

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА  
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГБОУ ВО «Кубанский государственный  
аграрный университет имени И. Т. Трубилина»

Учетно-финансовый факультет

Кафедра статистики и прикладной математики

## **Прикладная статистика**

### **Методические указания**

по самостоятельной работе для обучающихся по направлению  
подготовки 38.03.01 Экономика

Краснодар  
КубГАУ  
2021

*Составитель:* Н. Н. Яроменко

**Прикладная статистика** : метод. указания по самостоятельной работе / сост. Н. Н. Яроменко. – Краснодар : КубГАУ, 2021. – 41 с.

В методических указаниях сформулированы базовые требования в сфере прикладной статистики, а также содержатся тестовые, реферативные, практические задания и контрольные вопросы к экзамену.

Предназначены для обучающихся по направлению подготовки 38.03.01 Экономика для закрепления теоретических знаний и практических навыков в сфере экономического анализа.

Рассмотрено и одобрено методической комиссией учетно-финансового факультета Кубанского государственного аграрного университета, протокол № \_\_ от \_\_. \_\_. 2021.

Председатель  
методической комиссии

И.Н. Хромова

© Яроменко Н.Н.  
составление, 2021  
© ФГБОУ ВО «Кубанский  
государственный аграрный  
университет имени  
И. Т. Трубилина», 2021

## Оглавление

Введение.....	4
1 Цель и задачи дисциплины.....	5
2 Программа дисциплины.....	6
3 Реферативные задания.....	8
3.1 Примерная тематика реферативных работ .....	8
3.2 Методические указания к реферативным заданиям...	9
4 Тестовые задания по дисциплине.....	11
5 Практические задания по дисциплине.....	22
6 Перечень вопросов для изучения дисциплины.....	34
Список рекомендованной литературы.....	39

## Введение

Методические рекомендации по самостоятельной работе выполнены в соответствии с программой дисциплины «ПРИКЛАДНАЯ СТАТИСТИКА» и содержит тестовые задания по темам дисциплины, тематику рефератов и отдельные вопросы по курсу, вынесенных на самостоятельное более углубленное изучение.

Цель методических рекомендаций - закрепить теоретические знания и практические умения студентов в области прикладной статистики для обоснования оптимальных управленческих решений.

Задачами методических рекомендаций являются приобретение теоретических знаний, лежащих в основе методологии экономического анализа, формирование навыков аналитического мышления, позволяющих интерпретировать информацию, и на ее основе с учетом критериев социально-экономической эффективности обосновывать оптимальные управленческие решения.

Необходимо отметить, что подготовка рефератов на практических занятиях и студенческих научно-практических конференциях имеет большое значение в накоплении знаний и навыков, необходимых для освоения дисциплины и формировании навыков проведения научного исследования.

По своей структуре методические рекомендации включают в себя положения, имеющие обязательный и рекомендательный характер. Обязательный характер касается, в первую очередь, выполнения тестовых заданий и подготовки реферата по теме научного исследования. Такое положение как контрольные вопросы для самоконтроля носят рекомендательный характер.

## 1 Цель и задачи дисциплины

Целью освоения дисциплины «Прикладная статистика» является формирование комплекса знаний, умений и навыков эффективного инструментария аппарата статистического исследования в профессиональной деятельности.

Задачи дисциплины:

— освоение статистических и экономико-математических методов анализа экономической деятельности хозяйствующих субъектов, необходимой для принятия управленческих решений;

— овладение навыками интерпретации и использования полученных сведений при подготовке статистических отчетов по результатам бизнес-анализа и производственно-хозяйственной деятельности;

— формирование практических умений и навыков проведения анализа выявления связей для различных типов данных, с использованием современных информационных технологий и соответствующих общедоступных программных средств, характеризующих финансово-хозяйственную деятельность хозяйствующих субъектов;

— формирование навыков аналитического мышления, позволяющих интерпретировать информацию, полученную в результате проведенных исследований, и на ее основе обосновывать оптимальные управленческие решения.

## **2 Программа дисциплины**

### ***Тема 1. Предмет, метод, основные понятия и категории прикладной статистики***

Предмет, метод и задачи статистики. Основные понятия статистики. Выборочный метод в статистике. Репрезентативность и однородность выборки. Обзор программных продуктов, используемых в процессе изучения дисциплины: Microsoft Windows; Microsoft Office, (включает Word, Excel, PowerPoint)

### ***Тема 2. Статистическое оценивание многомерных случайных величин***

Многомерной средняя. Матрица ковариаций. Вероятностное оценивание. Робастное оценивание.

### ***Тема 3. Многомерный дисперсионный анализ***

Методы дискриминантного анализа. Непараметрические методы дискриминантного анализа. Параметрические методы дискриминантного анализа. Функции дискриминантного анализа.

### ***Тема 4. Проверка многомерных гипотез***

Простые и сложные гипотезы. Параметрические и непараметрические критерии. Понятие наилучшей критической области. Типичные задачи проверки гипотез о математических ожиданиях.

### ***Тема 5. Выявление связей между признаками.***

Выявление связей между качественными признаками. Коэффициенты контингенции и Крамера. Выявление связей для порядковых признаков. Коэффициенты Спирмена и Кэндела. Выявление связей для количественных признаков.

### ***Тема 6. Элементы корреляционного анализа***

Теоретическая и выборочная функция регрессии. Метод наименьших квадратов. Линейная выборочная регрессия. Типичные нелинейные регрессионные модели, сводящиеся к линейным. Оценка качества модели. Коэффициент детерминации. Анализ остатков. Значимость коэффициентов.

### ***Тема 7. Многомерные статистические методы***

Множественный корреляционный анализ. Парные, частные и множественные коэффициенты корреляции. Модель множественной регрессии. Теорема Гаусса-Маркова. Оценка качества модели. Исправленный коэффициент детерминации. Анализ остатков, оценка значимости коэффициентов. Мультиколлинеарность.

