8	7	90
11.	8	7	70
12.	2	2	85
13.	2	2	70
14.	7	6	75
15.	7	6	85
16.	5	5	80
17.	5	5	75
18.	5	5	85
19.	3	3	80
20.	3	3	90
21.	8	9	75
22.	8	9	60
23.	8	10	70
24.	8	10	65
25.	2	1	85
26.	2	1	90
27.	9	10	70
28.	9	10	90
29.	1	1	85
30.	3	4	90

Варианты 6.1, 6.2, 6.3, 6.4.

Процесс листопроката.

Процесс листопроката является одним из самых распространенных в металлургическом производстве. Хорошее качество листопроката зависит от многих параметров. Качество характеризуется несколькими показателями, в том числе толщиной листа Y (мм) с соблюдением допуска на толщину ε. В таблице приведены данные зависимости Y только от двух параметров: X_1 (скорость проката; и, обор./мин.) и X_2 (зазор между валками в последней клети; δ, мм). Значения X_1 и X_2 даны в относительных единицах (абсолютные значения: $v \in [40; 50]$; $\delta \in [1,6; 2,1]$), Y - в абсолютных.

Необходимо определить $Y = f(X_1, X_2)$ и установить значения X_1 и X_2, обеспечивающие заданный номинал $Y_{\text{ном.}} = 1,9; 2,0; 2,1; 2,2$. Определить, какая погрешность ε соответствует этим номиналам.

Таблица 40.6

№	X_{1i}	X_{2i}	Y_i
1.	10	1	1,85
2.	4	7	2,15
3.	4	7	2,10
4.	4	7	2,05
5.	9	1	1,90
6.	5	5	2,00
7.	5	5	1,85
8.	5	5	2,10
9.	8	2	1,80
10.	8	2	2,00
11.	3	6	2,15
12.	3	6	1,85
13.	2	8	2,15
14.	6	3	1,90
15.	6	3	1,95
16.	6	3	1,85
17.	2	9	1,95
18.	2	9	2,10
19.	2	9	1,90
20.	7	4	1,80
21.	7	4	2,10
22.	7	4	1,85
23.	9	2	1,75
24.	1	10	2,20
25.	1	10	2,15
26.	4	6	2,10
27.	3	9	2,00
28.	3	9	1,90
29.	1	9	2,25
30.	1	9	2,00

Варианты 7.1, 7.2, 7.3, 7.4.

Инвестирование.

Инвестиционная компания объявляет средний годовой доход $Y_{\text{ном.}}$ по акциям определенного производства. Средний годовой доход Y (измеряется в %) зависит от воздействия внешнего рынка X_1 (спрос растет в %) и внутреннего X_2 (конкуренция, спрос падает в %). Значения X_1 и X_2 даны в относительных единицах (абсолютные значения: $X_1 \in [1; 5]$ %, $X_2 \in [1; 1,5]$ %), Y - в абсолютных. Инвестор желает установить зависимость $Y = f(X_1, X_2)$ и прогнозировать с ее помощью возможность объявленного номинала $Y_{\text{ном.}}$=11; 11,5; 12,0; 12,5%, а также прогнозировать точность установления $Y_{\text{ном}}$.

Таблица 40.7

№	*X1i*	*X2i*	*Yi*
1.	6	4	11
2.	6	4	12,2
3.	1	10	10,05
4.	1	10	11
5.	10	2	11,75
6.	10	2	11,65
7.	5	5	11,25
8.	5	5	12
9.	5	5	10,25
10.	9	1	12,5
11.	9	1	12,1
12.	9	1	12
13.	4	7	11,5
14.	8	1	11,75
15.	8	1	11,8
16.	6	1	11,85
17.	3	7	10,75
18.	4	6	10,5
19.	4	6	9,8
20.	4	6	10,6
21.	1	9	10,75
22.	2	10	11
23.	2	10	10,9
24.	3	6	12
25.	3	6	11,9
26.	7	3	12
27.	2	9	11,25
28.	2	9	10,5
29.	9	2	11,25
30.	7	3	11,5

Варианты 8.1, 8.2, 8.3, 8.4.

Процесс трубосварки.

Трубосварочный цех металлургического завода выпускает стальные трубы различного диаметра Y (мм). Различный диаметр труб обеспечивается соответствующим технологическим процессом, в том числе установлением определенного зазора X_1 (δ, мм) между обжимными валками в последней клети установки, и скоростью проката X_2 (v, м/мин.). В таблице значения X_1 и X_2 даны в относительных единицах (абсолютные значения: $\delta \in [16; 21]$; $v \in [40; 50]$), Y - в абсолютных.

Необходимо определить зависимость $Y = f(X_1, X_2)$ и найти значения X_1 и X_2, обеспечивающие заданный номинал $Y_{ном.}$ = 18; 20; 21; 22 мм. С какой ошибкой устанавливаются $Y_{ном}$?

Таблица 40.8

№	X_{1i}	X_{2i}	Y_I
1.	10	1	22,0
2.	10	1	23,0
3.	10	1	22,5
4.	1	10	17,0
5.	10	2	20,5
6.	10	2	21,5
7.	2	8	18,0
8.	2	8	17,5
9.	5	5	19,5
10.	5	5	20,0
11.	5	5	20,0
12.	1	9	19,00
13.	1	9	21,00
14.	9	2	21,5
15.	9	2	19,5
16.	8	1	21,5
17.	8	1	20,5
18.	3	7	19,5
19.	3	7	18,5
20.	4	6	20,0
21.	4	6	21,0
22.	4	6	19,8
23.	7	3	20,5
24.	6	4	21,5
25.	6	4	19,5
26.	6	4	19,0

27.	4	7	18,0
28.	3	6	21,0
29.	3	6	20,5
30.	5	7	20,0

4. Контрольные вопросы

1) Определите понятие "уравнение регрессии".
2) Назовите ситуации, в которых может быть полезно линейное уравнение регрессии.
3) Что имеется в виду, когда говорится "регрессионная модель линейна"?
4) Что такое "порядок модели"?
5) Для каких целей может быть использованы уравнения регрессии?
6) Опишите процедуру оценивания "метод наименьших квадратов".
7) Что такое "система нормальных уравнений"?
8) Что является решением системы нормальных уравнений?
9) Что такое:
   - нескорректированная сумма квадратов Y-ов?
   - коррекция на среднее значение Y-ов?
   - скорректированная сумма квадратов Y-ов?
   - нескорректированная сумма смешанных произведений?
   - коррекция на среднее?
   - скорректированная сумма произведений X и Y?
10) Что такое "остатки"?
11) Чему равна сумма остатков?
12) Запишите формулы для определения коэффициентов b_0 и b_1 линейной регрессии от одного фактора.
13) К чему приводит исключение β_0 из модели?
14) Будет ли равна нулю сумма остатков, если из модели исключить β_0?
15) Укажите на графике уравнения регрессии, как определить коэффициенты b_0 и b_1 для линейной регрессии $Y = b_0 + b_1X$?

5. Требования к оформлению пояснительной записки

Пояснительная записка может быть представлена в тетради (объем около 12 листов) либо на листах формата А4.

Содержание пояснительной записки.

1. Введение.
   1.1. Характеристика регрессионного анализа, назначение регрессионных моделей, цель и задачи работы.
   1.2. Исходные данные, соответствующие конкретному варианту.

1.3. Описание содержания регрессионного анализа (формулировка проблемы, необходимые формулы).

2. Расчётная часть (расчёты, анализ результатов).
3. Заключение. (Краткое изложение результатов работы, выводы по результатам анализа модели, возможности ее использования при исследовании конкретного объекта. Ответы на контрольные вопросы).

Приложения (алгоритмы, программы).

Литература

II.I. Учебники

1. Гмурман В.Е. Теория вероятностей и математическая статистика. Изд. 6 - е, стер. – М.: Высш. шк., 1997. – 479 с.

На доступном уровне изложены основы теории вероятностей и математической статистики. В последние десятилетия – основной учебник для студентов экономических специальностей вузов.

2. Герасимович А.И., Матвеева Я.И. Математическая статистика. – Минск, Высш. шк., 1978. – 200 с.

Доступно изложены основные разделы математической статистики, соответствующие программе вуза.

3. Венецкий Н.Г., Кильдишев Г.Е. Теория вероятностей и математическая статистика. Изд. 3-е. М.: «Статистика», 1975. – 264 с.

Изложение характеризуется связью с практической деятельностью экономиста. Пособие рекомендовано для студентов экономических специальностей.

4. Демиденко Е. З. Линейная и нелинейная регрессия. – М.: Финансы и статистика, 1981. – 302 с., ил.

Изложены современные методы регрессионного анализа – МНК, робастные методы (устойчивые к возможным загрязнениям и разнородности наблюдений), ридж – оценки (применяющиеся при наличии мультиколлинеарностей) и т.д.

5. Дрейпер И., Смит Г. Прикладной регрессионный анализ: В 2-х кн. Пер. с англ. – 2-е изд. перераб. и доп. – М.: Финансы и статистика, 1986 Кн. 1. – 366 с., ил; 1987. Кн.2. – 351 с., ил.

Приводится описание линейной и нелинейной по параметрам регрессии. Методы выбора «наилучшего» уравнения регрессии и т.д.

6. Закс Л. Статистическое оценивание. Пер. с нем. В.Н. Варыгина. Под ред. Ю.П. Адлера, В. Г. Горского. М., «Статистика», 1976. – 598 с., ил.

На практических примерах рассматриваются основные методы статистического оценивания.

7. Йетс Ф. Выборочный метод в переписях и обследованиях. / Пер. с англ. Е.И. Арона. Под ред. А.Г. Волкова, М.: «Статистика», - 1965. – 434 с.

Содержит элементарное изложение основ выборочного метода, сопровождающиеся большим количеством примеров статистического исследования в сельском хозяйстве и при переписи населения.

8. Кендэл М. Временные ряды. / Пер. с англ. Ю.П. Лукашина. – М.: Финансы и статистика, 1981. – 199 с., ил.

Представлены основные идеи и методы анализа временных рядов. Элементарное изложение и достаточное количество примеров делает эту книгу незаменимой при практических исследованиях.

9. Кокрен У. Методы выборочного исследования. / Пер. с англ. И.М. Сонина. Под ред. А.Г. Волкова. М.: «Статистика», - 1976. – 440 с.

Книга является фундаментальным трудом, в котором систематически излагаются методы отбора, оценивания и вычисления ошибок выборки при выборочных обследованиях. Большое количество практических примеров позволяет использовать книгу при организации выборочного исследования.

10.Кремер Н.Ш. Теория вероятностей и математическая статистика : Учебное пособие для вузов.- М.:ЮНИТИ-ДАНА, 2000.-543с.

Учебник рекомендован для студентов экономических специальностей вузов. Изложение сопровождается большим количеством примеров применения вероятностных и математико-статистических методов в задачах: массового обслуживания, анализа временных рядов, изучения моделей финансового рынка.

11. Линник Ю. В. Метод наименьших квадратов и основы математико-статистической теории обработки наблюдений. 2-е изд. доп. и испр. – М.: Физматгиз, 1962. – 352 с.: ил.

Наиболее полно рассмотрены математико-статистические идеи МНК, при условии, что погрешности измерений можно рассматривать как случайные величины. Вводится понятие ортогональной регрессии, для случая ошибок во входных переменных.

12. Митропольский А.К. Техника статистических вычислений М., Наука, 1971., 576с.:ил.

В книге подробно изложены вычислительные аспекты статистических исследований (вычисление моментов, построение доверительных интервалов, изучение кривых Пирсона, корреляционная теория и т.д.).

13. Смирнов Н.В., Дунин – Барковский И.В. Курс теории вероятностей и математической статистики для технических приложений. М.: Наука, 1969. – 511 с.

Наряду с классическими вопросами теории вероятностей и математической статистики большое внимание уделено приложениям. Книга содержит, одно из немногих в отечественной литературе, изложение «вероятностно-статистических методов расчета, анализа и контроля точности» (расчет размерных цепей, при выполнении на производстве деталей машин; анализ точности и стабильности технологических процессов; контроль качества продукции и т. д.).

14. Пугачев В.С. Теории вероятностей и математическая статистика. – М.: Наука, 1979. – 496 с.,ил.

В учебнике дается строгое изложение основ теории вероятностей и математической статистики. Рассматривается взаимосвязь с теорией управления и идентификации сложных систем. Изложены методы построения основных классов статистических моделей для поддержки процесса принятия решения (регрессионных, разностных, факторных, распознавания и т. д.).

15. Пустыльник Е.И. Статистические методы анализа и обработки наблюдений. М.: Наука, 1968.-288 с.,ил.

На элементарном уровне рассматривается большое число понятий и методов классической математической статистики (основывающейся на предположении о нормальности распределения генеральной совокупности), применительно к задачам обработки наблюдений.

16. Хьюстон А. Дисперсионный анализ. Пер с англ. А.Г. Кругликова. – М.: «Статистика», 1971. – 88 с.

Дается элементарное изложение основных идей и моделей дисперсионного анализа.

17. Четыркин Е.М., Калихман И.Л. Вероятность и статистика. – М.: Финансы и статистика, 1982. – 319 с.:ил.

Рассматриваются основные идеи теории вероятностей и на их основе описываются методы математической статистики, имеющие наибольше применения на практике. Изложены основы анализа временных рядов.

18. Шеффе.Г. Дисперсионный анализ.-М.: Физматгиз, 1963.- 628 с., ил.

Классический труд по основам дисперсионного анализа.

II.II. Задачники и учебные пособия

19. Бондаренко П.С., Кацко И.А., Чуприна Н.В. Сборник задач по теории вероятностей и математической статистике. – Краснодар.: КГАУ, 1999. – 61 с.

20. Горелова Г.В., Карпова Е.А., Харчистов Б.Ф. Методические указания к выполнению домашнего задания №1 «Анализ статистических рядов». – Таганрог.: ТРТУ, 1998 – 21с.

