К. А. СОХТ, Е. И. ТРУБИЛИН, В. И. КОНОВАЛОВ

СТАТИСТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЙ ПРОЦЕССОВ И МАШИН В АГРОБИЗНЕСЕ

МИНИСТЕРСТВО СЕЛЬСКОГО ХОЗЯЙСТВА
РОССИЙСКОЙ ФЕДЕРАЦИИ

ФГБОУ ВПО «Кубанский государственный
аграрный университет»

К. А. СОХТ, Е. И. ТРУБИЛИН, В. И. КОНОВАЛОВ

СТАТИСТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЙ ПРОЦЕССОВ И МАШИН В АГРОБИЗНЕСЕ

Учебное пособие

Рекомендовано Учебно-методическим объединением вузов Российской Федерации по агроинженерному образованию в качестве учебного пособия для студентов высших учебных заведений, обучающихся по направлению «Агроинженерия»

Краснодар
КубГАУ
2016

УДК 519.24 (075.8)
ББК 40.7
С68

Р е ц е н з е н т ы :

В. Б. Рыков – зам. директора по научной работе ГНУ СКНИИМЭСК Россельхозакадемии, профессор кафедры Технологий и оборудования переработки продукции АПК, д-р техн. наук;

Е. И. Виневский – зав. лабораторией машинных агропромышленных технологий ГНУ ВНИИТТИ Россельхозакадемии, д-р техн. наук.

Сохт К. А.

С68 Статистические методы исследований процессов и машин в агробизнесе : учеб. пособие / К. А. Сохт, Е. И. Трубилин, В. И. Коновалов. – Краснодар : КубГАУ, 2016. – 217 с.

ISBN 978-5-94672-998-7

В учебном пособии рассмотрены основные вопросы по применению методов математической статистики при исследовании и испытании процессов и машин в агробизнсе. Приведены практические примеры, наглядно демонстрирующие применение предлагаемых методов. После каждой главы дан перечень вопросов для закрепления пройденного материала.

Предназначено для научных и инженерно-технических работников, а также студентов высших учебных заведений, обучающихся по направлениям 35.03.06 «Агроинженерия», 35.04.06 «Агроинженерия», 35.06.04 «Технологии, средства механизации и энергетическое оборудование в сельском, лесном и рыбном хозяйстве» и 23.05.01 «Наземные транспортно-технологические средства».

УДК 519.24 (075.8)
ББК 40.7

ISBN 978-5-94672-998-7

ВВЕДЕНИЕ

Сельское хозяйство является одной из отраслей народного хозяйства, которые наиболее насыщены непрерывно меняющимися условиями (факторами), определяющими функционирование выполняемых процессов сельскохозяйственными машинами и агрегатами. Это постоянно меняющиеся тип почвы, ее физико-механические свойства, влажность и твердость почвы, плодородие, урожайность сельскохозяйственных культур, рельеф поля и многое другое. Воздействие этих факторов на исследуемые процессы и машинно-тракторные агрегаты зачастую настолько превосходит влияние других активных факторов, регулируемых исследователем, что выделение степени их влияния иногда становится проблематичным. В таких достаточно сложных условиях получение достоверных результатов, как правило, связано с привлечением сложных методик, построенных на применении теории вероятностей и математической статистики. В настоящее время математическая статистика широко применяется в экспериментальных исследованиях. Нет, пожалуй, ни одного исследования в области сельского хозяйства без применения статистических методов.

Статистические методы позволяют выделить полезную искомую информацию (сигнал) из случайной наблюдаемой картины, повысить уровень достоверности получаемых результатов. К сожалению, часто встречаются случаи, например, когда при исследовании орудий для обработки почвы в связи с множеством неуправляемых факторов, получают описание невоспроизводимых ситуаций, что приводит лишь к частному случаю. В связи с вышеизложенным следует признать обучение студентов и специалистов статистическим методам исследований необходимым и вполне оправданным.

Статистические методы исследований являются мощным инструментом для каждого исследователя. К сожалению, иногда исследователь, не обладая достаточными знаниями в области планирования эксперимента и обработки ее данных, обращается к специалисту-теоретику по теории вероятностей и математической статистике, который не имеет специальных знаний в исследуемой области, что приводит тоже к серьезной путанице по причине недостаточного взаимопонимания. Поэтому для успешного применения статистических методов необходимы глубокие знания в области экспериментирования и одновременно достаточно хорошие знания по методам статистических исследований. Это позволит оптимально применять методы и лишь в исключительных случаях обращаться к специалисту по теории вероятностей и математической статистике.

1 ОСНОВНЫЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН И ИХ АНАЛИЗ

1.1 Множества. Элементы теории множеств

Современное изложение теории вероятностей и математической статистики в основном базируется на теории множеств.

Конечная или бесконечная совокупность объектов (элементов), которые имеют одно или больше общих свойств (признаков) называется множеством. Пример: множество тракторов, множество автомобилей, множество людей на планете Земля и т. д. С понятием множества связано понятие подмножество. **Если любой элемент множества *B* является элементом и множества *A*, то множество *B* называется подмножеством множества *A*.** Обозначается это так: $B \subset A$ (множество B принадлежит множеству A) или $A \supset B$ (множество A включает в себя множество B). Данное множество можно расчленить на два или более подмножеств по каким-либо дополнительным признакам. Например, множество тракторов всех видов можно разделить на два подмножества по виду ходовой части – колесные и гусеничные.

Множество, которое не содержит ни одного элемента, называется пустым и обозначается так Ø.

Множество, которое включает в себя все рассматриваемые элементы, называется множеством элементарных исходов.

Над множествами можно производить три основных действия (операции): объединение (операция «или»), пересечение (операция «и») и дополнение (операция «не»).

Объединением двух множеств *A* и *B* называется множество C_1, которое состоит из элементов, входящих в A или B или в оба множества. Другими словами множество C_1 охватывает все элементы A и B, но те из них, которые одинаковы для обоих множеств, в объединении C_1 участвуют только один раз.

Объединение C_1 множеств A и B обозначается так:

$$C_1 = A + B, \text{ или } C_1 = A \cup B.$$

Пример 1. Множество A включает в себя только четное число очков, появляющихся при бросании игральной кости, т. е. $A = \{2, 4, 6\}$, а B – число очков, превышающее 3, т. е. $B = \{4, 5, 6\}$. Определить объединение C_1 этих множеств.

Решение. Согласно определению объединения множеств A и B будет

$$C_1 = A + B = \{2, 4, 6\} \cup \{4, 5, 6\} = \{2, 4, 5, 6\}.$$

Аналогично определяется объединение множеств, число которых больше двух.

Пересечением двух множеств *A* и *B* называется множество C_2, элементы которого являются общими для *A* и *B* и обозначается:

$$C_2 = A \cap B, \text{ или } C_2 = AB, \text{ и читается так: } C_2 = A \text{ и } B.$$

Для множества в вышеприведенном примере общими будут только элементы 4 и 6, т. е.

$$C_2 = A \cap B = \{2, 4, 6\} \cap \{4, 5, 6\} = \{4, 6\}.$$

Если множества A и B не имеют одинаковых элементов, то их пересечение будет **пустым множеством**. Такие множества называются взаимно исключающимися.

Аналогично определяется пересечение множеств, число которых больше двух.

Дополнением множества *A* называется множество C_3, которое включает в себя все элементы множества элементарных исходов, которые не содержатся в *A*.

Дополнение обозначается так:

$$C_3 = \overline{A}.$$

Дополнение множества *A* из рассматриваемого примера будет включать в себя нечетные числа от 1 до 6, т. е.

$$C_3 = \overline{A} = \{1, 3, 5\}.$$

Операции, которые производят над множествами, удобно иллюстрировать с помощью так называемых диаграмм Вена.

1.2 Случайные величины и случайные события

На каждый исследуемый объект действует большое количество управляемых и неуправляемых случайных факторов, от которых зависит значение интересующего нас показателя. Значения показателя, получающиеся как результат прямого действия или взаимодействия этих факторов, называются **случайными событиями**.

Случайная величина – ряд значений случайных событий, получаемых в результате серии испытаний.

Другими словами случайная величина – значение интересующего показателя, получаемая в результате проведения опыта (испытания) с вероятностью менее 100 %. Тогда случайное событие – событие, которое в результате проведения опытов (испытаний) может произойти или не произойти с определенной вероятностью.

Случайные величины бывают трех видов: **дискретные, непрерывные и дискретно-непррывные.**

Дискретными называют такие случайные величины, которые могут принимать отдельные значения (случайные события) с определенной вероятностью. Число возможных значений дискретной случайной величины может быть **конечным** или **бесконечным** и они могут быть названы заранее.

Непрерывными случайными величинами называют такие, которые могут принимать любые численные значения в данном конечном или бесконечном интервале. Результат опыта при непрерывных случайных величинах получаются не подсчетом, а путем измерения. Например, глубина обработки почвы, глубина сева семян, потери при уборке и другие.

Случайные величины обозначаются прописными буквами X, Y, Z, а их конкретные значения – строчными буквами x, y, z с соответствующими индексами: 1, 2, 3, ..., n.

В математическом анализе зависимые переменные величины изменяют свое значение только при изменении условий испытания, т. е. от значений регулируемых переменных факторов. В отличие от этого случайные величины могут принимать различные значения даже при неизменных значениях основных и регулируемых факторов. Причина этого – не учитываемые и не поддающиеся экспериментатору случайные переменные факторы, количество которых даже невозможно подсчитать.

1.3 Генеральная совокупность. Выборка и выборочный метод

Наблюдения за интересующими нас показателями – зависимыми переменными, как правило, содержат бесконечное множество значений. И вполне понятно, что полностью охватить для обработки всю эту совокупность невозможно. Эти совокупности представляют бесконечные или очень многочисленные совокупности, называемые генеральной совокупностью. По сути же **генеральной совокупностью** называется множество всех возможных однородных элементов. Следовательно генеральной совокупностью являются и весь натуральный ряд чисел (бесконечная генеральная совокупность) и все коровы одного стада, все студенты одного и того же университета, все растения на данном конкретном поле, множество деталей, изготовленных на конкретном заводе за смену – конечные генеральные совокупности. Таких примеров можно привести много.

Обследовать бесконечные генеральные совокупности и получить необходимые данные чаще всего физически невозможно. Поэтому для исследования генеральной совокупности используют выборку из нее, т. е. выборочную совокупность. **Выборочная совокупность или выборка** – это случайно выбранные элементы генеральной совокупности, дающие представлеие о ней, и выбранные так, что каждый элемент генеральной совокупности имеет одинаковую вероятность попасть в выборку независимо от других элементов. **Исследование генеральных совокупностей с помощью выборок называется выборочным методом.**

Выборочный метод является основой математической статистики. Выборка обрабатывается с помощью методов математической статистики с целью получения закономерностей распределения генеральной совокупности и ее числовых характеристик. Выбор элементов выборки необходимо выполнить так, чтобы значения факторов не менялись от времени, температуры или координат места расположения и других факторов от которых зависят неконтролируемые факторы.

1.4 Основные законы распределения случайных величин

Все случайные величины подчиняются определенному закону распределения, согласно которому заранее можно вычислить при известных параметрах вероятность возникновения любого значения случайной величины. Закон распределения случайной величины считается заданным, если указана вся совокупность значений выборки, и указан способ количественного определения вероятности попадания случайной величины в любую область множества возможных значений случайной величины. Вероятность попадания в заданную область может быть определена выражением:

$$p_m = \lim_{n \to \infty} \frac{n_m}{n}, \tag{1.1}$$

где n_m – количество наблюдений случайной величины, оказавшихся в заданной области;

n – общее число наблюдений (частотное определение вероятности).

Для дискретных случайных величин закон распределения имеет две формы:

- ряд распределения;
- интегральная функция распределения.

Интегральная функция распределения и дифференциальная функция плотности распределения вероятностей значений случайной величины – формы законов распределения непрерывных случайных величин.

Ряд распределения содержит все возможные значения дискретной случайной величины и соответствующие им вероятности. В таблице 1.1 дан ряд распределения случайной величины X, которая принимает значения $x_1, x_2, \ldots, x_n$ и имеет соответствующие им вероятности $p_1, p_2, \ldots, p_n$

Таблица 1.1 – Ряд распределения дискретной случайной величины

X	x_1	x_2	x_3	…	…	x_n
p	p_1	p_2	p_3	…	…	p_n

Как уже отмечалось, отдельные численные значения $x_1, x_2, \ldots, x_n$ дискретной случайной величины можно рассматривать как случайные события. Но так как в любом отдельном опыте случайная величина может принимать одно и только одно из своих значений, то эти значения образуют полную группу попарно несовместимых событий. Следовательно, сумма вероятностей во втором ряду таблицы 1.1. будет равна:

$$p_1 + p_2 + ... + p_n = 1. \qquad (1.2)$$

Функция распределения *F(x)* выражает вероятность *P* того, что случайная величина *X* принимает значения меньше данного действительного числа *x*, т. е.:

$$F_x = P(X < x), \qquad (1.3)$$

и имеет следующие свойства:

1) численные значения *F(x)* принадлежат интервалу [0, 1]:

$$0 \leq F_x \leq 1;$$

2) функция *F(x)* неубывающая, т. е. если $x_2 > x_1$, то:

$$F(x_1) > F(x_2);$$

3) если возможные значения случайной величины *X* принадлежат интервалу [*a*, *b*], то:

$$F(x) = 0 \text{ при } x \leq a \text{ и } F(x) = 1 \text{ при } x \geq b.$$

Первое свойство связано с самим определением функции распределения как вероятности.

Второе свойство доказывается, если используются несовместимые события: случайная величина *X* принимает значение меньше x_1 и случайная величина *X* принимает значение большее или равное x_1 и меньше x_2. Тогда событие: «случайная величина *X* принимает значение меньше x_2» является суммой двух написанных выше несовместимых событий, поэтому можем записать равенство:

$$P(X < x_2) = P(X < x_1) + P(x_1 \le X < x_2), \qquad (1.4)$$

или

$$F(x_2) - F(x_1) = P(x_1 \le X < x_2). \qquad (1.5)$$

Так как вероятность всегда неотрицательное число, то:

$$F(x_2) - F(x_1) \ge 0, \text{ т. е } F(x_2) \ge F(x_1). \qquad (1.6)$$

Из $F(x_2) - F(x_1) = P(x_1 \le X < x_2)$, видно, что вероятность того, что случайная величина X попадает в интервал (x_1, x_2), равна приращению функции распределения в этом интервале.

Плотность распределения *f*(*x*) – вторая форма закона распределения непрерывных случайных величин является первой производной функции распределения:

$$f(x) = \frac{dF(x)}{dx} = F'(x). \qquad (1.7)$$

Поэтому *f(x)* называется еще дифференциальной функцией.

Если известна *f(x)*, то согласно формуле Ньютона – Лейбница:

$$F(x_2) - F(x_1) = \int_{x_1}^{x_2} F(x)\,dx = \int_{x_1}^{x_2} f(x)\,dx, \qquad (1.8)$$

или

$$P(x_1 < X < x_2) = \int_{x_1}^{x_2} f(x)\,dx, \tag{1.9}$$

т. е. вероятность попадания непрерывной случайной величины X в интервал (x_1, x_2) равна значению определенного интеграла от плотности распределения $f(x)$ вычисленному в пределах от x_1 до x_2. Геометрически это означает, что вероятность попадания X в интервал (x_1, x_2) равна площади криволинейной трапеции, ограниченной сверху графиком плотности распределения $f(x)$, снизу осью абсцисс x, а с боков вертикальными прямыми $x = x_1$ и $x = x_2$.

Если в последнюю формулу подставим $x_1 = \infty$, $x_2 = x$, то получим:

$$F(x) = \int_{-\infty}^{x} f(x)dx. \tag{1.10}$$

Отсюда вытекает второе название функции $F(x)$ – **интегральная функция**.

При достаточно малом dx вероятность попадания непрерывной случайной величины X в интервал $(x, x+dx)$ приблизительно равна площади прямоугольника с основанием dx и высотой $f(x)$.

1.5 Построение выборочных распределений

Многие исследования в области механизации сельскохозяйственного производства сопровождаются сбором обширного материала случайных величин, в целях получения характеристик обрабатываемого материала или других показателей, характеризующих исследуемый объект. В таблице 1.2 приведены данные о высоте прикрепления початков к стеблю кукурузы. Эти данные необходимы для проектирования кукурузной жатки – приставки к комбайну для уборки кукурузы.

Таблица 1.2 – Высота прикрепления початков кукурузы к стеблю

Номер п/п	Значение	Номер п/п	Значение	Номер п/п	Значение	Номер п/п	Значение	Номер п/п	Значение
1	**76**	21	**45**	41	**70**	61	**77**	81	**89**
2	**82**	22	**59**	42	**67**	62	**76**	82	**85**
3	**80**	23	**60**	43	**100**	63	**88**	83	**93**
4	**68**	24	**63**	44	**103**	64	**89**	84	**90**
5	**69**	25	**78**	45	**69**	65	**63**	85	**79**
6	**74**	26	**87**	46	**72**	66	**82**	86	**83**
7	**72**	27	**94**	47	**74**	67	**80**	87	**91**
8	**69**	28	**91**	48	**66**	68	**81**	88	**87**
9	**80**	29	**88**	49	**67**	69	**77**	89	**89**
10	**79**	30	**90**	50	**72**	70	**80**	90	**94**
11	**90**	31	**79**	51	**72**	71	**79**	91	**92**
12	**109**	32	**84**	52	**68**	72	**78**	92	**91**
13	**99**	33	**84**	53	**80**	73	**83**	93	**76**
14	**100**	34	**108**	54	**81**	74	**92**	94	**79**
15	**115**	35	**83**	55	**84**	75	**93**	95	**73**
16	**68**	36	**84**	56	**77**	76	**81**	96	**84**
17	**70**	37	**99**	57	**79**	77	**82**	97	**79**
18	**72**	38	**98**	58	**81**	78	**86**	98	**84**
19	**73**	39	**102**	59	**84**	79	**89**	99	**79**
20	**70**	40	**101**	60	**76**	80	**93**	100	**84**

Для построения графика интегральной функции распределения построим график ранжированного ряда 100 измерений высоты прикрепления початков кукурузы к стеблю (рисунок1.1), данные занумерованы от минимального значения измерений $x_{min} = 45$ до $x_{max} = 115$. Ранжированный ряд высоты прикрепления початков кукурузы к стеблю представлен в таблице 1.3.

Таблица 1.3 – Ранжированный ряд высоты прикрепления початков кукурузы к стеблю

Номер п/п	Значение	Номер п/п	Значение	Номер п/п	Значение	Номер п/п	Значение	Номер п/п	Значение
1	**45**	21	**72**	41	**79**	61	**84**	81	**91**
2	**59**	22	**72**	42	**79**	62	**84**	82	**91**
3	**60**	23	**73**	43	**79**	63	**84**	83	**92**
4	**63**	24	**73**	44	**80**	64	**84**	84	**92**
5	**63**	25	**74**	45	**80**	65	**84**	85	**93**
6	**66**	26	**74**	46	**80**	66	**84**	86	**93**
7	**67**	27	**76**	47	**80**	67	**85**	87	**93**
8	**67**	28	**76**	48	**80**	68	**86**	88	**94**
9	**68**	29	**76**	49	**81**	69	**87**	89	**94**
10	**68**	30	**76**	50	**81**	70	**87**	90	**98**
11	**68**	31	**77**	51	**81**	71	**88**	91	**99**
12	**69**	32	**77**	52	**81**	72	**88**	92	**99**
13	**69**	33	**77**	53	**82**	73	**89**	93	**100**
14	**69**	34	**78**	54	**82**	74	**89**	94	**100**
15	**70**	35	**78**	55	**82**	75	**89**	95	**101**
16	**70**	36	**79**	56	**83**	76	**89**	96	**102**
17	**70**	37	**79**	57	**83**	77	**90**	97	**103**
18	**72**	38	**79**	58	**83**	78	**90**	98	**108**
19	**72**	39	**79**	59	**84**	79	**90**	99	**109**
20	**72**	40	**79**	60	**84**	80	**91**	100	**115**

Сгруппировав значения измерений высоты прикрепления початков кукурузы к стеблю в заданном интервале в столбики пропорционально частотам, **получим гистограмму**. А соединив средние значения групп получим ломанную линию, которая называется **полигоном** (рисунок 1.2).

На графике ранжированных значений высоты прикрепления початков кукурузы к стеблю, данные, сгруппированные наиболее густо в середине ранжированного ряда, начинают постепенно становиться более редкими по бокам. Соответствие между графиком ранжированного ряда и гистограммой очевидно (рисунок 1.1).

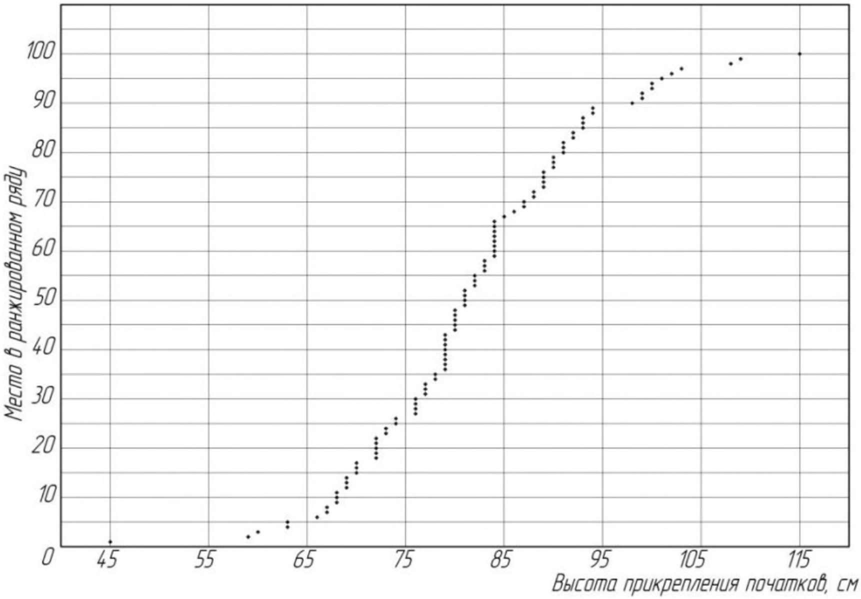


Рисунок 1.1 – Графическое изображение 100 ранжированных значений высоты прикрепления початков кукурузы к стеблю

Это значит, что график распределения выборочных данных на рисунке 1.1 по форме соответствует интегральной функции распределения, а она в свою очередь тоже соответствует гистограмме и полигону распределения, как прообразу дифференциальной функции плотности распределения.

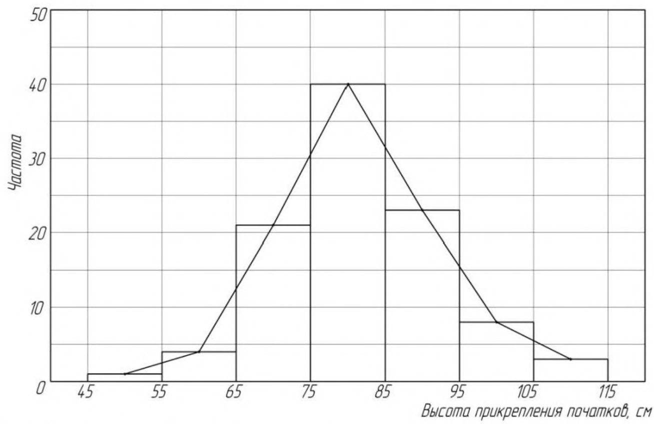


Рисунок 1.2 – Гистограмма и полигон высоты прикрепления початков кукурузы к стеблю

При построении приведенных графиков распределения следует внимательно проверить отдельно каждое наблюдение и в случае резкого отклонения отдельных наблюдений от общего поведения остальных следует внимательно еще раз изучить природу возникновения этого наблюдения и при обнаружении достоверных причин, недопустимых для проводимых измерений, исключить из общего массива.

1.6 Числовые характеристики случайной величины

При изучении какого-нибудь процесса необязательно знать подробно случайные величины, которые его описывают, а вполне достаточно знать некоторые их особые численные значения и характеристики. Такими характеристиками являются числовые характеристики. Основными числовыми статистическими характеристиками случайной величины являются математическое ожидание μ и дисперсия D для генеральной совокупности или среднее арифметическое $\bar{X}$ и стандартное отклонение S для выборки.

Математическое ожидание случайной величины совпадает с ее первым начальным моментом и оценивается средним арифметическим значением выборки.

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} x_i. \tag{1.11}$$

Математическое ожидание случайной величины называется ее первый начальный момент. Главным показателем рассеивания случайной величины вокруг математического ожидания является его **дисперсия**. Она совпадает со вторым центральным моментом

$$D[X] = \sigma^2[X] = \frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}, \tag{1.12}$$

где x_i – текущее значение случайной величины;

$\sigma^2[X]$ – квадрат среднеквадратического отклонения;

(n-1) – число степеней свободы, т. е. число свободно варьирующих величин.

Тогда среднеквадратическое отклонение $\sigma[X]$ равно:

$$\sigma[X] = \sqrt{D[X]}. \tag{1.13}$$

Дисперсия и среднеквадратическое отклонение генеральной совокупности оцениваются **стандартным отклонением от арифметического среднего значения случайной величины выборки *S***.

$$S = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{X})^2}{n-1}}. \tag{1.14}$$

Квадрат стандартного отклонения S^2 **равен**:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{X})^2}{n-1}. \qquad (1.15)$$

Коэффициент вариации V – стандартное отклонение, выраженное в процентах к средней арифметической данной выборки. Коэффициент вариации – относительный показатель изменчивости

$$V = \frac{S}{\bar{X}} 100\,\%. \qquad (1.16)$$

Ошибка выборочной средней или ошибка выборки $S_{\bar{X}}$ **является мерой отклонения выборочной средней** $\bar{X}$ **от средней всей (генеральной) совокупности** μ**.**

Величина ошибки средней арифметической зависит от стандартного отклонения S и объема выборки n и определяется из выражения:

$$S_{\bar{X}} = \frac{S}{\sqrt{n}}. \qquad (1.17)$$

Эта же величина, выраженная в процентах от соответствующей средней, называется **относительной ошибкой выборочной средней.**

$$S_{\bar{X}} = \frac{S_{\bar{X}}}{\bar{X}} 100\%. \qquad (1.18)$$

Модой M **случайной величины** X называют ее численное значение $M(X)$, которому соответствует самая большая вероятность для дискретных случайных величин и максимум плотности распределения *f(x)* – для непрерывных (рисунок 1.3).