### ***Тема 8. Кластерный анализ.***

Задачи кластерного анализа. Меры близости между объектами. Хеммингово расстояние.

### ***Тема 9. Элементы анализа временных рядов***

Понятие динамических рядов и их виды. Исчисление средних уровней в рядах динамики. Основные показатели анализа рядов динамики.

## **3 Реферативные задания**

### **3.1 Примерная тематика реферативных работ**

1. Методология и методы в статистике.
2. Классификации в статистике.
3. Статистические школы.
4. Статистика Англии.
5. Статистика США.
6. Статистика Германии.
7. Статистика Франции.
8. Статистика Японии
9. Статистические графики.
10. Ряды распределения.
11. Нормальное распределение.
12. Биномиальное распределение.
13. Распределение Пуассона.
14. Гамма распределение.
15. Бэ́та распределение.
16. Аналитическая группировка в Excel.
17. Виды дисперсии в группировке.
18. Дисперсионный анализ в группировки.
19. Показатели вариации.
20. Показатели вариации в Excel.
21. Коэффициент устойчивости.
22. Прогноз и ошибка прогноза.
23. Функциональные и корреляционные связи.

### 3. 2 Методические указания к выполнению реферативных заданий

Реферат – это краткое изложение в письменном виде содержания и результатов индивидуальной учебно-исследовательской деятельности, имеет регламентированную структуру, содержание и оформление. Его задачами являются:

- формирование умений самостоятельной работы студентов с источниками литературы, их систематизация;
- развитие навыков логического мышления;
- углубление теоретических знаний по проблеме исследования.

Текст реферата должен содержать аргументированное изложение определенной темы. Реферат должен быть структурирован (по главам, разделам, параграфам) и включать разделы:

- введение,
- основная часть,
- заключение,
- список используемых источников.

В зависимости от тематики реферата к нему могут быть оформлены приложения, содержащие документы, иллюстрации, таблицы, схемы и т. д.

Критериями оценки реферата являются: новизна текста, обоснованность выбора источников литературы, степень раскрытия сущности вопроса, соблюдения требований к оформлению.

Оценка *«отлично»* — выполнены все требования к написанию реферата: обозначена проблема и обоснована её актуальность; сделан анализ различных точек зрения на рассматриваемую проблему и логично изложена собственная позиция; сформулированы выводы, тема раскрыта полностью, выдержан объём; соблюдены требования к внешнему оформлению.

Оценка *«хорошо»* — основные требования к реферату выполнены, но при этом допущены недочёты. В частности, имеются

неточности в изложении материала; отсутствует логическая последовательность в суждениях; не выдержан объём реферата; имеются упущения в оформлении.

Оценка *«удовлетворительно»* — имеются существенные отступления от требований к реферированию. В частности: тема освещена лишь частично; допущены фактические ошибки в содержании реферата; отсутствуют выводы.

Оценка *«неудовлетворительно»* — тема реферата не раскрыта, обнаруживается существенное непонимание проблемы или реферат не представлен вовсе.

## 4 Тестовые задания по дисциплине

1. *Статистический анализ конкретных экономических данных проводится в рамках:*

- а) логистики
- б) эконометрики
- в) высшей математики
- г) математической статистики

2. *Вся совокупность объектов, характеризующая изучаемый признак, называется*

- а) точечной
- б) генеральной совокупностью
- в) объемом выборки
- г) выборочной совокупностью

3. *Временной ряд, для которого совместные функции распределения для любого числа моментов времени не меняются со временем, называется:*

- а) стационарным
- б) нестационарным
- в) непериодическим
- г) случайным

4. *"Размножение выборок" - это*

- а) бутстреп
- б) рандомизация
- в) байесовский подход
- г) подход случайного ножа

5. *В вероятностной теории статистических методов выборка обычно моделируется как конечная последовательность:*

- а) зависимых одинаково распределенных случайных величин или векторов

б) независимых экспоненциально распределенных случайных величин или векторов

в) независимых случайных векторов

г) независимых одинаково распределенных случайных величин или векторов

*6. К статистическим данным нечислового типа относятся:*

а) разбиения

б) толерантности

в) упорядочения

г) нечеткие множества

*7. Коэффициент эластичности спроса по цене, равный 2, показывает, что при изменении цены на 1 процент спрос:*

а) изменится на 2 единицы

б) изменится на 2 процента

в) спрос неэластичен по цене

*8. В модели случайной выборки данные рассматриваются как реализации*

а) независимых одинаково распределенных случайных величин

б) независимых случайных величин

в) зависимых одинаково распределенных случайных величин

г) одинаково распределенных случайных величин

*1. Оценка математического ожидания*

$\bar{x} = 50$  , выборочная дисперсия  $S_0^2 = 625$  ;  $n = 100$

Тогда 95%-ный доверительный интервал для математического ожидания

а) [45,1; 54, 9]

б) [54, 9; 64, 9]

в) [45, 9; 55, 9]

г) [25,1; 75,2]

2. *Статистика критерия согласия Колмогорова представляет собой:*

- а) интеграл квадрата эмпирического процесса по теоретической функции распределения
- б) интеграл квадрата теоретического процесса по теоретической плотности распределения
- в) супремум модуля эмпирического процесса
- г) инфимум эмпирического процесса

11. *Статистический анализ конкретных экономических данных проводится в рамках*

- а) логистики
- б) эконометрики
- в) высшей математики
- г) математической статистики

12. *Для порядковой шкалы допустимы*

- а) строго возрастающие преобразования
- б) тождественные преобразования
- в) только сравнения объектов

13. *Законы больших чисел позволяют описать поведение*

- а) произведений случайных величин
- б) сумм случайных величин
- в) отношений случайных величин
- г) отношений детерминированных величин
- д) сумм детерминированных величин

14. *Сравнивать выборки на основе среднего арифметического для данных, измеренных в порядковой шкале*