21. Горелова Г.В., Карпова Е.А. Методические указания к выполнению домашнего задания №2 «Построение и анализ регрессионных моделей». – Таганрог.: ТРТУ, 1998 – 17 с.

22. Гмурман Д.Е. Руководство к решению задач по теории вероятностей и математической статистике: Учебное пособие. – 4-е изд., стереотип. – М.: Высш. шк., 1998. – 400 с.

23. Теория статистики с основами теории вероятностей: Учеб. пособие для вузов/И.И. Елисеева, В.С. Князевский, Л.И. Ниворожкина, З.А. Морозова; Под ред. И,И, Елисеевой. – М.: ЮНИТИ-ДАНА, 2001. – 446 с.

Однажды, когда ночь покрыла небеса невидимою своею епанчою, знаменитый французский философ Декарт, у ступенек домашней лестницы своей сидевший и на мрачный горизонт с превеликим вниманием смотрящий, некий прохожий подступил к нему с вопросом: "Скажи, мудрец, сколько звезд на сем небе?" "Мерзавец! - ответствовал сей, - никто необъятного обнять не может!" Сии, с превеликим огнем произнесенные, слова возымели на прохожего желаемое действие.

Из гисторических материалов
Федотоа Кузьмича Пруткова (деда)

...Главный конструктор проекта уже заканчивает свой доклад: « Новый компьютер Ультроник... более 10^{17} логических ячеек. Это больше чем суммарное число нейронов у всех живущих в нашей стране. Уровень интеллекта невообразимо высок. Найдется в этой аудитории желающий инициировать нашу новую компьютерную систему Ультроник, задав ей первый вопрос?»

... Все в смятении, как будто чувствуя присутствие нового всемогущего разума. Адам поднимает руку.

«Ну вот, - говорит Главный конструктор, - парнишка в третьем ряду. У тебя есть вопрос к нашему... гм... новому другу?

..

«... СЕБЯ ЧУВСВУЕШЬ? О... весьма интересный вопрос, мой мальчик... э-э... я и сам хотел бы знать ответ», - сказал Главный конструктор. – «Давайте посмотрим, что может сказать наш друг об этом... странно... э-э... Ультроник говорит, что он не понимает, что... он не может понять, что ты имеешь в виду!» Отдельные смешки в аудитории переросли в громовой хохот. Адам чувствовал себя крайне неловко. Они могли отреагировать как угодно, но только не смеяться.

Р. Пенроуз. Новый ум короля: О компьютерах, мышлении и законах физики. 2003, (1989)

Заключение или эссе на тему принятия решений
(математический подход – концепции и размышлизмы…)

Что такое *принятие решения?* – по видимому вся известная нам сегодня история человечества (от индивидуума до государства) была историей постановки целей (задач) и принятия решений по поводу пути их достижения. Часто человек обладал знаниями для выбора наилучшего пути. «Гнозис (знание) во все времена основывался на естественных законах, а различные ветви математической науки – это просто способы выражения этих законов» (Е.П. Блаватская). Поэтому и история современной математики от египетских жрецов, Пифагора, Аполлония и Евклида до Эйлера, Гаусса, Лобачевского, Римана, Гильберта, Кантора, Колмогорова, фон Неймана, И. Пригожина и Л. Задэ была историей постановки и решения задач, имеющих чисто прикладное значение, соответствующее своему уровню науки и техники. (Правда есть ещё и «чистая математика» - этакая вещь в себе, которая чаще является игрой разума либо «возможно это будет нужно завтра».)

Сложно ответить на вопрос где и когда зародились математические знания, но наши предки знали, что «бог все время геометризирует» (Платон), может быть, поэтому до нас из древности дошли в основном геометрические работы. В первую очередь «Начала» Евклида, они на протяжении более чем 2000 лет были образцом математической строгости в математике. И вследствие этого основой математического описания был аксиоматический подход, основой которого являются субъективные индуктивные знания человека, подтверждающиеся его опытом взаимодействия с окружающим миром. Из аксиом с помощью рассуждений, основанных на индукции, аналогии и дедукции, выводят заключения. С древних времён люди понимали, что субъективный опыт человека основывается на чувственном восприятии окружающего мира, но насколько оно соответствует истине? Например, если существует не одна, а много геометрий (Евклида-параболическая, Лобачевского-гиперболическая, Римана-эллиптическая и удвоенная эллиптическая), которые в принципе согласуются с наблюдательными данными о структуре пространства. Но эти геометрии противоречат одна другой, следовательно, если и существуют фундаментальные законы природы, то они не вполне соответствуют математическим моделям, построенным человеком[1].

Да, без сомнения, законы природы сложнее, чем представляет себе человек, однако, наша техногенная цивилизация в большой степени основана на математике: атомные станции, телевизоры, компьютеры, самолёты, ракеты, заводы, фабрики, машины и т.д. – работают благодаря математике. Так, например, человек не знает физическую природу электромагнитных волн - их существование подтверждается только математикой и только она позволила создать радио и телевидение.

И. Кант считал, что, обладая лишь чувственными восприятиями природы мы не можем ее познать - мы видим в природе то, что позволяет наш ра-

---

[1] Клайн. М. Математика. Утрата определенности. -М:Мир,1984.-434с.

зум. Уайтхед и Браун, развивая идею Канта, назвали математику инструментом упорядочивающим наши ощущения, так как окружающий мир не дан нам объективно. Анри Пуанкаре считал, что человек подгоняет математику под физическую реальность. Давид Гильберт говорил, что существует объективный физический мир и человек стремится согласовать с ним свою математику, т.е. имеет место игра мышления и опыта, объясняющая эффективность математики. Математическое описание окружающего мира предполагает количественный подход: мы рассматриваем его с точки зрения меры, веса, продолжительности, стоимости, прибыли, убытка и тому подобных понятий, что является немалой ценой за универсальность математического подхода, например, что может говорить рост человека о человеке (его богатом и разнообразном опыте, интеллекте и т.д.). Следует отметить особую эффективность математики при описании простейших понятий и упрощенных явлений физического мира и в технике, между тем социальные и экономические процессы, увы… («Единственное, чему меня научила моя долгая жизнь: что вся наша наука перед лицом реальности выглядит примитивно и по детски наивно - и всё же это самое ценное, что у нас есть» А.Эйнштейн).

«Почему-то» в связи научно-технической революцией, особенно в связи с «информационными взрывами», математические модели построенные на аксиоматическом подходе (точнее на априорных предпосылках относительно изучаемых данных), как оказалось, работают далеко не всегда…. То есть совокупность с определенного момента перестаёт быть адекватна сумме своих составляющих − эту совокупность стали называть системой еще во времена Аристотеля), в диалектике этот эффект называется законом перехода количества в качество. Мы не будем касаться проблемы, почему это происходит.

P.S.–*Из-за такой малости! Из-за бабочки! - закричал Экельс. Она упала на пол - изящное маленькое создание, способное нарушить равновесие, повалились маленькие костяшки домино... большие костяшки... огромные костяшки, соединенные цепью неисчислимых лет, составляющих Время.* (Р. Бредбери. И грянул гром)

В определённый момент потенциальные (непроявленные) свойства составляющих системы «включаются и образуют взаимосвязи, которые мы принципиально не можем учесть, с современной точки зрения, образуется информационное поле» и объект живёт своей жизнью. Поэтому люди, работающие со сложными системами разговаривают с ними и говорят - помогает…, но здесь, скорее всего точнее могут объяснить другие специалисты….

В общем, проблемы принятия решений не являются принадлежностью одной только области знаний (математики или психологии) ибо разделение науки на части в высшей степени условно….

Краткий определитель наук:
Если оно зеленое или дергается – это биология.
Если воняет – это химия.
Если не работает – это физика.

Если непонятно – это математика.

Если это бессмысленно – это либо экономика, либо психология.

Из «законов Мерфи»

Ф.Энгельс писал, что личность человека характеризуется не только тем, что он делает, но и тем, как он это делает. В связи с этим важным становится умение принимать оптимальные решения, особенно в нестандартных ситуациях. При этом самое интересное заключается в том, что невозможно принять оптимальное решение в предметном знании…. Поэтому возникает необходимость подходить к проблеме принятия решения с точки зрения различных наук одновременно. При этом важную роль играют знания об изучаемой системе.

В XX веке появился так называемый «принцип системного анализа», который часто имел чисто теоретический (или описательный) характер, считалось, что прикладная сторона вопроса может быть решена в рамках дисциплины «Исследование операций». Акофф писал, что унция знания стоит фунта информации (И), а унция понимания (П) стоит фунта знаний (Зн), отсюда, коэффициент важности Зн=13,3×И, П=177,7×И.

Знания человека о системе или объекте, по-видимому, могут обеспечить построение адекватной модели. А насчет понимания …– иногда это спорно, ибо каждый сталкивался с ситуацией, когда «понять нельзя, но использовать можно». Ибо в данном случае у него не априорные (индуктивные) знания о составляющих системы, но дедуктивные знания об их взаимодействии. Но как строить дедуктивные модели? С нашей точки зрения основой построения таких моделей может быть когнитивный анализ.

Известный психолог Карл Юнг, занимаясь изучением сновидений, выявил, что человек часто обладает знаниями, которые при всей их неопределённости, призрачности и нелогичности являются «сообщениями подсознания, пытающегося говорить с нами на языке образов и символов», например, во сне.

Природа - древний храм, где строй живых колонн
Обрывки смутных фраз роняет временами.
Мы входим в этот храм в смятении, а за нами
Лес символов немых следит со всех сторон.

Бодлер

Откуда у человека понимание и знание, и почему оно на уровне подсознания? Ответ возможен такой. Мозг человека был идеологической основой при построении компьютеров. И, как известно, компьютер имеет оперативную память – ОЗУ и постоянную память – ПЗУ (аналоги сознания и подсознания человека). Если оперативной памяти не хватает, то соответственно…, однако в принципе существует возможность зарезервировать часть ПЗУ как оперативную память. Для человека это означает возможность обращаться к подсознанию так же быстро, как и к сознанию – это уровень эксперта, специалиста в предметной области (средний срок достижения уровня эксперта 5 лет, заметьте – и на работу людей охотней берут при наличии такого стажа…). Его знаниями можно пользоваться, несмотря на то, что они или плохо, или совсем неформализуемы. Примерно та же самая ситуация со всевозможными компьютерными реализациями математических когнитивных теорий, которые за несколько десятков лет борьбы с классическими математиками заслужили право на жизнь – получили название «мягких вычислений» и нашли применение в системах класса Data mining. Современное научное мировоззрение простирается от детерминизма и вероятности до «мягких вычислений», а что нужно в каждой конкретной ситуации – выбирать человеку….

Тебе бы опыт сделать не мешало –
Ведь он для вас источник всех наук…

Гете

Заключение

Ещё многое имею сказать вам,
но вы теперь не можете вместить

Евангелие от Иоанна , 16:12

В настоящем пособии рассмотрено введение в ряд классических разделов теории вероятностей и математической статистики, основывающейся на предпосылках нормальности распределения изучаемых случайных величин.

Успешная практическая деятельность человека все в большей мере зависит от организации сбора и обработки информации. Совершенствование технологий записи и хранения данных – создание баз данных и знаний во всех сферах деятельности человека предъявляет новые требования к уровню подготовки специалистов. Большой объем информации, которой сопровождается деятельность практически любого предприятия и учреждения, обычно содержит полезные сведения, благодаря которым, можно значительно повысить эффективность работы, совершенствуя технологию, управление и т.д. И завтра сегодняшним студентам новые идеи анализа данных будут очевидны и необходимы на практике.

Современные коммерческие организации интенсивно внедряют информационные хранилища (банки) данных и знаний, многие из которых содержат средства интеллектуального анализа данных или предполагают возможность их применения. Деловые люди осознали, что применение методов *анализа данных,* использующих вероятностную и геометрическую природу данных, может дать ощутимое преимущество в конкурентной борьбе.

Более подробно ознакомиться с современными методами прикладной статистики *и DATA MINING* можно, например, по электронным учебникам фирмы *StatSoft*, которые в основном соответствуют описанию статистических модулей пакета *STATISTICA*. Эти направления мы предлагаем для дальнейшего самостоятельного изучения и применения на практике, с помощью прикладных пакетов статистических и *DATA MINING* программ.

Следует отметить, что на современном этапе управление производством, фирмой, районом, регионом практически невозможно без системного подхода[2], разрабатывающего методики анализа целей, методы и модели совершенствования организационной структуры, управления функционированием социально-экономических объектов.

В зависимости от априорной информации об изучаемом объекте применяют следующие методы: мозговой атаки, построения сценариев, экспертной оценки, построения дерева целей, математической логики, теории множеств, теории игр, прикладной статистики, математического программирования, интеллектуального анализа данных и т.д. Разумеется, большинство методов пересекается. При этом рассматриваемые в книге вероятностные методы в

[2] Волкова В.Н., Денисов А.А. Основы теории систем и системного анализа. -СПб.:Изд. СПбГТУ,1997.510с.:ил.

рамках системного анализа являются одним из возможных подходов перевода вербального (словесного) описания модели изучаемого объекта в формальное, для решения задач управления и принятия решений.

Сегодня современные специалисты развивают системный подход в рамках необходимости изучения кроме механических и биологических систем мультиразумных (социальных), которые характеризуются взаимозависимыми переменными. Этот подход получил название *системного мышления*.

Безусловно, не существует абсолютных рецептов построения моделей сложных процессов. Моделирование в большей степени искусство, овладеть которым можно только, решая практические задачи. Целью настоящего изложения является не сама истина, которая со временем обычно изменяется, а возможный путь её поиска, двигаясь по которому можно приобрести неоценимый опыт.