Медианой случайной величины *X* называют численное значение $x_{1/2}$ случайной величины для которого выполнено условие:

$$P(X < x_{1/2}) = P(X > x_{1/2}). \tag{1.19}$$

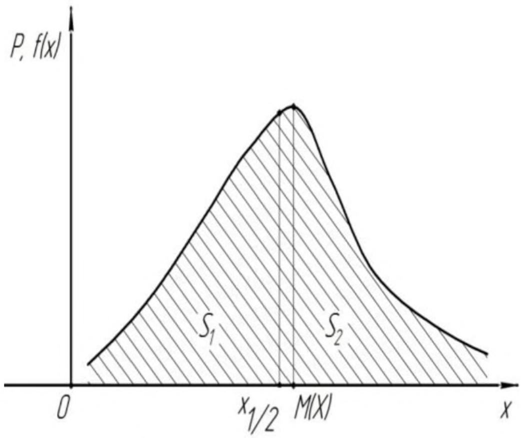


Рисунок 1.3 – Мода случайной величины *X*

Для непрерывных случайных величин медиана равна абсциссе точки, ордината которой делит площадь под кривой распределения на две равные части (рисунок 1.3).

Если распределение симметрично относительно арифметической средней, то медиана и мода совпадают.

Кроме характеристик положения – средних, типичных значений случайной величины, употребляется еще ряд характеристик, каждая из которых описывает то или иное свойство распределения. Понятия математического ожидания и дисперсии являются частными случаями более общего понятия для числовых характеристик случайных величин – **моментов распределения**. Чаще всего применяются на практике моменты двух видов: начальные и центральные.

Моменты относительно начала координат $x = 0$ называются **начальными моментами**. Для непрерывной случайной величины первым начальным моментом является интеграл:

$$M(X) = \int_{-\infty}^{\infty} xf(x)dx, \qquad (1.20)$$

где *f(x)* – плотность распределения.

Начальный момент *k*–го порядка, справедливое как для прерывных, так и для непрерывных величин определяется по выражению:

$$m_k(X) = \mu(X^k) = \sum_{i=1}^{n} x_i^k p_i, \qquad (1.21)$$

т. е. начальным моментом *k*-го порядка случайной величины X называется математическое ожидание *k*-й степени этой случайной величины. Начальный момент первого порядка есть центр распределения рассматриваемой случайной величины

Моменты относительно центра распределения $x = m$ называются **центральными моментами**. Центральный момент моментом *k*–го порядка определяется по выражению:

$$m_k(X) = \mu(X^k) = \mu\left[\left(X - \mu(X)\right)^k\right]. \qquad (1.22)$$

Таким образом, центральным моментом *k*-го порядка случайной величины X называется математическое ожидание *k*-й степени соответствующей центрированной случайной величины.

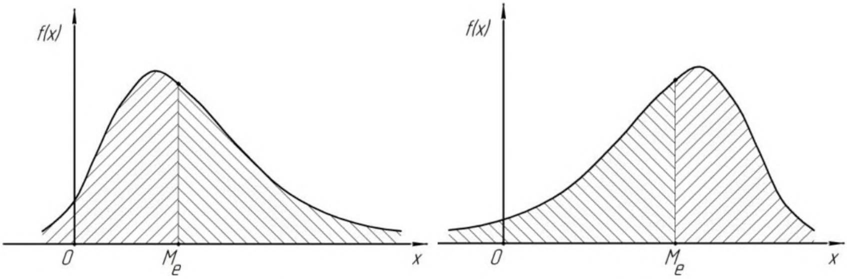
Центрирование случайной величины, равносильно переносу начала координат в среднюю, «центральную» точку, абсцисса которой равна математическому ожиданию. Для любой случайной величины X центральный момент первого порядка равен нулю. Помимо этого также видно, что дисперсия D есть не что иное, как центральный момент второго порядка.

Другими словами математическое ожидание – среднее вероятностное значение случайной величины в объеме выборки.

Показателями, характеризующими принадлежность распределения случайной величины к нормально распределенным, являются асимметрия и эксцесс. По отношению к эмпирическим распределениям асимметрия S_k выражается как отношение третьего центрального момента $m_з$ к кубу стандартного отклонения σ^3.

$$S_k = \frac{m_з}{\sigma^3}. \tag{1.23}$$

Для симметричного распределения сумма положительных кубов и сумма отрицательных кубов равны между собой, т. е. $S_k = 0$. У асимметрично распределенного распределения по одну сторону от центра группирования и моды расположена «длинная», а по другую сторону короткая часть распределения. Когда «длинная» часть распределения расположена справа от центра (рис 1.4 а), асимметрия считается положительной, так как сумма больших положительных отклонений превысит сумму кубов отрицательных отклонений. Отрицательной асимметрии соответствует форма распределения, при которой «длинная» часть расположена слева от центра (рисунок 1.4, *б*).



а) б)

Рисунок 1.4 – Распределение случайной величины с асимметрией:

а – положительная; *б* – отрицательная

В качестве характеристики заостренности вершины графика распределения по сравнению с графиком нормального распределения служит другой показатель, называемый **эксцессом** и определяемый выражением:

$$\varepsilon_k = \frac{m_4}{S^4} - 3, \qquad (1.24)$$

где m_4 – четвертый центральный момент.

В нормальном случае эксцесс равен нулю, так как для нормального распределения отношение $\dfrac{m_4}{S^4}$ равно 3.

Показатели асимметрии и эксцесса, отличные от нуля, указывают на отклонение рассматриваемого распределения по форме от нормального распределения.

Контрольные вопросы

1. Понятие случайного события.
2. Понятие случайной величины.
3. Что называется дискретной случайной величиной.
4. Что называется непрерывной случайной величиной.

5. Что называется генеральной совокупностью.

6. Что называется выборочной совокупностью (выборкой) случайной величины.

7. Что исследуется с помощью выборочного метода.

8. Понятие закона распределения случайной величины.

9. Какие существуют формы задания закона распределения для дискретных случайных величин.

10. Какие существуют формы задания закона распределения для непрерывных случайных величин.

11. Определение функции распределения случайной величины.

12. Как называется первая производная функции распределения случайной величины.

13. Как строят выборочное распределение случайной величины с помощью гистограммы и полигона.

14. Перечислить наиболее важные числовые характеристики случайной величины и формулы для их вычисления.

15 . Дать определение моды случайной величины.

16. Что называется медианой случайной величины.

2 ОСНОВНЫЕ ЗАКОНЫ РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

2.1 Нормальный закон распределения

Нормальный закон распределения играет важную роль при решении большинства задач, связанных с теорией вероятностей и математической статистики. Для этого распределения разработан весь спектр статистических задач.

Нормальная плотность вероятности распределения определяется выражением:

$$f(x;\mu;\sigma)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \qquad (2.1)$$

где x – случайная величина, $-\infty < x < +\infty$;

μ – математическое ожидание, $-\infty < \mu < +\infty$;

σ – среднеквадратичное отклонение, $\sigma > 0$

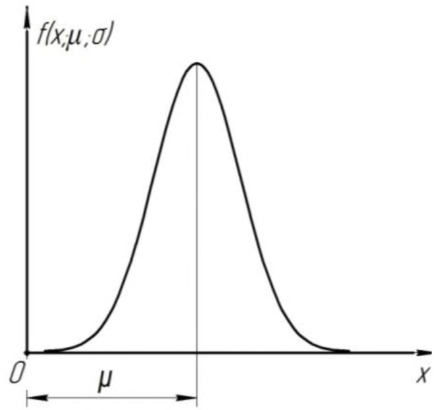


Рисунок 2.1 – График нормальной плотности вероятности

Соответствующая этой плотности дифференциальная кривая распределения показана на рисунке 2.1.

Из выражения (2.1) видно, что распределенная по нормальному закону случайная величина X может изменяться в пределах $-\infty < x < +\infty$. При любых значениях x и параметров μ и σ, всегда $f(x;\mu;\sigma)$ – положительно.

Интегральная функция распределения *F(x)* определяется выражением

$$F(x) = \int_{-\infty}^{x} f(x;\mu;\sigma)dx. \qquad (2.2)$$

Анализ показывает, что при любом *x*, каковы бы ни были параметры μ и σ, $f(x;\mu;\sigma) > 0$. С другой стороны, полная площадь под всей кривой равна 1. График интегральной функции распределения показан на рисунке 2.2.

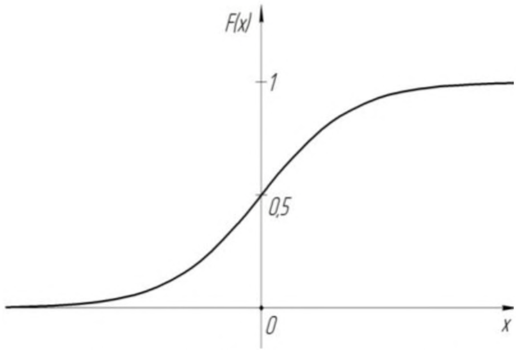


Рисунок 2.2 – График нормальной интегральной функции распределения

Нормальное распределение всегда симметрично относительно ординаты, отвечающей значению *x*, равному μ. Если изменять величину μ, то кривая $f(x;\mu;\sigma)$ будет перемещаться вдоль оси *x*, сохраняя свою форму (рисунок 2.3). При уменьшении величины σ график дифференциальной функции плотности распределения вероятности сплющивается и вытягивается в высоту, т. е. становится островершинной и, наоборот, при увеличении σ график становится плосковершинной, но площадь под графиком остается неизменной (рисунок 2.4).

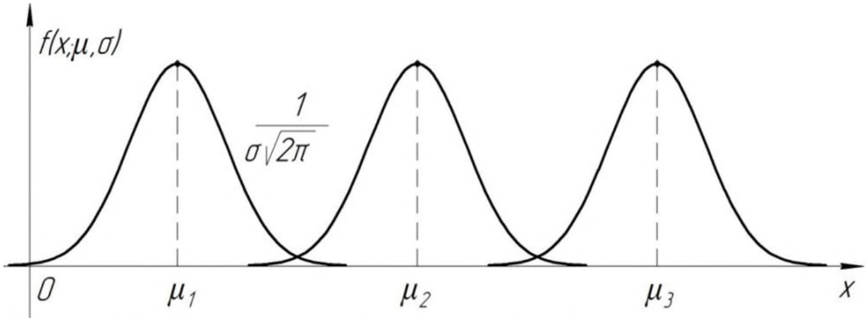


Рисунок 2.3 – Изменение расположения графика функции плотности распределения вероятности случайной величины в зависимости от изменения математического ожидания μ

2.2 Расчет вероятности нахождения случайной величины в заданном интервале при нормальном распределении

Для определения вероятности попадания в заданный интервал (x_1, x_2) случайной величины X с нормальным распределением, т. е. $P(x_1 \le X < x_2)$ необходимо вычислить определенный интеграл

$$P(x_1 \le X < x_2) = \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx. \quad (2.3)$$

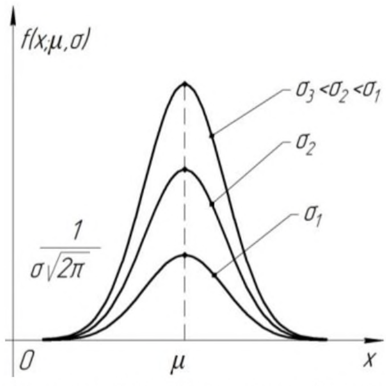


Рисунок 2.4 – Изменение формы кривой распределения плотности вероятности в зависимости от среднеквадратического отклонения σ

Приняв за новую переменную $\frac{(x-\mu)}{\sigma} = t$ и проинтегрировав по ней интеграл, получим:

$$\Phi_0(t) = \frac{1}{\sqrt{2\pi}} \int\limits_{\frac{(x_1-\mu)}{\sigma}}^{\frac{(x_2-\mu)}{\sigma}} e^{-\frac{t^2}{2}} dt. \tag{2.4}$$

Однако неопределенный интеграл вида:

$$\int e^{-\frac{t^2}{2}} dt, \tag{2.5}$$

не выражается через известные элементарные функции. Но вычисления его могут быть, например, произведены вычислением квадратуры площади.

Определенный интеграл с переменным верхним пределом вида:

$$\Phi_0(t) = \frac{1}{\sqrt{2\pi}} \int\limits_0^t e^{-\frac{t^2}{2}} dt, \tag{2.6}$$

выражающий площадь под кривой $f(t;0;1)$ в промежутке от 0 до t (рисунок 2.5) носит название нормированной функции Лапласа или просто функции Лапласа. Далее, подставив функцию Лапласа в выражение (2.3), получим:

$$P(x_1 < X < x_2) = \frac{1}{\sqrt{2\pi}} \int\limits_0^{\frac{(x_2-\mu)}{\sigma}} e^{-\frac{t^2}{2}} dt - \frac{1}{\sqrt{2\pi}} \int\limits_0^{\frac{(x_1-\mu)}{\sigma}} e^{-\frac{t^2}{2}} dt =$$

$$= \Phi_0\left(\frac{x_2-\mu}{\sigma}\right) - \Phi_0\left(\frac{x_1-\mu}{\sigma}\right),$$

(2.7) $$P(x_1 < X < x_2) = \Phi_0\left(\frac{x_2-\mu}{\sigma}\right) - \Phi_0\left(\frac{x_1-\mu}{\sigma}\right), \qquad (2.8)$$

где $\Phi_0\left(\frac{x_2-\mu}{\sigma}\right)$ и $\Phi_0\left(\frac{x_1-\mu}{\sigma}\right)$ – значения функции распределения

Функция $\Phi_0(t)$ имеет следующие свойства:

$$\Phi_0(0) = 0,$$

$$\Phi_0(-\infty) = -\frac{1}{2},$$

$$\Phi_0(+\infty) = \frac{1}{2},$$

$$\Phi_0(-t) = -\Phi_0(t),$$

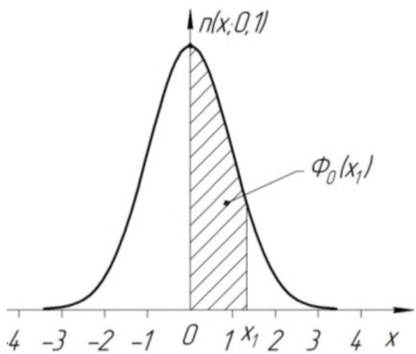


Рисунок 2.5 – Геометрическая интерпретация функции Лапласа

т. е. площадь в промежутке (0,-t) равна площади в промежутке (0,t), но считается отрицательной.

Пользуясь последним выражением (2.8) и таблицей нормированных функций Лапласа (таблица А1) легко можно определить вероятности попадания нормально распределенной случайной величины в интервалы $(\mu-\sigma; \mu+\sigma)$, $(\mu-2\sigma; \mu+2\sigma)$, $(\mu-3\sigma; \mu+3\sigma)$ (рисунок 2.6).

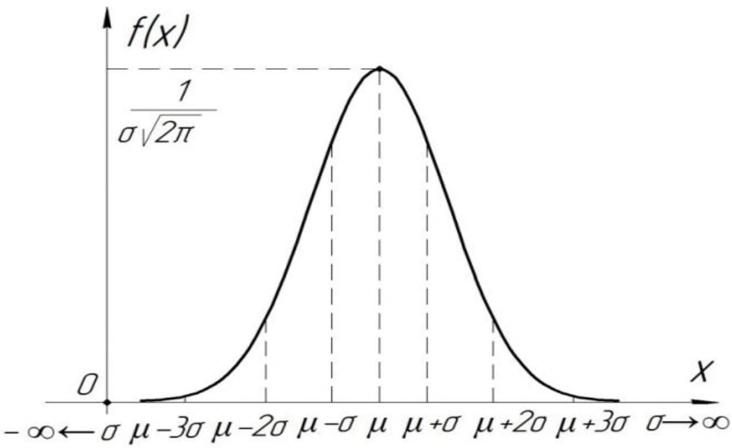


Рисунок 2.6 – Вероятности попадания нормально распределенной случайной величины в интервалы $(\mu-\sigma;\mu+\sigma)$, $(\mu-2\sigma;\mu+2\sigma)$, $(\mu-3\sigma;\mu+3\sigma)$

Пример. Определить вероятность попадания семян при посеве кукурузы в интервал заданной глубины $a \pm 1$ см, если заданная глубина $a = 7$ см, а по результатам испытания получили фактическую среднюю арифметическую глубину $\bar{a} = 7,5$ см и стандартное отклонение $S = 1,5$ см.

$$P(x_1 < X < x_2) = \Phi_0(\frac{8-7,5}{1,5}) - \Phi_0(\frac{6-7,5}{1,5}) =$$

$$= \Phi_0(0,333) - \Phi_0(-1) = 0,1293 + 0,3413 = 0,4706$$

или в заданный интервал 7 ± 1 см (6–8 см) с учетом смещения средней арифметической по фактическим результатам эксперимента до 7,5 см попадет 47,16 % высеянных семян.

2.3 *t*-распределение Стьюдента

Заметное место в экспериментальной работе и, в особенности, с малыми выборками, имеет открытое в 1908 г. *t*-распределение Стьюдента.

Распределение *t*-распределение Стьюдента (таблица Б1) для выборки определяется равенством:

$$t = \frac{\overline{x} - \mu}{\frac{S}{\sqrt{n}}} = \frac{\overline{x} - \mu}{S_{\overline{x}}}. \tag{2.9}$$

Числитель характеризует отклонение выборочной средней от математического ожидания генеральной совокупности, а знаменатель – стандартная ошибка математического ожидания генеральной совокупности.

t-распределение Стьюдента позволяет определить доверительный интервал, накрывающий математическое ожидание μ, и проверить ту или иную гипотезу относительно генеральной совокупности. Для этого достаточно знать параметры выборки – среднее арифметическое и стандартное отклонение.

2.4 F-распределение Фишера

Если две сравниваемые выборки являются случайными независимыми из общей генеральной совокупности с математическим ожиданием μ, то фактическое значение F не выйдет за определенные пределы и не превысит критическое для данных значений числа степеней свободы γ_1 и γ_2 теоретического значения F, т. е. $F_{факт} < F_{теор.}$. Теоретическое значение F для 5%-го и 1%-го уровня значимости табулированы и приводятся в таблицах (таблица В1). В этой таблице приводятся только правые критические точки для $F \geq 1$, так как всегда принято находить отношение большей дисперсии к меньшей, т. е.:

$$F = \frac{S_1^2}{S_2^2}, (S_1^2 > S_2^2), \tag{2.10}$$

где S_1^2 и S_2^2 – сравниваемые дисперсии.

Распределение F зависит только от числа степеней свободы. Сравнение нескольких дисперсий с различными числами степени свободы проводится с помощью критерия Бартлета.

Существует большое разнообразие случайных величин, в которых значения распределены по совершенно разным законам. Однако, как практика показывает, наибольшее распространение получил закон нормального распределения. Выбор закона распределения определяется не только сложившейся практикой в области исследования, но и предварительными исследованиями.

Контрольные вопросы

1. Написать и объяснить структуру нормального закона распределения.
2. Что будет изменяться с графиком нормального распределения при изменении величины математического ожидания и дисперсии случайной величины.
3. Какие еще кроме нормального распределения существуют законы распределения случайной величины.

3 СТАТИСТИЧЕСКАЯ ПРОВЕРКА ГИПОТЕЗ

Одной из важнейших задач математической статистики является проверка статистических гипотез. Согласно энциклопедического определения гипотеза (нем) – это предположение для объяснения какого-либо явления. Статистическая гипотеза – это обычно предположение о законе распределения и статистических параметрах исследуемых случайных величин. При этом в большинстве случаев гипотеза сводится к проверке отсутствия различия между фактическими и теоретически ожидаемыми наблюдениями. Такую гипотезу называют нулевой гипотезой и обозначают H_0. Если гипотеза H_0 не подтверждается расчетами, то принимается противоположная, так называемая альтернативная гипотеза H_1.

Перед проведением проверки следует оговорить уровень значимости принимаемого решения. Чем меньше уровень значимости, тем меньше вероятность отвергнуть нулевую гипотезу H_0, когда она верна, или, как говорят, совершить ошибку I рода, но тем больше вероятность совершить ошибку II рода, когда не отвергают H_0, в действительности неверную.

3.1 Проверка гипотезы о равенстве средней арифметической выборки заданному значению

Одна из часто встречающихся задач статистической проверки гипотез заключается в сравнении центров распределения двух или более нормально распределенных величин или о равенстве средней арифметической какому-либо постоянному значению C.

В качестве критерия в данном случае берется величина:

$$t = \sqrt{n} \cdot \left(\frac{\overline{x} - C}{S}\right), \tag{3.1}$$

где t – коэффициент Стьюдента при числе степеней свободы n–1;

c – постоянная величина, принятая для сравнения;

n – объем выборки;

S – стандартное отклонение средней арифметической;

$\bar{x}$ – среднее арифметическое выборки.

Если при этом вычисленное значение t не превышает критического значения $t_{кр}$, найденного по таблице значений критерия Стьюдента t (таблица Б1) при заданном уровне значимости и числу степеней свободы n–1, то исходная гипотеза принимается, в противном случае она отвергается.

Пример. При посеве пропашных культур получены следующие значения:

– количество измерений n = 100;
– стандартное отклонение от средней арифметической боковых отклонений S = 5 см;
– среднее арифметическое отклонений $\bar{x} = 3$ см.

Проверить гипотезу о равенстве нулю средней арифметической отклонений.

$$t = \sqrt{100} \cdot (\frac{3-0}{5}) = 6.$$

Что превышает величину $t_{кр}$ при уровне значимости 0,05 и числе степеней свободы n – 1 = 99. Таким образом, гипотеза должна быть отвергнута.

Также следует упомянуть о понятии наименьшей существенной разности (НСР). **Наименьшая существенная разность** – величина, указывающая наименьшую разность между среднимизначениями выборок, при привышении которой разница между этими значениями (выборками) считается стастически значимой.

3.2 Проверка гипотезы относительно вида закона распределения

В практике применения статистических методов исследований часто приходится проверять по данным выборки вид распределения, которому подчинена исследуемая случайная величина. Такие проверки, как правило, основаны на так называемых «критериях соответствия», на выборе определенной меры расхождения между теоретическим и эмпирическим распределениями. И, если такая принятая мера расхождения (т. е. критерий) для рассматриваемого случая превосходит установленный предел, то гипотеза бракуется, и наоборот.

Одним из наиболее распространенных критериев является критерий χ^2 (**критерий Пирсона,** таблица Г1). Например, если по данным выборки имеем вариационный ряд с количеством элементов n_m, попавших в m интервал.

Пусть p_m – вероятность попадания в m-й интервал с использованием гипотетического распределения. Тогда критерий Пирсона формируется так:

$$\chi^2 = \sum_{m=1}^{k} \frac{(n_m - n \cdot p_m)^2}{n \cdot p_m}, \qquad (3.2)$$

где k – количество интервалов;

n – объем выборки.

Критерий χ^2 имеет $n = k - l - 1$ степеней свободы, где l – количество оцениваемых параметров в законе распределения. Например, при нормальном законе распределения оцениваются математическое ожидание и дисперсия, т. е. $l = 2$.

3.3 Проверка гипотезы нормальности распределения случайной величины с помощью критерия Пирсона

Критерий χ^2 Пирсона предполагает сравнение количества случайных событий, попавших в m-интервал из выборки объема n с количеством событий, которое в среднем может попасть из выборки того же объема, если бы была верна гипотеза о нормальном распределении, т. е. сравнивается относительная частота попадания в интервал событий проверяемой выборки и теоретической (гипотетической) вероятностью попадания в него.

При использовании критерия Пирсона необходимо иметь ввиду, что интервалы с количеством элементов меньше 10 надо объединять с соседними. Безусловно, что при этом количество интервалов, а значит и число степеней свободы уменьшается. Так как мы имеем отношение с табличными данными с нормированным нормальным распределением вероятностей, то для применения критерия Пирсона исследуемая случайная величина X должна быть пронормирована.

$$t = \frac{X - \overline{x}_0}{S}. \tag{3.3}$$

По этой формуле определяются концы интервалов (t_{m-1}, t_m) нормированной случайной величины, при этом начало первого интервала - ∞, а конец последнего интервала + ∞.

Теоретическая вероятность p_m попадания в m-ый интервал рассчитывается по формуле:

$$P_m = \Phi_0(t_m) - \Phi_0(t_{m-1}). \tag{3.4}$$

Таблица А1 содержит значения $\Phi_0(t)$ для $t \geq 0$ и поскольку $\Phi_0(t)$ нечетная функция, то для $t < 0$

$$\Phi_0(-t) = -\Phi_0(t). \tag{3.5}$$

Далее по известному количеству k элементов, попавших в m-ый интервал и по теоретическим вероятностям p_n рассчитывается величина критерия по выражению 3.1.

По таблице Г1 находится граница $\chi^2_{кр}$ критической области для заданного уровня значимости критерия и количества степеней свободы n. Если при этом $\chi^2 < \chi^2_{кр}$, то выборка не противоречит гипотезе о нормальном распределении вероятностей генеральной совокупности.

Контрольные вопросы

1. Что такое статистическая гипотеза.
2. Дать определение нулевой гипотезе.
3. Что такое альтернативная гипотеза.
4. Что такое наименьшая существенная разница (НСР).
5. Какие задачи решают с помощью критерия Пирсона χ^2.
6. Какие случайные величины называются нормировнными.
7. Что такое гистограмма.
8. Число степеней свободы.

4 ДИСПЕРСИОННЫЙ АНАЛИЗ

4.1 Постановка задачи

При изучении статистических характеристик случайных чисел предполагалось, что случайный разброс связан лишь со случайными причинами (факторами), а все подконтрольные испытателю факторы поддерживались на одном и том же уровне. Здесь задача сводилась к определению точности применяемого метода исследования и ошибки воспроизводимости. В дисперсионном анализе речь идет об общей оценке действующего переменного фактора и его сравнении с другими факторами или уровнями этого же фактора. При этом считается, что случайные ошибки наблюдений имеют нормальное распределение. Влияние изучаемых факторов может быть двояким: они могут изменять как среднее значение наблюдений, так и дисперсию этих наблюдений. Также предполагается, что дисперсия наблюдений остается неизменной при всех изменениях факторов и их уровней. Если же однородность дисперсий вызывает сомнение, нужно провести специальное исследование с помощью критериев Бартлета, Кохрана или Фишера. В случае значимого изменения дисперсии в процессе наблюдений нужно попытаться ее стабилизировать, подобрав соответствующую преобразующую функцию.

Таким образом, в дисперсионном анализе рассматривается лишь влияние переменных факторов на генеральное среднее наблюдаемого распределения. Для того чтобы сравнивать влияние различных факторов, нужно найти какой-нибудь достаточно надежный и универсальный показатель этого влияния.