- а) некорректно
- б) можно в любом случае
- в) можно, если функция распределения одной выборки всегда лежит над другой

15. Точечной оценкой для медианы является

- а) мода
- б) среднее арифметическое
- в) выборочная медиана

16. Заполните пропуски в утверждении: "для эффективной работы специалиста по методу ЖОК желательно, чтобы общее число факторов не превышало \*\*\*, число непосредственных взаимосвязей - \*\*\*"

- а) 20 и 80
- б) 30 и 465
- в) 20 и 40
- г) 10 и 40

17. На первом этапе решения любой прикладной задачи математическими методами/методами прикладной статистики осуществляется

- а) сбор информации
- б) переход от математических выводов к практической проблеме
- в) внутриматематическое изучение и решение задачи
- г) переход от исходной проблемы до теоретической чисто математической задачи

18. Верно ли, что результаты измерений значений альтернативного признака - это

- а) данные в шкале разностей
- б) данные в шкале наименований
- в) дихотомические данные
- г) бинарные данные

4. К классическим статистическим технологиям относятся использование

- а) метода наименьших квадратов
- б) статистик типа Колмогорова, Смирнова, омега-квадрат

- в) непараметрического коэффициента корреляции Спирмена
- г) теории нечетких множеств

19. ОМП для математического ожидания случайной величины, распределенной по закону Лапласа, является

- а) мода
- б) среднее арифметическое
- в) выборочная медиана
- г) дециль

20. На плоскости заданы две точки:  $A(7;2)$  и  $B(3;5)$ . Тогда евклидово расстояние между ними равно

- а) 25
- б) 5
- в) 10
- г) 7

21. Мера расхождения сглаженного (регрессионного) и наблюдаемого значения называется

- а) остатком
- б) коэффициентом разности
- в) подвязкой
- г) триангуляцией

22. Экономико-математическая модель – это:

- а) модель, описывающая механизм функционирования экономики
- б) математическое описание экономического объекта или процесса с целью их исследования и управления ими
- в) экономическая модель
- г) модель реального явления

23. Этапы построения эконометрической модели:

- а) постановочный, априорный, параметризация

- б) постановочный, информационный, априорный
- в) постановочный, априорный, параметризация, информационный, идентификация модели, верификация модели
- в) параметризация, информационный, идентификация модели

24. *Фиктивная переменная – переменная, принимающая в каждом наблюдении*

- а) ряд значений от 0 до 1
- б) только отрицательные значения
- в) только два значения 0 или 1
- г) только положительные значения
- д) случайные

25. *Линеаризация нелинейной модели регрессии может быть достигнута*

- а) отбрасыванием нелинейных переменных
- б) перекрестной суперпозицией переменных
- в) преобразованием анализируемых переменных
- г) сглаживанием переменных

26. *Простая (парная) регрессия – это*

- а) зависимость среднего значения какой-либо величины
- б) модель вида  $Y_x = a + bx$
- в) модель, где среднее значение зависимой переменной  $Y$  рассматривается как функция одной независимой  $X$
- г) модель, где среднее значение зависимой переменной  $Y$  рассматривается как функция нескольких независимых переменных

27. *Число степеней свободы для уравнения  $t$ -мерной регрессии при достаточном числе наблюдений  $n$  составляет*

- а)  $n/t$
- б)  $n-t$
- в)  $n-t+1$
- г)  $n-t-1$

д)  $t-1$

28. Одним из известных способов проверки регрессионных остатков эконометрической модели на автокорреляцию является критерий

- а) Дарбина-Уотсона
- б) Марка-Шагала
- в) Куприна-Утрехта
- г) Айзека-Азимова

29. Параметры множественной регрессии  $\beta_1, \beta_2, \dots, \beta_m$  показывают \_\_\_\_\_ соответствующих экономических факторов

- а) степень влияния
- б) случайность
- в) уровень независимости
- г) непостоянство
- д) цикличность

30. Процесс выбора необходимых переменных для регрессии переменных и отбрасывание лишних переменных называется

- а) унификацией переменных
- б) моделированием
- в) спецификацией переменных
- г) прогнозированием
- д) подгонкой

31. Требованиями к факторам, включаемым в модель линейной множественной регрессии, относятся

- а) число факторов должно быть в 6 раз меньше объема совокупности
- б) факторы должны представлять временные ряды
- в) факторы должны иметь одинаковую размерность
- г) между факторами не должно быть высокой корреляции

32. Уравнение множественной регрессии в стандартизованном масштабе имеет вид  $t_y = 20 + 0,9t_{x_1} + 0,5t_{x_2} + \varepsilon$ . На результативный признак оказывает большее влияние

- а)  $x_1$
- б)  $x_1$  и  $x_2$
- в)  $x_2$
- г) нельзя сделать вывод

33. Тесноту совместного влияния факторов на результат в уравнении линейной множественной регрессии оценивает

- а) коэффициент парной корреляции
- б) коэффициент частной корреляции
- в) коэффициент множественной корреляции
- г) коэффициент множественной детерминации

34. Верными утверждениями относительно коэффициента множественной корреляции являются

- а) чем ближе значение к единице  $R_{yx_1 \dots x_p}$ , тем теснее связь результативного признака со всеми факторами
- б) чем ближе значение к нулю  $R_{yx_1 \dots x_p}$ , тем теснее связь результативного признака со всеми факторами
- в)  $R_{yx_1 \dots x_p}$  принимает значения из промежутка  $[0, 1]$
- г)  $R_{yx_1 \dots x_p}$  принимает значения из промежутка  $[-1, 1]$

35. Коэффициент множественной детерминации характеризует

- а) тесноту совместного влияния факторов на результат в уравнении линейной множественной регрессии
- б) тесноту связи между результатом и соответствующим фактором, при устранении влияния других факторов, включенных в модель

в) долю дисперсии результативного признака, объясненную регрессией в его общей дисперсии

*36. Для определения части вариации, обусловленной изменением величины изучаемого фактора, используется*