Литература

1. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Основы моделирования и первичная обработка данных. – М.: Финансы и статистика, 1985. – 472 с.
2. Айвазян С.А., Енюков И.С., Мешалкин Л.Д. Прикладная статистика. Исследование зависимостей. – М.: Финансы и статистика, 1985. – 488 с.
3. Айвазян С.А., Бухштабер В.М., Енюков И.С. и др. Прикладная статистика. Классификация и снижение размерности. – М.: Финансы и статистика, 1989. – 607 с.: ил.
4. Афифи А., Эйзен С. Статистический анализ: Подход с использованием ЭВМ. Пер. с англ. – М.: Мир, 1982. – 488 с., ил.
5. БорисовВ.В., И.А.Бычков и др. Компьютерная поддержка сложных организационно-технических систем..-М.:Горячая линия − Телеком, 2002. − 154 с.:ил.
6. Боровиков В.П., Боровиков И.П.. STATISTICA – Статистический анализ и обработка данных в среде Windows. Издание 2-е,стереотипное – М.: Информационно-издательский дом «Филинъ», 1998. – 608 с.
7. Боровиков В.П. STATISTICA. Искусство анализа данных на компьютере: Для профессионалов. 2-е изд. – СПб.: Питер, 2003. – 688с.:ил.
8. Горелова Г.В., Кацко И.А. Теория вероятностей и математическая статистика в примерах и задачах с применением EXCEL. Учеб. пособие для вузов. Издание 4-е испр. и доп. – Ростов н/Д: Феникс, 2006. – 475с.:ил.
9. Дюк В. Обработка данных на ПК в примерах – СПб.: Питер, 1997. – 240 с., ил.
10. Дюк В., Самойленко А. В. Data Mining: учебный курс. – СПб, 2001. – 368 с.:ил.

11. Дубров А.М., Мхитарян В.С., Трошин Л.И. Многомерные статистические методы: Учебник. – М.: Финансы и статистика, 1998. – 352 с.: ил.

12. Киселев М., Соломатин Е. Средства добычи знаний в бизнесе и финансах. - Открытые системы, № 4, 1997, с.41-44.

13. Орлов А.И. Современная прикладная статистика (обобщающая статья). Заводская лаборатория. 1998. т.64. №3. – 52-60 с.

14. Сошникова Л.А., Тамашевич В.Н., Уебе Г., Шеффер М. Многомерный статистичекий анализ в экономике: Учеб. пособие для вузов/Под ред.проф. В.Н. Тамашевича. – М.:ЮНИТИ-ДАНА, 1999. – 598 с.

15. Трубилин А.И., Шоль В.В. Эффективность производства высококачественного зерна озимой пшеницы в Краснодарском крае. – Краснодар, 2001. – 239с.

16. Тьюки Дж. Анализ результатов наблюдений. Разведочный анализ. – М.: Мир,1981. – 693 с.:ил.

17. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере / Под. ред. В.Э. Фигурнова. – 3-е изд., перераб. и доп. – М.: ИНФРА – М, 2003. – 544 с., ил.

18. StatSoft, Inc. (2012). Электронный учебник по статистике. Москва, StatSoft. WEB: http://www.statsoft.ru/home/textbook/default.htm.

19. SPSS: искусство обработки информации. Анализ статистических данных и восстановление скрытых закономерностей: Пер. с нем./Ахим Бююль, Петер Цёфель – СПб.: ООО «ДиаСофтЮП», 2002. – 608 с.

ОТВЕТЫ

Часть I

Контрольные задания 1.1

1. а) да; б) нет; в) да; г) нет; д) нет; е) нет.
2. а) да; б) нет; в) нет; г) да.
3. а) нет; б) да; в) нет.

Контрольные задания 1.3

1. а) 1/6; б) 1/2 в) 1/2; г) 1/3.
2. 1/90.
3. а) 1/6; б) 11/36; в) 1/18.
4. а) 0,78; б) 78%.
5. 6.
8. 13/25.
9. 0,708г.
10. $(2l)/(\pi L)$.
11. а) 1/4, б) 1/2, в) 1/3.

Контрольные задания 1.4

4. 362880
5. 450450
7. 720

Контрольные задания 1.5

1. а) 0,25; б) 0,375; в) 0,75.
2. 1/40320.
3. 1/120.
4. 0,25.
5. а) 4/9; б) 0,00033.
6. 1/120.
7. 0,25.
8. 0,26.
9. 0,5.
10. 0,079.
11. 0,38.
12. 0,05.
13. а) 0,087; б) 0,0043.
14. 0,348.
15. а) 1/376992; б) 0,0123.
16. 0,00077.
17. 0,3.
18. а) 0,6; б) 0,3; в) 0,9.

Контрольные задания 1.7

1. а) 0,5; б) 0,15.
2. 0,5.
3. а) 91/460; б) 7/46; в) 6/115.
4. 0,027.
5. а) 0,902; б) 0,098.
6. 0,6.
7. 0,763.
8. а) 0,222; б) 0,291.
9. а) 1/3; б) 8/15; в) 3/5; г) 7/15.
10. 0,271.
11. а) 0,648; б) 0,72.
12. а) 1/360; б) 1/180; в) 1/180.
13. а) 0,189; б) 0,027; в) 0,343; г) 0,216; д) 0,657.
14. а) 0,0105; б) 0,4265; в) 0,558.
15. 1/7.
16. 36/71 и 35/71.
17. 31/35.
18. Найдет.
19. 0,059.
20. а) 0,379; б) 0,621.
21. а) 0,558; б) 0,385; в) 0,615.
22. 37/64.
23. а) 0,392; б) 0,428; в) 0,904; г) 0,096.
24. а) 0,51; б) 0,94; в) 0,34.
26. а) 0,741; б) 0,241; в) 0,88(8).
27. 0,006; 0,092; 0,398; 0,504.
28. 0,288.
29. 0,4053.
30. 4.
31. а) 0,479; б) 0,333; в) 0,124.
32. а) 0,049; б) 0,267.
33. 0,72

34. 0,405
35. 0,2

Контрольные задания 1.9

1. а) 0,5075; б) 0,8276.
2. 1/6.
3. 0,676.
4. 0,643.
5. 0,27.
6. 0,1688.
7. а) 0,56; б) 0,857.
8. 0,25.
9. 0,02.
10. а) 0,79; б) 0,772.
11. 0,4.
12. а) 0,725; б) 0,276.
14. 0,0715
15. 0,2857

Контрольные задания 1.11

1. а) 0,116; б) 0,52.
2. а) 0,328; б) 0,738; в) 0,0067.
3. 0,630; 0,315; 0,052; 0,003.
4. а) 0,31; б) 0,5; в) 0,5; г) 0,625.
5. 124 или 125.
6. От 299 до 305.
7. а) Одну из двух; б) не менее двух из четырех.
8. 0,045.
9. 0,999.
10. 0,737
11. $(1/4^n)\cdot C_{2n}^n$.
12. а) 0,1488 б) 0,1869 в) 0,43

0,0238095

Контрольные задания 1.12

1. а) 0,061; б) 0,92; в) 0,264.
2. 0,09.
3. 0,387; 0,42; 0,368.
4. 0,212.
5. 15; 0,115.
6. 2;0,271.
7. 0,993.
9. а) 0,9938; б) 0,9937.
10. 0,0007.
11. 444.
12. 56; 0,119.
13. а) 0,992; б) 0,988.
14. От 19 до 21.
15. 69.
17. 0,999.
19. От 792 до 828.
20. 8100.
21. От 382 до 394.
22. 3162.
23. 0,002.

Контрольные задания 2.4

1. $M(X)$=3,6; $D(X)$=0,36; $\sigma(X)$=0,6.
2. $M(X)$=2,06; $D(X)$=0,999; $\sigma(X)$=1,0.
3. $M(X)$=1,2; $D(X)$=0,72; $\sigma(X)$=0,85.
4. $M(X)$=1,8.
7. x_0=1.
8. x_0=1.
9. а)$M(X)$=3,1216;б) $M(X)$=4,091.
10. $M(X)$=2,4264.
11. $M(X)$=39,221.
12. $M(X)$=2,4.
13. $M(X)$=1,87.
15. Третьего; второго.
16. $M(X)$=42; $D(X)$=35; $\sigma(X)$=5,92.
17. а) 4; б)14; в) 20; г) 35.
18. а) 8; б)8; в) 72; г) 32.
19. а) 14 и 88; б)-2 и 112; в) 30 и 186.
20. а) 34 и 96; б)15 и 161; в) 13 и 49.
21. а) 0,6768; б)0,8647.
22. $M(Z)$=7,4; $D(Z)$=2,2; $\sigma(Z)$=1,48; $M(V)$=13,68; $D(V)$=29,3376; $\sigma(V)$=5,4164.
23. $M(X)$=4,2; $D(X)$=0,64; $\sigma(X)$=0,8.

25. 34,1;150,19; 12,26.

26. x_2=2,6.

27. $p(x=2)=0,1$; $p(x=3)=0,2$.

28. $M(X)=2$; σ(x)=1,4.

29. p=0,1.

30. $M(X)=7$.

31. $x_1=4$; $p_1=0,4$; $x_2=5$; $p_2=0,6$.

32. $x_1=0$; $x_2=1$;$p_2=0,15$.

33. $x_2=2$; $x_3=5$;$p_2=0,3$.

34. $p_2=0,3$; $p_3=0,4$; $p_4=0,2$.

36. а) 0,857375;0,1355375;0,007125;0,000125; в) 0,00725

37. 0,0238095 $M(X)$=100\$, $D(X)$=16667\$

Контрольные задания 2.7

6. $M(X)=2,1$.

7. а) 0,5; б) 0,75; в) 0,25;г) 0,5.

8 0,29663.

9. б) $M(X)=2\frac{1}{6}$; $D(X)=\frac{11}{36}$; $\sigma(X)=\frac{\sqrt{11}}{6}\approx 0,553$; в) $\frac{3}{8}$.

10. а)0,5а; б) $M(X)=2-2a$; $D(X)=\frac{4}{3}$; $\sigma(X)=\frac{2\sqrt{3}}{3}$.

11. $A=0$; $B=2$; $M(X)=1,5$; $D(X)=0,15$; $\sigma(X)=0,387$.

12. б) 0,847.

13. б) 0,599; в) $M(X)=2,566$; $D(X)=0,079$; $\sigma(X)=0,28$.

14. б) 7/36.

15. б) 0,221.

16. б) 0,1875.

17. б) 3/16; в) $M(X)=0,8\sqrt{a}$; $D(X)=\frac{2}{75}a$; $\sigma(X)=\frac{\sqrt{6}}{15}a$.

18. б) 0,6321 ; в) $M(X)=1/3$; $D(X)=1/9$; $\sigma(X)=1/3$.

19. а) $a=0,25$; б) $M(X)=2\frac{1}{3}$; в) 0,25.

20. $M(X)=0$; $D(X)=\frac{\pi^2}{4}-2$.

21. а) $c=48$; б) $M(X)=199/64$; $D(X)=0,463$.

22. $c=1/2\pi$.

Контрольные задания 2.8

2. а) $M(X)=8$; $D(X)=3$; б) $M(X)=1$; $D(X)=5\frac{1}{3}$.

3. $M(X)=a$; $D(X)=5\frac{1}{3}$; $\sigma(X)=\frac{4\sqrt{3}}{3}$.

4. в) $\frac{5}{16}$; г) $M(X)=\frac{a}{3}$; $D(X)=\frac{a^2}{18}$; $\sigma(X)=\frac{a\sqrt{2}}{6}$.

5. в) 0,75; г) $M(X)=0$; $D(X)=\frac{a^2}{6}$; $\sigma(X)=\frac{a\sqrt{6}}{6}$.

6. а) 0,8413; б) 0,9544.

7. а) 0,7258; б) 0,9995; в) 0,9082; г) 0,8164.

8. а) 0,8413; б) 0,5328.

9. 0,9997.

10. (56,1; 63,9).

11. 10,0.

12. (240; 360).

13. а) 0,5328; б) $M_0=5$; $M_e=5$.

14. (-2; 2).

15. 0,8185.

17. а) 0,865; б) 0,018.

19. а) 0,5465; б) 0,4036; в) 0,9502; г) 0,0498.

21. а) 0,134; б) 0,9826; в) 0,9975.

22. а) 0,134; б) 0,9379.

23. а) $\frac{1}{\pi} arctgx + \frac{1}{2}$; б) $\frac{2}{3}$.

25. а) $x_0=1{,}25^{\frac{1}{\beta}} a$;

в) $M(X)=\frac{\beta \cdot a}{\beta - 1}$ при $\beta > 1$.

24. б) 0,4712.

в) $M(X)=a \cdot \sqrt{\frac{\pi}{2}}$; $D(X)= a^2 \cdot (2-\frac{\pi}{2})$;

$\sigma(X)=a \cdot \sqrt{2-\frac{\pi}{2}}$.

26. 13158,6

27. 5,83 и 2,41

28. 13158,6

Контрольные задания 2.10.

2. $M(X)$=7,5; $D(X)$=56,25; $M(Y)$=5,5; $D(Y)$=24,75.

3. $f(x,y)=\begin{cases} \ln^2 3 \cdot 3^{-x-y} & при\ x \ge 0,\ y \ge 0 \\ 0 & при\ x < 0,\ y < 0 \end{cases}$.

4. 0,2926.

5. $\frac{3}{128}$

6. $f(x)=8 \cdot e^{-4x-2y}$ $при\ x \ge 0,\ y \ge 0$;

$f(x)=0$ при $x < 0,\ y < 0$.

12. а) $f(x,y)=\frac{1}{18}$ внутри треугольника,

$f(x,y)$=0 вне треугольника;

б) $f_1(x)=\frac{1}{3}-\frac{1}{18}x$, $0 < x < 6$; $f_2(y)= \frac{1}{3}-\frac{1}{18}y$, $0 < y < 6$; $f(x/y)= \frac{1}{6-y}$, $0 < y < 6$; $f(y/x)=\frac{1}{6-x}$, $0 < x < 6$

.

9. a=0,5; $M(X)=M(y)=\frac{\pi}{4}$;

$D(X)=D(Y)=(\pi^2+8\pi-32)/16$.