Допустим, изучаемый фактор A на различных уровнях привел к серии результатов: $x_1, x_2, x_3, \ldots, x_{k-1}, x_k$.

В качестве показателя влияния фактора A можно рассматривать число:

$$\sigma_A^2 = \frac{1}{k}\sum_{i=1}^{k}(x_i - \overline{x})^2, \qquad (4.1)$$

где $\overline{x}$ – среднее арифметическое чисел x_i;

σ_A^2 – дисперсия от влияния фактора A на уровнях 1, 2, 3,…, k-1, k.

Числа $x_1, x_2, x_3,..., x_{k-1}, x_k$ не являются случайными, поэтому σ_A^2 не связана ни с какой случайной величиной.

Выбор σ_A^2 удобен по двум причинам. Во-первых, дисперсия является простейшей мерой рассеивания. Во-вторых, показатель влияния фактора A определен теперь аналогично показателю влияния случайного фактора, т. е. обычной дисперсии σ^2. Это позволит непосредственно сравнивать фактор A с эффектом случайности. **Изучение переменных факторов по их дисперсиям называется дисперсионным анализом.**

Рассмотрим самый простой случай, когда дисперсия наблюдений σ^2 известна заранее и исследуется один переменный фактор A.

Пусть при изменении фактора A получились результаты наблюдений $x_1, x_2, x_3,..., x_{k-1}, x_k$. Найдем выборочную дисперсию:

$$S^2 = \frac{1}{k-1}\sum_{i=1}^{k}(x_i - \overline{x})^2 = \frac{1}{k-1}\left[\sum_{i=1}^{k} x_i^2 - \frac{(\sum_{i=1}^{k} x_i)^2}{k}\right]. \qquad (4.2)$$

Сравним эту дисперсию, имеющую k – 1 степеней свободы с генеральной дисперсией σ^2. Если S^2 от σ^2 отличается незначимо, то и влияние фактора A нужно признать незначимым, так как он не сумел существенно увеличить случайный разброс наблюдений. Если же S^2 отличается значимо от σ^2, то это может быть вызвано только влиянием фактора A, которое теперь нужно признать значимым. Для того, чтобы оценить σ_A^2, воспользуемся тем, что дисперсия суммы двух независимых случайных величин равна сумме их дисперсий. В нашем случае складываются эффект случайности σ^2 и эффект воздействия фактора A, которые являются независимыми случайными величинами. Поэтому общая дисперсия наблюдений должна быть равна $\sigma^2+\sigma_A^2$. А величина S^2 является оценкой этой общей дисперсии. Следовательно:

$$\sigma_A^2 \approx S^2-\sigma^2. \tag{4.3}$$

Пример. Изучается сопротивление плуга в зависимости от глубины вспашки в пределах от 10 до 35 см. Делая замеры сопротивления плуга через каждые 5 см глубины, получили следующие данные (в кН): 0,5; 0,9; 1,5; 2,2; 2,9; 3,6. Допустим, что ошибка воспроизводимости известна из предыдущих опытов и равна $\sigma=0,2$. Находим выборочную дисперсию $S^2=1,42$. Тогда критерий Фишера $F=1,42\,/\,0,04=35,45$. По таблице распределения Фишера находим табличное значение $F_{0,95}(5,\infty)=2,2$. Найденное нами выше $F=35,45>0,2$. Следовательно, влияние глубины вспашки на тяговое сопротивление плуга следует считать значимым.

Следует также заметить, что мы оцениваем лишь влияние фактора в целом и не выясняем количественных соотношений. Поэтому замеры сопротивления плуга можно было делать и не через равные промежутки глубины. Можно было вообще и не отмечать глубину, достаточно было бы, чтобы она изменялась в заданных пределах.

4.2 Однофакторный дисперсионный анализ

Рассмотрим случай, когда неизвестно значение дисперсии генеральной совокупности σ^2. Нужно найти такую схему анализа, которая позволила бы одновременно дать оценку и дисперсии от изучаемых факторов, и дисперсии σ^2.

Проведем наблюдения за исследуемым объектом на нескольких уровнях одного фактора *A*, т. е. на уровнях $A_1, A_2, A_3, ..., A_{k-1}, A_k$. Для того, чтобы дать оценку неизвестной дисперсии воспроизводимости σ^2 нужно обязательно иметь дублирующие наблюдения при каждом уровне фактора *A*. Для этого на одном из уровней фактора *A* надо провести достаточно много наблюдений, чтобы сразу определить дисперсию σ^2 и использовать найденное значение для изучения остальных уровней. Лучше, однако, равномерно повторить наблюдения на всех уровнях, так как при этом появляется возможность контроля за постоянством дисперсии σ^2. При одинаковом количестве наблюдений также получаются наиболее простые расчеты.

Обозначим серию наблюдений на уровне A_i через $x_{i1}, x_{i2}, x_{i3}, ..., x_{i(n-1)}, x_{in}$ (n – число повторных наблюдений на каждом уровне). Через $\overline{x_i}$ обозначим среднее значение наблюдений на i-ом уровне:

$$\overline{x_i} = \frac{x_{i1} + x_{i2} + x_{i3} +, ..., + x_{i(n-1)} + x_{in}}{n}. \qquad (4.4)$$

Кроме того нам понадобится среднее всех наблюдений по всем уровням:

$$\bar{\bar{x}} = \frac{1}{k \cdot n} \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij} = \frac{1}{k} \sum_{i=1}^{k} \bar{x}_i \qquad (4.5)$$

Общее число всех наблюдений равно kn.

Определяем общую выборочную дисперсию всех наблюдений.

$$S^2 = \frac{1}{kn-1} \sum_{i=1}^{k} \sum_{j=1}^{n} (x_{ij} - \bar{\bar{x}})^2 = \frac{1}{kn-1} \left[\sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}{}^2 - \frac{(\sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij})^2}{kn} \right]. \qquad (4.6)$$

Эта дисперсия образовалась в результате действия на исследуемый показатель как фактора A, так и случайных факторов. Основная задача, которую решает дисперсионный анализ – это разложение общей дисперсии S^2 на составляющие, которые характеризовали бы фактор A и фактор случайности в отдельности.

Фактор случайности оценить несложно благодаря наличию повторных наблюдений на каждом уровне, Для уровня A_i выборочная дисперсия S_i^2 равна:

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2 = \frac{1}{n-1} \left[\sum_{j=1}^{n} x_{ij}{}^2 - \frac{(\sum_{j=1}^{n} x_{ij})^2}{n} \right]. \qquad (4.7)$$

Пользуясь последним выражением, получаем серию выборочных дисперсий S_i^2, характеризующих фактор случайности на всех уровнях A_i. Если у нас нет предварительной уверенности в том, что генеральная дисперсия воспроизводимости σ^2 одинакова на всех уровнях, то все дисперсии S_i^2 нужно сравнить методом критерия Фишера или Кохрана.

Если между дисперсиями S_i^2 нет значимых (существенных различий на заданном уровне значимости), то их все можно использовать для оценки генеральной дисперсии σ^2 по принципу «текущих измерений» и получим:

$$S_0^2 = \frac{1}{k}\sum_{i=1}^{k} S_i^2 = \frac{1}{k(n-1)}\left[\sum_{i=1}^{k}\sum_{j=1}^{n} x_{ij}^2 - \frac{1}{n}\sum_{i=1}^{k}\Big(\sum_{j=1}^{n} x_{ij}\Big)^2\right], \qquad (4.8)$$

имеющую $k(n-1)$ степеней свободы.

После того как найдена дисперсия S_0^2 связанная **со случайностью**, можно уже дать приближенную оценку для дисперсии фактора A:

$$\sigma_A^2 \approx S^2 - S_0^2. \qquad (4.9)$$

Однако эта оценка слишком груба из-за погрешностей величин S^2 и S_0^2. Более точную оценку для σ_A^2 находят из следующих соображений. Влияние фактора A наиболее заметно на изменении средних $\overline{x}_i$ по отдельным уровням. Действительно, дисперсия случайности для средних значений в n раз меньше, чем для отдельных наблюдений. Поэтому

$$\frac{1}{k-1}\sum_{i=1}^{k}(\overline{x}_i-\overline{\overline{x}})^2 \approx \sigma_A^2+\frac{\sigma^2}{n}\approx \sigma_A^2+\frac{S_0^2}{n}. \qquad (4.10)$$

Отсюда легко определить дисперсию фактора A:

$$\sigma_A^2 \approx \frac{1}{k-1}\sum_{i=1}^{k}(\overline{x}_i-\overline{\overline{x}})^2-\frac{S_0^2}{n}, \qquad (4.11)$$

причем точность этой оценки выше, чем у предыдущей, благодаря тому, что дисперсия S_0^2 входит в нее, деленная на n.

Введем обозначение:

$$S_A^2 \approx \frac{n}{k-1}\sum_{i=1}^{k}(\overline{x}_i-\overline{\overline{x}})^2 \approx n\sigma_A^2+S_0^2. \qquad (4.12)$$

Эта дисперсия имеет k – 1 степень свободы. Далее рассуждения ведем так.

1. Для того чтобы влияние фактора A было значимым, необходимо и достаточно чтобы дисперсия S_A^2 значимо отличалась от S_0^2. Сравнение S_A^2 и S_0^2 проводится по критерию Фишера, т. е. влияние фактора A признается значимым на уровне значимости p, если

$$\frac{S_A^2}{S_0^2} > F_{1-P}, \qquad (4.13)$$

где F-распределение рассматривается с $f_1 = k-1$ и $f_2 = k(n-1)$ степенями свободы.

2. Если же $\frac{S_A^2}{S_0^2} \le F_{1-P}$, то влияние фактора A нужно считать незначимым. В этом случае общая дисперсия S^2 связана только с фактором случайности и поэтому $\sigma^2 \approx S^2$. Последняя оценка дисперсии случайности лучше, чем S_0^2, так как имеет $kn - 1$ степень свободы (на $k - 1$ степень больше, чем у S_0^2).

Проведение дисперсионного анализа связано с большими вычислениями, поэтому желательно применять вычислительную технику. Перед началом вычислений все данные можно уменьшить на одну и ту же величину – на дисперсиях это не отразится.

Вычисления ведут по следующим этапам:

1. Все данные располагаем в таблицу.
2. Находим сумму квадратов всех наблюдений:

$$Q_1 = \sum_{i=1}^{k} \sum_{j=1}^{n} x_{ij}^2. \qquad (4.14)$$

3. Находим сумму квадратов итогов по столбцам, деленный на число параллельных наблюдений:

$$Q_2 = \frac{1}{n} \sum_{i=1}^{k} x_i^2 = \frac{1}{n} \sum_{i=1}^{k} \left(\sum_{j=1}^{n} x_{ij}\right)^2. \qquad (4.15)$$

4. Находим квадрат общего итога, деленный на число всех наблюдений:

$$Q_3 = \frac{1}{kn} \left(\sum_{i=1}^{k} x_i\right)^2 = \frac{1}{kn} \left[\sum_{i=1}^{k} \left(\sum_{j=1}^{n} x_{ij}\right) \right]^2. \qquad (4.16)$$

5. Вычисляем дисперсии S_A^2 и S_0^2 по формулам:

$$S_0^2 = \frac{Q_1 - Q_2}{k(n-1)}, \qquad (4.17)$$

$$S_A^2 = \frac{Q_2 - Q_3}{k-1}, \qquad (4.18)$$

Проделав указанные вычисления, можно перейти к сравнению дисперсий S_A^2 и S_0^2. Если их различие оказывается незначимым, то получаем оценку генеральной дисперсии

$$S^2 = \frac{Q_1 - Q_3}{kn-1}, \qquad (4.19)$$

имеющую $kn - 1$ степень свободы. Если же различие S_A^2 и S_0^2 оказывается значимым, то находим оценку влияния фактора *A:*

$$\sigma_A^2 = \frac{S_A^2 - S_0^2}{n}, \qquad (4.20)$$

Пример. Исследовалось влияние дозы удобрений на величину урожайности (т/га) подсолнечника. Обозначая дозы удобрений через $A_1, A_2, A_3, \ldots, A_{k-1}, A_k$. Полученные данные занесены в таблицу 4.1.

Таблица 4.1 – Результаты замеров влияния дозы удобрений на величину урожайности подсолнечника, (т/га)

Номер наблюдения	Доза удобрений				
	A_1	A_2	A_3	A_4	A_5
1	3,2	2,6	2,9	3,7	3,0
2	3,1	3,1	2,6	3,4	3,4
3	3,1	2,7	3,0	3,2	3,2
4	2,8	2,9	3,1	3,3	3,5
5	3,3	2,7	3,0	3,5	2,9
6	3,0	2,8	2,8	3,3	3,1
Итоги	18,5	16,8	17,4	20,4	19,1

Вычисления по выражениям (4.16; 4.15; 4.16) дали следующие результаты.

$$Q_1 = 285{,}6;\ Q_2 = 284{,}7;\ Q_3 = 283{,}4.$$

Определяем дисперсию от фактора A S_A^2 и дисперсию от случайных факторов S_0^2.

$$S_0^2 = \frac{285{,}6 - 284{,}7}{25} = 0{,}036;\ S_A^2 = \frac{284{,}7 - 283{,}4}{4} = 0{,}325.$$

Найденные дисперсии сравним по критерию Фишера:

$$F = \frac{0{,}325}{0{,}036} = 9{,}03.$$

По таблице Г4 теоретическое (табличное) значение F при уровне $F_{0,95}(4, 25) > 2{,}8$. Поэтому различие доз удобрений следует признать значимым.

$$\sigma_A^2 = \frac{0{,}325 - 0{,}06}{6} = 0{,}048.$$

4.3 Двухфакторный дисперсионный анализ

Пусть изучаются одновременно два фактора A и B на уровнях $A_1, A_2,\ldots, A_k$ и $B_1, B_2,\ldots, B_m$. Результаты наблюдений занесем в таблицу 4.2. Через X_i обозначим итоги данных по столбцам, через X'_j – итоги по строкам, т. е.

$$\bar{x}_i = \frac{X_i}{m}, \qquad \bar{x}'_j = \frac{X'_j}{k}, \tag{4.21}$$

Здесь важно заметить из каких составляющих состоит рассеивание средних по строкам и столбцам.

Таблица 4.2 – Схема двухфакторного дисперсионного анализа

B	A				Итоги
	A_1	A_2	…	A_k	
B_1	x_{11}	X_{21}	…	x_{k1}	X'_1
B_2	x_{12}	X_{22}	…	x_{k2}	X'_{25}
…	…	…	…	…	…
B_m	x_{1m}	x_{2m}	…	x_{km}	X'_m
Итоги	X_{1i}	X_{2i}	…	X_{3i}	

Очевидно, на каждое такое рассеивание оказывает влияние лишь один из факторов A и B, так как все уровни второго фактора усреднены. Так, рассеивание $\bar{x}_i$ (средние по столбцам) не зависит от фактора B, рассеивание $\bar{x}_j$ (средние по строкам) не зависит от фактора A. Кроме того, на всех рассеиваниях сказывается влияние фактора случайности с дисперсией $\frac{\sigma^2}{m}$ для $\bar{x}_i$ и $\frac{\sigma^2}{k}$ для $\bar{x}_j$. Окончательно получаем, что

$$\frac{1}{k-1}\sum_{i=1}^{k}(\bar{x}_i - \bar{\bar{x}})^2 \approx \sigma_A^2 + \frac{\sigma^2}{m}, \tag{4.22}$$

$$\frac{1}{m-1}\sum_{j=1}^{m}(\overline{x}'_j-\overline{\overline{x}})^2 \approx \sigma_B^2+\frac{\sigma^2}{k}, \qquad (4.23)$$

где через $\overline{\overline{x}}$ обозначено среднее всех данных таблицы.

Получившиеся равенства позволяют оценить дисперсии факторов A и B, если будет известна оценка генеральной дисперсии наблюдений σ^2. Казалось бы, для оценки последней дисперсии нет никаких предпосылок, так как полностью отсутствуют параллельные наблюдения. И тем не менее дисперсию σ^2 удается оценить, сравнивая рассеивание средних с рассеиванием самих наблюдений.

Найдем дисперсию наблюдений по i-му столбцу:

$$S_i^2=\frac{1}{m-1}\sum_{j=1}^{m}(x_{ij}-\overline{x}_i)^2. \qquad (4.24)$$

Очевидно, что эта дисперсия возникла под влиянием фактора B и фактора случайности, поэтому $S_i^2 \approx \sigma_B^2+\sigma^2$. Равенство станет более точным, если S_i^2 заменить средневзвешенной дисперсией по всем столбцам, т. е.

$$\sigma_B^2+\sigma^2 \approx \frac{1}{k}\sum_{i=1}^{k}S_i^2=\frac{1}{k(m-1)}\sum_{i=1}^{k}\sum_{j=1}^{m}(x_{ij}-\overline{x}_i)^2. \qquad (4.25)$$

Полученная оценка содержит одновременно обе дисперсии $\sigma_B^2+\sigma^2$. Ранее у нас была получена еще одна оценка, содержащая эти же дисперсии – это равенства 4.22 и 4.23. Вычитая 4.23 из 4.25, получим, что:

$$\sigma^2 - \frac{\sigma^2}{k} \approx \frac{1}{k(m-1)} \sum_{i=1}^{k} \sum_{j=1}^{m} (x_{ij} - \bar{x}_i)^2 - \frac{1}{m-1} \sum_{j=1}^{m} (\bar{x}'_j - \bar{\bar{x}})^2. \quad (4.26)$$

Откуда

$$\sigma^2 \approx \frac{1}{(k-1)(m-1)} \left[\sum_{i=1}^{k} \sum_{j=1}^{m} (x_{ij} - \bar{x}_i)^2 - k \sum_{j=1}^{m} (\bar{x}'_j - \bar{\bar{x}})^2 \right]. \quad (4.27)$$

Полученное равенство дает оценку искомой дисперсии σ^2 через некоторую величину, зависящую от всех наблюдений. По своей природе эта величина является выборочной дисперсией с $(k-1)(m-1)$ степенями свободы; будем эту дисперсию обозначать через S_0^2. Кроме того, введем обозначения:

$$S_A^2 = \frac{m}{k-1} \sum_{i=1}^{k} (\bar{x}_i - \bar{\bar{x}})^2 \approx m\sigma_A^2 + S_0^2, \quad (4.28)$$

$$S_B^2 = \frac{k}{m-1} \sum_{i=1}^{k} (\bar{x}'_j - \bar{\bar{x}})^2 \approx k\sigma_B^2 + S_0^2 \quad (4.29)$$

Величины S_A^2 и S_B^2 также можно считать выборочными дисперсиями с $k-1$ и $m-1$ степенями свободы соответственно.

Окончательную схему расчетов можно представить следующим образом:

1. Находим сумму квадратов всех наблюдений:

$$Q_1 = \sum_{i=1}^{k} \sum_{j=1}^{m} x_{ij}^2. \quad (4.30)$$

2. Находим сумму квадратов итогов по столбцам, деленную на число наблюдений в столбце:

$$Q_2 = \frac{1}{m}\sum_{i=1}^{k} X_i^2. \tag{4.31}$$

3. Находим сумму квадратов итогов по строкам, деленную на число наблюдений в строке:

$$Q_3 = \frac{1}{k}\sum_{j=1}^{m} X_j'^2. \tag{4.32}$$

4. Находим квадрат общего итога, деленный на число всех наблюдений:

$$Q_4 = \frac{1}{mk}\left(\sum_{i=1}^{k} X_i\right)^2 = \frac{1}{mk}\left(\sum_{j=1}^{m} X_j'\right)^2, \tag{4.33}$$

(наличие двойной формулы можно использовать для проверки правильности вычислений).

5. Вычисляют дисперсии S_0^2, S_A^2 и S_B^2 по формулам:

$$S_0^2 = \frac{Q_1 + Q_4 - Q_2 - Q_3}{(k-1)(m-1)}, \tag{4.34}$$

$$S_A^2 = \frac{Q_2 - Q_4}{k-1}, \tag{4.35}$$

$$S_B^2 = \frac{Q_3 - Q_4}{m-1}. \tag{4.36}$$

После того, как будут проделаны все необходимые вычисления, можно приступить к непосредственному анализу влияния факторов A и B. Для того, чтобы влияние фактора A можно было признать значимым, нужно , чтобы S_A^2 значимо отличалось от S_0^2, то же самое справедливо для фактора B и дисперсии S_B^2. Сравнение дисперсий, как обычно, проводится по критерию Фишера, т. е. вычисляются отношения $\frac{S_A^2}{S_0^2}$ и $\frac{S_B^2}{S_0^2}$ и сравниваются с табличными значениями F-распределения.

Допустим, что выбран уровень значимости p. Влияние фактора A признается значимым, если:

$$\frac{S_A^2}{S_0^2} > F_{1-P}, \tag{4.37}$$

где в F-распределении берутся $f_1 = k-1$, $f_2 = (k-1)(m-1)$ степеней.

Дисперсия фактора A оценивается в этом случае равенством:

$$\sigma_A^2 \approx \frac{S_A^2 - S_0^2}{m}. \tag{4.38}$$

Аналогично, влияние фактора B считается значимым, если

$$\frac{S_B^2}{S_0^2} > F_{1-P}, \tag{4.39}$$

где на этот раз берутся $f_1 = m - 1$, $f_2 = (k-1)(m-1)$ степеней свободы. Дисперсия фактора *B* оценивается при этом равенством:

$$\sigma_B^2 \approx \frac{S_B^2 - S_0^2}{k}. \tag{4.40}$$

Если $\frac{S_A^2}{S_0^2} \le F_{1-p}$, то влияние фактора *A* нужно признать незначимым. В этом случае обе дисперсии S_A^2 и S_0^2 можно использовать для оценки генеральной дисперсии σ^2. Это приведет к равенству:

$$\sigma^2 \approx \frac{(k-1)S_A^2 + (k-1)(m-1)S_0^2}{(k-1)+(k-1)(m-1)} = \frac{Q_1 - Q_3}{m(k-1)}, \tag{4.41}$$

которое дает более точную оценку для σ^2, чем S_0^2, ибо у этой оценки $m(k–1)$ степеней свободы (на $k–1$ степень больше, чем у S_0^2).

Аналогично, влияние фактора *B* признается незначимым, если $\frac{S_B^2}{S_0^2} \le F_{1-p}$, и это дает оценку:

$$\sigma^2 \approx \frac{(m-1)S_B^2 + (k-1)(m-1)S_0^2}{(m-1)+(k-1)(m-1)} = \frac{Q_1 - Q_2}{k(m-1)}. \tag{4.42}$$

Если незначимыми оказываются влияния обоих факторов одновременно, то для дисперсии σ^2 получается самая лучшая оценка:

$$\sigma^2 \approx \frac{(k-1)S_A^2+(m-1)S_B^2+(k-1)(m-1)S_0^2}{(k-1)+(m-1)+(k-1)(m-1)} = \frac{Q_1-Q_4}{mk-1}, \quad (4.43)$$

имеющая $mk - 1$ степень свободы. Точно такую же оценку мы получили бы, если бы отвлеклись от факторов A и B и все наблюдения рассматривали как единую выборку.

Изложенный анализ является исчерпывающим, если факторы A и B независимы. Если же между ними существует взаимодействие, то этому взаимодействию как фактору присуща своя дисперсия σ_{AB}^2. В указанной выше схеме расчетов дисперсия σ_{AB}^2 как составная часть войдет в дисперсию S_0^2. Выделить эту часть из дисперсии S_0^2, к сожалению, невозможно без наличия параллельных наблюдений.

Допустим теперь, что при каждом сочетании уровней факторов A и B проводится n параллельных наблюдений. Тогда при любых i и j на пересечении i-го столбца и j-ой строки появится целая серия наблюдений $x_{ij_1}, x_{ij_2}, x_{ij_3}, \ldots, x_{ij_{(n-1)}}, x_{ij_n}$. Сохраним обозначение x_{ij} за средним каждой такой серии. Каждая серия имеет свою выборочную дисперсию:

$$S_{ij}^2 = \frac{1}{n-1}\sum_{\gamma}^{n}(x_{ij\gamma} - ij)^2, \quad (4.44)$$

которой соответствует $n - 1$ степень свободы.

Поскольку на протяжении серии факторы A и B остаются неизменными, то S_{ij}^2 является оценкой дисперсии воспроизводимости σ^2. Еще лучшей оценкой будет средневзвешенная дисперсия по всем сериям x_{ij}, то поэтому она складывается из дисперсии взаимодействия:

$$S^2 = \frac{1}{mk}\sum_{i=1}^{k}\sum_{j=1}^{m} S_{ij}^2, \tag{4.45}$$

которой соответствует уже $mk(n-1)$ степеней свободы.

Для того чтобы выделить теперь дисперсию взаимодействия, найдем, в каких пропорциях входят σ_{AB}^2 и S^2 в дисперсию S_0^2.

Дисперсия S_0^2 соответствует рассеянию средних, поэтому она складывается из дисперсии взаимодействия σ_{AB}^2 и дисперсии случайности, которая для средних x_{ij} равна $\frac{\sigma^2}{n}$. Таким образом:

$$S_0^2 \approx \sigma_{AB}^2 + \frac{\sigma^2}{n} \approx \sigma_{AB}^2 + \frac{S^2}{n}, \text{ откуда } \sigma_{AB}^2 = S_0^2 - \frac{S^2}{n}. \tag{4.46}$$

Разумеется, эффект взаимодействия факторов A и B может оказаться незначимым. Для проверки этого составляют отношение $\frac{nS_0^2}{S^2}$ и сравнивают его с табличным значением F_{1-P}, где берутся $f_1 = (k-1)(m-1)$, $f_2 = mk(n-1)$ степеней свободы.

Наличие параллельных наблюдений используется только для отделения эффекта взаимодействия от дисперсии случайности σ^2. Поэтому общая схема вычислений здесь лишь слегка дополняет ту схему, что приводилась выше. Вначале вычисляем сумму квадратов всех наблюдений:

$$Q_5 = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{\gamma=1}^{n} S_{ij\gamma}^2. \quad (4.47)$$

После этого отдельные наблюдения серий нам уже не нужны, и в таблицу можно занести только средние серий x_{ij}, так что получится вновь таблица 4.2. Над числами x_{ij} проводятся теперь все вычисления прежней схемы, т. е. вычисляются Q_1, Q_2, Q_3, Q_4, а по ним S_A^2, S_B^2 и S_0^2. Так же, как описано выше, производится оценка значимости факторов A и B и вычисляются, в случае необходимости, дисперсии σ_A^2 и σ_B^2.