- а) коэффициент вариации
- б) коэффициент корреляции
- в) коэффициент детерминации
- г) коэффициент эластичности

*37. Качество подбора линейного уравнения регрессии можно охарактеризовать на основе*

- а) индекса детерминации
- б) средней ошибки аппроксимации
- в) коэффициента эластичности
- г) коэффициента регрессии

*38. Если наблюдаемое значение F-критерия Фишера больше критического, то можно делать вывод о*

- а) статистической незначимости построенной модели
- б) статистической значимости построенной модели
- в) незначимости (несущественности) моделируемой зависимости
- г) отсутствии связи между изучаемыми переменными

*39. Множественный коэффициент детерминации определяет*

- а) долю дисперсии факторов, объясненную регрессией
- б) долю дисперсии результативного признака, объясненную регрессией
- в) долю дисперсии факторов, не объясненную регрессией
- г) долю дисперсии результативного признака, не объясненную регрессией

*40. Множественный коэффициент корреляции равен 0,8. Значит, множественный коэффициент детерминации составит*

- а) 0,8
- б) 0,2
- в) 0,64
- г) 0,36

*41. Множественный коэффициент эластичности  $\varepsilon_j$  показывает*

а) на сколько единиц в среднем изменится результативный признак, при увеличении  $j$ -ого фактора на единицу, если остальные факторы закреплены на постоянном уровне

б) на сколько процентов в среднем изменится результативный признак, при увеличении  $j$ -ого фактора на единицу, если остальные факторы закреплены на постоянном уровне

в) на сколько процентов в среднем изменится результативный признак, при увеличении  $j$ -ого фактора на один процент, если остальные факторы закреплены на постоянном уровне

г) силу связи между результативным признаком и  $j$ -ым фактором

*42. Значимость множественного коэффициента корреляции проверяется с помощью*

- а) нормального закона распределения
- б)  $\chi^2$  – критерия Пирсона
- в)  $t$  – критерия Стьюдента
- г)  $F$  – критерия Фишера

*43. Создаваемый с целью получения и/или хранения информации специфический объект, отражающий свойства, характеристики и связи объекта-оригинала произвольной природы, существенные для решаемой субъектом задачи, - это*

- а) датчик
- б) метод
- в) модель

*44. Верно, что толерантность*

- а) предполагает выполнение свойства транзитивности

- б) не предполагает выполнение свойства транзитивности
- в) это симметричное бинарное отношение
- г) это рефлексивное бинарное отношение
- д) это несимметричное бинарное отношение

## 5 Практические задания по дисциплине

### Задание 1

При создании торгового предприятия была запланирована еженедельная прибыль в 3,5 тыс. долларов. За прошедшие 20 недель доходы Вашего предприятия приведены в выборке с номером N. Можно ли утверждать, что доход от Вашего предприятия равен запланированному?

### Задание 2

Дана выборка с номером N. Объем выборки равен 20. Построить вариационный ряд, нарисовать гистограмму и полигон, найти выборочное среднее, моду, медиану, выборочную дисперсию, стандартное отклонение, коэффициент вариации и размах.

### Задание 3

Объемы продаж холодильников за 20 недель в двух магазинах, работающих в разных районах города, приведены в выборках с номерами N и N+1. Можно ли считать, что объем продаж не зависит от расположения магазинов? Будут ли выборки однородными?

### Задание 4

Из генеральной совокупности извлечена выборка объема  $n = 10$ :

$x_i$	-2	1	2	3	4	5
$n_i$	2	1	2	2	2	1

Оцените с надежностью 0,95 математическое  $\sigma$  нормально распределенного признака по выборочному среднему с помощью доверительного интервала.

### Задание 5

Исправленное среднее квадратическое отклонение ежесуточного

дохода случайно выбранных 10 киосков некоторой фирмы оказалось равно 100 ден.ед. Постройте доверительный интервал для среднего квадратического отклонения с надежностью 0,9. Предполагается, что доход – это нормально распределенная величина.

### **Задание 6**

Предполагается, выполнение некоторой работы занимает случайное время с равномерным распределением на отрезке  $[a, b]$ . Хронометраж 20 испытаний дал среднее время работы 30 мин и исправленную дисперсию 24 мин<sup>2</sup>. Оцените параметры  $a$  и  $b$  методом моментов. Оцените, за какое время работа будет выполнена с вероятностью 0,98.

### **Задание 7**

Имеется выборка по 77 сельскохозяйственным организациям центральной агроклиматической зоны Краснодарского края. Переменные:  $Y$  – выручка на 1 га пашни, тыс. руб.;  $X_1$  – основные средства на 1 га пашни тыс. руб.,  $X_2$  – численность работников на 1 га пашни, чел.;  $X_3$  – годовая заработная плата на среднегодового работника, тыс. руб.

С использованием пакета Gretl построить модели оценки влияния на выручку указанных факторов с использованием медианной, квантильной регрессии и непараметрических моделей ядерного сглаживания: loess, Надарая-Ватсона.

### **Задание 8**

Характеристика данных. Для анализа выбран срез социально-демографических групп, отражающий, согласно опросу, получение информации из соответствующих СМИ.

$V_1$  – центральное телевидение;  $V_2$  – интернет-новостные, аналитические, официальные сайты;  $V_3$  – региональное, местное телевидение;  $V_4$  – интернет – социальные сети и блоги;  $V_5$  – центральная пресса;  $V_6$  – региональная, местная пресса;  $V_7$  –

центральное радио; V8 – региональное, местное радио; V9 – зарубежные сми.

Методами непараметрической статистики получить значимые результаты без использования классических предположений параметрической статистики

### Задание 9

Имеется пять пар наблюдений переменных  $x$  и  $y$ :

$x_i$	1	3	3	5	7
$y_i$	0	2	5	6	6

Найти параметры модели и дать интерпретацию по ним.