13. а) $f(x,y)=\frac{1}{a^2}$ внутри квадрата;

$f(x,y)$=0 вне квадрата;

б) $f(x)=\begin{cases} \frac{a\sqrt{2}-2|x|}{a^2} & при\ |x| \le \frac{a\sqrt{2}}{2}, \\ 0 & при |x| > \frac{a\sqrt{2}}{2}; \end{cases}$

$f(y)=\begin{cases} \frac{a\sqrt{2}-2|y|}{a^2} & при\ |y| \le \frac{a\sqrt{2}}{2}, \\ 0 & при |y| > \frac{a\sqrt{2}}{2}; \end{cases}$

$f(x/y)=\begin{cases} \frac{1}{a\sqrt{2}-2|y|} & при\ |x \pm y| \le \frac{a\sqrt{2}}{2}, \\ 0 & при |x \pm y| > \frac{a\sqrt{2}}{2}; \end{cases}$

$f(y/x)=\begin{cases} \frac{1}{a\sqrt{2}-2|x|} & при\ |x \pm y| \le \frac{a\sqrt{2}}{2}, \\ 0 & при |x \pm y| > \frac{a\sqrt{2}}{2}. \end{cases}$

14. $M(X)=M(Y)=\frac{\sqrt{3\pi}}{6}$; $D(X)=D(Y)=$

15. $M(X)=M(Y)=\frac{a}{3}$; $D(X)=D(Y)=\frac{a^2}{18}$;

$\frac{4-\pi}{12}$.

$$K(X,Y)=-\frac{a^2}{36}.$$

16. а) $f(x,y)=\begin{cases}0 & \text{при } x\le 0,\ y\le 0,\\ 10e^{-(2x+5y)} & \text{при } x>0,\ y>0;\end{cases}$

б) $F(x,y)=\begin{cases}0 & \text{при } x\le 0,\ y\le 0,\\ (1-e^{-2x})\cdot(1-e^{-5y}) & \text{при } x>0,\ y>0.\end{cases}$

Контрольные задания 2.13.

3. а) $M(Y)=10$; $D(Y)=36$; $\sigma(y)=6$;
б) $M(Y)=14{,}5$; $D(Y)=95{,}85$; $\sigma(Y)=9{,}8$.

5. а) $g(y)=\frac{1}{\pi\sqrt{1-y^2}}$ при $y\in(-1;1)$;
б) $g(y)=\frac{2}{\pi\sqrt{1-y^2}}$ при $y\in(0;1)$.

6. а) $g(y)=\frac{1}{2\sqrt{2\pi}}e^{-\frac{(y-10)^2}{8}}$,

б) $g(y)=\frac{1}{3\sqrt{2\pi}\cdot\sqrt[3]{y^2}}e^{\frac{-(\sqrt[3]{y}-2)^2}{2}}$.

7. $g(S)=-\ln S$, где $0<S<1$.

8. а) $g(y)=\frac{3}{\pi(9+y^2)}$

б) $g(y)=\frac{1}{3\pi(y^{\frac{2}{3}}+y^{\frac{4}{3}})}$.

9. $$g(z)=\begin{cases}0 & \text{при } z<0,\\ \frac{z}{20} & \text{при } 0<z\le 2,\\ \frac{1}{10} & \text{при } 2<z\le 10,\\ \frac{12-z}{20} & \text{при } 10<z\le 12,\\ 0 & \text{при } z>12.\end{cases}$$

10. $$g(z)=\begin{cases}0 & \text{при } z\le -3,\\ \frac{z+3}{25} & \text{при } -3<z\le 2,\\ \frac{7-z}{25} & \text{при } 2<z\le 7,\\ 0 & \text{при } z>7.\end{cases}$$

11. $$g(z)=\begin{cases}0 & \text{при } z\le 0,\\ e^{-\frac{z}{2}}(1-e^{-\frac{z}{2}}) & \text{при } z>0\end{cases}$$

12. $M(Z)=0$, $D(Z)=2$,

$$g(z)=\frac{1}{\sqrt{2\pi}\sqrt{2}}\cdot e^{-\frac{Z^2}{4}}.$$

13. $$g(x)=\frac{1}{x\sigma\sqrt{2\pi}}\cdot e^{-\frac{(\ln x-a)^2}{2\sigma^2}}$$
при $x>0$.

Контрольные задания 2.15.

1. $P(x>10)\le\frac{6}{10}$.
2. а) $P(x>8)\le\frac{1}{2}$, б) $P(x\le 6)\ge\frac{1}{3}$.
3. $P\ge 0{,}9872$.
4. $P_1\ge 0{,}936$, $P_2=0{,}99996$.
5. $P\ge 0{,}64$.
6. $P\le\frac{2}{3}$.
7. $P\ge 0{,}5456$.
8. а) $P\ge 0{,}98$, б) $P\ge 0{,}955$.
9. а) $P\ge\frac{7}{9}$, б) $\frac{35}{36}$.
10. а) $P\ge\frac{7}{9}$, б) $\frac{35}{36}$.
11. 400.
12. 3125.
13. $P>0{,}96$
14. $P\ge 0{,}92$

Контрольные задания 2.16.

2. $0\le P\le 1,\quad 0\le S\le 1$.
3. $Pi,\quad i+1=\frac{5-i}{5};\quad Pi,\quad i-1=\frac{i}{5}$,

   $Pij=0$, для остальных i,j;

   $Pij(2)=\frac{5+10i-2i^2}{25}$, $P_{i,\,i-2}(2)=\frac{i(i-1)}{25}$,

   $P_{i,i+2}(2)=\frac{(5-i)(4-i)}{25}$,

   остальные $Pij=0$.
4. 1/6.
5. б) 0; 0;

   $(\frac{1}{2})^2\cdot(\frac{1}{2})^2=\frac{1}{108}$; 0;

   $(\frac{1}{2})^3\cdot(\frac{1}{3})^2=\frac{1}{72}$.

Часть II

Контрольные задания 3.2.

1. $\overline{X}=3{,}49;\quad M_0=4;\quad M_e=4;$
   $\sigma^*=1{,}418;$
2. $\overline{X}=4{,}24;\quad M_0=4;\quad M_e=4;$
   $\sigma^*=1{,}436.$
4. $\overline{X}=11{,}167;\quad \sigma^*=9{,}728;\ V^*=87{,}1\%$
5. $\overline{X}=7{,}323;\quad \sigma^*=1{,}750;\ V^*=23{,}9\%$
   $Sk^*=-0{,}817;\quad Ex^*=0{,}868.$

Контрольные задания 3.3.

1. (1,702; 1,898); (1089;1215).
2. а) (0,474; 0,526); б) (0,47; 0,53).
3. (142,75; 157,25); (107062,1; 117937,9).
4. а) 3,49; 2,030; 1,4;
   б) (3,24; 3,74); в) 0,8294; г) 180.
5. а) 4,24; 2,083; 1,443;
   б) (4,027; 4,453); в) 250.
7. (243,3; 256,7), (34,5; 45,5), 651.
8. а) 16%; 6,5%; 2,55%;
   б) (15; 17).
9. 385.
10. (983,83; 1170,01); (0,373; 0,535).
11. (0,662; 0,799).
14. 0,9503
15. (2.34; 2,86)

16. 0,9319

Контрольные задания 4.4.

1. y=1,178 x – 7,139;
   r = 0,912; r^2 = 0,832.
2. y=0,339 x + 5;
   r = 0,886; r^2 = 0,785.
3. y= 0,446x+3,357;
   r = 0,932.

Значения функций

$$\varphi(x)=\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{1}{2}x^2} \text{ и } \Phi(x)=\frac{1}{\sqrt{2\pi}}\int_0^x e^{-\frac{1}{2}t^2}dt.$$

x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$
0,00	0,3989	0,0000	0,40	0,3683	0,1554	0,80	0,2897	0,2881
01	3989	0040	41	3668	1591	81	2874	2910
02	3989	0080	42	3652	1628	82	2850	2939
03	3988	0120	43	3637	1664	83	2827	2967
04	3986	0160	44	3621	1700	84	2803	2995
05	3984	0199	45	3605	1736	85	2780	3023
06	3982	0239	46	3589	1772	86	2756	3051
07	3980	0279	47	3572	1808	87	2732	3078
08	3977	0319	48	3555	1844	88	2709	3106
09	3973	0359	49	3538	1879	89	2685	3133
0,10	0,3970	0,0398	0,50	0,3521	0,1915	0,90	0,2661	0,3159
11	3965	0438	51	3503	1950	91	2637	3186
12	3961	0478	52	3485	1985	92	2613	3212
13	3956	0517	53	3467	2019	93	2589	3238
14	3951	0557	54	3448	2054	94	2565	3264
15	3945	0596	55	3429	2088	95	2541	3289
16	3939	0636	56	3410	2123	96	2516	3315
17	3932	0675	57	3391	2157	97	2492	3340
18	3925	0714	58	3372	2190	98	2468	3365
19	3918	0753	59	3352	2224	99	2444	3389
0,20	0,3910	0,0793	0,60	0,3332	0,2257	1,00	0,2420	0,3413
21	3902	0832	61	3312	2291	01	2396	3438
22	3894	0871	62	3292	2324	02	2371	3461
23	3885	0910	63	3271	2357	03	2347	3485
24	3876	0948	64	3251	2389	04	2323	3508
25	3867	0987	65	3230	2422	05	2299	3531
26	3857	1026	66	3209	2454	06	2275	3554
27	3847	1064	67	3187	2486	07	2251	3577
28	3836	1103	68	3166	2517	08	2227	3599
29	3825	1141	69	3144	2549	09	2203	3621
0,30	0,3814	0,1179	0,70	0,3123	0,2580	1,10	0,2179	0,3643
31	3802	1217	71	3101	2611	11	2155	3665
32	3790	1255	72	3079	2642	12	2131	3686
33	3778	1293	73	3056	2673	13	2107	3708
34	3765	1331	74	3034	2703	14	2083	3729
35	3752	1368	75	3011	2734	15	2059	3749
36	3739	1406	76	2989	2764	16	2036	3770
37	3726	1443	77	2966	2794	17	2012	3790
38	3712	1480	78	2943	2823	18	1989	3810
39	3697	1517	79	2920	2852	19	1965	3830

Продолжение приложения 1

x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$	x	$\varphi(x)$	$\Phi(x)$
1,20	0,1942	0,3849	1,70	0,0940	0,4554	2,40	0,0224	0,4918
21	1919	3869	71	0925	4564	42	0213	4922
22	1895	3883	72	0909	4573	44	0203	4927
23	1872	3907	73	0893	4582	46	0194	4931
24	1849	3925	74	0878	4591	48	0184	4934
25	1826	3944	75	0863	4599	50	0175	4938
26	1804	3962	76	0848	4608	52	0167	4941
27	1781	3980	77	0833	4616	54	0158	4945
28	1758	3997	78	0818	4625	56	0151	4948
29	1736	4015	79	0804	4633	58	0143	4951
1,30	0,1714	0,4032	1,80	0,0790	0,4641	2,60	0,0136	0,4953
31	1691	4049	81	0775	4649	62	0129	4956
32	1669	4066	82	0761	4656	64	0122	4959
33	1647	4082	83	0748	4664	66	0116	4961
34	1626	4099	84	0734	4671	68	0110	4963
35	1604	4115	85	0721	4678	70	0104	4965
36	1582	4131	86	0707	4686	72	0099	4967
37	1561	4147	87	0694	4693	74	0093	4969
38	1539	4162	88	0681	4699	76	0088	4971
39	1518	4177	89	0669	4706	78	0084	4973
1,40	0,1497	0,4192	1,90	0,0656	0,4713	2,80	0,0079	0,4974
41	1476	4207	91	0644	4719	82	0075	4976
42	1456	4222	92	0632	4726	84	0071	4977
43	1435	4236	93	0620	4732	86	0067	4979
44	1415	4251	94	0608	4738	88	0063	4980
45	1394	4265	95	0596	4744	90	0060	4981
46	1374	4279	96	0584	4750	92	0056	4982
47	1354	4292	97	0573	4756	94	0053	4984
48	1334	4306	98	0562	4761	96	0050	4985
49	1315	4319	99	0551	4767	98	0047	4986
1,50	0,1295	0,4332	2,00	0,0540	0,4772	3,00	0,00443	0,49865
51	1276	4345	02	0519	4783			
52	1257	4357	04	0498	4793	3,10	00327	49903
53	1238	4370	06	0478	4803	3,20	00238	49931
54	1219	4382	08	0459	4812			
55	1200	4394	10	0440	4821	3,30	00172	49952
56	1182	4406	12	0422	4830	3,40	00123	49966
57	1163	4418	14	0404	4838			
58	1145	4429	16	0387	4846	3,50	00087	49977
59	1127	4441	18	0371	4854			
1,60	0,1109	0,4452	2,20	0,0355	0,4861	3,60	00061	499841
61	1092	4463	22	0339	4868	3,70	00042	49989
62	1074	4474	24	0325	4875	3,80	00029	499928
63	1057	4484	26	0310	4881			
64	1040	4495	28	0297	4887	3,90	00020	49995
65	1023	4505	30	0283	4893	4,00	0,0001338	499968
66	1006	4515	32	0270	4898			
67	0989	4525	34	0258	4904	4,50	0000160	499997
68	0973	4535	36	0246	4909	5,00	0000015	49999997
69	0957	4545	38	0235	4913			