Новые вычисления появляются лишь при оценке взаимодействия факторов A и B. Для этого находим:

$$S^2 = \frac{Q_5 - nQ_1}{mk(n-1)}. \quad (4.48)$$

Затем сравниваем nS_0^2 и S^2 по F-критерию. Если отличие этих дисперсий окажется значимым, то нужно признать значимым эффект взаимодействия факторов A и B, причем:

$$\sigma_{AB}^2 \approx \frac{nS_0^2 - S^2}{n}. \quad (4.49)$$

Если же отличие оказывается незначимым, то взаимодействием факторов можно пренебречь и оценки S_0^2 и S^2 объединить в одну:

$$\sigma^2 \approx \frac{(k-1)(m-1)S_0^2 + mk(n-1)S^2}{(k-1)(m-1) + mk(n-1)} = \frac{Q_4 + Q_5 - Q_2 - Q_3}{mkn - m - k + 1}. \quad (4.50)$$

Пример двухфакторного дисперсионного анализа.

Исследовалась зависимость тягового сопротивления 4-корпусного навесного плуга от глубины вспашки и скорости пахотного агрегата. Для этого провели измерения тягового сопротивления на 3-х разных глубинах (фактор A) и 3-х скоростях агрегата в трехкратной повторности. Полученные данные приведены в таблице 4.3

Преобразуем таблицу 4.3 в таблицу 4.4 только со средними параллельных наблюдений.

В нашем примере $k = m = 3$. Обработку полученной таблицы ведем по выражениям, изложенным выше.

Таблица 4.3 – Экспериментальные наблюдения исследования зависимости тягового сопротивления плуга от глубины вспашки (фактор A) и скорости вспашки (фактор B), кН

Фактор B	Фактор A								
	A_1			A_2			A_3		
B_1	3,6	3,8	4,0	3,0	3,2	3,1	2,8	2,4	2,6
B_2	4,0	4,2	4,1	3,3	3,2	2,8	3,0	3,2	2,8
B_3	3,9	4,4	4,9	3,8	3,6	3,4	3,8	4,0	4,2

Таблица 4.4 – Преобразованная таблица со средними параллельных наблюдений

Фактор B	Фактор A			Итоги
	A_1	A_2	A_3	
B_1	3,8	3,1	2,6	9,5
B_2	4,1	3,1	3,0	10,2
B_3	4,4	3,6	4,0	12,0
Итоги	12,3	9,8	9,6	31,7

$$Q_1 = \sum_{i=1}^{k}\sum_{j=1}^{m} x_{ij}^2 = 3,8^2 + 3,1^2 + 2,6^2 + 4,1^2 +$$

$$+3,1^2 + 3^2 + 4,4^2 + 3,6^2 + 4^2 = 114,55$$

$$Q_2 = \frac{1}{m}\sum_{i=1}^{k} X_i^2 = \frac{12,3^2 + 9,8^2 + 9,6^2}{3} = 113,16$$

$$Q_3 = \frac{9,5^2 + 10,2^2 + 12^2}{3} = 112,76$$

$$Q_4 = \frac{31,7^2}{9} = 111,65$$

Находим сумму квадратов всех наблюдений

$$Q_5 = \sum_{i=1}^{k}\sum_{j=1}^{m}\sum_{\gamma=1}^{n} S_{ij\gamma}^2 = 3,6^2 + 3,8^2 + 4^2 + 3^2 + 3,2^2 + 3,1^2 + 2,8^2 + 2,4^2 +$$
$$+ 2,6^2 + 4^2 + 4,2^2 + 4,1^2 + 3,3^2 + 3,2^2 + 2,8^2 + 3^2 + 3,2^2 + 2,8^2 + 3,9^2 +$$
$$+ 4,4^2 + 4,9^2 + 3,8^2 + 3,6^2 + 3,4^2 + 3,8^2 + 4^2 + 4,2^2 = 12,96 + 14,44 +$$
$$+ 16 + 9 + 10,24 + 9,61 + 7,84 + 5,76 + 6,76 + 16 + 17,64 + 16,81 +$$
$$+ 10,89 + 10,24 + 7,84 + 9 + 10,24 + 7,84 + 15,21 + 19,36 + 24,01 +$$
$$+ 14,44 + 12,96 + 11,56 + 14,44 + 16 + 17,64 = 344,73$$

Находим дисперсию воспроизводимости S^2:

$$S^2 = \frac{Q_5 - nQ_1}{mk(n-1)},$$

$$S^2 = \frac{344,73 - 3\cdot 114,55}{3\cdot 3(3-1)} = 0,06.$$

Находим совместную дисперсию воспроизводимости и взаимодействия:

$$S_0^2 = \frac{Q_1 + Q_4 - Q_2 - Q_3}{(k-1)(m-1)},$$

$$S_0^2 = \frac{114,55 + 111,65 - 113,16 - 112,76}{(3-1)(3-1)} = 0,07.$$

Определяем влияние факторов A и B

$$S_A^2 = \frac{Q_2 - Q_4}{k-1}, \qquad S_B^2 = \frac{Q_3 + Q_4}{m-1},$$

$$S_A^2 = \frac{113{,}16 - 111{,}65}{3-1} = 0{,}76, \quad S_B^2 = \frac{112{,}76 - 111{,}65}{3-1} = 0{,}56\,.$$

Для проверки значимости эффекта взаимодействия сравним дисперсии $nS_0^2 = 3 \cdot 0{,}7 = 2{,}1$ и $S^2 = 0{,}06$ по критерию Фишера:

$$\frac{nS_0^2}{S^2} = \frac{2{,}1}{0{,}06} = 35$$

При этом табличное значение $F_{0,95}$ со степенями свободы $f_1 = 4$, $f_2 = 18$ равно 3,0. Таким образом, эффект взаимодействия факторов скорости и глубины пахоты оказывается значимым и оценивается дисперсией:

$$\sigma_{AB}^2 = S_0^2 - \frac{S^2}{n} = 0{,}07 - \frac{0{,}06}{3} = 0{,}01\,.$$

Сравнив по критерию Фишера S_A^2 и S_B^2 с S_0^2, находим:

$$\frac{S_A^2}{S_0^2} = \frac{0{,}76}{0{,}07} = 10{,}9\,,$$

а табличное значение $F_{0,95}$ со степенями свободы $f_1 = 2$, $f_2 = 4$, равно 6,9. Следовательно влияние фактора A надо признать значимым.

Отношение:

$$\frac{S_B^2}{S_0^2} = \frac{0,56}{0,07} = 8,$$

при том же значении $F_{0,95}$ = 6,9. Значит действие и фактора B на тяговое сопротивление тоже в пределах изменения факторов существенно.

Проведенный анализ показывает, что факторы глубины и скорости вспашки и их взаимодействие значимо влияют на сопротивление плуга.

4.4 Трехфакторный дисперсионный анализ

Применительно к различным направлениям исследований разработано большое количество дисперсионных планов и способов их обработки. В исследованиях процессов и машин в агробизнесе широкое применение нашли дисперсионные планы на основе расщепленных делянок. Метод расщепленных делянок обычно используют для многофакторных полевых опытов при исследовании цепочки различных технологических операций или испытании сравниваемых машин с обязательной подготовкой делянок по различным технологиям. На примере зависимости степени очистки початков кукурузы от оберточных листьев в зависимости от скорости движения кукурузоуборочного агрегата V (фактор A), угла наклона русел початкоотделяющего аппарата α (фактор B) и частоты вращения протягивающих вальцев ω (фактор C) рассмотрим трехфакторный дисперсионный анализ (таблица 4.5), в которую занесены результаты наблюдений. Каждый фактор варьировался на двух уровнях. В таблице определяем суммы всех наблюдений $\sum X$, суммы вариантов $\sum V$ и повторений $\sum P$ и их средние.

Составим схему дисперсионного анализа. Для этого разложим общую сумму квадратов отклонений C_Y на компоненты:

$$C_Y = C_V + C_P + C_Z, \qquad (4.51)$$

где C_V, C_P и C_Z – соответственно суммы квадратов отклонений для вариантов, повторений и случайного варьирования (ошибки – остатка).

В свою очередь сумма квадратов отклонений C_V образуется из следующих компонентов:

$$C_Y = C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC} + C_{ABC}, \qquad (4.52)$$

где $C_A, C_B, C_C, C_{AB}, C_{AC}, C_{BC}, C_{ABC}$ – соответственно суммы квадратов отклонений от действия факторов A, B, C (главных эффектов), парных взаимодействий AB, AC, BC и тройного взаимодействия ABC.

Для определения сумм квадратов отклонений используем следующее выражение:

$$\sum (X - \bar{X})^2 = \sum X^2 - \frac{(\sum X)^2}{n}. \qquad (4.53)$$

Таблица 4.5 – Влияние изучаемых факторов на степень очистки початков от оберточных листьев, %

Факторы			Повторения			Сумма, *V*	Средние
A	*B*	*C*	1	2	3		
V = 3 км/ч	α = 23°	ω_{min} = 700 мин$^{-1}$	38,2	44,7	26,2	109,1	36,4
		ω_{max} = 900 мин$^{-1}$	51,7	57,8	51,2	160,7	53,6
	α = 43°	ω_{min} = 700 мин$^{-1}$	47,7	61,8	57,9	167,4	55,8
		ω_{max} = 900 мин$^{-1}$	48,8	50	50,8	149,6	50
V = 9 км/ч	α = 43°	ω_{min} = 700 мин$^{-1}$	32,2	33,3	33,3	98,8	32,9
		ω_{max} = 900 мин$^{-1}$	62,3	50	58	170,3	56,8
	α = 43°	ω_{min} = 700 мин$^{-1}$	44,9	50	40	134,9	45
		ω_{max} = 900 мин$^{-1}$	62	52,5	62	176,5	58,8
Суммы *P*			387,8	400,1	379,4	$\sum X = 1167{,}3$	$\sum X = 48{,}7$

Определяем корректирующий фактор

$$C = \frac{(\sum X)^2}{N}, \qquad (4.54)$$

где N – общее количество наблюдений в эксперименте.

$$N = m \cdot l \cdot k \cdot n, \qquad (4.55)$$

где m, l, k – количество уровней соответственно по факторам A, B, C;

n – количество повторностей.

$m = l = k = 2$; $n = 3$ – количество повторностей.

$$C = \frac{1167,3^2}{24} = 56774,55\,.$$

Определяем общую сумму квадратов

$$C_Y = \sum X^2 - C, \qquad (4.56)$$

$$C_Y = 38,2^2 + 44,7^2 + 26,2^2 + 51,7^2 + 57,8^2 + 51,2^2 + 47,7^2 + 61,8^2 +$$
$$+ 57,9^2 + 48,8^2 + 50^2 + 50,8 + 32,2^2 + 33,3^2 + 33,3^2 + 62,3^2 + 50^2 +$$
$$+ 58^2 + 44,9^2 + 50^2 + 40^2 + 62^2 + 52,5^2 + 62^2 - 56774,55 = 2473,58$$

Число степеней свободы N равно:

$$N - 1 = 24 - 1 = 23\,.$$

Определяем сумму квадратов отклонений для повторений

$$C_P = \frac{\sum P^2}{m \cdot l \cdot k} - C, \qquad (4.57)$$

$$C_P = \frac{387,8^2 + 400,1^2 + 379,4^2}{2 \cdot 2 \cdot 2} - 56774,55 = 27,10.$$

Число степеней свободы равно (n–1) = (3–1) = 2.

Определяем сумму квадратов отклонений для вариантов:

$$C_V = \frac{\sum V^2}{n} - C, \qquad (4.58)$$

$$C_V = \frac{109,1^2 + 160,7^2 + 167,4^2 + 149,6^2 + 98,8^2 + 170,3^2 + 134,9^2 + 176,5^2}{3} -$$

$$- 56774,55 = 1973,45.$$

Число степеней свободы равно $m \cdot l \cdot k - 1 = 2 \cdot 2 \cdot 2 - 1 = 7$

Сумму квадратов отклонений для остатка (случайного варьирования) определяем из выражения:

$$C_Z = C_Y - C_V - C_Z, \qquad (4.59)$$

$$C_Z = 2473,58 - 27,10 - 1973,45 = 473,03.$$

Определяем суммы квадратов отклонений для факторов *A*, *B*, *C* (главных эффектов) и парных взаимодействий *AB*, *AC*, *BC*. Для этого на основании таблицы 4.6 составим вспомогательную таблицу 4.7, выпишем суммы по вариантам и определим суммы сумм для главных эффектов взаимодействий.

Таблиц 4.6 – Расчетные формулы для определения главных эффектов и взаимодействия

<table>
<tr><th colspan="4">Суммы по вариантам</th><th colspan="5">Суммы сумм по факторам и взаимодействиям</th></tr>
<tr><th rowspan="2">A</th><th rowspan="2">B</th><th colspan="2">C</th><th rowspan="2">A</th><th rowspan="2">B</th><th rowspan="2">AB</th><th rowspan="2">AC</th><th rowspan="2">BC</th></tr>
<tr><th>0</th><th>1</th></tr>
<tr><td rowspan="2">0</td><td>0</td><td>X_1</td><td>X_2</td><td rowspan="2">A_0
X_1+X_2+
$+X_3+X_4$</td><td rowspan="2">B_0
X_1+X_2+
$+X_5+X_6$</td><td>A_0B_0
X_1+X_2</td><td>A_0C_0
X_1+X_3</td><td>B_0C_0
X_1+X_5</td></tr>
<tr><td>1</td><td>X_3</td><td>X_4</td><td>A_0B_1
X_3+X_4</td><td>A_0C_1
X_2+X_4</td><td>B_0C_1
X_2+X_6</td></tr>
<tr><td rowspan="2">1</td><td>0</td><td>X_5</td><td>X_6</td><td rowspan="2">A_1
X_5+X_6+
$+X_7+X_8$</td><td rowspan="2">B_1
X_3+X_4+
$+X_7+X_8$</td><td>A_1B_0
X_5+X_6</td><td>A_1C_0
X_5+X_7</td><td>B_1C_0
X_3+X_7</td></tr>
<tr><td>2</td><td>X_7</td><td>X_8</td><td>A_1B_1
X_7+X_8</td><td>A_1C_1
X_6+X_8</td><td>B_1C_1
X_4+X_8</td></tr>
<tr><td colspan="2">Суммы сумм С</td><td>C_0
X_1+X_3+
$+X_5+X_7$</td><td>C_1
X_2+X_4+
$+X_6+X_8$</td><td>-</td><td>-</td><td>-</td><td>-</td><td>-</td></tr>
<tr><td colspan="2">$\sum X$ проверка</td><td colspan="2">C_0+C_1</td><td>A_0+A_1</td><td>B_0+B_1</td><td>$A_0B_0+A_0B_1+$
$+A_1B_0+A_1B_1$</td><td>$A_0C_0+A_0C_1+$
$+A_1C_0+A_1C_1$</td><td>$B_0C_0+B_0C_1+$
$+B_1C_0+B_1C_1$</td></tr>
<tr><td colspan="2"></td><td colspan="2">$\sum C$</td><td>$\sum A$</td><td>$\sum B$</td><td>$\sum AB$</td><td>$\sum AC$</td><td>$\sum BC$</td></tr>
</table>

Таблица 4.7 – Определение главных эффектов и взаимодействий

Суммы по вариантам				Суммы сумм по факторам и взаимодействиям				
A	*B*	*C*		*A*	*B*	*AB*	*AC*	*BC*
		0	1					
0	0	109,1	160,7	A_0 586,8	B_0 538,9	A_0B_0 269,8	A_0C_0 276,5	B_0C_0 207,9
	1	167,4	149,6			A_0B_1 317	A_0C_1 310,3	B_0C_1 331
1	0	98,9	170,3	A_1 580,5	B_1 628,4	A_1B_0 269,1	A_1C_0 233,7	B_1C_0 302,3
	2	134,9	176,5			A_1B_1 311,4	A_1C_1 346,8	B_1C_1 326,1
Суммы сумм С		C_0 510,2	C_1 657,1	–	–	–	–	–
$\sum X$ провер-ка		1167,3		1167,3	1167,3	1167,3	1167,3	1167,3
		$\sum C$		$\sum A$	$\sum B$	$\sum AB$	$\sum AC$	$\sum BC$

$$C_A = \frac{\sum A^2}{l \cdot k \cdot n} - C, \qquad (4.60)$$

$$C_A = \frac{586,8^2 + 580,5^2}{2 \cdot 2 \cdot 3} - 56774,55 = 1,66,$$

при $f_m = m - 1 = 2 - 1 = 1$ степень свободы.

$$C_B = \frac{\sum B^2}{m \cdot k \cdot n} - C, \qquad (4.61)$$

$$C_B = \frac{538,9^2 + 628,4^2}{2 \cdot 2 \cdot 3} - 56774,55 = 333,76,$$

при $f_l = l - 1 = 2 - 1 = 1$ степень свободы.

$$C_C = \frac{\sum C^2}{m \cdot l \cdot n} - C, \qquad (4.62)$$

$$C_C = \frac{510,2^2 + 657,1^2}{2 \cdot 2 \cdot 3} - 56774,55 = 899,15\,,$$

при $f_k = k - 1 = 2 - 1 = 1$ степень свободы.

$$C_{AB} = \frac{\sum AB^2}{k \cdot n} - C_A - C_B - C, \qquad (4.63)$$

$$C_{AB} = \frac{269,8^2 + 317^2 + 269,1^2 + 311,4^2}{2 \cdot 3} - 1,66 - 333,76 - 56774,55 = 1,$$

при $f_{ml} = (m - 1)(l - 1) = (2 - 1)(2 - 1) = 1$ степень свободы.

$$C_{AC} = \frac{\sum AC^2}{l \cdot n} - C_A - C_C - C, \qquad (4.64)$$

$$C_{AC} = \frac{276,5^2 + 310,3^2 + 233,7^2 + 346,8^2}{2 \cdot 3} - 1,66 - 899,15 - 56774,55 = 262,02,$$

при $f_{mk} = (m - 1)(k - 1) = (2 - 1)(2 - 1) = 1$ степень свободы.

$$C_{BC} = \frac{\sum BC^2}{k \cdot n} - C_B - C_C - C, \qquad (4.65)$$

$$C_{BC} = \frac{207,9^2 + 331^2 + 302,3^2 + 326,1^2}{2 \cdot 3} - 333,76 - 899,15 - 56774,55 = 410,86,$$

при $f_{lk} = (l - 1)(k - 1) = (2 - 1)(2 - 1) = 1$ степени свободы.

$$C_{ABC} = C_V - (C_A + C_B + C_C + C_{AB} + C_{AC} + C_{BC}), \qquad (4.66)$$

$$C_{ABC} = 1373,45 - (1,66 + 333,76 + 899,15 + 1 + 262,02 + 410,86) = 65$$

при f_{mlk} = $(m-1)(l-1)(k-1)=(2-1)(2-1)(2-1) = 1$ числе степеней свободы.

Фактические и теоретические значения $F_ф$ и F_{05}, а также числа степеней свободы для всех факторов и их взаимодействий приведены в таблице 4.8

Определяем значения НСР. Значение t-критерия при 5%-м уровне значимости и 14 степенях свободы равно 2,15 (таблица А1).

Таблица 4.8 – Результаты дисперсионного анализа

Дисперсия	Суммы квадратов	Степени свободы	Средний квадрат	$F_ф$	F_{05}
Общая	2473,58	23	–	–	–
Повторений	27,10	2	–	–	–
Скорости A	1,66	1	1,66	0,05	4,60
Угла B	333,76	1	333,76	9,88	4,60
Частоты вращения C	899,15	1	899,15	26,41	4,60
Взаимодействия AB	1,00	1	1,00	0,03	4,60
Взаимодействия AC	262,02	1	262,02	7,75	4,60
Взаимодействия BC	410,86	1	410,86	12,16	4,60
Взаимодействия ABC	65,00	1	65,00	1,92	4,60
Остаток (ошибка)	473,03	14	33,79	–	–

Для частных различий:

$$HCP_{05} = t_{05}\sqrt{\frac{2S^2}{n}}100\,\%, \qquad (4.67)$$

$$HCP_{05} = 2,15\sqrt{\frac{2 \cdot 33,79}{3}}100\,\% = 10,2\,\%.$$

Для главных эффектов:

$$HCP_{05} = t_{05}\sqrt{\frac{2S^2}{m \cdot l \cdot n}}100\,\%, \qquad (4.68)$$

$$HCP_{05} = 2{,}15\sqrt{\frac{2 \cdot 33{,}79}{2 \cdot 2 \cdot 3}}100\,\% = 5{,}1\,\%.$$

Для парных взаимодействий:

$$HCP_{05} = t_{05}\sqrt{\frac{2S^2}{m \cdot n}}100\,\%, \qquad (4.69)$$

$$HCP_{05} = t_{05}\sqrt{\frac{2 \cdot 33{,}79}{2 \cdot 3}}100\,\% = 7{,}2\,\%.$$

Из сравнения дисперсий (таблица 4.8) следует (при условии $F_{факт} > F_{05}$), что существенно влияют на степень очистки початков:

– угол наклона русел;
– частота вращения вальцов;
– взаимодействие скорости и частоты вращения вальцов;
– угла наклона русел и частоты вращения вальцов.

При этом влияние факторов составило:

$$\eta_V^2 = \frac{C_V}{C_Y}100\,\%, \qquad (4.70)$$

$$\eta_V^2 = \frac{1373{,}45}{2473{,}58}100\,\% = 79{,}8\,\%.$$

Влияние повторений:

$$\eta_P^2 = \frac{C_P}{C_Y} 100\,\%, \tag{4.71}$$

$$\eta_P^2 = \frac{27{,}1}{2473{,}58} 100\,\% = 1{,}1\,\%.$$

Влияние случайных факторов:

$$\eta_X^2 = \frac{C_Z}{C_Y} 100\,\%, \tag{4.72}$$

$$\eta_X^2 = \frac{473{,}03}{2473{,}58} 100\,\% = 19{,}1\,\%.$$

В конечном итоге дисперсионный анализ имеет ряд существенных преимуществ перед методом попарных сравнений по критерию Стьюдента:

1. Более высокая точность оценки сравниваемых вариантов за счет общего большего числа наблюдений;

2. Особое преимущество дисперсионного анализа заметно при исследовании многофакторных опытов.

В то же время дисперсионный анализ требует строго соблюдать все требования к обеспечению условий для проведения подобных опытов; принцип рандомизации (случайности), независимость наблюдений, однородность дисперсий, нормальное распределение случайных величин.

Контрольные вопросы

1. Что называется дисперсионным анализом.

2. Структура однофакторного дисперсионного анализа.

3. Структура математической модели двух и трехфакторного дисперсионного анализа.

4. Однородность дисперсий.

5. Преобразование случайных величин для достижения однородности дисперсий.

6. Критерий Фишера.

7. Взаимодействие факторов – двойное и тройное.

8. Определение корректирующего фактора.

5 КОРРЕЛЯЦИОННЫЙ И РЕГРЕССИОННЫЙ АНАЛИЗЫ

5.1 Зависимые и независимые случайные величины. Основы корреляционного анализа

При изучении объектов, на которые действуют случайные величины необходимо всегда оценивать характер и степень их зависимости, т. е. действующих факторов на оцениваемый показатель и самих действующих факторов между собой.

Понятие независимости случайных величин является одним из важнейших понятий в теории вероятностей.

Случайная величина *Y* считается независимой от случайной величины *X*, если закон распределения случайной величины *Y* не зависит от значения случайной величины *X*.

Зависимость или независимость случайных величин всегда взаимны, т. е., если случайная величина *Y* не зависит от случайной величины *X*, то можно утверждать, что и величина *X* не зависит от величины *Y*.

Из последнего утверждения можно уточнить, **что случайные величины *X* и *Y* называются независимыми, если закон распределения каждой из них не зависит от того, какое значение приняла другая. В противном случае случайные величины *X* и *Y* называются зависимыми.**

Понятие статистической зависимости случайных величин отличается от понятия функциональной зависимости, используемых при описании законов естественных наук. Функциональная зависимость представляет собой жесткую зависимость, когда каждому значению одной из величин соответствует одно строго вычисленное по функциональной зависимости значение другой величины.

В случае со случайными величинами мы имеем отношение с более общим случаем, когда одному и тому же значению одной случайной величины может соответствовать несколько значений другой случайной величины, в том числе и случай с функциональной зависимостью. В общем случае при изучении случайных величин знание одной случайной величины может позволить установить лишь закон распределения второй случайной величины. Зависимость между случайными величинами указывает не только величину (тесноту) связи, но и направление изменения зависимой случайной величины.

При исследовании взаимной зависимости случайных величин используется понятие **корреляция** – статистическая взаимосвязь двух и более случайных величин. Общую картину их изменчивости можно получить, изобразив на координатной плоскости все пары точек случайных величин выборки (x_i, y_i). Такое изображение называется **корреляционным полем или корреляционной решеткой**. Уже по виду корреляционного поля можно делать вывод о характере связи между случайными величинами. Например, на рисунке (5.1, *а*) связь можно назвать полной положительной зависимостью.

Для определения существования связи двух случайных величин и ее силы используем свойство дисперсии суммы независимых случайных величин. Известно, что дисперсия суммы двух независимых случайных величин X и Y равна сумме дисперсий этих величин, т. е.:

$$D(X+Y) = D(X) + D(Y). \tag{5.1}$$

Но, если для исследуемых случайных величин не будет соблюдаться это равенство, т. е.:

$$D(X+Y) \neq D(X) + D(Y), \tag{5.2}$$

то это служит доказательством существования зависимости между случайными величинами X и Y.