### Задание 10

На заводе разработаны две новые технологии  $T_1, T_2$ . Чтобы оценить, как изменится дневная производительность при переводе на новые технологии, завод в течение 10 дней работал по каждой, включая существующую  $T_0$ . Дневная производительность в условных единицах приводится в таблице. Проверить гипотезу об отсутствии влияния технологии на производительность.

№	$T_0$	$T_1$	$T_2$	№	$T_0$	$T_1$	$T_2$
1	46	74	52	6	44	68	70
2	48	82	63	7	66	76	78
3	73	64	72	8	46	88	68
4	52	72	64	9	60	70	70
5	72	84	48	10	48	60	54

### Задание 11

Имеются следующие данные по 10 сельскохозяйственным предприятиям Краснодарского края: себестоимость 1 центнера зерна, руб. ( $Y$ ); урожайность зерновых культур, ц с 1 га ( $X$ ).

№ п.п.	1	2	3	4	5	6	7	8	9	10
У	345	278	284	235	250	267	356	327	236	425
Х	58,6	69,0	67,1	74,7	68,7	66,9	54,9	63,3	71,0	48,8

*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 каждого уравнения с помощью средней ошибки аппроксимации.

4. Найти коэффициент эластичности.

5. Оценить тесноту связи между переменными с помощью показателей корреляции и детерминации.

### ***Задание 12.***

Имеются следующие данные по 9 сельскохозяйственным предприятиям Краснодарского края: себестоимость 1 центнера подсолнечника, руб. (У); урожайность подсолнечника, ц с 1 га (Х).

№ п.п.	1	2	3	4	5	6	7	8	9
У	483	275	554	463	652	322	412	351	382
Х	24,7	30,0	16,1	28,3	26,6	29,8	30,3	33,2	28,3

1. Рассчитать параметры степенного уравнения регрессии.

2. Оценить тесноту связи с помощью показателей корреляции и детерминации.

3. Определить среднюю ошибку аппроксимации.

4. Оценить значимость уравнения связи с помощью F-критерия Фишера.

5. Сделать выводы по полученным результатам.

### ***Задание 13.***

Имеются следующие данные по 9 сельскохозяйственным предприятиям Краснодарского края: выручка от реализации продукции на 1 гектар сельскохозяйственных угодий, тыс. руб. (Y); основные фонды на 1 гектар сельскохозяйственных угодий, тыс. руб. (X).

№ п.п.	1	2	3	4	5	6	7	8	9
Y	23,5	29,9	32,3	36,3	36,9	33,5	29,3	30,7	27,3
X	11,9	19,5	12,8	39,8	34,2	23,1	18,6	19,5	18,8

1. Рассчитать параметры степенного уравнения регрессии.
2. Оценить тесноту связи с помощью показателей корреляции и детерминации.
3. Определить среднюю ошибку аппроксимации.
4. Оценить значимость уравнения связи с помощью F-критерия Фишера.
5. Сделать выводы по полученным результатам.

### ***Задание 14***

По 35 сельскохозяйственным организациям провести регрессионный анализ влияния факторов ( $X_1$  и  $X_2$ ) на изменение результативного признака (Y).

Y – производственная себестоимость 1 ц молока, руб.;

$X_1$  – надой молока на среднегодовую корову, ц;

$X_2$  – удельный вес молока в выручке от реализации продукции животноводства, %.

Y		X1		X2	
Среднее значение	1550	Среднее значение	50,5	Среднее значение	69,4
Стандартная ошибка среднего	68	Стандартная ошибка среднего	3,1	Стандартная ошибка среднего	3,9
Медиана	1532	Медиана	52,5	Медиана	71,9
Среднее квадратическое отклонение	295	Среднее квадратическое отклонение	13,4	Среднее квадратическое отклонение	17
Дисперсия	86850	Дисперсия	178,9	Дисперсия	287,9

выборки		выборки		выборки	
Эксцесс	1,92	Эксцесс	-0,72	Эксцесс	-0,5
Асимметричность	1,07	Асимметричность	-0,11	Асимметричность	-0,68

Парные коэффициенты корреляции:

$$r_{yx_1} = -0.562 ; r_{yx_2} = -0.441; r_{x_1x_2} = 0.487.$$

1. Составить матрицу парных коэффициентов корреляции между тремя переменными.

2. Определить параметры множественного уравнения регрессии в стандартизированной и естественной форме.

3. Рассчитать частные коэффициенты эластичности.

4. Рассчитать частные и множественный коэффициенты корреляции и детерминации.

5. Оценить значимость множественного уравнения регрессии с помощью  $F$ -критерия Фишера, для чего составить таблицу дисперсионного анализа.

6. С помощью частных  $F$ -критериев Фишера оценить целесообразность включения фактора  $x_1$  после  $x_2$  и фактора  $x_2$  после  $x_1$ .

7. Оценить значимость множественных коэффициентов регрессии с помощью  $t$ -критерия Стьюдента.

8. Написать выводы по представленным данным и результатам расчетов.

### Задание 15

По 40 сельскохозяйственным организациям провести регрессионный анализ влияния факторов ( $X_1$  и  $X_2$ ) на изменение результативного признака ( $Y$ ).

$Y$  – производственная себестоимость 1 ц молока, руб.;

$X_1$  – надой молока на среднегодовую корову, ц;

$X_2$  – среднегодовое поголовье коров, голов.