Приложение 2

Критические точки χ^2 -распределения Пирсона

ν \ α	0,20	0,10	0,05	0,02	0,01	0,001
1	1,642	2,706	3,841	5,412	6,635	10,827
2	3,219	4,605	5,991	7,824	9,210	13,815
3	4,642	6,251	7,815	9,837	11,345	16,266
4	5,989	7,779	9,488	11,668	13,277	18,467
5	7,289	9,236	11,070	13,388	15,086	20,515
6	8,558	10,645	12,592	15,033	16,812	22,457
7	9,803	12,017	14,067	16,622	18,475	24,322
8	11,030	13,362	15,507	18,168	20,090	26,125
9	12,242	14,684	16,919	19.679	21,666	27,877
10	13,442	15,987	18,307	21,161	23,209	29,588
11	14,631	17,275	19,675	22,618	24,725	31,264
12	15,812	18,549	21,026	24,054	26,217	32,909
13	16,985	19,812	22,362	25,472	27,688	34,528
14	18,151	21,064	23,685	26,783	29,141	36,123
15	19,311	22,307	24,996	28,259	30,578	37,697
16	20,465	23,542	26,296	29,633	32,000	39,252
17	21,615	24,769	27,587	30,995	33,409	40,790
18	22,760	25,989	28,869	32,346	34,805	42,312
19	23,900	27,204	30,144	33,687	36,191	43,820
20	25,038	28,412	31,410	35,020	37,566	45.315
21	26,171	29,615	32,671	36,343	38,932	46,797
22	27,301	30,813	33,924	37,659	40,289	48,268
23	28,429	32,007	35,172	38,968	41,638	49,728
24	29,553	33,196	36,415	40,270	42,980	51,179
25	30,675	34,382	37,652	41,566	44,314	52,620
26	31,795	35,563	38,885	42,856	45,642	54,052
27	32,912	36,741	40,113	44,140	46,963	55,476
28	34,027	37,916	41,337	45,419	48,278	56,893
29	35,139	39,087	42,557	46,693	49,588	58,302
30	36,250	40,256	43,773	47,962	50,892	59,703

Критические точки *t* - распределения Стьюдента

Число степеней свободы *k*	Уровень значимости α (двусторонняя критическая область)					
	0,10	0,05	0,02	0,01	0,002	0,001
1	6,31	12,71	31,82	63,66	318,31	636,62
2	2,92	4,30	6,97	9,93	22,33	31,60
3	2,35	3,18	4,54	5,84	10,22	12,92
4	2,13	2,78	3,75	4,60	7,17	8,61
5	2,01	2,57	3,37	4,03	5,89	6,87
6	1,94	2,45	3,14	3,71	5,21	5,96
7	1,90	2,37	3,00	3,50	4,79	5,41
8	1,86	2,31	2,90	3,36	4,50	5,04
9	1,83	2,26	2,82	3,25	4,30	4,78
10	1,81	2,23	2,76	3,17	4,14	4,59
11	1,80	2,20	2,72	3,11	4,03	4,44
12	1,78	2,18	2,68	3,06	3,93	4,32
13	1,77	2,16	2,65	3,01	3,85	4,22
14	1,76	2,15	2,62	2,98	3,79	4,14
15	1,75	2,13	2,60	2,95	3,73	4,07
16	1,75	2,12	2,58	2,92	3,69	4,02
17	1,74	2,11	2,57	2,90	3,65	3,97
18	1,73	2,10	2,55	2,88	3,61	3,92
19	1,73	2,09	2,54	2,86	3,58	3,88
20	1,73	2,09	2,53	2,85	3,55	3,85
21	1,72	2,08	2,52	2,83	3,53	3,82
22	1,72	2,07	2,51	2,82	3,51	3,79
23	1,71	2,07	2,50	2,81	3,49	3,77
24	1,71	2,06	2,49	2,80	3,47	3,75
25	1,71	2,06	2,49	2,79	3,45	3,73
26	1,71	2,06	2,48	2,78	3,44	3,71
27	1,70	2,05	2,47	2,77	3,42	3,69
28	1,70	2,05	2,47	2,76	3,41	3,67
29	1,70	2,05	2,46	2,76	3,40	3,66
30	1,70	2,04	2,46	2,75	3,39	3,65
40	1,68	2,02	2,42	2,70	3,31	3,55
60	1,67	2,00	2,39	2,66	3,23	3,46
120	1,66	1,98	2,36	2,62	3,16	3,37
∞	1,65	1,96	2,33	2,58	3,09	3,29
	0,05	0,025	0,01	0,005	0,001	0,0005
	Уровень значимости α (односторонняя критическая область)					

Критические точки распределения F Фишера – Снедекора

(ν_1 – число степеней свободы большей дисперсии, ν_2 – число степеней свободы меньшей дисперсии), уровень значимости α=0,05

$\nu_2 \backslash \nu_1$	1	2	3	4	5	6	7	8	9	10	12	16	20	24	30	50	100	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	252	253	254
2	18,51	19,00	19,16	19,25	19,30	19,33	19,36	19,37	19,38	19,39	19,41	19,43	19,44	19,45	19,46	19,47	19,49	19,50
3	10,13	9,55	9,28	9,12	9,01	8,94	8,88	8,84	8,81	8,78	8,74	8,69	8,66	8,64	8,62	8,58	8,56	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,84	5,80	5,77	5,74	5,70	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,78	4,74	4,68	4,60	4,56	4,53	4,50	4,44	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,92	3,87	3,84	3,81	3,75	3,71	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,63	3,57	3,49	3,44	3,41	3,38	3,32	3,28	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,34	3,28	3,20	3,15	3,12	3,08	3,03	2,98	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,13	3,07	2,98	2,93	2,90	2,86	2,80	2,76	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,97	2,91	2,82	2,77	2,74	2,70	2,64	2,59	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,86	2,79	2,70	2,65	2,61	2,57	2,50	2,45	2,40
12	4,75	3,88	3,49	3,26	3,11	3,00	2,92	2,85	2,80	2,76	2,69	2,60	2,54	2,50	2,46	2,40	2,35	2,30
13	4,67	3,80	3,41	3,18	3,02	2,92	2,84	2,77	2,72	2,67	2,60	2,51	2,46	2,42	2,38	2,32	2,26	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,77	2,70	2,65	2,60	2,53	2,44	2,39	2,35	2,31	2,24	2,19	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,70	2,64	2,59	2,55	2,48	2,39	2,33	2,29	2,25	2,18	2,12	2,67
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,33	2,28	2,24	2,20	2,13	2,07	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,62	2,55	2,50	2,45	2,38	2,29	2,23	2,19	2,15	2,08	2,02	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,25	2,19	2,15	2,11	2,04	1,98	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,55	2,48	2,43	2,38	2,31	2,21	2,15	2.11	2,07	2,00	1,94	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,52	2,45	2,40	2,35	2,28	2,18	2,12	2,08	2,04	1,96	1,90	1,84
22	4,30	3,44	3,05	2,82	2,66	2,55	2,47	2,40	2,35	2,30	2,23	2,13	2,07	2,03	1,98	1,91	1,84	1,78
24	4,26	3,40	3,01	2,78	2,62	2,51	2,43	2,36	2,30	2,26	2,18	2,09	2,02	1,98	1,94	1,86	1,80	1,73
26	4,22	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,05	1,99	1,95	1,90	1,82	1,76	1,69
28	4,20	3,34	2,95	2,71	2,56	2,44	2,36	2,29	2,24	2,19	2,12	2,02	1,96	1,91	1,87	1,78	1,72	1,65
32	4,15	3,30	2,90	2,67	2,51	2,40	2,32	2,25	2,19	2,14	2,07	1,97	1,91	1,86	1,82	1,74	1,67	1,59
36	4,11	3,26	2,86	2.63	2,48	2,36	2,28	2,21	2,15	2,10	2,03	1,93	1,87	1,82	1,78	1,69	1,62	1,55
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,07	2,00	1,90	1,84	1,79	1,74	1,66	1,59	1,51
60	4,00	3,15	2.76	2,52	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,81	1,75	1,70	1,65	1,56	1,48	1,39
100	3,94	3,09	2,70	2,46	2,30	2,19	2,10	2,03	1,97	1,92	1,85	1,75	1,68	1,63	1,57	1,48	1,39	1,28
200	3,89	3,04	2,65	2,41	2.26	2,14	2,05	1,98	1,92	1,87	1,80	1,69	1,62	1,57	1,52	1,42	1,32	1,19
∞	3,84	2,99	2,60	2,37	2,21	2,09	2,01	1,94	1,88	1,83	1,75	1,64	1,57	1,52	1,46	1,35	1,24	1,00

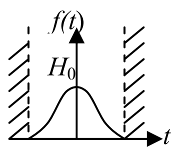
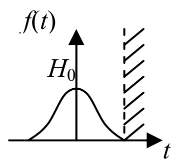
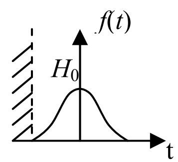
Приложение 5

Сравнение дисперсий

<table>
<tr><td colspan="3">1 - выборка</td><td colspan="2">2 - выборки</td><td colspan="2">K - выборок</td></tr>
<tr><td colspan="3">H_0:$\sigma^2=\sigma_0^2$ (σ_0^2 генеральная дисперсия)
H_1: $\sigma^2<\sigma_0^2$</td><td colspan="2">H_0: $\sigma_1^2=\sigma_2^2$
H_1: $\sigma_1^2>\sigma_2^2$</td><td>H_0: $\sigma_1^2=\sigma_2^2=\ldots=\sigma_k^2$
H_1: $\sigma_1^2>\sigma_2^2>\ldots>\sigma_k^2$</td><td>$H_0$: $\sigma_1^2=\sigma_2^2=\ldots=\sigma_k^2$
H_1: $\sigma_1^2>\sigma_2^2>\ldots>\sigma_k^2$</td></tr>
<tr><td colspan="3">Дано: n, S^2, σ_0^2</td><td colspan="2">Дано: n_1, n_2, S_1^2, S_2^2</td><td>Дано: $n_1=n_2=\ldots=n_k$</td><td>Дано: $n_1\neq n_2\neq\ldots\neq n_k$,
$n_i\geq 4$</td></tr>
<tr><td colspan="3">$K_{расч.}=\chi^2_{расч.}=\dfrac{(n-1)S^2}{\sigma_0^2}, \nu=n-1.$
а) $n<30$, $\chi^2_{кр.}$ при α, ν - по таблице распределения Пирсона;
б) $n>30$, $\chi^2_{кр.}$ при α, ν: $\chi_{кр.}^2=\nu\left(1-\dfrac{2}{9\nu}+u_\alpha\sqrt{\dfrac{2}{9\nu}}\right)^3$.
u_α-по таблице распределения Лапласа</td><td colspan="2">$K_{расч.}=F_{расч.}=\dfrac{S_1^2}{S_2^2}$,
при ν_1, ν_2.
$S_1^2>S_2^2$,
$\nu_1=n_1-1$,
$\nu_2=n_2-1$</td><td>$K_{расч.}=\dfrac{S^2\max}{\sum_{i=1}^{k}S_i^2}$,
$S^2max=max\{S_i^2\}$</td><td rowspan="3">$K_{расч.}=B_{расч.}=\dfrac{V}{C}$
$V=2{,}303[\lg S^2-\sum_{i=1}^{k}\nu_i\lg S_i^2]$,
$C=1+\dfrac{1}{3(k-1)}[\sum\dfrac{1}{\nu_i}-\dfrac{1}{\nu}]$,
где $\nu_i=n_i-1$, $\nu=\sum\nu_i$,
$S_k^2=\dfrac{\sum\nu_i S_i^2}{\nu}$,
$\sum=\sum_{i=1}^{k}$</td></tr>
<tr><td>Двусторонний критерий</td><td colspan="2">Односторонний критерий</td><td>Двусторонний критерий</td><td>Односторонний критерий</td><td rowspan="2">Односторонний критерий</td></tr>
<tr><td>H_1: $\sigma^2\neq\sigma_0^2$</td><td>H_1: $\sigma^2>\sigma_0^2$</td><td>H_1: $\sigma^2<\sigma_0^2$</td><td>H_1: $\sigma_1^2\neq\sigma_2^2$</td><td>H_1: $\sigma_1^2>\sigma_2^2$</td></tr>
<tr><td rowspan="2">$\chi^2_{кр.\,лев.}\leq\chi^2_{расч.}\leq\chi^2_{кр.прав.}$ Тогда H_0 принимается.
$\chi^2_{кр.\,лев.}$ при $(1-\alpha/2)$,ν
$\chi^2_{кр.\,прав.}$ при $\alpha/2$,ν
$\Phi(u_{\alpha/2})=(1-\alpha)/2$</td><td>$\chi^2_{расч.}<\chi^2_{кр.прав.\,\alpha,\,\nu}$</td><td>$\chi^2_{расч.}>\chi^2_{кр.лев.\,1-\alpha,\nu}$</td><td>$F_{расч.}<F_{кр.}$
при ν_1, ν_2, $\dfrac{\alpha}{2}$</td><td>$F_{расч.}<F_{кр.прав.}$
при ν_1, ν_2, α</td><td rowspan="2">По критерию Кокрена. $K_{кр.}$, α, ν, k.
H_0 - принимается, если $K_{расч.}<K_{кр.\alpha,\nu,k}$. Если H_0, принята то в качестве оценки дисперсии генеральной совокупности принимают
$S^2=\dfrac{\sum S_i^2}{k}$</td><td>Односторонний критерий</td></tr>
<tr><td colspan="2">$\Phi(u_\alpha)=(1-2\alpha)/2$</td><td colspan="2">$F_{кр.}$ - определяется по таблице распределения Фишера - Снедекора</td><td>H_0 принимается, если $B_{расч.}<\chi^2_{кр.прав.}$при α,ν.
По таблице распределения Пирсона</td></tr>
</table>

Приложение 6

1 Выборка. Сравнение выборочной средней с генеральной средней (гипотетической) H_0: $\overline{X} = \overline{X}_0$ или H_0: $\overline{X} - \overline{X}_0 = 0$