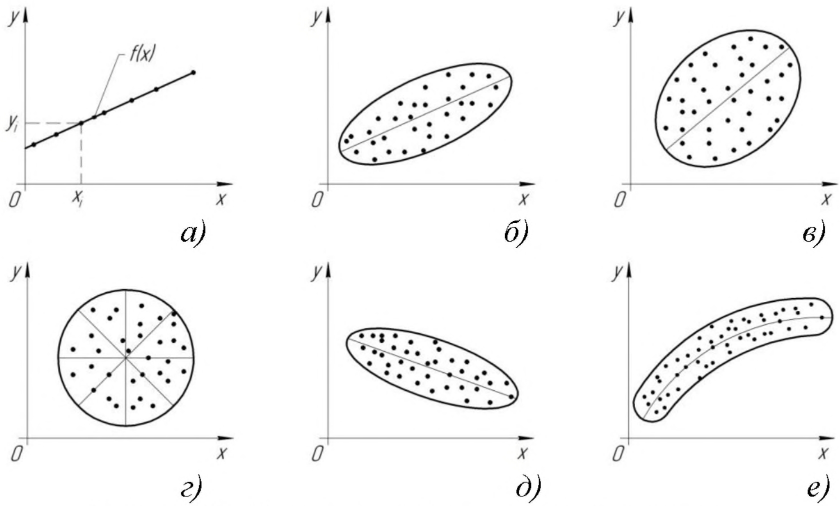


Рисунок 5.1 – Точечные диаграммы с соответствующими им коэффициентами корреляции и названиями корреляции:

а – полная положительная зависимость $r \approx +1$; *б* – сильная положительная $r \approx +0,8$; *в* – слабая положительная $r \approx +0,1$; *г* – нулевая зависимость $r \approx 0$; *д* – средняя отрицательная $r \approx -0,5$; *е* – нелинейная зависимость

Из свойств дисперсии и математического ожидания имеем:

$$D(X+Y) = M\left[X+Y-M(X+Y)\right]^2 =$$
$$= M\left[(X-MX)^2 + 2(X-MX)(Y-MX) + M(Y-MY)^2\right] =$$
$$= M(X-MX)^2 + 2M\left[(X-MX)(Y-MY)\right] + M(Y-MY)^2.$$

Так как

$$M(X-MX)^2 = D(X), \text{ и } M(Y-MY)^2 = D(Y),$$

получим

$$D(X+Y)-(D(X)+D(Y))=2M\left[(X-MX)(Y-MY)\right].$$

Следовательно, зависимость $D(X+Y)-D(X)+D(Y)\neq 0$.

Тогда надо признать, что именно величина $M\left[(X-MX)(Y-MY)\right]$ и есть корреляция.

Величину $M\left[(X-MX)(Y-MY)\right]$ называют **корреляционным моментом**, который зависит от единиц измерения величин X и Y. Поэтому на практике чаще используется безразмерная величина:

$$\rho=\frac{M\left[(X-MX)(Y-MY)\right]}{\sqrt{DXDY}}. \tag{5.3}$$

Данная величина называется **коэффициентом корреляции**.

Свойства коэффициента корреляции:

1. Коэффициент корреляции независимых случайных величин равен нулю.

2. Коэффициент корреляции не меняется от прибавления к случайным величинам X и Y постоянных слагаемых или от умножения X и Y на положительные числа.

3. Если одну из величин X или Y, не меняя другой, умножить на –1, то на –1 умножится и коэффициент корреляции.

4. Значениям коэффициента корреляции отвечает неравенство

$$-1\leq\rho\leq 1.$$

Из этого следует, что максимальная корреляция соответствует значениям $\rho=\pm 1$, что возможно только в случае, когда между значениями случайных величин X и Y существует строгая функциональная связь.

5. Если $\rho > 0$, то величины X и Y с точностью до случайных погрешностей одновременно возрастают или убывают, если же $\rho < 0$, то с возрастанием одной величины другая убывает.

Учитывая вышеизложенное, можно считать, **что коэффициент корреляции есть показатель того, насколько связь между случайными величинами близка к строгой линейной зависимости. Он одинаково отмечает и слишком большую долю случайности, и слишком большую криволинейность этой связи.**

Выборочный коэффициент корреляции r вычисляется по такой же формуле, что и генеральный коэффициент ρ, но только здесь берутся выборочные математические показатели – средние арифметические и стандартные отклонения. Если через $\overline{x}$ и $\overline{y}$ обозначить средние значения для x_i и y_i:

$$\overline{x} = \frac{1}{m}\sum x_i, \quad \overline{y} = \frac{1}{m}\sum y_i.$$

Откуда:

$$r = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{(m-1)S_x S_y},$$

где через S_x и S_y обозначены выборочные дисперсии

$$S_x^2 = \frac{1}{m-1}\sum (x_i - \overline{x})^2, \quad S_y^2 = \frac{1}{m-1}\sum (y_i - \overline{y})^2.$$

При вычислении удобно пользоваться формулами:

$$\sum (x_i - \bar{x})(y_i - \bar{y}) = \sum x_i y_i - \frac{\sum x_i \sum y_i}{m},$$

$$(m-1)S_x^2 = \sum x_i^2 - \frac{1}{m}(x_i)^2,$$

$$(m-1)S_y^2 = \sum y_i^2 - \frac{1}{m}(y_i)^2.$$

При достаточно большом объеме выборки m выборочный коэффициент корреляции r приближенно равен генеральному коэффициенту ρ, однако оценить возникающую при этом погрешность очень трудно. Но это не обязательно, так как точное знание ρ в расчетах почти не используется и нужно нам лишь как показатель силы связи.

В итоге заметим, что **оценка зависимости случайных величин по выборочному коэффициенту корреляции называется корреляционным анализом.**

Для вычисления коэффициента корреляции составляют таблицу 5.1.

Таблица 5.1 – Форма для расчета коэффициента корреляции

№ пары наблюдения	Значения признаков		x	y	xy
	x	y			
1					
	$\sum x$	$\sum y$	$\sum x^2$	$\sum y^2$	$\sum xy$

Квадрат коэффициента корреляции r^2 называется **коэффициентом детерминации** и обозначается d_{xy}. Он показывает долю (%) тех изменений, которые в данной вариации зависят от случайного фактора.

Стандартная ошибка коэффициента корреляции определяется по формуле:

$$S_r = \sqrt{\frac{1-r^2}{n-2}}, \qquad (5.4)$$

где S_r – ошибка коэффициента корреляции;

r – коэффициент корреляции;

n – число пар значений, по которым вычислен выборочный коэффициент корреляции.

Критерий существенности коэффициента корреляции рассчитывается по формуле:

$$t_r = \frac{r}{S_r}. \qquad (5.5)$$

Если $t_{r факт} \geq t_{теор}$ – связь существенна, а при $t_{r факт} \leq t_{теор}$ – несущественна.

Число степеней свободы равно n–2 (n – число пар).

5.2 Регрессионный анализ

Основная задача регрессионного анализа – определение формулы корреляционной зависимости, т. е. уравнения прямой линии.

Уравнение прямолинейной регрессии y по x имеет вид:

$$Y = \overline{y} - b_{XY}(x - \overline{x}), \qquad (5.6)$$

где $\overline{x}$ и $\overline{y}$ – средние арифметические для рядов y и x;

b_{yx} – коэффициент регрессии y по x.

Коэффициент регрессии вычисляют по выражениям:

$$b_{YX} = \frac{(X - \bar{x})(Y - \bar{y})}{\sum (X - \bar{x})^2}, \quad \text{и} \quad b_{XY} = \frac{(X - \bar{x})(Y - \bar{y})}{\sum (Y - \bar{y})^2}, \qquad (5.7)$$

Коэффициент регрессии b_{YX} показывает, как изменяется Y при изменении X на единицу измерения, и выражается в единицах Y, а b_{XY} указывает регрессию X на Y и выражается в единицах X.

При исследовании односторонней зависимости, например корреляции между урожаями Y и количеством выпавших осадков X, вычисляют только один коэффициент регрессии результативного признака Y на факториальный X, т. е. значение b_{XY}, так как регрессия X на Y лишена в подобных случаях логического смысла.

Таким образом, **коэффициентом линейной регрессии называется число, показывающее, в каком направлении и на какую величину изменяется в среднем признак Y (функция) при изменении признака X (аргумента) на единицу измерения. Коэффициенты регрессии имеют знак коэффициента корреляции.**

Произведение коэффициентов регрессии равно квадрату коэффициента корреляции:

$$b_{YX} b_{XY} = r^2. \qquad (5.8)$$

Этой формулой можно пользоваться как проверочной при вычислении коэффициентов регрессии.

Ошибку коэффициента регрессии вычисляют по выражениям:

$$S_{b_{YX}} = S_r \sqrt{\frac{\sum (Y - \bar{y})^2}{\sum (X - \bar{x})^2}}, \quad \text{и}\, S_{b_{XY}} = S_r \sqrt{\frac{\sum (X - \bar{x})^2}{\sum (Y - \bar{y})^2}}. \qquad (5.9)$$

Критерий существенности коэффициента регрессии определяют по формуле:

$$t_b = \frac{b}{S_b}. \qquad (5.10)$$

Если определен критерий существенности для коэффициента корреляции, он может быть использован для оценки значимости коэффициента регрессии, так как:

$$t_b = t_r. \qquad (5.11)$$

Существенность коэффициента регрессии оценивают по таблице критерия значимости Стьюдента при принятом уровне значимости. При этом число степеней свободы принимают равным $n - 2$.

Корреляция может быть изображена графически в виде линии регрессии. Для построения графика по оси абсцисс откладывают значения признака X, по оси ординат – значения признака Y и каждое наблюдение над двумя переменными отмечают точкой с координатами (X, Y). Такой график называется «точечной диаграммой» или «корреляционным полем».

5.3 Частная и множественная линейные корреляции и регрессии

Корреляция называется множественной, если на величину результативного признака одновременно влияют несколько факториальных.

Наиболее простой формой множественной связи является линейная зависимость между тремя признаками, когда один из них, например тяговое сопротивление, рассматривается как функция (Y), а два другие – как аргументы (X и Z). В качестве тесноты связи линейной связи трех признаков используют частные коэффициенты корреляции, обозначаемые

$r_{XY \cdot Z}, r_{XZ \cdot Y}, r_{YZ \cdot X}$ и множественные коэффициенты корреляции, обозначаемые символами, $R_{X \cdot YZ}, R_{Y \cdot XZ}, R_{Z \cdot XY}$.

Частный коэффициент корреляции – это показатель, измеряющий степень сопряженности двух признаков при постоянном значении третьего.

Математическая статистика позволяет установить корреляцию между двумя признаками при постоянном значении третьего, не ставя специального эксперимента, а используя парные коэффициенты корреляции r_{XY}, r_{XZ}, r_{YZ}. Частные коэффициенты корреляции рассчитывают по формулам:

$$r_{XY \cdot Z} = \frac{r_{XY} - r_{XZ} \cdot r_{YZ}}{\sqrt{(1 - r_{XY}^2)(1 - r_{YZ}^2)}},$$
$$r_{XZ \cdot Y} = \frac{r_{XZ} - r_{XY} \cdot r_{ZY}}{\sqrt{(1 - r_{XY}^2)(1 - r_{ZY}^2)}}, \qquad (5.12)$$
$$r_{YZ \cdot X} = \frac{r_{YZ} - r_{XY} \cdot r_{XZ}}{\sqrt{(1 - r_{XY}^2)(1 - r_{ZX}^2)}}.$$

Здесь в индексах буквы перед точкой указывают, между какими признаками изучается зависимость, а буква после точки – влияние какого признака исключается (элиминируется). Ошибку и критерий значимости частной корреляции определяют по тем же формулам, что и парной корреляции:

$$S_{r_{XY \cdot Z}} = \sqrt{\frac{1 - S_{r_{XY \cdot Z}}^2}{n - 2}}, \quad t = \frac{r}{S_r}. \qquad (5.13)$$

Теоретические значения t берут из таблиц Стьюдента для принятого уровня значимости и $n - 3$ степеней свободы.

Подобно парным коэффициентам корреляции частные коэффициенты могут принимать значения, заключенные между – 1 и +1. Частные коэффициенты детерминации находят путем возведения в квадрат частных коэффициентов корреляции:

$$d_{XY \cdot Z} = r^2_{XY \cdot Z}, \quad d_{XZ \cdot Y} = r^2_{XZ \cdot Y}, \quad d_{YZ \cdot X} = r^2_{YZ \cdot X}. \tag{5.14}$$

Определение степени частного воздействия отдельных переменных на результативный признак при исключении связи его с другими признаками, искажающими эту корреляцию, часто представляет большой интерес. Например, тесноту связи урожаев с осадками может сильно искажать варьирование температуры, и поэтому целесообразно изучить связь между первыми двумя признаками при постоянных значениях третьего. При этом третий признак удерживается на постоянном уровне, а другие признаки варьируют и находятся в корреляционном отношении друг с другом.

Множественный коэффициент корреляции трех переменных – это показатель тесноты линейной связи между одним из признаков (буква индекса перед точкой) и совокупностью двух других признаков (буквы индекса после точки):

$$\begin{gathered} R_{X \cdot YZ} = \sqrt{\frac{r^2_{XY} + r^2_{XZ} - 2r_{XY} r_{XZ} r_{YZ}}{1 - r^2_{YZ}}}, \\ R_{Y \cdot XZ} = \sqrt{\frac{r^2_{XY} + r^2_{YZ} - 2r_{XY} r_{XZ} r_{YZ}}{1 - r^2_{XZ}}}, \\ R_{Z \cdot XY} = \sqrt{\frac{r^2_{XZ} + r^2_{YZ} - 2r_{XY} r_{XZ} r_{YZ}}{1 - r^2_{XY}}}. \end{gathered} \tag{5.15}$$

Эти формулы позволяют легко вычислить множественные коэффициенты корреляции при известных значениях коэффициентов парной корреляции r_{XY}, r_{XZ}, r_{YZ}.

Коэффициент R не отрицателен и всегда находится в пределах от 0 до 1. При приближении R к единице степень линейной связи трех признаков увеличивается. Между коэффициентом множественной корреляции, например $R_{Y \cdot XZ}$, и двумя коэффициентами парной корреляции r_{XY} и r_{YZ} существует следующее соотношение: каждый из парных коэффициентов не может превышать по абсолютной величине $R_{Y \cdot XZ}$.

Квадрат коэффициента множественной корреляции R^2 называется коэффициентом множественной детерминации. Он показывает долю вариации зависимой переменной под воздействием изучаемых факторов.

Значимость множественной корреляции оценивается по F-критерию:

$$F = \frac{R^2}{1 - R^2} \left(\frac{n-k}{k-1}\right), \qquad (5.16)$$

где n – объем выборки;

k – число признаков; в нашем случае $k = 3$.

Теоретическое значение F-критерия находят по таблице критерия Фишера чисел степеней свободы $k - 1$ и $n - k$ и принятого уровня значимости. Нулевая гипотеза о равенстве множественного коэффициента корреляции в совокупности нулю (H_0: $R = 0$) принимается, если $F_{факт} < F_{теор}$ и отвергается, если $F_{факт} \geq F_{теор}$.

Математическое уравнение для прямолинейной зависимости между тремя переменными называется линейным уравнением плоскости регрессии. Оно имеет следующий общий вид:

$$Y = a + b_1 X + b_2 Z, \tag{5.17}$$

где Y – зависимая переменная;

X и Z – независимые переменные;

a – общее начало отсчета;

b_1 и b_2 – коэффициенты частной регрессии.

Коэффициент b_1 показывает, на какую величину увеличивается Y при увеличении на одну единицу X при постоянном значении Z. Коэффициент b_2 указывает, на какую величину увеличивается Y при увеличении Z на единицу при постоянном значении X. Поэтому часто используют обозначения $b_1 = b_{YX \cdot Z}$ и $b_2 = b_{YZ \cdot X}$, принятые для частных коэффициентов корреляции. При этом параметры a, b_1 и b_2 вычисляют методом наименьших квадратов, который позволяет найти такое положение плоскости регрессии в пространстве, когда сумма квадратов отклонений эмпирических точек от нее является минимальной.

$$b_1 = \frac{\sum (Z-\overline{z})^2 \cdot (X-\overline{x})(Y-\overline{y}) - \sum (X-\overline{x})(Z-\overline{z}) \cdot (Y-\overline{y})(Z-\overline{z})}{\sum (X-\overline{x})^2 \sum (Z-\overline{z})^2 - \left[\sum (X-\overline{x})(Z-\overline{z})\right]^2}, \tag{5.18}$$

$$b_2 = \frac{\sum (X-\overline{x})^2 \cdot (Y-\overline{y})(Z-\overline{z}) - \sum (X-\overline{x})(Z-\overline{z}) \cdot (X-\overline{x})(Y-\overline{y})}{\sum (X-\overline{x})^2 \sum (Z-\overline{z})^2 - \left[\sum (X-\overline{x})(Z-\overline{z})\right]^2}, \tag{5.19}$$

$$a = \overline{y} - b_1 \overline{x} - b_2 \overline{z}. \tag{5.20}$$

И, наконец, необходимо подчеркнуть, что уравнения регрессии имеют смысл только в области фактических значений X, Y, и Z тогда, когда корреляционная связь значимо отличается от нуля.

5.4 Криволинейная корреляция и регрессия

Если связь между изучаемыми явлениями существенно отличается от линейной, что легко установить по точечному графику, то коэффициент корреляции непригоден в качестве меры связи. Он может указать на отсутствие сопряженности там, где налицо сильная криволинейная зависимость. Поэтому необходим новый показатель, который правильно измерял бы степень криволинейной зависимости. Таким показателем является **корреляционное отношение**, обозначаемое греческой буквой η (эта). Оно измеряет степень корреляции при любой ее форме.

Корреляционное отношение при малом числе наблюдений вычисляют по формуле:

$$\eta_{yx} = \sqrt{\frac{\sum(Y-\bar{y})^2 - \sum(Y-\bar{y}_x)^2}{\sum(Y-\bar{y})^2}}, \tag{5.21}$$

где $\sum(Y-\bar{y})^2$ – сумма квадратов отклонений индивидуальных значений Y от общей средней арифметической $\bar{y}$;

$\sum(Y-\bar{y}_x)^2$ – сумма квадратов отклонений вариантов от частных средних $\bar{y}_x$, соответствующих определенным, фиксированным значениям независимой переменной X.

Для вычисления корреляционного отношения значений независимого признака X располагают по ранжиру в возрастающем порядке и разбивают весь ряд наблюдений на 4–7 групп с таким расчетом, чтобы в каждой группе по ряду X было не менее двух наблюдений. Затем определяют общую среднюю $\bar{y}$, групповые средние $\bar{y}_x$, соответствующие каждой фиксированной группе X, и суммы квадратов отклонений для общего $\sum(Y-\bar{y})^2$ и группового $\sum(Y-\bar{y}_x)^2$ варьирования признака Y.

При большом объеме наблюдений ($n > 30$) обработка материала для вычисления корреляционного отношения проводится в корреляционной таблице. После группировки и разноски дат определяют сумму квадратов отклонений группового варьирования $\sum f(\bar{y}_x-\bar{y})^2$,сумму квадратов отклонений общего варьирования $\sum f(Y-\bar{y})^2$ и вычисляют корреляционное отношение по формуле:

$$\eta_{yx}=\sqrt{\frac{\sum f(\bar{y}_x-\bar{y})^2}{\sum f(Y-\bar{y})^2}}. \qquad (5.22)$$

Сумма квадратов отклонений групповых средних $\bar{y}_x$ от общей средней $\bar{y}$ (групповое среднее) характеризует ту часть варьирования признака Y, которая связана с изменчивостью признака X. Сумма квадратов разностей между каждой датой и общей средней $\bar{y}$, т. е. $\sum f(Y-\bar{y})^2$, характеризует общее варьирование признака Y.

При функциональной зависимости Y от X корреляционное отношение равно единице; если оно равно нулю, то показывает некоррелированность Y от X; при промежуточном характере корреляционной зависимости корреляционное отношение заключено в пределах:

$$0<\eta_{yx}<1.$$

Чем ближе η_{yx} к единице, тем сильнее, ближе функциональная зависимость Y от X, и, наоборот, чем ближе η_{yx} к нулю, тем слабее выражена эта зависимость.

Отношение сумм квадратов группового варьирования к общему, т. е. η^2_{yx}, имеет самостоятельное значение. Оно показывает ту долю варьирования признака Y, которая обусловлена степенью колебания признака X. Эта величина, называемая **индексом детерминации**, определяет процент вариации Y под влиянием X.

Ошибку и критерий существенности корреляционного отношения рассчитывают по формулам:

$$S_\eta = \sqrt{\frac{1-\eta^2}{n-2}}, \qquad t_\eta = \frac{\eta}{S_\eta}. \tag{5.23}$$

Теоретическое значение критерия t для 5%-го или 1%-го уровня значимости находят по таблице квантилей распределения Стьюдента; число степеней свободы принимают равным $n-2$.

При обработке экспериментального материала методом дисперсионного анализа значение η_{yx} определяется как отношение суммы квадратов отклонений для вариантов C_V к общей сумме квадратов C_Y.

$$\eta^2_{yx} = \frac{C_V}{C_Y}, \qquad \eta_{yx} = \sqrt{\frac{C_V}{C_Y}}. \tag{5.24}$$

5.5 Критерий линейности корреляции

Для определения степени приближения криволинейной зависимости к прямолинейной используется критерий F, вычисленный по выражению:

$$F = \frac{(\eta^2 - r^2)(n-k)}{(1-\eta^2)(k_x - 2)}, \tag{5.25}$$

где η^2 – квадрат корреляционного отношения Y по X;

r^2 – квадрат коэффициента линейной корреляции;

n – объем выборки;

k_x – число групп по ряду X.

Связь можно практически принять за линейную, если $F_{факт} < F_{теор}$, и определять показатели для прямолинейной корреляции и регрессии. Корреляция нелинейная, если $F_{факт} \geq F_{теор}$. Теоретическое значение F приведены в таблицах квантилей распределения Фишера для числа степеней свободы $\gamma_1 = k_x - 2$ и $\gamma_2 = n - 2$.

Гипотеза о линейности отвергается ($F_{факт} > F_{теор}$) и пользоваться линейной корреляцией и регрессией нельзя на выбранном уровне значимости.

Криволинейные зависимости между двумя переменными могут быть выражены в виде кривых линий регрессии.

Криволинейная регрессия – это такая зависимость, когда при одинаковых приращениях независимой переменной *X* зависимая переменная *Y* имеет неодинаковые приращения.

Эмпирические точки корреляционного поля при криволинейной корреляции раполагаются около кривых различного типа – парабол, гипербол, логарифмических кривых и т. п.

Контрольные вопросы

1. Понятия зависимых и независимых случайных величин.
2. Понятие корреляционного поля или корреляционной решетки.

3. Понятие корреляции.
4. Свойства корреляции.
5. Предельные значения коэффициента корреляции.
6. Сущность корреляционного анализа.
7. Определение коэффициента детерминации.
8. Основная задача регрессионного анализа.
9. Определение коэффициентов регрессии уравнения прямолинейной регрессии.
10. Что называется множественной корреляцией.
11. Понятие частного коэффициента корреляции.
12. Как определяется честный коэффициент детерминации.
13. Как определяется коэффициент множественной детерминации.
14. Понятие криволинейной корреляции и регрессии.
15. Корреляционное отношение и предельные его значения.
16. Критерий линейности корреляции и его определение.

6 АППРОКСИМАЦИЯ ЭКСПЕРИМЕНТАЛЬНЫХ ДАННЫХ

При проведении экспериментальных исследований всегда появляется необходимость нахождения математической зависимости, определяемой результатами эксперимента в точках факторного пространства, в которых проводились опыты. Эксперимент в научных исследованиях проводится с целью уточнения и корректировки функциональных зависимостей, полученных на основе физических законов, полученных ранее при проведении теоретических исследований. Получение регрессионных зависимостей по экспериментальным точкам проводится методами аппроксимации. **Аппроксимация (от лат. approximo – приближаюсь) в смысле приближения исследователя из всех известных алгебраических многочленов к наиболее простой зависимости для математического описания функционирования исследуемого объекта.** Математические зависимости, получаемые из эксперимента – регрессионные уравнения применяются практически во всех областях исследований и в первую очередь при исследовании процессов и машин в сельском хозяйстве. Ни в теоретических, ни в экспериментальных исследованиях невозможно полностью учесть влияние всех факторов, определяющих поведение интересующих исследователя параметров. Более того зачастую исследователь не владеет о существовании некоторых факторов. В связи с вышеизложенным ожидать точного описания исследуемых объектов в исследованиях невозможно. В этих условиях необходимо максимально полно учесть все действующие факторы и организовать все измерения с требуемой точностью. Это позволит получить максимально приближенные к истине уравнения регрессии, которые необходимы при решении многих задач:

– определении параметров переменных факторов, при которых достигается определенный (желаемый) результат. Например, при каком значении числа оборотов почвенной фрезы и ее поступательной скорости можно достичь получение необходимого по агротехническим требованиям качества крошения почвы;

– аппроксимация экспериментальных данных с целью получения оптимальных значений переменных факторов;

– выделение участков линейной зависимости исследуемых параметров (зависимой переменной) от независимых переменных факторов и во многих других случаях.

Одним из наиболее универсальных и точных методов аппроксимации, хотя и более повышенным вычислительным объемом, является метод наименьших квадратов.

6.1 Теория метода наименьших квадратов

Метод наименьших квадратов (МНК) – статистический метод, с помощью которого неизвестные параметры модели оцениваются по минимуму квадратов отклонений эмпирических значений от теоретических.

Допустим, что провели n независимых опытов при значениях независимой переменной $X = x_1, x_2, ..., x_n$, причем измерения зависимой переменной Y при этих значениях дали результаты $y_1, y_2, ..., y_n$ (таблица 6.1). Нанесем на координатную сетку точки $(x_i; y_i)$ (рисунок 6.1).

Таблица 6.1 – Таблица измерений независимых и зависимых переменных

X	x_1	x_2	...	x_{n-1}	x_n
Y	y_1	y_2	...	y_{n-1}	y_n

Предположим, что зависимая переменная Y зависит в «среднем» линейно от независимых переменных X:

$$y = a + bx. \qquad (6.1)$$

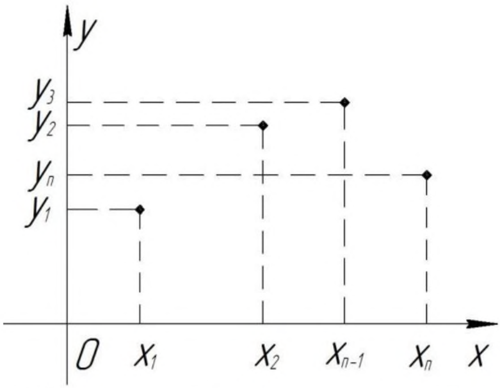


Рисунок 6.1 – Экспериментальные точки на координатной сетке

Если повторить опыт при одном и том же значении x_i, то мы получим соответствующую серию значений y. Благодаря воздействию на испытуемый объект случайных неучтенных факторов получим серию значений y при каждом значении x_i. Предполагая линейную зависимость между y и x для каждой серии опытов при одном и том же значении x, получим значения параметров a и b как арифметические средние из этой серии опытов. Ели же предположить, что для линейной зависимости проведены хотя бы две серии опытов соответственно при двух значения независимой переменной x и в каждом измерении присутствует действие неучтенных факторов для каждого измерения имеем:

$$y_i = a + bx + \delta, \tag{6.2}$$

где δ – отклонения от линейности, вызванные действием неучтенных случайных факторов.