$Y$		$X_1$		$X_2$	
Среднее значение	1550	Среднее значение	50,5	Среднее значение	955

Стандартная ошибка среднего	68	Стандартная ошибка среднего	3,1	Стандартная ошибка среднего	144
Медиана	1532	Медиана	52,5	Медиана	712
Среднее квадратическое отклонение	295	Среднее квадратическое отклонение	13,4	Среднее квадратическое отклонение	62717
Дисперсия выборки	86850	Дисперсия выборки	178,9	Дисперсия выборки	392816
Эксцесс	1,92	Эксцесс	-0,72	Эксцесс	1,26
Асимметричность	1,07	Асимметричность	-0,11	Асимметричность	1,38

Парные коэффициенты корреляции:

$$r_{yx_1} = -0.562 ; r_{yx_2} = -0.478; r_{x_1x_2} = 0.439.$$

1. Составить матрицу парных коэффициентов корреляции между тремя переменными.
2. Определить параметры множественного уравнения регрессии в стандартизированной и естественной форме.
3. Рассчитать частные коэффициенты эластичности.
4. Рассчитать частные и множественный коэффициенты корреляции и детерминации.
5. Оценить значимость множественного уравнения регрессии с помощью  $F$ -критерия Фишера, для чего составить таблицу дисперсионного анализа.
6. С помощью частных  $F$ -критериев Фишера оценить целесообразность включения фактора  $x_1$  после  $x_2$  и фактора  $x_2$  после  $x_1$ .
7. Оценить значимость множественных коэффициентов регрессии с помощью  $t$ -критерия Стьюдента.
8. Написать выводы по представленным данным и результатам расчетов.

### **Задание 16**

По 40 сельскохозяйственным организациям провести регрессионный анализ влияния факторов ( $X_1$  и  $X_2$ ) на изменение результативного признака ( $Y$ ).

$Y$  – производственная себестоимость 1 ц молока, руб.;

$X_1$  – надой молока на среднегодовую корову, ц;

$X_2$  – затраты на корма на 1 ц молока, руб.

$Y$		$X_1$		$X_2$	
Среднее значение	1550	Среднее значение	50,5	Среднее значение	727
Стандартная ошибка среднего	68	Стандартная ошибка среднего	3,1	Стандартная ошибка среднего	50
Медиана	1532	Медиана	52,5	Медиана	663
Среднее квадратическое отклонение	295	Среднее квадратическое отклонение	13,4	Среднее квадратическое отклонение	216
Дисперсия выборки	86850	Дисперсия выборки	178,9	Дисперсия выборки	46658
Эксцесс	1,92	Эксцесс	-0,72	Эксцесс	1,43
Асимметричность	1,07	Асимметричность	-0,11	Асимметричность	1,47

Парные коэффициенты корреляции:

$$r_{yx_1} = -0.562 ; r_{yx_2} = 0,773; r_{x_1x_2} = -0,452.$$

1. Составить матрицу парных коэффициентов корреляции между тремя переменными.

2. Определить параметры множественного уравнения регрессии в стандартизированной и естественной форме.

3. Рассчитать частные коэффициенты эластичности.

4. Рассчитать частные и множественный коэффициенты корреляции и детерминации.

5. Оценить значимость множественного уравнения регрессии с помощью  $F$ -критерия Фишера, для чего составить таблицу дисперсионного анализа.

6. С помощью частных  $F$ -критериев Фишера оценить целесообразность включения фактора  $x_1$  после  $x_2$  и фактора  $x_2$  после  $x_1$ .

7. Оценить значимость множественных коэффициентов регрессии с помощью  $t$ -критерия Стьюдента.

8. Написать выводы по представленным данным и результатам расчетов.

### Задание 17

В системе *Statistica* создать файл с именем *Reklama.sta*. Сохранить файл в своей папке. Внести исходную информацию о длине и ширине рекламных баннеров (таблица 1). Провести анализ зависимости цены рекламы от длины при фиксированной ширине. Представить данные в виде диаграммы рассеяния с соответствующими заголовками осей и уравнением.

Таблица 1 – Данные для нахождения площади рекламных баннеров и цены рекламы

Ширина	Длина	Площадь	Цена
47	35		
47	73		
47	111		
47	149		
47	187		
47	225		
47	263		
47	301		

### Задание 18

Из базы данных кафедры статистики и прикладной математики в *Statistica* загрузить файл *Excel Nedvig.xls*. Провести иерархическую классификацию недвижимости, используя правило объединения (метод) Варда и Евклидову меру близости. В качестве переменных для анализа выбрать только количественные переменные для однокомнатных квартир: общую площадь, жилую площадь, площадь кухни, цену (V5, V6, V7, V8).

### Задание 19

Загрузить файл с данными о стоимости жилья в г. Краснодаре Nedvig.xls (база данных кафедры статистики и прикладной математики). Провести иерархическую классификацию недвижимости, используя правило объединения (метод) Варда и Евклидову меру близости по вариантам, указанным в таблице 1.

Таблица 1 – Исходная информация для выполнения заданий

№ варианта	Общая площадь, м <sup>2</sup>	Число комнат
1, 12	Менее 54	1
2, 13	Менее 40	1
3, 14	Менее 45	1
4, 15	38–48	1
5, 16	Не менее 35	1
6, 17	45–70	2
7, 18	33–75	2
8, 19	37–98	2
9, 20	30–130	3
10, 21	65–110	3
11, 22	От 50 до 70	3

### **Задание 20**

Сельскохозяйственное предприятие может реализовать некоторую продукцию:

- A1) сразу после уборки;
- A2) в зимние месяцы;
- A3) в весенние месяцы.

Прибыль зависит от цены реализации в данный период времени, затратами на хранение и возможных потерь. Размер прибыли, рассчитанный для разных состояний-соотношений дохода и издержек ( $S1$ ,  $S2$  и  $S3$ ), в течение всего периода реализации, представлен в виде матрицы (млн. руб.)

	$S1$	$S2$	$S3$
$A1$	2	-3	7
$A2$	-1	5	4
$A3$	-7	13	-3

Определить наиболее выгодную стратегию по всем критериям (критерий Байеса, критерий Лапласа, максиминный критерий Вальда), если вероятности состояний спроса: 0,2; 0,5; 0,3; коэффициент пессимизма  $C = 0,4$ ; коэффициент достоверности информации о состояниях спроса  $u = 0,6$ .