<table>
<tr><td></td><td colspan="3">Большой объем выборки, n</td><td colspan="3">Малый объем выборки, n</td></tr>
<tr><td rowspan="6">Генеральная дисперсия известна $D[X]=\sigma^2$</td><td colspan="3">Задача № 1
$K_{расч.} = \dfrac{(\overline{X} - \overline{X}_0)\cdot\sqrt{n}}{\sigma}$, H_0: $\overline{X} = \overline{X}_0$; H_0 принимается если:</td><td colspan="3" rowspan="6"></td></tr>
<tr><td>Двусторонний критерий</td><td colspan="2">Односторонний критерий</td></tr>
<tr><td>H_1: $\overline{X} \neq \overline{X}_0$</td><td>$H_1$: $\overline{X} > \overline{X}_0$</td><td>$H_1$: $\overline{X} < \overline{X}_0$</td></tr>
<tr><td>$\lvert K_{расч.}\rvert < K_{кр.\alpha/2}$</td><td>$K_{расч.} < K_{кр.\alpha}$</td><td>$K_{расч.} > -K_{кр.\alpha}$</td></tr>
<tr><td>$\Phi(K_{кр.,\alpha/2}) = (1-\alpha)/2$</td><td colspan="2">$\Phi(K_{кр.,\alpha}) = (1-2\alpha)/2$</td></tr>
<tr><td colspan="3">$K_{кр.}$- по таблице значений интеграла Лапласа, Ф(x)</td></tr>
<tr><td rowspan="6">Генеральная дисперсия не известна</td><td colspan="3" rowspan="6"></td><td colspan="3">Задача № 2
$K_{расч.} = t_{расч.} = \dfrac{(\overline{X} - \overline{X}_0)\cdot\sqrt{n}}{S}$
$S^2 = \dfrac{\sum X_j^2 \cdot n_j - \dfrac{(\sum X_j n_j)^2}{n}}{n-1} = \dfrac{\sum (\overline{X}_j - \overline{X})^2}{n-1}$
H_0 принимается, если:</td></tr>
<tr><td>Двусторонний критерий</td><td colspan="2">Односторонний критерий</td></tr>
<tr><td>H_1 : $\overline{X} \neq \overline{X}_0$</td><td>$H_1$: $\overline{X} > \overline{X}_0$</td><td>$H_1$: $\overline{X} < \overline{X}_0$</td></tr>
<tr><td>$\lvert t_{расч.}\rvert < t_{кр.\ \nu.\ \alpha/2}$</td><td>$t_{расч} < t_{кр.\ \nu.\alpha}$</td><td>$t_{расч.} > -\ t_{кр.\nu.\alpha}$</td></tr>
<tr><td>
$-t_{кр.} < t_{расч.} < +t_{кр.}$</td><td>
$t_{расч.} < +t_{кр.}$</td><td>
$t_{расч.} > -t_{кр.}$</td></tr>
<tr><td colspan="3">$t_{кр}$ - по таблице распределения Стьюдента</td></tr>
</table>

2 выборки. Сравнение двух средних 2 генеральных совокупностей. H_0: $M[X_1]=M[X_2]$ или H_0: $M[X_1]-M[X_2]=0$

<table>
<tr><td rowspan="2"></td><td colspan="3">2 выборки большого объема</td><td colspan="6">2 выборки малого объема</td></tr>
<tr><td colspan="3">Независимые n_1 и n_2</td><td colspan="3">Независимые n_1 и n_2</td><td colspan="3">Зависимые: $n_1=n_2=n$</td></tr>
<tr><td rowspan="6">Генеральные дисперсии известны D[X1] и D[X2]</td><td colspan="3">Задача №3
$K_{расч.} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\left(\frac{D[X_1]}{n_1}+\frac{D[X_2]}{n_2}\right)}}$
H_0: $M[X_1]=M[X_2]$, H_0 принимается, если:</td><td colspan="3" rowspan="6"></td><td colspan="3" rowspan="6"></td></tr>
<tr><td>Двустор. критерий</td><td colspan="2">Односторонний критерий</td></tr>
<tr><td>H_1:$M[X_1]\neq M[X_2]$</td><td>H_1:$M[X_1]>M[X_2]$</td><td>H_1:$M[X_1]<M[X_2]$</td></tr>
<tr><td>$|K_{расч.}|<K_{кр.\alpha/2}$</td><td>$K_{расч.}<K_{кр.\alpha}$</td><td>$K_{расч.} > -K_{кр.\alpha}$</td></tr>
<tr><td>$\Phi(K_{кр.,\alpha/2})=(1-\alpha)/2$</td><td colspan="2">$\Phi(K_{кр.,\alpha})=(1-2\alpha)/2$</td></tr>
<tr><td colspan="3">$K_{кр.}$ - по таблице интеграла Лапласа, $\Phi(x)$</td></tr>
<tr><td rowspan="5">Генеральные дисперсии не известны</td><td colspan="3" rowspan="5"></td><td colspan="3">Задача №4
$D[X_1]\approx D[X_2]$
$K_{расч.}=t_{расч.}=$
$=\frac{\bar{X}_1-\bar{X}_2}{\sqrt{(n_1-1)\cdot S_1^2+(n_2-1)\cdot S_2^2}}\cdot\sqrt{\frac{n_1 n_2(n_1+n_2-2)}{n_1+n_2}}$,
$\nu=n_1+n_2-2$.
H_0: $M[X_1]=M[X_2]$
H_0 принимает если:</td><td colspan="3">Задача №5
$K_{расч.}=t_{расч.}=\frac{\bar{d}\cdot\sqrt{n}}{S_d}$, где
$\bar{d}=\frac{\sum d_j}{n}$; $d_j = X_{1j}-X_{2j}$;
$S_d=\sqrt{\frac{\sum d_j^2-\frac{\left(\sum d_j\right)^2}{n}}{n-1}}$; $\nu=n-1$
H_0 принимается если:</td></tr>
<tr><td>Двустор. критерий</td><td colspan="2">Односторонний критерий</td><td>Двустор. критерий</td><td colspan="2">Односторонний критерий</td></tr>
<tr><td>H_1:$M[X_1]\neq M[X_2]$</td><td>H_1:$M[X_1]>M[X_2]$</td><td>H_1:$M[X_1]<M[X_2]$</td><td>H_1:$M[X_1]\neq M[X_2]$</td><td>H_1:$M[X_1]>M[X_2]$</td><td>H_1:$M[X_1]<M[X_2]$</td></tr>
<tr><td>$|t_{расч.}|<t_{кр.\ \alpha/2,\nu}$</td><td>$t_{расч}<t_{кр.\ \alpha,\nu}$</td><td>$t_{расч.}> -t_{кр.\ \alpha,\nu}$</td><td>$|t_{расч.}|<t_{кр.\ \alpha/2,\nu}$</td><td>$t_{расч.}<t_{кр.\ \alpha,\nu}$</td><td>$t_{расч.}>-t_{кр.\ \alpha,\nu}$</td></tr>
<tr><td colspan="3">$t_{кр.}$ по таблице распределения Стьюдента</td><td colspan="3">$t_{кр}$ по таблице распределения Стьюдента</td></tr>
</table>

Приложение 8

Основные показатели производства в сельскохозяйственных предприятиях Краснодарского края									
№ п.п.	Среднегодовая численность работников, чел.	Численность тракторов, эт. ед.	Площадь сельскохозяйственных угодий, га	Энергетические мощности, л.с.	Основные фонды сельскохозяйственного назначения, тыс. руб.	Затраты на производство валовой продукции , тыс. руб.	Затраты на производство реализованной продукции, тыс. руб.	Валовая продукция, тыс. руб.	Реализованная продукция, тыс. руб.
1	2	3	4	5	6	7	8	9	10
1	591	102	12139	34503	74171	111276	80946	120456	90126
2	334	54	6773	14698	64382	30960	25670	31362	26072
3	335	45	8698	15506	69721	38056	29209	50375	41528
4	657	102	12926	32885	52744	63272	38176	78800	53704
5	541	75	11135	32901	93277	82953	68145	98897	84089
6	864	113	12135	36032	174537	83600	54719	92718	63837
7	370	68	7105	27849	62482	62289	56879	83151	77741
8	437	54	6530	22851	116405	46774	36995	45309	35530
9	410	76	7154	24693	79399	55942	49226	63354	56638
10	552	68	9083	24027	94116	61685	60013	88644	86972
11	246	48	4474	10782	74385	34126	29769	41407	37050
12	492	104	13735	28253	103326	75099	54292	67383	46576
13	217	53	4501	13596	77558	26284	19065	26981	19762
14	603	98	7465	25200	99567	74367	70913	99974	96520
15	400	58	6270	19798	64488	47618	25379	51983	29744
16	602	121	10550	33420	88935	83584	60564	104487	81467
17	389	89	8753	26936	117937	79097	54599	111868	87370
18	435	45	10830	20598	97580	57820	47952	70245	60377
19	422	82	9646	24645	44073	51076	41828	61868	52620
20	100	17	4034	6485	13777	14988	12304	19681	16997
21	234	52	4680	18271	26591	16755	14897	20923	19065
22	395	79	9074	17540	58851	45593	36571	58057	49035
23	618	124	12730	39763	106264	70006	63853	90189	84036
24	308	77	8059	25510	75552	31741	26238	38159	32656
25	421	51	9912	18970	120849	66304	49863	80708	64267
26	615	75	10131	24848	108571	81290	55057	86152	59919
27	590	101	11576	24604	202088	90212	56178	81232	47198
28	230	43	6425	10061	50281	43276	36242	46181	39147
29	961	100	10533	41544	181137	128388	94954	129971	96537
30	414	42	6990	20498	61996	38719	26757	43039	31077
31	247	52	8160	15656	78086	55440	41135	61574	47269
32	605	100	11345	28422	137723	84156	67322	84577	67743
33	434	68	7671	26512	85796	63608	53064	77256	66712
34	741	82	10154	18016	103640	86587	61015	95941	70369
35	319	82	7740	26535	77899	54788	39825	58821	43858

1	2	3	4	5	6	7	8	9	10
36	441	58	5566	21576	103400	51722	42155	51716	42149
37	855	129	10276	43163	191108	114683	117335	171615	174267
38	1385	246	16816	72473	336464	227368	158245	285753	216630
39	320	78	7203	25810	87836	59059	33775	62608	37324
40	654	156	13313	56635	277550	146888	126933	175841	155886
41	1137	179	14800	67952	365315	192684	146848	243738	197902
42	650	102	11175	34110	159956	70307	57192	73950	60835
43	354	55	5931	22306	130280	74955	47242	78175	50462
44	649	113	11093	45756	207540	110701	90913	145212	125424
45	358	68	4968	10470	90959	57390	34531	59258	36399
46	922	108	15294	22725	144854	82418	55884	98967	72433
47	433	59	5761	22647	76179	73604	58443	78686	63525
48	486	48	4767	13138	55665	68175	58911	76802	67538
49	218	52	4789	15454	78164	33864	31961	41700	39797
50	189	22	4776	12747	40077	26161	16736	25950	16525
51	911	126	13325	45990	317847	103085	78571	110162	85648
52	513	83	9132	25214	40053	44710	37169	47502	39961
53	312	57	5217	12896	68042	28615	21611	25202	18198
54	117	22	2788	10238	26645	14237	11837	17034	14634
55	284	43	6638	15757	31237	43087	41277	44903	43093
56	304	32	5133	12985	22407	35222	23856	39517	28151
57	377	70	6084	20010	81056	39135	28504	41985	31354
58	341	67	7213	19137	73079	39635	29687	56662	46714
59	155	28	3678	9832	41464	19184	14300	17382	12498
60	525	74	12211	24850	62348	100930	58596	84462	42128
61	383	57	10916	16780	94416	43249	28381	44137	29269
62	372	73	8332	19700	95584	47875	33253	44900	30278
63	519	69	8491	30238	142772	64781	45353	70270	50842

Приложение 9

Динамика урожайности с/х культур в учхозе "Кубань" КГАУ ,ц/га

№ п/п	Годы	Озимые зерновые	Кукуруза	Подсолнечник	Картофель	Овощи
1	1965	37,5	70,3	25,2	103,4	248
2	1966	33,8	44,2	15,3	127,5	182,1
3	1967	37,9	48,4	23,1	162,8	206,4
4	1968	36,8	50,9	27,4	84,8	160
5	1969	39,2	40,3	26,5	96,2	178,1
6	1970	40,8	22,5	23,5	110,8	211,8
7	1971	56,2	20	25,9	127,8	207,2
8	1972	44,5	16,3	20,9	56,5	156,4
9	1973	39,6	34,6	22,7	110,9	158,3
10	1974	50,6	56,9	27,4	137,8	237,9
11	1975	30,2	20,5	20,2	96	199,2
12	1976	54,2	65,9	29,1	141	228,6
13	1977	49,8	61,7	27,5	138	206,3
14	1978	49,1	41,6	29	150,3	283,9
15	1979	47	47,2	28,5	81,2	276,4
16	1980	39,1	41,9	26,8	90,9	249,4
17	1981	49,2	33,7	24	58,2	221,6
18	1982	51,8	64,4	11,3	202,8	360,8
19	1983	47,9	77,8	28,3	164,9	327,8
20	1984	54,8	72,6	20,4	201	320,9
21	1985	42,8	61,2	28	173	319
22	1986	58	52	20,5	199	127,3
23	1987	48,9	76,8	20,3	43,1	124,8
24	1988	48,9	43,2	16,1	127,3	170,8
25	1989	56,7	55,9	8,2	188,9	189
26	1990	66	69,5	27	94,4	214,3
27	1991	46,9	41,2	14	107,7	160,8
28	1992	48,6	32,4	17,1	144,8	138,5
29	1993	49,4	52,7	19,3	90,6	30
30	1994	45,7	11,9	14,4	65,5	81,9
31	1995	43,4	30,1	10,3	8,6	111
32	1996	31,5	32,9	9,8	41,6	81,5
33	1997	44,4	35,7	9	66,5	91,4
34	1998	53,4	6,2	12,7	57,9	111,4
35	1999	59,5	24,6	18,9	45,7	148,3

35	2000	62,3	15,0	13,4	-	-
36	2001	78,8	-	12,8	-	94,8
37	2002	76,3	35,0	17,7	-	-
38	2003	48,8	21,9	20,7	-	67,8
39	2004	61,8	61,7	26,6	-	56,6
40	2005	66,9	34,7	21,4	-	74,0
41	2006	57,7	61,5	26,2	-	73,2
42	2007	47,8	36,2	21,8	-	16,0
43	2008	76,9	65,7	31,2	-	-
44	2009	54,8	45,9	22,8	-	-
45	2010	67,3	34,5	25,9	-	-