Но учитывая, что математическое ожидание величин $M\delta_i$ в каждой серии опытов равно нулю, то дисперсия $D\delta_i = \sigma^2$, которая не зависит от x_i. В конечном итоге наша задача заключается в оценке не только a и b, но и σ^2. Тогда плотность распределения наблюдаемых значений y_i выглядит так:

$$f(x_1,...,x_n,a,b,\sigma^2)=\frac{1}{\sigma^2\sqrt{2\pi}}e^{-\frac{\sum_{i=1}^{n}(y_i-a-bx_i)^2}{2\sigma^2}}. \qquad (6.3)$$

Задача заключается в оценке параметров a и b методом наименьших квадратов. Для этого находим частные производные показателя выражения $\sum_{i=1}^{n}(y_i-a-bx_i)^2$ по a и b.

$$\frac{\partial\sum_{i=1}^{n}(y_i-a-bx_i)^2}{\partial a}=-2\sum_{i=1}^{n}(y_i-a-bx_i), \qquad (6.4)$$

$$\frac{\partial\sum_{i=1}^{n}(y_i-a-bx_i)^2}{\partial b}=-2\sum_{i=1}^{n}x_i(y_i-a-bx_i). \qquad (6.5)$$

Приравняв полученные производные нулю после несложных преобразований, получим систему уравнений:

$$b\sum_{i=1}^{n}x_i^2+a\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}x_iy_i=0, \qquad (6.6)$$

$$b\sum_{i=1}^{n}x_i+an-\sum_{i=1}^{n}y_i=0. \qquad (6.7)$$

Для вычисления коэффициентов a и b удобно пользоваться вспомогательной таблицей 6.2.

Пример. Зависимость величины продольного сопротивления рыхлительной лапы $R_{прод}$ от глубины обработки почвы b. Экспериментальные данные приведены в таблице 6.3.

Таблица 6.2 – Вспомогательная таблица для вычисления коэффициентов a и b

i	x_i	y_i	x_i^2	$x_i y_i$	$d_i = (y_i - a - bx_i)$	d_i^2
1	x_1	y_1	x_1^2	$x_1 y_1$	$d_1 = (y_1 - a - bx_1)$	d_1^2
2	x_2	y_2	x_2^2	$x_2 y_2$	$d_2 = (y_2 - a - bx_2)$	d_2^2
…	…	…	…	…	…	…
n	x_n	y_n	x_n^2	$x_n y_n$	$d_n = (y_n - a - bx_n)$	d_n^2
$\sum$	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n} x_i^2$	$\sum_{i=1}^{n} x_i y_i$	$\sum_{i=1}^{n} d_i$	$\sum_{i=1}^{n} d_i^2$

Таблица 6.3 – Зависимость величины продольного сопротивления рыхлительной лапы $R_{прод}$ от глубины обработки почвы «a» по экспериментальным данным

b, см	12	15	17	20	30
$R_{прод}$, Н	500	700	850	1050	1450

Результаты расчета коэффициентов приведены в таблице 6.4.

Таблица 6.4 – Расчет коэффициентов уравнения регрессии

i	x_i	y_i	x_i^2	$x_i y_i$	d_i	d_i^2
1	12	500	144	6000	-56,89	3236,47
2	15	700	225	10500	-12,68	160,78
3	17	850	289	14450	33,5	1122,25
4	20	1050	400	21000	77,67	6032,63
5	30	1450	900	43500	-41,6	1730,56
$\sum$	94	4550	1958	95450	0	12282,69

Подставляя полученные в таблице результаты в уравнения 6.6 и 6.7 имеем:

$$1958b + 94a - 95450 = 0,$$
$$94b + 5a - 4550 = 0.$$

Решив полученную систему получим:

$$b = 51,93;$$
$$a = -66,27.$$

Таким образом, искомая регрессионная зависимость продольного сопротивления полольной лапы Y от глубины обработки почвы x можно выразить уравнением регрессии:

$$Y = 51,93x - 66,27.$$

Сумма квадратов отклонений экспериментальных данных от расчетных равна $d^2 = 12282,69H^2$.

Итак, подводя итоги, следует заметить, что построение эмпирической формулы проводится по следующим этапам:

- выбор формулы (теоретической зависимости) определенного типа (линейная, квадратичная, гиперболическая, параболическая, логарифмическая и т. д.);
- определение параметров выбранной формулы, т. е. значений различных постоянных величин, входящих в формулу;
- оценка результатов аппроксимации (степени приближения) экспериментальных данных эмпирической формулой.

Далее, пользуясь вышеприведенной методикой, рассмотрим нахождение параметров нескольких достаточно распространенных функциональных зависимостей в качестве аппроксимирующих уравнений регрессии экспериментальных данных.

6.2 Гиперболическая зависимость

Задание. Найти параметры a и b аппроксимирующей гиперболической функции $y = a + \frac{b}{x}$ по экспериментальным данным зависимости коэффициента вариации глубины обработки почвы у от ее средней арифметической x (таблица 6.5).

Таблица 6.5 – Зависимость коэффициента вариации глубины обработки почвы у от ее средней арифметической x

x, см	2	4	6	12
y, %	8,00	5,25	3,50	3,25

Определим сумму квадратов отклонений экспериментальных значений y_i от расчетных:

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n} (a + \frac{b}{x_i} - y_i)^2. \qquad (6.8)$$

Для этого сначала находим частные производные выражения (6.8) по параметрам a и b.

$$\frac{\partial \sum_{i=1}^{n} (a + \frac{b}{x_i} - y_i)^2}{\partial a} = 2\sum_{i=1}^{n} (a + \frac{b}{x_i} - y_i),$$
$$\frac{\partial \sum_{i=1}^{n} (a + \frac{b}{x_i} - y_i)^2}{\partial b} = 2\sum_{i=1}^{n} \frac{1}{x_i}(a + \frac{b}{x_i} - y_i). \qquad (6.9)$$

Приравняв нулю выражения (6.9) и раскрыв скобки, получим

$$\begin{cases} na + b\sum_{i=1}^{n}\frac{1}{x_i} - \sum_{i=1}^{n} y_i = 0, \\ a\sum_{i=1}^{n}\frac{1}{x_i} + b\sum_{i=1}^{n}\frac{1}{x_i^2} - \sum_{i=1}^{n}\frac{y_i}{x_i} = 0. \end{cases} \tag{6.10}$$

Составим расчетную таблицу на основании уравнений (таблица 6.6)

Таблица 6.6 – Расчет коэффициентов гиперболической зависимости

i	x_i	$\frac{1}{x_i}$	$\frac{1}{x_i^2}$	y_i	$\frac{y_i}{x_i}$
1	x_1	$\frac{1}{x_1}$	$\frac{1}{x_1^2}$	y_1	$\frac{y_1}{x_1}$
…	…	…	…	…	…
n	x_n	$\frac{1}{x_n}$	$\frac{1}{x_n^2}$	y_n	$\frac{y_n}{x_n}$
$\sum$	$\sum_{i=1}^{n} x_i$	$\sum_{i=1}^{n}\frac{1}{x_i}$	$\sum_{i=1}^{n}\frac{1}{x_i^2}$	$\sum_{i=1}^{n} y_i$	$\sum_{i=1}^{n}\frac{y_i}{x_i}$

Таблица 6.7 – Результаты расчета коэффициентов гиперболической зависимости

i	x_i	$\frac{1}{x_i}$	$\frac{1}{x_i^2}$	y_i	$\frac{y_i}{x_i}$	d_i	d_i^2
1	2	0,5	0,25	8,00	4	-0,58	0,3364
2	4	0,25	0,06	5,25	1,31	-0,25	0,625
3	6	0,17	0,03	3,50	0,58	0,69	0,4761
4	12	0,08	0,01	3,25	0,27	0,14	0,0196
$\sum$	24	1,00	0,37	20	6,16	0	1,4571

Подставив, полученные результаты (таблица 6.7) в выражение (6.10), получим систему искомых уравнений, решив которые найдем параметры регрессионных уравнений.

$$4a + b - 20 = 0;$$
$$a + 0,37b - 6,16 = 0;$$
$$a = 2,58; \quad b = 9,67.$$

Уравнение регрессии, полученное методом МНК выглядит так:

$$y = 2,58 + \frac{9,67}{x}.$$

Сумма квадратов отклонений экспериментальных данных от расчетных равно $d^2 = 1,4571$

График полученной зависимости показан на рисунке 6.2.

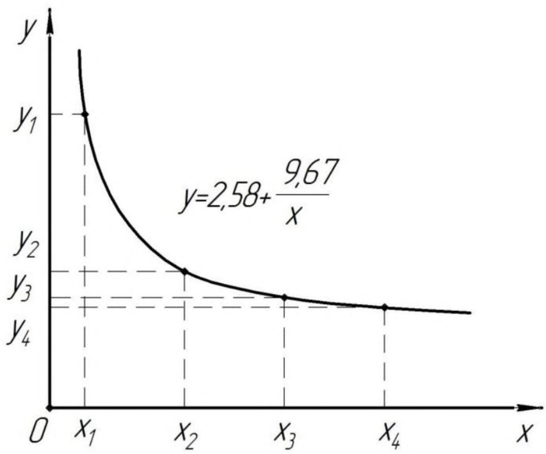


Рисунок 6.2 – График гиперболической зависимости с полученными параметрами и заданными экспериментальными точками

6.3 Квадратичная функция

Задание. Найти параметры a, b и c при аппроксимации экспериментальных данных, полученных при исследовании зависимости удельного сопротивления почвы R от ее абсолютной влажности W_A (данные по П.У. Бахтину) квадратичной функцией вида ax^2+bx+c (таблица 6.8).

Аппроксимирующая функция – квадратичная функция

$$y = ax^2 + bx + c.$$

Сумма квадратов отклонений экспериментальных значений y_i от расчетных

$$\sum_{i=1}^{n} d_i^2 = \sum_{i=1}^{n}\left[y_i - (ax_i^2 + bx_i + c)\right]^2. \qquad (6.11)$$

Таблица 6.8 – Экспериментальные данные, полученные при исследовании зависимости удельного сопротивления почвы y (R) от ее влажности x (W)

x, %	7	8	10	15	18	20	25	30
y, кг/см^2	0,58	0,54	0,51	0,46	0,45	0,46	0,50	0,60

Находим частные производные выражения (6.10) по параметрам a и b.

$$\frac{\partial \sum_{i=1}^{n}\left[y_i - (ax_i^2 + bx_i + c)\right]^2}{\partial a} = -2\sum_{i=1}^{n} x_i^2 \left[y_i - (ax_i^2 + bx_i + c)\right],$$

$$\frac{\partial \sum_{i=1}^{n}\left[y_i - (ax_i^2 + bx_i + c)\right]^2}{\partial b} = -2\sum_{i=1}^{n} x_i \left[y_i - (ax_i^2 + bx_i + c)\right],$$

$$\frac{\partial \sum_{i=1}^{n}\left[y_i-(ax_i^2+bx_i+c)\right]^2}{\partial b}=-2\sum_{i=1}^{n}\left[y_i-(ax_i^2+bx_i+c)\right].$$

Раскрыв скобки и приравняв производные нулю, получим:

$$a\sum_{i=1}^{n}x_i^4+b\sum_{i=1}^{n}x_i^3+c\sum_{i=1}^{n}x_i^2-\sum_{i=1}^{n}x_i^2y_i=0,$$
$$a\sum_{i=1}^{n}x_i^3+b\sum_{i=1}^{n}x_i^2+c\sum_{i=1}^{n}x_i-\sum_{i=1}^{n}x_iy_i=0,$$
$$a\sum_{i=1}^{n}x_i^2+b\sum_{i=1}^{n}x_i+cn-\sum_{i=1}^{n}y_i=0.$$

Для удобства составления системы уравнений сведем данные уравнения в таблицу 6.9.

Таблица 6.9 – Данные к составлению уравнений регрессии

i	x_i	x_i^2	x_i^3	x_i^4	y_i	$x_i y_i$	$x_i^2 y_i$
1	x_1	x_1^2	x_1^3	x_1^4	y_1	$x_1 y_1$	$x_1^2 y_1$
…	…	…	…	…	…	…	…
n	x_n	x_n^2	x_n^3	x_n^4	y_n	$x_n y_n$	$x_n^2 y_n$
$\sum$	$\sum_{i=1}^{n}x_i$	$\sum_{i=1}^{n}x_i^2$	$\sum_{i=1}^{n}x_i^3$	$\sum_{i=1}^{n}x_i^4$	$\sum_{i=1}^{n}y_i$	$\sum_{i=1}^{n}x_iy_i$	$\sum_{i=1}^{n}x_i^2y_i$

Далее рассчитываем необходимые данные для составления системы уравнений (таблица 6.10.)

Таблица 6.10 – Расчеты к составлению системы уравнений

i	x_i	x_i^2	x_i^3	x_i^4	y_i	$x_i y_i$	$x_i^2 y_i$	d_i	d_i^2
1	7	49	343	2401	0,58	4,06	28,42	0,076169	0,005802
2	8	64	512	4096	0,54	4,32	34,56	0,040805	0,001665
3	10	100	1000	10000	0,51	5,10	51,00	0,01814	0,000329
4	15	225	3375	50625	0,46	6,90	103,50	-0,02483	0,000616
5	18	324	5832	104976	0,45	8,10	145,80	-0,03836	0,001472
6	20	400	8000	160000	0,46	9,20	184,00	-0,03395	0,001152
7	25	625	15625	390625	0,50	12,50	312,50	-0,01921	0,000369
8	30	900	27000	810000	0,60	18,00	540,00	0,039371	0,00155
$\sum$	133	2687	61687	1532723	4,1	68,18	1399,78	0,058138	0,012955

$$1532723a + 61687b + 2687c - 1399,78 = 0;$$
$$61687a + 2687b + 133c - 68,18 = 0;$$
$$2867a + 133b + 8c - 4,1 = 0.$$

Решение последней системы уравнений и дает параметры:

$$a = 0,000322989; \quad b = -0,009481085; \quad c = 0,554371953;$$

Квадратическое уравнение с искомыми параметрами, полученное методом МНК выглядит так:

$$y = 0,000322989x^2 - 0,009481085x + 0,554371953.$$

В приложении Д приведены наиболее распространенные функциональные зависимости и их графические изображения.

Контрольные вопросы

1. Объяснить смысл аппроксимации.
2. Понятие метода наименьших квадратов.
3. Порядок выполнения аппроксимации методом наименьших квадратов.

7 ПЛАНИРОВАНИЕ ЭКСПЕРИМЕНТА. ОБЩИЕ ИДЕИ И СТРУКТУРА

Раньше математические методы применялись только на последнем этапе исследования – при математической обработке результатов наблюдений. В последнее время математические методы находят все более широкое применение во всех аспектах постановки эксперимента: при обобщении и формализации результатов ранее выполненных исследований перед постановкой собственных экспериментов, при планировании эксперимента, обработке полученной информации и, наконец, при принятии решения. Однако какие бы успехи не были достигнуты в области планирования эксперимента, следует признать, что оно создавалось многими поколениями ученых. Отрадно, что исходя из практических потребностей сельскохозяйственного опытного дела серьезные задачи в области планирования эксперимента в полевых опытах были выполнены Р. Фишером на Рочемстедской агробиологической станции в Англии. На сегодняшний день планирование эксперимента сформировалось в отдельную разветвленную дисциплину с широким спектром эффективно выполненных исследований при решении практических задач. Данное пособие позволит на приводимых практических задачах легче освоить и применять основы планирования эксперимента специалистам в области агроинженерии.

Прежде всего следует ответить на вопрос что же дает планирование эксперимента.

1. Рассмотрение неконтролируемых факторов как случайные величины, что возможно только при рандомизации. Это дает возможность получать несмещенные характеристики исследуемых процессов.

2. При планировании эксперимента резко повышается эффективность эксперимента. Разработана концепция многофакторного эксперимента, что позволяет повысить точность определения параметров, интересующих исследователя.

3. Планирование эксперимента позволяет проводить эксперимент на более обоснованном логически последовательном уровне. Например, экстремальные эксперименты, отсеивающие эксперименты и последовательные симплекс – методы существенно быстрей продвигают исследователя к оптимальным условиям функционирования исследуемого объекта.

4. Планирование эксперимента позволяет практически стандартизировать опыты по определенным планам, что дает возможность получать сопоставимые результаты у всех исследователей независимо от места постановки эксперимента.

Часто экспериментаторы стремятся к тому, чтобы подогнать получаемые данные под прогнозируемую теорию и сделать некоторые соответствующие выводы, не обращая при этом особого внимания на то, каким образом эти данные получены. Излюбленным приемом некоторых экспериментаторов является проверка с помощью критерия Стьюдента разности средних, не учитывая при этом происхождение выборок. Экспериментатор должен обладать глубокими и разносторонними знаниями. Он должен на высоком уровне достигнутый по рассматриваемому вопросу теоретический уровень, знать методы планирования эксперимента, обработки полученного материала и анализа результатов.

Эксперимент в первую очередь **включает постановку** вопроса, а это в свою очередь требует глубокого понимания достигнутого уровня. Правильное определение постановки вопроса во многом предопределяет успех в решении поставленной задачи. Следующим важным шагом в эксперименте является правильный выбор **независимых переменных**, которые будут исследоваться. Требуется установить можно ли их **измерять** и с **какой точностью**. Если невозможно измерить напрямую в числовом выражении, то надо найти другой метод выражения результатов опытов. Это может быть выражено изменением цвета, ответом на вопрос «да» или «нет» и др. Необходимо также предварительно по возможности установить какому **закону распределения** могут подчиняться исследуемые показатели. Далее необходимо установить, **какие факторы**, т. е. независимые переменные, могут влиять на отклики, т. е. зависимые переменные. Требуется также выяснить какие факторы из принятых необходимо поддерживать на постоянном уровне, а какие должны устанавливаться на определенных уровнях, как изменяются переменные факторы – количественно или качественно. Все перечисленные требования составляют важную часть в подготовке эксперимента.

Следующим важным моментом во всей цепочке экспериментальных исследований является собственно **планирование эксперимента**. В первую очередь необходимо определить **размер выборки**, возможный размах результатов измерений. Если нет такой предварительной информации, то нужно планировать возможно большие выборки в зависимости от трудоемкости их выполнения. Чаще всего размер выборки бывает произволен. В этом случае в ходе эксперимента на основании первых результатов измерений необходимо уточнять размер выборки. Для будущих расчетов необходимо обсуждать и устанавливать степень риска принимаемых решений.

Как уже отмечалось, важнейшим условием в процессе планирования эксперимента является **рандомизация**, так как в эксперименте, как правило, всегда присутствуют факторы, которые невозможно проконтролировать (более подробно процедура рандомизации изложена в разделе 7.10).И, наконец, необходимо построить математическую модель, описывающую исследуемый объект.

Заключительный этап эксперимента – анализ, включает в себя сбор данных, вычисление запланированных характеристик, проверка гипотез, интерпретация результатов, построение графиков и составление таблиц. Изложенные выше положения эксперимента, планирования и анализа можно свести к следующему перечню.

ЭКСПЕРИМЕНТ

1. Постановка задачи.

2. Выбор отклика, или зависимой переменной. Параметр оптимизации

3. Выбор переменных (варьируемых) факторов.

4. Выбор уровней переменных факторов.

ПЛАНИРОВАНИЕ

1. Определение необходимого числа наблюдений.
2. Выбор математической модели.
3. Выбор плана проведения эксперимента.
4. Рандомизациия.

АНАЛИЗ

1. Сбор и обработка данных.
2. Вычисление статистических характеристик и проверка гипотез.
3. Интерпретация полученных результатов.

В каждом конкретном случае в зависимости от поставленной цели и условий проведения исследований необходимо применять соответствующую методику построения плана эксперимента. В этом смысле планирование эксперимента при поиске оптимальных условий является одной из наиболее распространенных научно-технических задач. Для успешного проведения экспериментальных исследований, прежде всего, требуются глубокие знания в избранной области, основах теории вероятностей и математической статистике и, безусловно, планирования эксперимента. Есть ряд и других условий, только обеспечение которых может позволить получение достоверных результатов. Это и наличие соответствующего экспериментального оборудования и вычислительной техники и подбор условий проведения опытов, соответствующих реальным условиям функционирования исследуемого объекта, наконец, в ходе проведения исследований часто бывает необходимым корректировать параметры эксперимента. Иногда, при проведении исследований под словами экспериментатор и статистик подразумевают разных специалистов. Цель первого из них состоит в возможно более полном и точном проведении эксперимента с разумной экономией времени и средств, а другой дает специальные консультации и оказывает содействие по количественным аспектам при планировании и интерпретации. Однако лучше всего эти две роли объединить, так как неполное их взаимопонимание может погубить эксперимент.

Иногда эксперимент, который проводится для оптимизации параметров называют экстремальным. Действительно такое совпадение бывает. Это тот случай, когда одновременно экстремальное значение отклика и его оптимальное значение совпадают. Однако это не всегда бывает, особенно, при экспериментировании с многокритериальными объектами. Поэтому при постановке задачи на этот момент надо обращать внимание.

Постановка задачи перед экспериментом является важным шагом на пути к ее решению. Задачи, которые ставятся перед экспериментом выявляются при теоретическом рассмотрении исследуемого вопроса. Результаты теоретических исследований являются необходимым предварительным этапом не только для постановки выработки плана эксперимента, но и на всем протяжении его выполнения, включая этап объяснения полученных результатов. К, сожалению, на этом этапе допускаются нередко непоправимые ошибки. Поэтому нельзя приступать к планированию эксперимента без четкого представления достигнутого уровня теоретических и экспериментальных исследований, проведенных другими исследователями и выяснения необходимости продолжения этих работ или исправления ошибок, допущенных ими. Отсюда и вытекают цели и задачи собственных теоретических и экспериментальных исследований. В целом постановка вопроса по выбранной проблеме и определение, в частности, целей и задач эксперимента возможно только в результате полного анализа всех доступных теоретических и экспериментальных исследований и выработки физических связей между всеми факторами, действующими в исследуемом объекте.

7.1 Выбор отклика и параметра оптимизации

В зависимости от объекта, поставленных целей и задач его исследований параметр оптимизации может быть самый разнообразный и дать какую-то общую методику его определения невозможно. Он может быть одним из множества экономических, технико-экономических, технологических и других показателей. Конкретный показатель из этого множества – параметр оптимизации определяется потребностями производства, ситуации в экономике и многими другими условиями. Здесь тоже нельзя предложить рецептуру поведения экспериментатора. В конечном итоге параметр оптимизации должен быть сформирован при постановке эксперимента.

Важным моментом являются требования к параметру оптимизации. Параметр должен быть измеряемым – **количественным** или **качественным**. Если параметр является количественным, мы должны уметь его замерять при любом сочетании переменных факторов, действующих на данный объект. Если же невозможно количественно замерять параметр оптимизации, значения параметра нужно выражать **ранжированием** по условленным заранее балам. Однако если есть возможность, то надо отдавать предпочтение количественным измерениям. Желательно, чтобы параметр оптимизации был **универсальным**. Это значит, что параметр должен быть обобщающим множество показателей объекта, от значения каждого из которых в отдельности не зависит в достаточной степени эффективность функционирования исследуемого объекта. Наиболее удачным считается, если параметр оптимизации имеет физический смысл. Это позволяет легче объяснять смысл результатов эксперимента. В итоге следует заметить, что выбор параметра оптимизации конкретен для каждого эксперимента и подменять его каким-то другим, более легким или удобным по каким – то соображениям нельзя. Иногда встречаются ошибки такого рода, когда экспериментатор в качестве параметра оптимизации выбирают один из зависимых переменных (откликов) и находят его экстремальное значение. Довольно часто встречаются задачи с несколькими выходными параметрами. Особенно часто такие случаи бывают при оценке качества протекания технологического процесса. Так, например, качество обработки почвы оценивается по степени ее крошения, равномерности глубины обработки, глыбистости, крошению почвы. Кроме того, важными показателями эффективности почвообрабатывающего орудия являются его производительность, эксплуатационные и приведенные затраты и другие показатели в зависимости от конкретного агрегата. В таких случаях, если какой-нибудь показатель из всего набора можно выделить как самый главный, то его принимают за параметр оптимизации, а остальные оценивают с ограничениями. С целью сокращения количества выходных показателей необходимо провести корреляционный анализ, вычислив при этом все возможные варианты парных корреляций. И если при этом окажутся коррелированными на условленном заранее уровне риска какие-то пары, то только один из них оставляется, и как правило, оставляется тот показатель, который на взгляд экспериментатора более важен, легче измеряется. Такой методический подход может существенно уменьшить количество параметров. За последние годы широкое распространение получила методика **обобщенного параметра оптимизации**, которая основана на том, что переводит все параметры с физическим смыслом на единую безразмерную шкалу, с которыми можно проводить необходимые математические действия, позволяющие получать один искусственный параметр и таким образом сравнивать эффективность функционирования исследуемого объекта при различных совокупностях действующих факторов.

7.2 Обобщенный параметр оптимизации

Анализ результатов исследований часто сводится к сравнению двух и более вариантов по нескольким параметрам. Так, при отборе сортов сравниваются их урожайность, морозоустойчивость, засухоустойчивость, устойчивость к полеганию и т. д. Сельскохозяйственные машины оцениваются по их агротехническим, технико-экономическим и экономическим показателям. Аналогичные задачи решаются и при разработке системы машин и выборе технологической схемы агрегата, при отборе пород животных и при оценке качества зерна. При этом характерно, что в каждом конкретном случае набор и количество показателей определяются поставленной целью и исследуемым объектом. Общей же их характеристикой является отличие сравниваемых показателей не только по физическому смыслу, размерности, но также и по их относительной важности, что создает некоторые сложности и приводит к неопределенности выводов.

Решение подобных задач связано с приведением всех сравниваемых показателей к единому количественному показателю, то есть с созданием обобщенного параметра оптимизации (обобщенного критерия оценки).

Одним из наиболее признанных и, действительно, эффективных способов получения обобщенного параметра оптимизации является **функция желательности Харрингтона**. Ее основа – преобразование натуральных значений зависимой переменной (отклика) в безразмерную шкалу желательности. Это преобразование предложено выразить формулой:

$$D = \sqrt[n]{\prod_{i=1}^{n} d_i}, \qquad (7.1)$$

где D – обобщенная функция желательности Харрингтона;

d_i – частная функция желательности i-го отдельного параметра;

n – количество обобщаемых параметров.

Частные функции желательности при этом предложено рассчитывать по формуле:

$$d_i = \exp\left[-\exp(-y')\right], \qquad (7.2)$$

где y' – кодированное значение отдельного параметра по шкале желательности.

Для получения шкалы желательности рекомендуется использовать таблицу 7.1. соответствия между отношениями предпочтения в эмпирической и числовой системах.