### **Задание 21**

Сельскохозяйственное предприятие может реализовать некоторую продукцию:

$A1$ ) сразу после уборки;

$A2$ ) в зимние месяцы;

$A3$ ) в весенние месяцы.

Прибыль зависит от цены реализации в данный период времени, затратами на хранение и возможных потерь. Размер прибыли, рассчитанный для разных состояний-соотношений дохода и издержек ( $S1$ ,  $S2$  и  $S3$ ), в течение всего периода реализации, представлен в виде матрицы (млн. руб.)

	$S1$	$S2$	$S3$
$A1$	2	-3	7
$A2$	-1	5	4
$A3$	-7	13	-3

Определить наиболее выгодную стратегию по следующим критериям

(Критерий пессимизма-оптимизма Гурвица), если вероятности состояний спроса: 0,2; 0,5; 0,3; коэффициент пессимизма  $C = 0,4$ ; коэффициент достоверности информации о состояниях спроса  $u = 0,6$ .

## Задание 22

Сельскохозяйственное предприятие может реализовать некоторую продукцию:

A1) сразу после уборки;

A2) в зимние месяцы;

A3) в весенние месяцы.

Прибыль зависит от цены реализации в данный период времени, затратами на хранение и возможных потерь. Размер прибыли, рассчитанный для разных состояний-соотношений дохода и издержек ( $S1$ ,  $S2$  и  $S3$ ), в течение всего периода реализации, представлен в виде матрицы (млн. руб.)

	$S1$	$S2$	$S3$
$A1$	2	-3	7
$A2$	-1	5	4
$A3$	-7	13	-3

Определить наиболее выгодную стратегию по следующим критериям

(Критерий Ходжа-Лемана, критерий минимаксного риска Сэвиджа), если вероятности состояний спроса: 0,2; 0,5; 0,3; коэффициент пессимизма  $C = 0,4$ ; коэффициент достоверности информации о состояниях спроса  $u = 0,6$ .

## 6 Перечень вопросов для изучения дисциплины

1. Что Вы понимаете под репрезентативностью выборки?
2. Что такое гистограмма частостей, статистическим аналогом чего она является?
3. Что такое кумулята частостей, статистическим аналогом чего она является?
4. Как записывается выборочное среднее для не сгруппированных данных?
5. Как записывается выборочное среднее для сгруппированных данных?
6. Как записывается несмещенная выборочная дисперсия для не сгруппированных данных ?
7. Что такое выборочная мода (можно на примере)? Оценкой какого параметра она является?
8. Что такое выборочная медиана (можно на примере)? Оценкой какого параметра она является?
9. Что характеризуют асимметрия и эксцесс? Как записываются выборочные асимметрия и эксцесс?
10. Для чего используется коэффициент вариации?
11. Каков содержательный смысл распределения Бернулли? Приведите пример сл.в., имеющей распределение Бернулли.
12. Каков содержательный смысл распределения равномерного распределения? В какой типичной ситуации оно появляется?
13. Что такое нормальное распределение? В какой типичной ситуации оно появляется?
14. Что происходит с графиком плотности нормального распределения если увеличивать мат.ожидание? Дисперсию?
15. Что такое распределение Стюдента? Что происходит с графиком плотности распределения Стюдента при увеличении числа степеней свободы?

16. Что такое распределение  $\chi^2$ ? Что происходит с графиком плотности распределения  $\chi^2$  при увеличении числа степеней свободы?

17. Что такое распределение Фишера?

18. Что такое доверительный интервал? Для чего он нужен?

19. Какое распределение используется при построении доверительного интервала для матожидания? Как записывается доверительный интервал для матожидания?

20. Во сколько раз следует увеличить объем выборки, чтобы на порядок уменьшить длину доверительного интервала для матожидания?

21. В каком случае при построении доверительного интервала требование нормальности существенно?

22. Какое распределение используется при построении доверительного интервала для дисперсии?

23. Что происходит с длиной доверительного интервала при увеличении доверительной вероятности?

24. Что такое статистическая гипотеза?

25. Что такое параметрическая гипотеза? Приведите пример.

26. Что такое непараметрическая гипотеза? Приведите пример.

27. Что такое простая гипотеза? сложная гипотеза?

28. Что такое критическая область?

29. Что такое наилучшая критическая область (область принятия решения)?

30. Что такое ошибка первого рода? второго рода при проверке статистических гипотез?

31. Что происходит с вероятностью ошибки второго рода при уменьшении вероятности ошибки первого рода?

32. Что такое критерии согласия?

33. Какая гипотеза проверяется с помощью критерия согласия  $\chi^2$  Пирсона? Как следует группировать данные для применения этого критерия?

34. Параметрические или непараметрические гипотезы проверяются с помощью критерия Пирсона? Обоснуйте ответ.

35. В чем «идея» критерия знаков?
36. В чем «идея» критерия знаковых ранговых сумм?
37. В чем разница между парными и независимыми наблюдениями? Приведите примеры.
38. Какие критерии проверки однородности Вы знаете для парных наблюдений?
39. Какие критерии проверки однородности Вы знаете для независимых (непарных) наблюдений?
40. В чем состоят основная и альтернативная гипотезы в однофакторном дисперсионном анализе?
41. Каким условиям должны удовлетворять выборки, чтобы можно было воспользоваться однофакторным дисперсионным анализом?
42. Что дают критерии Барлетта и Кочрена для однофакторного анализа?
43. Таблицы какого распределения используются для принятия решения в одно-(много) факторном дисперсионном анализе?
44. Каким критерием следует воспользоваться, если при однофакторном анализе Вы обнаружили, что нет нормальности?
45. С помощью какого критерия можно выявить связь между двумя качественными признаками?
46. Что характеризует коэффициент Крамера?
47. Что Вы понимаете под порядковым признаком?
48. С помощью какого критерия можно выявить связь между двумя порядковыми признаками?
49. Для чего используются коэффициенты Спирмена и Кэнделла?
50. Что характеризует выборочный коэффициент корреляции?
51. С помощью какого критерия можно выявить связь между двумя количественными признаками?
52. Что такое «ложная корреляция»? Приведите пример.
53. Что характеризует частный коэффициент корреляции?
54. Что такое функция регрессии? Выборочная функция регрессии?