Приложение 10

Задания к модулю 1 уровня А

№ варианта	номера задач									
1	4	8	23	24	26	44	54	61	66	73
2	3	10	17	22	28	35	37	41	49	59
3	5	20	35	36	39	42	43	62	65	67
4	1	14	15	18	45	51	62	63	64	75
5	3	8	28	29	38	41	52	75	79	80
6	4	7	16	17	18	26	33	41	42	59
7	8	22	35	48	56	65	69	70	77	78
8	6	19	27	31	32	37	59	71	73	80
9	25	29	30	32	34	36	45	46	53	65
10	3	7	12	13	16	40	50	56	64	74
11	12	33	44	47	49	53	57	59	68	66
12	2	5	13	28	31	55	58	60	70	75
13	9	12	14	29	30	42	56	69	71	74
14	1	3	27	39	49	51	61	64	66	78
15	11	18	21	22	40	52	73	70	78	79
16	24	25	36	44	47	48	52	54	69	72
17	10	13	20	31	37	57	61	72	73	74
18	5	6	15	18	21	50	62	63	67	68
19	1	14	26	33	34	41	50	58	70	76
20	5	11	21	23	25	32	40	43	53	78
21	6	19	31	33	52	58	62	71	72	80
22	9	13	20	23	25	38	56	60	67	68
23	2	4	7	16	30	32	34	38	63	76
24	4	11	23	24	28	39	51	54	76	77
25	1	6	9	13	19	40	46	48	57	58
26	2	16	17	24	47	57	64	71	74	77
27	8	10	20	35	37	44	53	54	63	69
28	2	7	15	19	22	26	45	66	72	80
29	11	12	27	28	30	39	55	61	67	79
30	9	15	17	27	34	43	49	50	51	55
31	10	21	36	38	46	55	60	65	68	75

Приложение 11

Исходные данные к контрольным заданиям модуля 1 уровня В

81							
№ варианта	*n1*	*n2*	*n3*	*p1*	*p2*	*p3*	*k*
1	1	4	5	0,8	0,6	0,5	3
2	2	5	3	0,7	0,8	0,6	4
3	4	3	2	0,6	0,5	0,4	5
4	6	2	2	0,5	0,4	0,8	6
5	3	5	2	0,7	0,9	0,4	5
6	1	2	7	0,8	0,3	0,5	4
7	2	3	5	0,4	0,8	0,4	7
8	3	1	6	0,5	0,8	0,7	6
9	1	3	6	0,7	0,4	0,4	5
10	2	7	1	0,4	0,5	0,6	7
11	4	2	6	0,5	0,6	0,5	3
12	9	9	10	0,2	0,8	0,7	7
13	1	5	9	0,3	0,3	0,6	6
14	2	3	1	0,6	0,8	0,5	4
15	1	2	3	0,6	0,8	0,4	6
16	1	4	4	0,8	0,1	0,8	3
17	6	4	4	0,3	0,8	0,3	4
18	4	9	5	0,1	0,3	0,4	5
19	5	4	10	0,4	0,5	0,5	8
20	8	10	3	0,8	0,9	0,3	3
21	10	1	7	0,6	0,4	0,4	7
22	8	10	5	0,2	0,1	0,4	9
23	4	8	4	0,7	0,6	0,4	6
24	8	2	3	0,6	0,5	0,4	7
25	2	4	5	0,2	0,6	0,8	7
26	6	4	10	0,6	0,3	0,3	7
27	1	3	1	0,5	0,2	0,3	6
28	9	2	3	0,9	0,4	0,8	3
29	8	7	8	0,2	0,2	0,9	3
30	8	2	2	0,7	0,7	0,7	4
31	8	7	3	0,4	0,5	0,3	9

Продолжение приложения 11

№	82	83								84				85		86
	N	n_1	n_2	n_3	n_4	m_1	m_2	m_3	m_4	*n*	*l*	*m*	*k*	*k*	*n*	*k*
1	3	1	2	3	4	1	1	2	3	10	2	4	6	6	4	4
2	4	2	2	4	2	1	1	1	2	10	2	3	6	7	4	5
3	5	2	3	4	1	1	2	3	1	10	3	5	7	8	5	6
4	6	1	4	2	3	1	2	1	2	10	3	5	6	9	5	5
5	7	4	2	2	2	3	1	2	1	11	2	5	7	10	6	6
6	8	3	2	3	2	2	1	3	1	11	3	4	8	11	4	7
7	9	5	1	2	2	3	1	1	1	11	3	5	7	12	4	6
8	10	2	5	2	1	1	3	1	1	12	3	8	5	13	3	7
9	3	4	2	3	2	2	1	2	1	12	2	8	3	14	3	8
10	4	3	3	4	1	2	1	2	1	12	2	5	4	13	4	7
11	5	2	3	3	3	1	2	3	1	9	2	4	6	12	3	8
12	6	1	3	4	3	1	2	2	1	9	3	5	6	11	3	5
13	7	2	3	4	2	1	2	3	1	9	2	3	7	10	4	6
14	8	1	2	3	5	1	1	2	3	8	2	4	5	9	4	7
15	9	2	3	4	2	1	2	2	1	8	2	5	4	8	3	8
16	10	3	2	2	4	2	1	1	1	8	3	4	5	7	3	9
17	11	4	3	2	3	2	1	2	1	10	4	6	5	6	4	8
18	12	3	3	4	2	2	1	2	2	10	5	7	7	7	4	7
19	13	2	4	5	1	2	2	3	1	10	4	6	7	8	5	6
20	14	3	4	3	2	2	2	3	2	12	4	8	6	9	5	5
21	15	2	5	2	3	1	3	1	2	8	2	3	4	10	6	4
22	16	4	4	2	2	2	2	2	1	8	2	3	5	11	4	4
23	17	2	7	2	1	1	5	2	1	8	2	4	3	12	4	5
24	18	3	1	6	2	2	1	3	1	8	3	5	4	13	3	6
25	19	2	2	2	3	1	1	1	2	8	1	4	2	14	3	7
26	20	1	3	3	2	1	3	1	1	9	2	3	5	12	3	8
27	3	1	4	2	2	0	2	1	1	9	3	4	4	11	3	9
28	4	2	3	1	3	1	2	0	1	9	2	6	3	10	4	10
29	5	3	1	2	3	0	1	1	2	9	4	5	5	9	4	9
30	6	3	2	3	1	2	2	2	0	9	3	5	4	8	3	8
31	8	2	3	1	3	2	1	0	2	9	2	3	6	7	3	7

№	87			88			89		90				91
	T_1	T_2	t	R	S_1	S_2	k_1	k_2	p_1	p_2	n_1	n_2	k
1	900	1000	10	11	2,25	3,52	71	47	0,61	0,55	2	3	4
2	900	1100	20	12	2,37	3,52	78	39	0,62	0,54	3	2	5
3	1000	1100	10	13	2,49	3,52	87	31	0,63	0,53	2	3	6
4	1000	1200	20	14	2,55	1,57	72	46	0,64	0,52	3	2	7
5	1100	1200	15	11	2,27	5,57	79	38	0,65	0,51	2	3	8
6	1100	1300	15	12	2,39	5,57	86	32	0,66	0,49	3	2	9
7	900	930	10	13	2,51	1,57	73	45	0,67	0,48	2	3	10
8	900	1130	20	14	2,57	3,52	81	37	0,68	0,47	3	2	11
9	1000	1030	15	11	2,29	3,52	85	33	0,69	0,46	2	3	4
10	1000	1130	15	12	2,41	3,52	74	44	0,71	0,45	3	2	5
11	1100	1130	5	13	2,53	3,52	82	36	0,72	0,44	2	3	6
12	1100	1230	5	14	2,59	5,57	84	34	0,73	0,43	3	2	7
13	1200	1300	5	15	2,5	8,7	75	43	0,74	0,42	2	3	8
14	1200	1230	10	16	2,6	8,5	83	35	0,75	0,41	3	2	9
15	1200	1330	5	11	2,2	3,5	76	42	0,76	0,39	2	3	10
16	1300	1400	10	12	2,4	3,5	77	41	0,77	0,38	3	2	12
17	1800	1900	10	13	2,5	3,5	47	71	0,78	0,37	2	3	5
18	1800	2000	20	14	2,6	1,8	39	78	0,39	0,45	3	2	6
19	1700	1800	10	15	2,7	7,9	31	87	0,38	0,46	2	3	7
20	1700	1900	20	16	2,7	8,2	72	46	0,37	0,47	3	2	8
21	1900	2000	15	11	2,3	3,5	38	79	0,36	0,48	2	3	9
22	1900	2100	15	12	2,4	3,5	32	86	0,35	0,49	3	2	10
23	1700	1730	10	13	2,5	3,5	73	45	0,34	0,51	2	3	11
24	1700	1830	20	14	2,6	5,6	81	37	0,33	0,52	3	2	4
25	1600	1630	15	15	2,5	8,7	33	85	0,32	0,53	2	3	5
26	1600	1730	15	11	2,3	5,6	44	74	0,31	0,54	3	2	6
27	1700	1730	5	12	2,4	5,6	36	82	0,29	0,55	2	3	7
28	1700	1830	5	13	2,5	3,5	84	34	0,28	0,56	3	2	8
29	1600	1700	5	14	2,6	5,6	75	43	0,27	0,57	2	3	9
30	1600	1630	10	15	2,7	7,9	83	35	0,26	0,58	3	2	10
31	1600	1730	5	12	2,25	3,52	76	42	0,25	0,59	2	3	11

Продолжение приложения 11

№	92	93		94					95				96						
	M	n_1	n_2	N_1	M_1	N_2	M_2	K	k	l	m	n	m_1	m_2	m_3	n_1	n_2	n_3	j
1	12	100	250	4	1	2	5	3	8	10	3	2	50	30	20	70	80	90	1
2	8	430	180	7	3	5	1	4	7	6	2	3	50	30	20	70	80	90	2
3	5	170	540	2	3	5	4	1	6	8	3	1	50	30	20	70	80	90	3
4	11	520	390	8	2	3	2	5	12	5	3	2	60	20	20	70	80	90	1
5	7	360	600	6	4	1	7	2	13	11	2	4	60	20	20	70	80	90	2
6	10	700	90	3	2	4	4	2	11	8	2	5	60	20	20	70	80	90	3
7	6	240	610	5	5	4	10	4	12	7	2	4	40	30	30	80	80	90	1
8	9	80	710	13	12	4	6	3	9	6	2	3	40	30	30	80	80	90	2
9	3	630	230	1	9	3	3	4	10	7	4	1	40	30	30	80	80	90	3
10	8	500	320	3	7	5	2	3	11	7	4	4	40	20	40	90	90	80	1
11	5	810	70	4	6	7	8	5	13	8	5	2	40	20	40	90	90	80	2
12	10	450	280	2	3	7	1	2	8	7	3	3	40	20	40	90	90	80	3
13	10	270	640	2	2	3	1	1	12	10	4	2	70	20	10	70	80	90	1
14	9	380	470	2	8	3	1	3	9	6	1	3	70	20	10	70	80	90	2
15	4	640	80	6	4	3	3	4	6	8	3	2	70	20	10	70	80	80	3
16	7	160	570	5	5	4	3	3	14	13	3	3	60	10	30	80	90	80	1
17	5	590	200	25	3	25	2	2	11	10	4	5	60	10	30	80	90	80	2
18	8	620	190	20	1	40	7	3	7	5	2	2	60	10	30	80	90	80	3
19	9	730	100	20	4	25	5	1	15	9	4	3	50	20	30	90	80	90	1
20	6	540	200	50	8	20	6	2	8	10	3	3	50	20	30	90	80	90	2
21	12	90	690	40	8	10	2	3	12	5	2	2	50	20	30	90	80	90	3
22	8	220	550	25	2	20	4	1	14	11	3	5	30	30	40	70	70	80	1
23	10	290	700	20	1	40	5	5	6	7	2	2	30	30	40	70	70	80	2
24	7	350	440	25	2	25	6	4	13	9	4	4	30	30	40	70	70	80	3
25	3	470	360	10	3	50	11	2	9	6	3	3	20	40	40	90	70	80	1
26	6	680	230	20	1	20	4	1	11	10	2	5	20	40	40	90	70	80	2
27	9	710	160	25	3	25	7	3	7	8	4	3	20	40	40	90	70	80	3
28	4	180	270	40	5	50	8	2	12	11	5	4	10	50	40	70	90	80	1
29	7	260	620	40	8	20	4	2	8	3	2	2	10	50	40	70	90	80	2
30	5	650	140	25	3	40	2	4	6	6	1	2	10	50	40	70	90	80	3
31	8	230	480	20	1	50	6	1	10	8	3	3	20	30	50	70	70	90	1

Продолжение приложения 11

	97		98		99				
№	*n*	*m*	*p*	*n*	*n*	n_1	n_2	p_1	p_2
1	3	2	0,3	10	15	1	2	0,1	0,2
2	7	3	0,3	14	15	2	1	0,15	0,15
3	4	7	0,3	13	15	2	2	0,15	0,15
4	4	3	0,3	12	15	1	1	0,1	0,15
5	3	6	0,3	11	15	3	2	0,2	0,25
6	6	5	0,3	15	15	2	2	0,15	0,2
7	3	5	0,4	11	15	3	1	0,2	0,15
8	8	3	0,4	13	15	1	2	0,13	0,17
9	6	4	0,4	14	15	2	1	0,14	0,16
10	4	5	0,4	10	15	1	3	0,16	0,24
11	2	7	0,4	12	15	3	2	0,17	0,23
12	5	4	0,4	15	15	3	1	0,18	0,12
13	8	6	0,5	12	15	4	1	0,19	0,11
14	2	6	0,4	12	15	3	3	0,2	0,26
15	2	3	0,5	11	14	1	3	0,09	0,21
16	4	2	0,5	13	14	1	4	0,1	0,21
17	7	6	0,5	14	14	2	2	0,11	0,2
18	5	3	0,5	15	14	2	4	0,12	0,2
19	4	6	0,6	13	14	3	3	0,15	0,2
20	8	5	0,6	11	14	2	3	0,2	0,2
21	6	3	0,6	12	14	3	4	0,3	0,2
22	5	2	0,6	10	14	2	3	0,1	0,2
23	3	7	0,6	15	14	3	4	0,2	0,25
24	6	8	0,6	14	14	5	4	0,25	0,35
25	5	6	0,7	14	14	4	4	0,21	0,39
26	7	4	0,7	10	14	4	3	0,1	0,3
27	5	7	0,5	11	14	2	2	0,25	0,35
28	6	2	0,6	12	14	1	2	0,1	0,15
29	7	5	0,7	12	14	1	1	0,05	0,15
30	8	4	0,7	13	14	1	2	0,1	0,1
31	7	2	0,3	13	14	2	2	0,05	0,05