Таблица 7.1 – Стандартные оценки на шкале желательности

Желательность	Отметки на шкале желательности
Очень хорошо	1,00…0,8
Хорошо	0,8…0,63
Удовлетворительно	0,63…0,37
Плохо	0,37…0,20
Очень плохо	0,20…

На графике функции желательности (рисунок 7.1), которая задается уравнением:

$$d_i = \exp\left[-\exp(-y')\right],$$

где отмечены значения d в соответствии с таблицей 7.1.

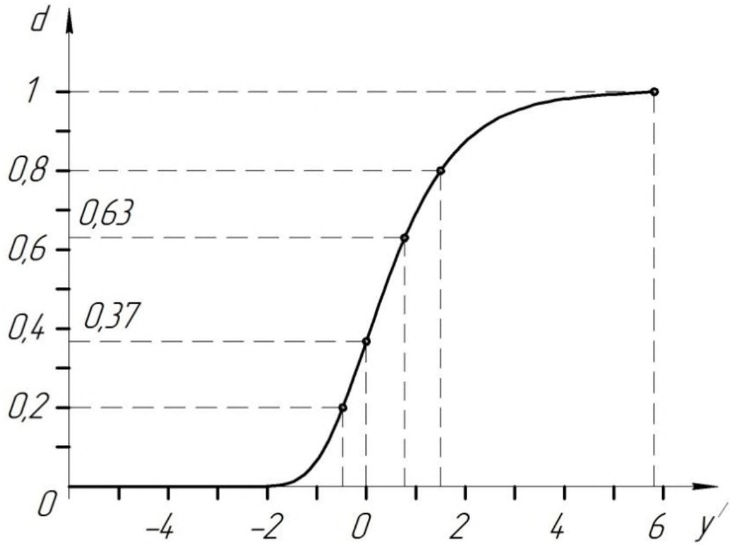


Рисунок 7.1 – Функция желательности

На оси абсцисс отмечены значения отклика в условном масштабе. Началом отсчета по этой оси выбрана точка, соответствующая желательности 0,37. Далее необходимо выполнить преобразование частных откликов в частные функции желательности по представлению экспериментатора или группы специалистов в избранном направлении исследований.

Важным моментом в нахождении обобщенного параметра оптимизации является нахождение **значимости** (весомости) k_i отдельных обобщаемых показателей d_i. Это вызвано тем, что даже при одинаковом уровне выполнения отдельных показателей, они будут по разному влиять на степень оптимальности их совокупности. Например, качество обработки почвы оценивают по степени крошения почвы, равномерности глубины обработки, глыбистости и гребнистости поверхности поля после ее обработки. Однако при этом надо заметить, что вариант с отличной оценкой степени крошения и удовлетворительной оценкой гребнистости поверхности поля значительно предпочтительней, чем вариант с противоположными оценками этих показателей, т. е. чем вариант с удовлетворительным крошением и абсолютно ровной поверхностью поля. Следовательно, необходимо проранжировать все обобщаемые показатели по степени их важности (весомости). С учетом важности k_i каждого показателя обобщенный параметр оптимизации примет вид.

$$D = \sqrt[n]{\prod_{i=1}^{n} d_i^{k_i}}. \tag{7.3}$$

Весомости показателей k_i определяются методом экспертной оценки. При этом члены экспертной комиссии подбираются из специалистов по избранной теме исследований. Степень согласованности мнений членов экспертной комиссии проверяется по коэффициенту χ^2.

$$\chi^2 = \frac{S}{\frac{1}{12}mn(n-1) - \frac{1}{(n-1)}\sum_{j=1}^{m} T_j}, \tag{7.4}$$

где S – сумма квадратов отклонений средней суммы рангов от суммы рангов каждого показателя;

m – число членов экспертной комиссии;

n – количество показателей.

$$T_j = \frac{1}{12}\sum_{j_1}^{j_m}(t_j^3 - t_j), \tag{7.5}$$

где t_j – число одинаковых рангов в j-м ряду.

Обобщенная функция желательности нашла широкое применение для оценки в исследованиях новых технологических процессов, сельскохозяйственных машин и агрегатов.

7.3 Выбор переменных факторов

На каждый объект исследования, представляемую как кибернетическую систему, действуют независимые управляемые факторы $x_1, x_2, x_3, \ldots, x_k$, различные сочетания значений которых образуют факторное пространство. В результате действия этих факторов на выходе системы имеем параметры – зависимые переменные, подлежащие впоследствии оптимизации $y_1, y_2, y_3, \ldots, y_m$ (рисунок 7.2).

В любом эксперименте всегда присутствуют управляемые и неуправляемые переменные факторы. В активный эксперимент должны быть включены все значимые управляемые факторы. И если хоть один значимый фактор не будет включен в эксперимент и он в процессе эксперимента менял произвольно свое значение, это может привести к существенному повышению ошибки опыта, а если этот неучтенный и к тому же значимый фактор будет в процессе эксперимента поддерживаться на одном уровне, то найденное значение оптимума не будет соответствовать его истинному значению. А все включенные в эксперимент факторы должны отвечать вполне определенным требованиям.

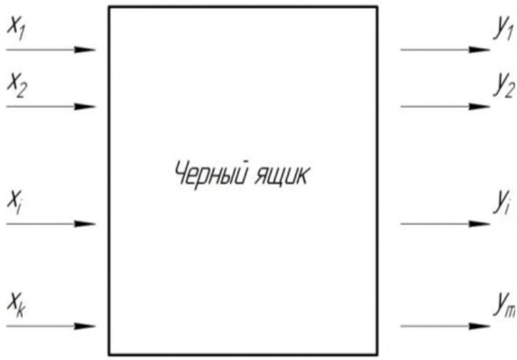


Рисунок 7.2 – Схема кибернетической системы исследуемого объекта

При планировании активного эксперимента **факторы должны быть управляемыми**. Это значит, что заранее выставленное значение фактора возможно поддерживать в течение опыта неизменным, а при переходе на другой опыт – выставить новое значение согласно плана эксперимента.

Факторы **не должны** быть функционально **зависимы** или линейно коррелированны. Это значит, что установление фактора на определенный согласно плана уровень не должно зависеть от значения другого фактора.

Должно соблюдаться требование совместимости всех запланированных в эксперименте значений факторов. Например, обороты почвообрабатывающей фрезы и ее поступательная скорость. Совместимость уровней различных факторов должна соблюдаться и с точки зрения безопасности (взрыв, поломки и др.).

Значения факторов должны быть измерены с точностью, требуемой для данного эксперимента и чувствительности выходного параметра к изменениям данного фактора.

Желательно, чтобы факторы имели однозначное управление. Это значит, что значения каждого фактора можно устанавливать напрямую без участия других факторов. Например, обороты почвофрезы не должны в эксперименте зависеть от ее поступательной скорости. Эти два фактора должны регулироваться независимо друг от друга. Если это требование не удается выполнить, то необходимо вводить новые переменные факторы в виде их соотношения. Например, в случае с почвофрезой отношение окружной и поступательной скоростей фрезы.

В итоге следует заметить, что выбор факторов сложный и ответственный процесс и к нему нужно относиться со всей серьезностью.

7.4 Выбор области эксперимента и уровней переменных факторов

Правильный выбор области эксперимента и уровней переменных факторов является весьма важным условием обеспечения успеха в наикратчайшем достижении оптимальных параметров и запланированной точности. Но при этом необходимо учесть ряд ограничений значениям факторов, которые нельзя нарушать. Во-первых, это - **принципиальные ограничения**, основанные на фундаментальных законах. Например, скорость нельзя планировать выше скорости света. Во-вторых – это **технико- экономические** ограничения. Например, нельзя планировать стоимость удобрений или сельскохозяйственных машин, запланированных в эксперименте ниже уже установившихся цен. И, наконец, значения факторов ограничены **реальными условиями** эксперимента. Например, нельзя планировать скорости обработки почвы выше максимальной скорости применяемого при этом энергетического средства. Необходимо **хорошо проанализировать всю полученную ранее информацию** другими исследователями с предварительным прогнозом параметров и области факторов, а также ряда других характеристик условий эксперимента.

Результатом анализа априорной информации может быть также точка с наилучшим выходом интересующего нас параметра, которой соответствует ряд комбинаций переменных факторов. Именно эту точку в многомерном факторном пространстве можно взять как исходную точку для построения плана. Эта точка называется **основным уровнем** и все остальные точки плана выбираются симметрично относительно этой точки. В практике встречаются случаи, когда может потребоваться корректировка положения основного уровня: когда основной уровень расположен близко к одной из границ области определения; когда имеется несколько эквивалентных точек оптимального параметра; когда есть только область определения оптимальной точки, а самой оптимальной точки нет. Во всех этих и других возможных случаях необходимо принимать обоснованные решения по корректировке и выбору оптимальной точки факторного пространства, соответствующей оптимальным параметрам. Чем ближе удастся с первого раза определить из априорной информации, в том числе из предварительного теоретического рассмотрения функционирования исследуемого объекта, значение основного уровня к экстремуму выходного показателя, тем быстрее экспериментатор доберется до оптимальных параметров.

Следующим шагом общей процедуры планирования эксперимента является **выбор интервалов варьирования факторов**. Интервалы варьирования получаются путем прибавления или вычитания от основного уровня некоторой величины фактора p_i.

$$\begin{aligned} x_i^в &= x_i^о + p_i, \\ x_i^н &= x_i^о - p_i, \end{aligned} \tag{7.6}$$

где $x_i^в$ и $x_i^н$ – вверхний и нижний пределы варьирования i-го фактора в натуральном выражении;

$x_i^о$ – основной уровень i-го фактора в натуральном выражении;

p_i – интервал варьирования i-го фактора в натуральном выражении

Для перехода на упрощенную запись плана эксперимента и обработки экспериментальных данных масштабы по осям факторов выбираются так, чтобы верхней уровень соответствовал +1, нижний уровень соответствовал –1 и нулевой уровень - 0. Это можно сделать с помощью выражения для преобразования:

$$x_j^{в} = \frac{x_i^{в} - x_i^{о}}{p_i},$$
$$x_j^{н} = \frac{x_i^{н} - x_i^{о}}{p_i}, \qquad (7.7)$$

где $x_j^{в}$ и $x_j^{н}$ – кодирование значения верхнего и нижнего пределов *i*-го фактора $x_i^{в}$ и $x_i^{н}$;

От правильного выбора величины интервала варьирования зависит дальнейший ход эксперимента. Во-первых, при занижении этого параметра плана могут оказаться часть или все коэффициенты полученной модели несущественными из-за соизмеримости интервала варьирования и точности измерения параметров, участвующих в эксперименте. При завышении же интервала варьирования точность определенного экстремального значения выходного показателя далеко не оптимальным. В этом случае говорят, что перешагнули экстремальную точку и придется уменьшить интервал варьирования и повторить эксперимент. В связи с вышеизложенным интервал варьирования не может быть меньше точности измерения факторов. Невозможно дать исчерпывающую для всех экспериментов методику выбора интервалов варьирования факторов, но общие полезные советы наиболее полно приведены в литературе.

7.5 Определение необходимого числа наблюдений

Здесь речь идет о необходимости выяснения, сколько параллельных наблюдений проводить для каждого экспериментального условия, т. е. опыта. Фактически это повторение наблюдений при одной и той же настройке исследуемого объекта, при одних и тех же значениях переменных факторов. В исследованиях в области сельского хозяйства, как правило, присутствуют неуправляемые факторы, многие из которых которые оказывают влияние на выходные оптимизируемые параметры влияние. Доля этих факторов в формировании выходного показателя иногда доходит до 50 % и более. Поэтому вряд ли возможно провести эксперимент без параллельных опытов. Чем больше повторностей, тем точнее определяются статистические параметры выходных показателей. Однако надо учитывать, что при увеличении повторностей повышаются затраты труда и время на проведение эксперимента. Поэтому необходимо всегда руководствоваться минимальным, но достаточным количеством параллельных опытов, обеспечивающих запланированную точность определения параметров. При планировании эксперимента важно показать, что разброс выходного оцениваемого параметра не выходит за пределы запланированной точности. **Известно, что для нормально распределенных случайных величин все ее значения лежат в пределах $\pm 3\sigma$. Поэтому, если неизвестно с какой точностью надо определять измеряемые параметры, то можно брать в пределах $\pm 3\sigma$.** Если известна заранее необходимая точность определения измеряемого параметра, то ее можно выразить в долях δ стандартного отклонения σ. Из общей теории ошибок известно, что необходимое и достаточное количество ошибок зависит от величины стандартного отклонения измерений σ и желаемой надежности принимаемого решения, т. е. доверительной вероятности α (Приложение Б). Стандартное отклонение σ заранее устанавливается или из априорной информации или из специального предпланировочного эксперимента. Такая процедура проводится по всем выходным параметрам. Для упрощения обработки экспериментального материала из всех найденных чисел повторностей для разных параметров назначается самое большое число.

Если исследователь располагает двумя дисперсиями двух малых выборок S_1^2 и S_2^2, доверительной вероятностью α и необходимой точностью δ, то объем выборки можно подсчитать по формуле:

$$n = \frac{t^2 S_0^2}{\delta^2}, \tag{7.8}$$

где t – величина, отвечающая значению нормированной функции Лапласа для нормального распределения (таблица А1).

$$\Phi_0(t) = \frac{1}{\sqrt{2\pi}} \int_0^t e^{-\frac{t^2}{2}} dt, \quad \Phi_0(t) = \frac{a}{2}, \quad S_0^2 = \frac{S_1^2 + S_2^2}{2}. \tag{7.9}$$

Пример. Две небольшие выборки тягового сопротивления плуга имеют дисперсии $S_1^2 = 250$ кг $S_2^2 = 350$ кг. Зададимся доверительной вероятностью α=0,9. Необходимая точность определения тягового сопротивления 50 кг. Надо найти минимально допустимое число измерений выходных показателей n.

Таблица 7.2 – Необходимое (минимально допустимое) количество повторностей опыта

Ошибка δ в долях стандарта	Доверительная вероятность, α					
	0,7	0,8	0,9	0,95	0,99	0,999
3,0	1	1	2	3	4	5
2,0	1	2	3	4	5	7
1,0	3	4	5	7	11	17
0,5	6	9	13	18	31	50
0,4	8	12	19	27	46	74
0,3	13	20	32	46	78	127
0,2	29	43	70	99	171	277
0,1	169	266	273	387	668	1089
0,05	431	659	1084	1540	2659	4338
0,01	10732	16436	27161	38416	66348	108307

$$n = \frac{1,6 \cdot 300^2}{50^2} = 92,16.$$

Следовательно, необходимо выполнить 93 измерения. Чем больше доверительная вероятность, что получаемые усредненные показатели находятся в пределах заданной точности, тем большее количество измерений необходимо выполнить. Например, при $\alpha = 0,95$ при всех остальных параметрах вышеприведенного примера необходимый объем повторностей $n = 139$. Такое же влияние на число измерений оказывает и повышение задаваемой точности измерений. Для повышения точности измерений, безусловно, необходимо уделить внимание измерительным приборам и квалификации специалистов.

7.6 Выбор математической модели

Выбрать математическую модель объекта означает выбор функции поверхности отклика. В общем виде модель любого объекта выражается в виде функции отклика:

$$y = \varphi(x_1, x_2, \ldots, x_{k-1}, x_k), \tag{7.10}$$

где $x_1, x_2, \ldots, x_{k-1}, x_k$ – действующие на объект переменные факторы;

y – функция отклика.

Если принять, что каждому x_i-му фактору из x_k факторов соответствует своя координатная ось, то образованное ими пространство называют $k+1$– мерным факторным пространством с учетом координаты отклика. Для одного фактора это пространство выглядит в виде линии на плоскости, для двух факторов – в виде поверхности любой конфигурации в пространстве с одним или несколькими экстремальными точками. Значения переменных факторов, которые соответствуют этим экстремальным значениям (максимальные или минимальные) в зависимости от постановки вопроса называются оптимальными параметрами.

Для решения поставленных перед экспериментом задач используют разложение искомой функции отклика (7.10) в ряд Тейлора:

$$\varphi(x_1, x_2, ..., x_{k-1}, x_k) = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{i,j}^{k} b_{ij} x_i x_j + \sum_{i=1}^{k} b_{ii} x_i^2 + ..., \quad (7.11)$$

где $i, j = 1,2,3,\ldots,k{-}1$, k; $i \neq j$,

$$b_i = \frac{d\varphi}{dx_i}, \qquad b_{ij} = \frac{d^2\varphi}{dx_i dx_j}, \qquad b_{ii} = \frac{1}{2}\frac{d^2\varphi}{dx_i^2}\ldots \quad (7.12)$$

В качестве математической модели используют только часть ряда (7.12), в которой определяют коэффициенты при каждом члене многочлена b_i, b_{ij}, b_{ii}, определяются также число членов многочлена и конкретно какие члены определяются при планировании эксперимента. Надо стремиться выбрать наиболее простые модели с малым количеством членов модели, возможно линейной моделью. Такие модели чаще встречаются в задачах по оптимизации. В интерполяционных задачах необходимо увеличить и количество членов модели и количество опытов в целях более точного определения коэффициентов и более точного описания поверхности отклика. Таким образом, план эксперимента – это совокупность числа и условий проведения опытов в избранной области определения факторов. В целом же в вопросе выбора математической модели необходимо учитывать результаты ранее выполненных исследований другими исследователями и рекогносцировочных собственных исследований.

7.7 Выбор плана проведения эксперимента

После того, когда уже выбраны неформализованные параметры эксперимента, выбраны математическая модель, т. е. выбран вид функции $y = \varphi(x_1, x_2, \ldots, x_{k-1}, x_k)$ факторное пространство и число повторностей каждого опыта и некоторые, присущие только данному конкретному эксперименту, дополнительные условия, остается выбрать план эксперимента и провести его для оценки численных значений коэффициентов уравнения. Разработано уже довольно большое количество планов эксперимента с присущими каждому из них свойствами. Рассмотрим некоторые из них, которые нашли наибольшее распространение в исследованиях не только по сельскому хозяйству, но практически и во всех остальных отраслях.

7.8 Метод полного факторного эксперимента

Метод полного факторного эксперимента (ПФЭ), основанный на варьировании переменных факторов на двух уровнях применяется для получения линейных моделей исследуемых объектов или суммы линейной модели и членов полинома, содержащих произведения факторов первой степени:

$$y = b_0 + \sum_{i=1}^{k} b_i x_i + \sum_{\substack{i=1,\, j=1 \\ i<j}}^{k} b_{ij} x_i x_j + \sum_{i=1,\, j=1,\, l=1}^{k} b_{ijl} x_i x_j x_l + \ldots, \quad (7.13)$$

где b_0 – свободный член модели;

$\sum_{i=1}^{k} b_i x_i$ – линейные члены модели;

$\sum_{\substack{i=1,\, j=1 \\ i<j}}^{k} b_{ij} x_i x_j$ – взаимодействия первого порядка по два члена модели;

$\sum\limits_{i=1,\,j=1,\,l=1}^{k} b_{ijl}x_i x_j x_l$ – взаимодействия первого порядка по три члена модели.

Для линейных моделей переменные факторы варьируются на двух уровнях и обозначаются для k факторов как ПФЭ 2^k. Так как все k факторы варьируются на двух уровнях, то общее число опытов для полного факторного эксперимента ПФЭ 2^k определяется комбинациями уровней факторов:

$$N = 2^k. \tag{7.14}$$

Как уже упоминалось выше для упрощения вычислений коэффициентов b_0, b_i, b_j, b_{ij} факторы в натуральной размерности $x_i^в$ – на верхнем уровне и $x_i^н$ – на нажнем уровне кодируют соответственно как +1 и – 1.

Геометрически полный факторный эксперимент при двух и при трех факторах представляет собой совокупность точек, соответствующих опытам в вершинах соответственно квадрата и куба, а при количестве факторов больше двух в вершинах многогранника (рисунок 7.3).

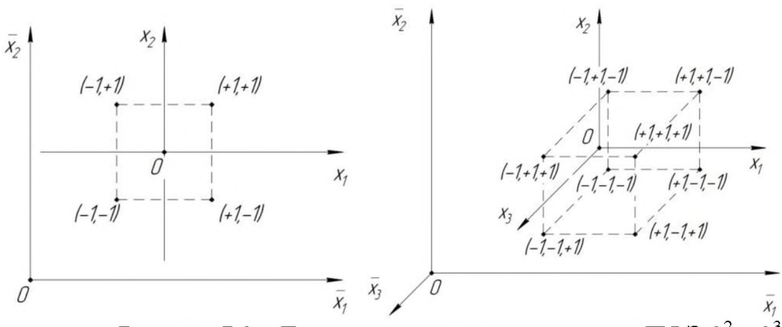


Рисунок 7.3 – Геометрическая интерпретация ПФЭ 2^2 и 2^3

Рассмотрим построение плана ПФЭ. Для этого требуется написать все сочетания уровней факторов в эксперименте. Для этого составим таблицу, которую впоследствии будем называть **матрицей планирования эксперимента**. Для полного двух и трех факторного эксперимента матрица планирования приведена соответственно в таблицах 7.3 и 7.4 Часть таблицы, обведенная пунктирной линией, соответствует собственно плану (матрица плана). Запись матрицы плана, особенно, при большом количестве факторов, занимает много времени и места. Поэтому для ее сокращения применяют условные буквенные обозначения строк. Для этого порядковый номер фактора ставится в соответствие строчной букве латинского алфавита: $x_1 - a$, $x_2 - b$, $x_3 - c$. Тогда, если для какой-нибудь строки матрицы выписать латинские буквы для факторов, находящихся на верхнем уровне, то условия задачи будут заданы однозначно. Опыт со всеми факторами на нижних уровнях принято обозначать (1).

Таблиц 7.3 – Матрица планирования эксперимента ПФЭ 2^2

Номер опыта	x_0	x_1	x_2	Буквенные обозначения строк	x_1x_2	Функция отклика, y
1	+	-	-	(1)	+	y_1
2	+	+	-	a	-	y_2
3	+	-	+	b	-	y_3
4	+	+	+	ab	+	y_4

Для построения матрицы плана применяют разные методы, облегчающие процедуру ее заполнения. Например, в любой матрице плана столбец первого фактора заполняется чередованием –1 и +1, во втором столбце для второго фактора чередованием знака -1 в первой и второй строчках, со знаком +1 в третьей и четвертой строчках и т. д. Третий столбец для третьего фактора чередованием заполнения –1 в первых четырех строчках третьего столбца с последующим заполнением

+1 в последующих четырех строчках третьего столбца , т. е. в строчках 5, 6, 7, 8 и т.д. Часто используют метод достраивания матрицы меньшего размера до матрицы на один фактор больше. Так, в таблице 7.4 для полного факторного эксперимента 2^3 первые четыре строчки – это план 2^2 при фиксированном значении третьего фактора x_3.

Таблица 7.4 – Матрица планирования эксперимента ПФЭ 2^3

№	x_0	x_1	x_2	x_3	Буквенные обозначения строк	x_1x_2	x_1x_3	x_2x_3	$x_1x_2x_3$	Функция отклика, y
1	+	-	-	-	(1)	+	+	+	-	y_1
2	+	+	-	-	*a*	-	-	+	+	y_2
3	+	-	+	-	*b*	-	+	-	+	y_3
4	+	+	+	-	*ab*	+	-	-	-	y_4
5	+	-	-	+	*c*	+	-	-	+	y_5
6	+	+	-	+	*ac*	-	+	-	-	y_6
7	+	-	+	+	*bc*	-	-	+	-	y_7
8	+	+	+	+	*abc*	+	+	+	+	y_8

И теперь, если к плану 2^2 с фиксированным значением третьего фактора со знаком -1 пристроить такой же план 2^2, но с фиксированным значением третьего фактора +1, получается ПФЭ 2^3. Знаки взаимодействий первого порядка $x_1\,x_2$, $x_1\,x_3$, $x_2\,x_3$, $x_1\,x_2\,x_3$, получаются перемножением знаков основных факторов, входящих во взаимодействие.

Матрица планирования полного факторного эксперимента обладает следующими свойствами:

$$\sum_{u=1}^{N} x_{iu}x_{ju} = 0,\ \ i \neq j,\quad j = 0,1,2,...m, \tag{7.15}$$

$$\sum_{u=1}^{N} x_{i\,u} = 0, \quad i = 0,1,2,...m, \tag{7.16}$$

$$\sum_{u=1}^{N} x_{i\,u}^{2} = N \quad i = 0,1,2,...m, \tag{7.17}$$

где m – номер последнего столбца в планах ПФЭ.

Первое свойство (7.15) выражает условие ортогональности матрицы ПФЭ. Свойство (7.16) означает симметричность плана. Условие (7.17) отражает равенство сумм квадратов всех столбцов числу опытов N.

7.9 Дробный факторный эксперимент

Полный факторный эксперимент применяется при числе факторов не более 5 и при этом получают линейную модель или неполную квадратичную модель, т. е. линейную с членами полинома в виде взаимодействия первой степени. В связи с тем, что при дальнейшем повышении количества факторов эксперимент становится громоздким появляется желание сократить количество опытов за счет той части информации, которая не существенна при построении линейных моделей. Для этой цели применяется так называемый дробный факторный эксперимент (ДФЭ).

Дробный факторный эксперимент является частью полного факторного эксперимента большей размерности или иначе говоря дробной репликой. ДФЭ обладают теми же свойствами, что и ПФЭ и математическая обработка проводится по тем же формулам.

Для наглядного объяснения построения плана дробного факторного эксперимента еще раз обратимся к таблице (7.3) полного факторного эксперимента ПФЭ 2^2. Если в матрице этого полного двухфакторного эксперимента взаимодействие двух переменных факторов x_1 и x_2 стремится к нулю, т. е. $x_1\, x_2 \to 0$, то его можно принять за третий фактор x_3. И тогда четырьмя опытами можно составить план дробного факторного эксперимента 2^3 и вычислить четыре коэффициента b_0, b_1, b_2, b_3 в уравнении:

$$y = b_0 + b_1x_1 + b_2x_2 + b_3x_3. \tag{7.18}$$

Планирование эксперимента, при котором с целью сокращения числа опытов часть факторов приравнивают к произведению нескольких переменных факторов, называют дробным планированием или планированием со смешиванием. Дробный факторный эксперимент обозначают 2^{k-p}, где k – общее число факторов, p – число факторов, приравненных к произведениям. Согласно вышеизложенному матрица дробного трехфакторного эксперимента $2^{3\text{-}1}$ выглядит так (таблица 7.5).