55. В каких случаях выборочную функцию регрессии следует искать в виде линейной функции?
56. Какой метод используется для нахождения коэффициентов линейной выборочной функции регрессии?
57. Что такое остаточная дисперсия? Что она характеризует?
58. Что можно сказать про остаточную дисперсию, если выборочный коэффициент корреляции близок к 1? ; к -1 ?
59. Что характеризует коэффициент детерминации  $R^2$ ?
60. Что происходит с коэффициентом детерминации, если в модели увеличивается число независимых переменных?
61. Как оценить качество выбранной регрессионной модели?
62. Что следует проверить при анализе остатков?
63. В чем состоит задача кластерного анализа?
64. В каких случаях в качестве меры близости между объектами используется обычное евклидово расстояние, а в каких — нормализованное евклидово?
65. Для каких признаков обычно используется Хеммингово расстояние?
66. Что можно использовать в качестве расстояния между признаками (не объектами)?
67. Как записывается расстояние между двумя кластерами по принципу «ближнего соседа»?
68. Как записывается расстояние между двумя кластерами по принципу «дальнего соседа»?
69. Как записывается расстояние между двумя кластерами с использованием расстояния «по центрам тяжести»?
70. В чем состоит идея агломерационных методов кластерного анализа?
71. В чем состоит идея метода Варда?
72. Что такое дендрограмма (можно на примере) ?
73. Как выбираются векторы главных компонент в  $k$ -мерном пространстве?
74. Для чего используется метод главных компонент?

75. Как связаны собственные значения и собственные вектора ковариационной матрицы с главными компонентами?
76. Как выбрать количество оставляемых главных компонент?
77. Почему метод главных компонент можно использовать как средство борьбы с мультиколлинеарностью? Каким образом?
78. В чем состоит цель факторного анализа?
79. Какая задача решается методами дискриминантного анализа?
80. Чем различаются задачи, решаемые методами дискриминантного и кластерного анализа?
81. Что Вы понимаете под непараметрическими методами дискриминантного анализа?
82. Что Вы понимаете под параметрическими методами дискриминантного анализа?
83. Какие функции используются в качестве дискриминантных в параметрическом методе?
84. При каком предположении в параметрическом методе дискриминантные функции получаются линейными?

## Список рекомендуемой литературы

1. Аскеров, П. Ф. Общая и прикладная статистика : учебник для студентов высшего профессионального образования / П.Ф. Аскеров, Р.Н. Пахунова, А.В. Пахунов ; под общ. ред. Р.Н. Пахуновой. — Москва : ИНФРА-М, 2019. — 272 с. + Доп. материалы [Электронный ресурс; Режим доступа <http://www.znanium.com>]. — (Высшее образование: Бакалавриат). — [www.dx.doi.org/10.12737/748](http://www.dx.doi.org/10.12737/748). - ISBN 978-5-16-006669-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1008000> – Режим доступа: по подписке.

2. Введение в математическое моделирование : учебное пособие / В. Н. Ашихмин, М. Б. Гитман, И. Э. Келлер [и др.] ; под ред. П. В. Трусова. - Москва : Логос, 2020. - 440 с. - ISBN 978-5-98704-637-1. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1211604> – Режим доступа: по подписке.

3. Гармаш, А. Н. Математические методы в управлении: Учебное пособие / А.Н. Гармаш, И.В. Орлова. - М.: Вузовский учебник: НИЦ Инфра-М, 2018. - 272 с. ISBN 978-5-9558-0200-8. - Текст : электронный. - URL: <https://znanium.com/catalog/product/934346> – Режим доступа: по подписке.

4. Григорьев, А. А. Методы и алгоритмы обработки данных : учебное пособие / А. А. Григорьев, Е. А. Исаев. — 2-е изд., перераб. и доп. — Москва : ИНФРА-М, 2021. — 383 с. — (Высшее образование: Бакалавриат). - ISBN 978-5-16-015581-4. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1032305> – Режим доступа: по подписке.

5. Информационные системы и цифровые технологии. Практикум : учебное пособие. Часть 1 / под общ. ред. проф. В.В. Трофимова, доц. М.И. Барабановой. — Москва : ИНФРА-М, 2021. — 212 с. — (Высшее образование: Бакалавриат). - ISBN 978-5-16-109660-4. - Текст : электронный. - URL:

<https://znanium.com/catalog/product/1731904> – Режим доступа: по подписке.

6. Козлов, А. Ю. Статистический анализ данных в MS Excel : учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. — Москва : ИНФРА-М, 2021. — 320 с. — (Высшее образование: Бакалавриат). — DOI 10.12737/2842. - ISBN 978-5-16-004579-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1684740> – Режим доступа: по подписке.

7. Статистика : учебник / В.В. Глинский, В.Г. Ионин, Л.К. Серга [и др.] ; под ред. В.Г. Ионина. — 4-е изд., перераб. и доп. — Москва : ИНФРА-М, 2021. — 355 с. — (Высшее образование: Бакалавриат). — DOI 10.12737/25127. - ISBN 978-5-16-012070-6. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1228803> – Режим доступа: по подписке.

8. Статистика : учебное пособие / А. М. Восковых, Т. А. Журкина, С. Л. Закупнев [и др.] ; под редакцией И. М. Сурков. — Воронеж : Воронежский Государственный Аграрный Университет им. Императора Петра Первого, 2017. — 244 с. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <https://www.iprbookshop.ru/72755.html> — Режим доступа: для авторизир. пользователей

# **ПРИКЛАДНАЯ СТАТИСТИКА**

*Методические указания*

Составитель: **Яроменко** Наталья Николаевна

Усл. печ. л. – 2,38.