Продолжение приложения 11

№	100			101			
	m	*n*	*p*	*N*	*p*	k_1	k_2
1	7	500	0,003	100	0,8	80	90
2	7	500	0,004	100	0,8	85	95
3	8	500	0,008	100	0,8	70	95
4	6	600	0,009	100	0,7	83	93
5	10	600	0,003	100	0,7	50	60
6	7	600	0,009	100	0,7	65	75
7	8	700	0,005	100	0,7	70	80
8	6	700	0,007	100	0,6	40	50
9	5	700	0,004	100	0,75	65	80
10	7	800	0,002	100	0,75	70	85
11	6	800	0,006	100	0,75	68	78
12	10	800	0,001	100	0,7	60	-
13	9	900	0,004	100	0,7	70	-
14	8	900	0,007	100	0,7	80	-
15	6	900	0,002	100	0,6	65	-
16	6	1000	0,005	100	0,6	75	-
17	8	1000	0,002	100	0,6	50	-
18	9	500	0,005	100	0,8	70	-
19	7	500	0,009	100	0,8	80	-
20	9	500	0,004	100	0,8	90	-
21	6	600	0,003	100	0,8	95	-
22	5	600	0,006	100	0,3	-	20
23	9	600	0,004	100	0,3	-	30
24	8	700	0,003	100	0,3	-	40
25	8	700	0,001	200	0,4	-	80
26	9	700	0,003	200	0,4	-	90
27	9	800	0,001	200	0,4	-	100
28	7	800	0,009	300	0,8	-	250
29	10	800	0,005	400	0,6	-	270
30	7	900	0,003	400	0,7	-	290
31	7	900	0,009	400	0,8	-	300

Приложение 12

Задания к модулю 2 уровня А

№ варианта	номера задач								
1	4	8	23	24	26	44	45	46	51	71
2	3	10	17	22	28	35	37	41	52	72
3	5	20	35	36	39	42	43	44	53	73
4	1	14	15	18	39	40	45	47	54	74
5	3	8	28	29	34	35	36	37	55	75
6	4	7	16	17	18	26	33	41	56	76
7	8	9	10	11	12	17	22	50	57	77
8	6	20	26	28	30	31	44	49	58	78
9	25	29	30	32	36	45	46	50	59	79
10	3	7	12	13	16	40	42	45	60	80
11	12	33	34	36	38	39	42	47	61	81
12	2	5	14	29	31	32	38	43	62	82
13	9	12	14	15	17	23	27	29	63	83
14	1	3	4	9	16	18	19	22	64	84
15	11	18	21	22	40	43	44	48	65	85
16	24	25	36	37	38	39	43	48	66	86
17	10	13	14	16	18	22	26	49	67	87
18	5	6	15	18	21	23	24	28	68	88
19	1	14	26	27	31	33	34	42	69	89
20	5	11	21	23	25	32	40	43	70	90
21	20	21	41	44	49	50	51	52	55	71
22	9	13	20	23	25	31	41	52	55	72
23	2	4	7	16	30	32	34	38	53	73
24	4	11	23	24	28	41	42	47	54	74
25	1	6	9	13	19	40	46	48	55	75
26	2	16	17	24	25	29	31	33	56	76
27	8	10	20	22	25	29	35	46	57	77
28	2	7	15	19	26	45	47	47	58	78
29	11	12	27	28	30	39	40	41	59	79
30	9	15	17	27	34	43	49	50	60	80
31	10	21	35	36	37	38	48	49	61	81

Приложение 13

Исходные данные к контрольным заданиям модуля 2 уровня сложности В

№	91				92			93, 94		
	a	*b*	x_1	x_2	*n*	*p*	*a*	*m*	*a*	*b*
1	2,5	4	3	3,3	5	0,37	-	4	0	2
2	1,5	3	2	2,6	14	0,28	-	3	1	2
3	1,5	2,5	2	2,3	6	0,53	-	2	2	3
4	1	3,5	2	2,8	9	0,46	-	5	2	3
5	-1	2	-0,7	1,1	7	0,18	-	4	3	4
6	-2	1	-1.5	0,3	3	0,67	-	3	1	5
7	-3	5	2	2	8	0,32	-	2	0	2
8	-1,5	2,5	-1	0	10	0,87	-	3	2	3
9	1	1.8	1.3	1,6	4	0,25	-	4	3	4
10	1	2,4	1,5	2	12	0,41	-	2	4	5
11	2	3,5	2,5	3	-	-	0,68	5	0	6
12	2	2,8	2.1	2,5	-	-	0,35	4	1	2
13	1	2,8	-1	3	-	-	0,21	3	2	4
14	1	2,6	1.5	3	-	-	0,89	5	2	3
15	2	3	1	3	-	-	0,72	4	3	5
16	2	4,8	4,5	5	-	-	0,43	4	0	4
17	-4	-2	-1	0	-	-	0,17	3	1	2
18	-3	-1	9	0	-	-	0,95	2	2	3
19	2	4	0	3	-	-	0,38	4	3	4
20	1	3	0	2	-	-	0,63	3	4	5
21	1	1,5	0	0,5	-	-	0,026	4	0	5
22	-1	1,5	0	1	-	-	0,38	5	1	6
23	-1,5	-1	-1	3	-	-	0,033	6	3	4
24	-1,5	1	-1	1	-	-	0,218	5	2	3
25	0,5	1	0	3	-	-	0,65	4	3	5
26	0.2	2	0	4	-	-	0,816	3	0	2
27	0,5	3	0	0,5	-	-	0,74	4	1	3
28	0,4	4	1	5	-	-	0,015	5	2	3
29	1/4	1	0	3	-	-	0,671	4	3	5
30	0,02	2	0	3	-	-	0,324	3	4	6
31	0,05	4	0	10	-	-	0,57	2	3	2

Исходные данные к задаче 95

№	*P(x)*	*y=φ(x)*	№	*P(x)*	*y=φ(x)*
95.1	$\frac{1}{\pi(1+x^2)}$	$y=\|x\|$	95.2	$\begin{cases}\frac{2}{\pi}\cos^2 x, x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right),\\ 0, x\overline{\in}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\end{cases}$	$y=e^{-x^2}$
95.3	$\begin{cases}\frac{1}{\pi}, x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right),\\ 0, x\overline{\in}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\end{cases}$	$y=\|x\|$	95.4	$\frac{1}{\pi\,\mathrm{ch}x}$	$y=e^{-x^2}$
95.5	$\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$	$y=\|x\|$	95.6	$\frac{1}{\pi(1+x^2)}$	$y=e^{-x^2}$
95.7	$\begin{cases}\frac{2}{\pi}\cos^2 x, x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right),\\ 0, x\overline{\in}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\end{cases}$	$y=\|x\|$	95.8	$\begin{cases}\frac{1}{\pi}, x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right),\\ 0, x\overline{\in}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\end{cases}$	$y=e^{-x^2}$
95.9	$\frac{1}{\pi\,\mathrm{ch}x}$	$y=\|x\|$	95.10	$\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$	$y=e^{-x^2}$
95.11	$\frac{1}{\pi(1+x^2)}$	$y=e^{-x^2}$	95.12	$\begin{cases}\frac{2}{\pi}\cos^2 x, x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right),\\ 0, x\overline{\in}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\end{cases}$	$y=e^{-x^2}$
95.13	$\begin{cases}\frac{1}{\pi}, x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right),\\ 0, x\overline{\in}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\end{cases}$	$y=e^{-x^2}$	95.14	$\frac{1}{\pi\,\mathrm{ch}x}$	$y=e^{-x^2}$
95.15	$\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$	$y=e^{-x^3}$	95.16	$\begin{cases}\frac{1}{\pi}, x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right),\\ 0, x\overline{\in}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\end{cases}$	*y=2x+3*
95.17	$\frac{1}{\pi(1+x^2)}$	*y=2x+3*	95.18	$\frac{1}{\pi\,\mathrm{ch}x}$	*y=4x+5*
95.19	$\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$	*y=2x+3*	95.20	$\frac{1}{\pi(1+x^2)}$	*y=6x+4*
95.21	$\begin{cases}\frac{2}{\pi}\cos^2 x, x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right),\\ 0, x\overline{\in}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\end{cases}$	*y=2x+3*	95.22	$\begin{cases}\frac{1}{\pi}, x\in\left(-\frac{\pi}{2},\frac{\pi}{2}\right),\\ 0, x\overline{\in}\left(-\frac{\pi}{2},\frac{\pi}{2}\right)\end{cases}$	*y=6x+4*

Продолжение приложения 13

№	$P_{\xi}(x)$	$y=\varphi(x)$	№	$P_{\xi}(x)$	$y=\varphi(x)$
95.23	$\dfrac{1}{\pi\,\mathrm{ch}x}$	$y=2x+3$	95.24	$\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$y=6x+4$
95.25	$\dfrac{1}{\pi(1+x^2)}$	$y=4x+5$	95.26	$\begin{cases}\dfrac{2}{\pi}\cos^2 x, x\in\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right),\\ 0, x\notin\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)\end{cases}$	$y=6x+4$
95.27	$\begin{cases}\dfrac{1}{\pi}, x\in\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right),\\ 0, x\notin\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)\end{cases}$	$y=4x+5$	95.28	$\dfrac{1}{\pi\,\mathrm{ch}x}$	$y=6x+4$
95.29	$\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$y=4x+5$	95.30	$\dfrac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$	$y=8x+1$
95.31	$\begin{cases}\dfrac{2}{\pi}\cos^2 x, x\in\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right),\\ 0, x\notin\left(-\dfrac{\pi}{2},\dfrac{\pi}{2}\right)\end{cases}$	$y=4x+5$			

Продолжение приложения 13

№	96			97						99		100			
	a	b	n	x_A	y_A	x_B	y_B	x_C	y_C	α_1	α_2	α	n	x_1	x_2
1	1	2	I	0	0	1	1	1	-1	-1	0,1	1/3	108	17	20
2	2	1	2	0	0	-1	1	1	1	-2	0,2	1/4	162	22	26
3	3	2	3	0	о	-1	1	-1	-1	-3	0,3	1/5	300	28	33
4	2	3	2	0	0	-1	-1	1	-1	-4	0,4	1/6	432	35	38
5	4	2	4	0	о	2	2	2	-2	-0,5	0,05	1/7	584	40	44
6	2	4	2	0	о	-2	2	2	2	-1,5	0,15	1/5	768	46	51
7	3	4	2	0	о	-2	2	-2	-2	-2,5	0,06	1/9	972	53	56
8	4	3	2	0	0	-2	-2	2	-2	-3,5	0,07	1/10	1200	58	62
9	5	1	2	1	1	1	-1	0	0	-5	0,08	1/11	1432	64	69
10	5	2	5	-1	1	1	1	0	0	-6	0,25	1/12	1728	71	74
11	5	3	4	-1	1	-1	-1	0	0	-7	0,26	1/13	2028	76	80
12	3	5	2	-1	-1	1	-1	0	0	0	0,27	2/3	108	34	40
13	2	2	3	2	2	2	-2	0	0	-4,5	0,01	1/2	162	44	52
14	3	7	3	-2	2	2	2	0	0	-5,5	0,02	2/5	300	56	66
15	7	4	2	-2	2	-2	-2	0	0	-6,5	0,03	2/7	584	80	88
16	4	5	1	-2	-2	2	-2	0	0	-7,5	0,04	2/9	972	106	112
17	1	2	1	-1	0	0	1	0	-1	-9	0,06	2/11	1454	128	138
18	2	1	2	-1	0	0	2	0	-2	0	0,09	2/13	2028	152	160
19	3	2	3	-1	0	0	-1	0	1	-11	0,31	1	108	51	60
20	2	3	2	-1	0	0	-2	0	2	-12	0,32	3/4	162	66	78
21	4	2	4	-1	0	1	1	1	-1	-8,5	0,33	3/5	300	74	99
22	2	4	2	-1	0	1	2	1	-2	-9,5	0,34	3/7	584	120	152
23	3	4	2	-1	0	1	-1	1	1	-10,5	0,36	3/8	768	138	153
24	4	3	2	-1	0	1	-2	1	2	-11,5	0,37	3/10	1200	174	186
25	5	1	2	0	-1	-1	0	1	0	-13	0,49	3/11	1452	192	207
26	5	2	5	0	-1	-2	0	2	0	-14	0,48	3/13	2028	228	240
27	5	3	4	0	-1	1	0	-1	0	-15	0,47	1/14	584	20	22
28	3	5	2	0	-1	2	0	-2	0	-16	0,46	1/20	1200	29	31
29	2	2	3	0	0	1	1	1	-1	-1,4	0,44	1/26	2028	38	40
30	3	7	3	0	0	-1	1	1	1	-1,6	0,43	2	108	102	120
31	7	4	2	0	0	-1	-1	1	-1	-1,8	0,42	3/2	162	132	156

Учебное издание

Бондаренко Петр Сергеевич
Горелова Галина Викторовна
Кацко Игорь Александрович

ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Учебное пособие для бакалавров

В авторской редакции

Подписано в печать 09.10.13. Формат 60×84 $^1/_{16}$.
Тираж 500 экз. Усл. печ. л. – 21,2. Уч.-изд. л. – 19,7
Заказ № 650

Типография Кубанского государственного аграрного университета.
350044, Краснодар, ул. Калинина, 13