Таблица 7.5 – Варианты матрицы дробного факторного эксперимента $2^{3\text{-}1}$

Номер опыта	I. $x_3 = x_1x_2$				Номер опыта	II. $x_3 = -x_1x_2$			
	x_1	x_2	x_3	$x_1x_2x_3$		x_1	x_2	x_3	$x_1x_2x_3$
1	+	+	+	+	1	+	+	–	–
2	–	–	+	+	2	–	–	–	–
3	+	–	–	+	3	+	–	+	–
4	–	+	–	+	4	–	+	+	–

Любую из двух полуреплик полного факторного эксперимента ПФЭ 2^3 можно взять как ДФЭ $2^{3\text{-}1}$, приравняв один фактор, например x_3 , взаимодействию $x_1 x_2$ со знаком (+) или со знаком (–), т. е. I. $x_3 = x_1 x_2$ или II. $x_3 = - x_1 x_2$.

Как показывает практика исследований дробные факторные эксперименты находят достаточно широкое распространение. И это объясняется тем, что повышается размерность факторного пространства, т. е. увеличивается количество переменных факторов, а, следовательно, повышается и объем опытов. При построении дробных реплик важным моментом является выбор генерирующего соотношения и порядка смешивания. Вопросы построения дробных реплик в достаточно полном объеме и доступно изложены в литературы. Заметим лишь, что по общей рекомендации при замене части переменных факторов прямого действия на взаимодействия нужно стараться выбрать взаимодействия большего количества факторов, так как вероятность существенности их влияния на выходной параметр падает с ростом числа факторов.

7.10 Рандомизация

План эксперимента должен быть таким, чтобы влияние мешающих факторов в виде неоднородности испытываемого материала или условий испытания не искажали получаемые результаты. Это достаточно ощутимо в исследованиях по сельскому хозяйству и выражается в различии плодородия почвы даже на небольшой площади, отведенной под эксперимент, в постоянном изменении твердости и влажности почвы и других физико-механических свойств почвы, в различии высоты и диаметра стебля растений одного и того же сорта, уклона поля и многих других характеристик, определяемых условиями проведения эксперимента. Вряд ли можно найти какую-нибудь другую область, где экспериментатора ожидает так часто и так жестко необходимость проведения эксперимента в условиях неопределенностей. Изменяющиеся условия проведения эксперимента приводят к смещению значений получаемых коэффициентов уравнений регрессии. Устранить полностью влияние неопределенностей условий эксперимента невозможно, смягчить однобокое влияние на одни опыты сильнее, чем на другие, позволяет прием, который называется рандомизацией, что в переводе с английского языка *random* - случайный. **Рандомизация – это умышленное внесение случайности в виде случайного размещения вариантов сочетания (опытов) по площади, в пространстве или во времени проведения эксперимента.** В нерандомизированном эксперименте непонятна причина различия выходного показателя в сравниваемых опытах: или это влияние конкретных сочетаний факторов, или влияние неуправляемого фактора неоднородности. Рандомизация помогает экспериментатору контролировать переменные факторы, которыми невозможно управлять. Рандомизация размещения опытов обеспечивает корректность применения аппарата математической статистики. Рандомизация тем самым обеспечивает более равномерное «размазывание» влияния неуправляемого фактора неоднородности условий проведения эксперимента, обеспечивая тем самым несмещенность получаемых результатов. Разработанных способов размещения вариантов и повторений множество. Применение того или другого способа размещения (рандомизации) зависит от природы и основных статистических показателей неуправляемых переменных факторов, из-за которых собственно и проводится рандомизация.

В пределах каждого повторения варианты эксперимента располагают в случайном порядке, используя таблицу случайных чисел (приложение Е). Эта процедура проводится по всем повторениям. Таким образом, получаем столько разных рандомизаций вариантов эксперимента, сколько повторений опытов будет проведено.

Метод неорганизованных повторений или полной рандомизации. Этот метод предусматривает по сравнению с рандомизацией в пределах каждого повторения отдельно дополнительно еще рандомизацию и самих вариантов. Например, если имеем четыре опыта по три повторения, то рандомизации будет подлежать 12 «делянок». Существуют также метод латинских квадратов и прямоугольников, парный метод и др.

7.11 Сбор и обработка опытных данных

Перед тем, как начать реализацию эксперимента, необходимо еще раз тщательно проверить и убедиться в правильности выбора измерительных приборов, их исправности, наличии всех журналов наблюдений и т. д. Согласно плану эксперимента необходимо выполнять программу исследований.

При этом необходимо периодически проверять состояние измерительных приборов, и ориентировочно просматривать получаемый, пусть еще не обработанный, исследовательский материал. В случае каких-либо подозрений в неправдоподобности получаемых результатов необходимо срочно остановить испытания и выяснить возможную ошибку и исправить ее. При этом нельзя допускать необоснованных исправлений, каким бы грубым ни показался результат.

В первую очередь определим ошибки параллельных опытов. Находим их среднюю арифметическую $\bar{y}$:

$$\bar{y} = \frac{\sum_{1}^{n} y_q}{n}, \tag{7.19}$$

где y_q – результат отдельного опыта;

n – количество параллельных опытов

Изменчивость средней арифметической можно выразить через ее дисперсию S^2.

$$S^2 = \frac{\sum_{1}^{n} (y_q - \bar{y})^2}{n-1}, \tag{7.20}$$

где n – 1 – число степеней свободы.

В результатах параллельных опытов могут по разным неизвестным экспериментатору причинам встречаться значения резко отличающиеся от основной массы результатов измеряемого параметра. Эти значения можно исключить только после статистической проверки с помощью критерия Стьюдента t.

$$\frac{y-\overline{y}}{S} \geq t, \qquad (7.21)$$

где y – испытываемое значение параметра.

Значение t находят по таблице t-распределения Стьюдента (Приложение Б). Выше мы рассмотрели, как подсчитать дисперсию среднего арифметического значений параллельных измерений отдельных опытов плана эксперимента. Для подсчета дисперсии параметра оптимизации по всему эксперименту из Nn измерений воспользуемся выражением:

$$S_{\overline{y}_i}^2 = \frac{\sum_{1}^{N}\sum_{1}^{n}(y_{iq}-\overline{y}_i)^2}{N(n-1)}, \qquad (7.22)$$

где N – количество опытов (сочетаний факторов) в опыте;

y_{iq} – значение оптимизируемого параметра q-ой повторности в i-м опыте;

$\overline{y}_i$ – среднее арифметическое по всему эксперименту из Nn.

Следующим этапом в подготовке к обработке результатов эксперимента является проверка однородности дисперсий параметра оптимизации, рассчитанным отдельно по каждому опыту из его значений по повторностям.

Проверка однородности дисперсий проводится по критерию Фишера (F-критерий), если количество сравниваемых дисперсий всего два. Критерий Фишера представляет собой отношение большей дисперсии к меньшей. При этом, если найденное отношение больше табличного (приложение В) для числа степеней каждой дисперсии соответственно, то сравниваемые дисперсии существенно отличаются, т. е. дисперсии не однородны.

Если же количество сравниваемых дисперсий больше двух, то проверка их однородности проводится по критерию Кохрена G. Критерий Кохрена – непараметрический критерий для проверки значимости различия двух и более воздействий на группы. Критерий Кохрена является дихотомической переменной, т. е. принимает два значения 0 или 1 (да/нет). Таким образом критерий Кохрена используется для проверки равенства дисперсий двух и более выборок. Для этого находится отношение максимальной дисперсии $S_{\max}^2$ из всех найденных по каждому опыту плана эксперимента к сумме всех дисперсий, найденных по каждому опыту.

$$G = \frac{S_{\max}^2}{\sum\limits_{1}^{N} S_i^2}. \tag{7.23}$$

Найденное расчетное значение G сравниваем с табличным значением. (Приложение Ж). Если окажется, что найденное значение не больше табличного, то дисперсии однородны и можно их усреднять. После выполнения всей программы расчетно-подготовительной работы, изложенной выше, необходимо приступить к непосредственному расчету коэффициентов уравнения регрессии. В силу свойств (ортогональности) матрицы планированного эксперимента коэффициенты уравнения регрессии можно вычислить, например, для однофакторной линейной модели по следующим выражениям:

$$b_0 = \frac{\sum\limits_{i=1}^{N} y_i}{N}, \qquad b_1 = \frac{\sum\limits_{i=1}^{N} y_i x_{1i}}{\sum\limits_{i=1}^{N} x_{1i}^2}, \tag{7.24}$$

где $\sum_{i=1}^{N} x_{1i}^2 = N.$

Правило здесь простое. Так как каждый фактор имеет два уровня, вычисления сводятся к тому, что приписывают значениям выходного параметра y знаки соответствующего данному коэффициенту столбца знаки +1 или – 1 и сложению полученных значений. А непосредственно коэффициент находится делением полученной суммы на число опытов в эксперименте. Также вычисляют коэффициенты и при эффектах взаимодействия. После вычисления всех коэффициентов уравнения регрессии приступают к проверке его адекватности и значимости коэффициентов.

В качестве меры адекватности уравнения регрессии принимается дисперсия адекватности $S_{ад}^2$.

$$S_{ад}^2 = \frac{1}{N-i}\sum_{i}^{N}(\bar{y}_i - \hat{y}_i)^2 \qquad (7.25)$$

где $\bar{y}_i$ и $\hat{y}_i$ – соответственно средние арифметические выходного параметра i-го опыта и расчетное значение выходного параметра по полученному уравнению регрессии по совокупности факторов по этому же опыту.

Дисперсия адекватности $S_{Ад}^2$ характеризует вариацию экспериментальных результатов в сравнении с посчитанными с помощью уравнения регрессии. Выражение $\sum_{i}^{N}(\bar{y}_i - \hat{y}_i)^2$ в формуле (7.25) определяет сумму квадратов отклонений средних значений откликов по опытам от предсказываемых уравнением регрессии. Величина $S_{ад}^2$ связана с числом степеней свободы:

$$f_{ад} = N - i, \qquad (7.26)$$

где i – число коэффициентов, входящих в уравнение регрессии после отбрасывания незначимых коэффициентов.

Для адекватной модели оценки дисперсий воспроизводимости $S_{\bar{y}_i}^2$ и адекватности $S_{ад}^2$ должны быть одинаковы. Следовательно, проверка адекватности полученной модели заключается в установлении однородности этих двух дисперсий с помощью критерия Фишера:

$$F_э = \frac{S_{ад}^2}{S_{\bar{y}_i}^2}. \qquad (7.27)$$

Если при этом $F_э$ соответсвующее экпериментальным данным меньше табличного (приложение В) при выбранном уровне значимости и соответствующих сравниваемым дисперсиям числам степеней свободы, то можно утверждать, что уравнение регрессии адекватно описывает результаты эксперимента. Если же дисперсия $S_{ад}^2 < S_{\bar{y}_i}^2$, то вывод об адекватности модели может быть сделан и без проверки условия $F_Э < F_T$.

Проверку значимости коэффициентов регрессии можно выполнить с помощью t-критерия Стьюдента или построения доверительного интервала.

В первую очередь необходимо определить дисперсию коэффициента регрессии $S_{b_i}^2$:

$$S_{b_i}^2 = \frac{S_{\bar{y}_i}^2}{N}. \qquad (7.28)$$

Определяем фактический t-критерий Стьюдента и сравниваем с табличным при заданном α и соответствующим числом степеней свободы. При интервальном способе определения значимости коэффициентов регрессии находим доверительный интервал $\Delta b_i = \pm t S_{b_i}$.

Если абсолютная величина коэффициента больше доверительного интервала, то коэффициент значим.

Контрольные вопросы

1. Преимущество многофакторных спланированных экспериментов в сравнении с однофакторными.
2. Перечень основных положений планирования и анализа эксперимента.
3. Как определять размер выборки.
4. Понятие рандомизации.
5. Выбор и обоснование переменных факторов.
6. Обоснование и выбор уровней факторов.
7. Выбор вида математической модели.
8. Выбор параметра оптимизации.
9. Выбор плана эксперимента.
10. Понятие обобщенного параметра оптимизации.
11. Что такое функция желательности Харрингтона.
12. Выбор области эксперимента и уровней переменных факторов.
13. Основной уровень и интервалы варьирования переменных факторов.
14. Полный факторный эксперимент.
15. Дробный факторный эксперимент.

8 ОБОСНОВАНИЕ РАБОЧИХ ОРГАНОВ И ТЕХНОЛОГИЧЕСКОЙ СХЕМЫ КОМБИНИРОВАННОГО АГРЕГАТА ДЛЯ ПОДГОТОВКИ ПОЧВЫ К ПОСЕВУ ОЗИМЫХ КОЛОСОВЫХ КУЛЬТУР

Анализ известных комбинированных агрегатов для подготовки почвы под озимые колосовые по пропашным предшественникам показывает, что наиболее распространенными рабочими органами являются дисковые рабочие органы, плоскорезы, лапы различного назначения, катки-дробители, измельчители почвы, волокуши-выравниватели и др., в основном заимствованные из других однооперационных машин. Комбинированные же агрегаты с активными рабочими органами нами не рассматриваются в силу ряда недостатков, присущих им по сравнению с агрегатами на основе пассивных рабочих органов: высокая энергоемкость, низкая производительность и надежность агрегатов.

Большинство известных комбинированных агрегатов с рабочими органами пассивного действия состоят из одного ряда дисковых батарей и плоскорезов, за которыми уже идут дробящие рабочие органы типа катков различной конструкции. При этом агрегаты отличаются чаще всего диаметром дисков, шириной захвата плоскорезов и конструкцией дробящих и прикатывающих рабочих органов. Однако, как показали многочисленные исследования, эти агрегаты в отдельные засушливые годы в условиях повышенной твердости почвы и наличия на поверхности почвы значительного количества пожнивных остатков не обеспечивают требуемого качества обработки почвы. Глубина обработки почвы в таких условиях сильно варьирует и составляет порой всего лишь 5–6 см. Степень крошения почвы составляет 40–50 %, что не соответствует агротехническим требованиям.

У всех агрегатов односледной обработкой дисковыми боронами степень измельчения растительных остатков ниже требуемой величины. Это может приводить к выглублению при наезде на сверху лежащий стебель дискового сошника. Вместе с тем степень измельчения растительных остатков у серийной дисковой бороны БДТ-3 значительно выше при условии заглубления ее рабочих органов больше, чем величина выреза диска по радиусу. Следовательно, в технологическую схему агрегата необходимо вводить дисковые батареи в два следа.

Если в составе агрегата отсутствуют плоскорезы, при работе на твердых почвах дисковые рабочие органы мало заглубляются в почву и из-за этого не прорабатывается почва в междисковом пространстве. Это приводит к огрехам. Поэтому введение в структуру агрегата плоскореза, подрезающего все сорняки на всю ширину захвата агрегата и обрабатывающего почву на заданную глубину необходимо.

После прохода основных рабочих органов (дисковых батарей и плоскорезов) на поверхности почвы остаются глыбы более 50 см, которые необходимо дополнительно крошить. Для этого и для прикатывания взрыхленной почвы, а также выравнивания ее поверхности в технологическую схему агрегата необходимо вводить кольчато-шпоровые катки или другие дробящие и прикатывающие устройства.

В целом технологическая схема агрегата представлена на рисунке 8.1. Она предусматривает использование принципа послойной обработки почвы и включает два ряда дисковых батарей с вырезными дисками 1, плоскорежущие рабочие органы 2 и кольчато-шпоровый каток 3. Дисковые и плоскорежущие рабочие органы размещены на раме агрегата на разных уровнях по глубине обработки со смещением каждого последующего ряда по высоте, т. е. весь обрабатываемый слой почвы 8–12 см делится на три яруса. Причем, второй ряд дисковых батарей установлен со сдвигом относительно первого ряда на величину h_2, а плоскорезы относительно второго ряда батарей на величину h_1 и находятся на заданной глубине обработки «a». Такой схемой расстановки преследуется цель получения лучшего качества крошения почвы. Для оценки гипотезы и целесообразности выбранной технологической схемы и рабочих органов сравнивались агротехнические показатели ряда серийных и экспериментальных образцов орудий и агрегатов, для чего был использован обобщенный показатель качества.

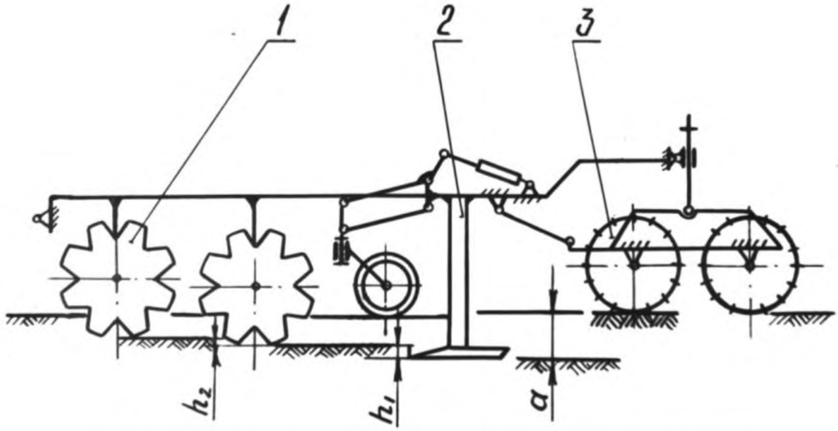


Рисунок 8.1 – Технологическая схема комбинированного почвообрабатывающего агрегата.

Анализ результатов исследований часто сводится к сравнению двух и более вариантов по нескольким показателям. Так, при отборе сортов сравниваются их урожайность, качество зерна, устойчивость к полеганию, засухоустойчивость, морозоустойчивость и т. д. Сельскохозяйственные машины оцениваются по их агротехническим, техническим, эксплуатационным и экономическим показателям. Аналогичные задачи решаются при разработке системы машин и выборе технологической схемы агрегата, при отборе пород животных и при оценке качества зерна. При этом характерно, что в каждом конкретном случае набор и количество показателей различны и определяются поставленной целью. Общей же их характеристикой является отличие сравниваемых показателей не только по физическому смыслу, размерности, но также и по относительной их важности, что создает некоторые сложности и приводит к неопределенности выводов.

Решение подобных задач связано с приведением всех сравниваемых показателей к единому количественному показателю, то есть с созданием обобщенного критерия оценки.

Для сравнения были выбраны серийно выпускаемые дисковая борона БДТ-2,5, фреза ФБ-1,0, культиватор-плоскорез КПП-2,2, а также комбинированные агрегаты, составленные из различных рабочих органов. Агротехническая оценка качества обработки почвы проводилась по равномерности глубины обработки, степени крошения, глыбистости и гребнистости поверхности поля. Исследования вели на слабовыщелоченном сверхмощном черноземе, влажность и твердость которого на глубине 5–10 см составляли в среднем соответственно 12 % и 3,25 МПа.

Для характеристики показателей качества обработки почвы использовалась оценочная шкала, предложенная П.У. Бахтиным, и соответствующие значения желательностей по Харрингтону (таблица 8.1). При этом уровень желательности 0,37 соответствует нижнему пределу допускаемых значений показателей по агротехническим требованиям или показателям серийных машин.

Обобщенный показатель качества обработки почвы находили как среднее геометрическое желательностей отдельных показателей по выражению 7.3.

Желательности отдельных показателей d_i (рисунок 8.2) определяли по выражению 7.2.

Таблица 8.1 – Показатели качества обработки почвы при различных уровнях желательности

Показатели	Обозначение	Желательность, d				
		1.0	0.8	0.63	0.37	0.2
		величина показателя				
Равномерность глубины обработки ($\sigma_{доп.}$), см	Y_1	0	0,72	1,45	2,1	2,5
Глыбистость, %	Y_2	0…15	16…30	31…45	46…60	Более 60
Степень крошения, %	Y_3	90 и более	70…89	50…69	30…49	0…29
Гребнистость, см	Y_4	0…3	4…6	7…9	10…12	Более 12

Кодированные значения Y_i' рассчитывались путем перевода в безразмерную шкалу Y' натуральных величин показателей Y_i с помощью формулы перехода:

$$Y_i' = a_0 + a_1 Y_1 + a_2 Y_1^2. \quad (8.3)$$

Коэффициенты a_0, a_1 и a_2 определяли по трем базовым точкам, соответствующим желательностям 0,37; 0,63 и 0,8 (таблица 8.2). Значимость коэффициентов проверялась по критерию Стьюдента.

Таблица 8.2 – Значения коэффициентов a_0, a_1 и a_2 для различных показателей

Показатель Y_1	Коэффициент		
	a_0	a_1	a_2
Y_1	2,143148	– 0,805724	– 0,102299
Y_2	2,219982	– 0,046333	$-0{,}66676 \cdot 10^{-4}$
Y_3	-2,056268	0,043000	$-0{,}375500 \cdot 10^{-4}$
Y_4	2,219987	– 0,231665	– 0,001666

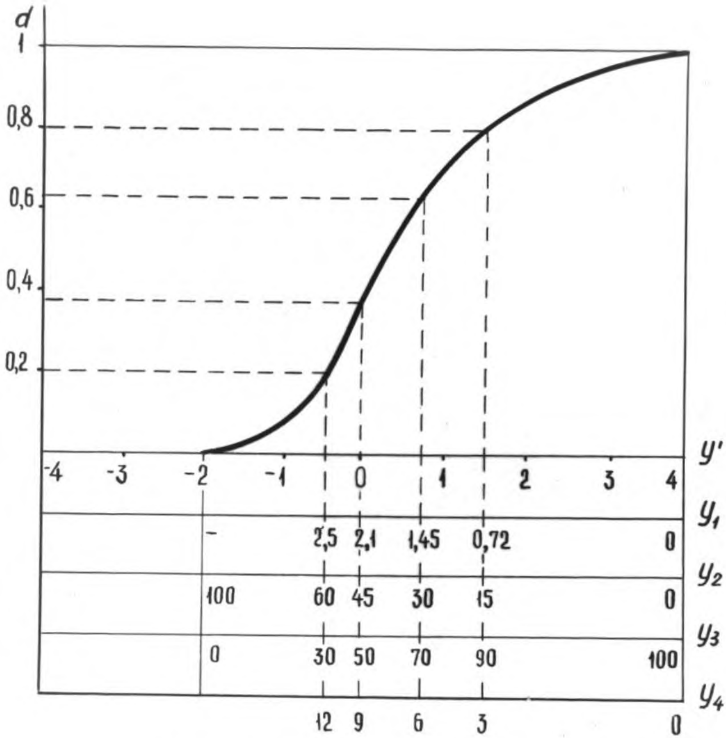


Рисунок 8.2 – Функция желательности

Весомости показателей k_i определялись методом экспертной оценки. При этом члены экспертной комиссии представляли специалистов по земледелию и агротехнике возделывания озимой пшеницы, а также специалистов по разработке орудий для обработки почвы и посева (таблица 8.3.).

Степень согласованности мнений членов экспертной комиссии проверялась по коэффициенту χ^2 по выражению 7.4.

Таблица 8.3 – Результаты экспертной оценки

Эксперты	Показатели			
	Y_1	Y_2	Y_3	Y_4
1	1/0,50	3/0,15	2/0,25	4/0,10
2	1/0,45	2/0,40	3/0,10	4/0,05
3	4/0,05	2,5/0,10	1/0,75	2,5/0,10
4	3/0,20	2/0,30	1/0,40	4/0,10
5	3/0,20	1,5/0,35	1,5/0,35	4/0,10
6	2,5/0,25	1/0,40	2,5/0,25	4/0,10
7	2,5/0,20	2,5/0,20	1/0,50	4/0,10
Сумма рангов	17/0,264	14,5/0,272	12/0,372	26,5/0,092
Отклонение от средней суммы рангов	– 0,5	– 3	– 5,5	9
Квадраты отклонений	0,25	9	30,25	81

Примечание. В числителе – ранг фактора Y_{ij}, присвоенный i-му показателю j-ом экспертом, в знаменателе – весомость k_{ij}.

Полученное значение χ^2 (10,620) больше табличного значения (7,815) при 5 %-м уровне существенности. Это значит, что имеется неслучайная согласованность мнений экспертов. По результатам опросов в порядке убывания важности показатели расположились в следующем порядке: степень крошения почвы (k_3=0,372); глыбистость поверхности поля (k_2=0,272); равномерность глубины обработки почвы(k_1=0,264); гребнистость (k_4=0,092).

В таблице 8.4. приведены натуральные значения показателей, полученных экспериментальным путем, а также их желательности и обобщенные показатели качества обработки почвы по сравниваемым технологическим схемам.

Таблица 8.4 – Натуральные значения показателей Y_i, их желательности d_i и обобщенные показатели качества D сравниваемых технологических схем

Номер п/п	Варианты	Y_1/d_1	Y_2/d_2	Y_3/d_3	Y_4/d_4	D
1	Борона дисковая БДТ-2,5	2,1/ 0,37	70,2/ 0,08	76,3/ 0,69	5,45/ 0,65	0,753
2	Фреза ФБ-1,0	2,1/ 0,37	58,5/ 0,21	76,6/ 0,69	2,78/ 0,84	0,809
3	Плоскорез КПП-2,2	2,7/ 0,12	43,7/ 0,38	57,4/ 0,47	4,25/ 0,73	0,753
4	Сферические диски ∅450 мм в один ряд и плоскорезы	2,5/ 0,2	37,2/ 0,51	67,5/ 0,59	5,1/ 0,69	0,811
5	Сферические диски ∅450 мм в два ряда и плоскорезы	2,1/ 0,37	38,2/ 0,49	71,6/ 0,64	5,7/ 0,64	0,847
6	Сферические диски ∅450 мм в два ряда, плоскорезы и двурядный кольчато-шпоровый каток	2,1/ 0,37	13,3/ 0,85	78,7/ 0,70	4,25/ 0,73	0,889
7	Плоскорезы, ротационный измельчитель и двурядный кольчато-шпоровый каток	2,1/ 0,37	18,9/ 0,77	80,7/ 0,71	2,25/ 0,88	0,888

Надо отметить, что ни одно из серийно выпускаемых в настоящее время орудий за один проход не обеспечивает даже удовлетворительного качества по всем показателям агротехнической оценки при данных условиях. Если дисковая борона БДТ-2,5 и фреза ФБ-1,0 создают глыбистую поверхность, то культиватор-плоскорез КПП-2,2 не обеспечивает требуемой равномерности глубины обработки. Это указывает на неизбежность многократных проходов по полю серийными машинами для полной подготовки почвы к посеву.

Установка перед плоскорезом сферических дисков в два ряда (вариант 5) стабилизировала устойчивость глубины обработки, а введение в техническую схему дополнительно и двухрядного кольчато-шпорового катка существенно снизило глыбистость поверхности поля и повысило степень крошения почвы (вариант 6).

Комбинированные агрегаты по технологическим схемам, включающим сферические диски ∅450 мм в два ряда, плоскорезы и двухрядный кольчато-шпоровый каток (вариант 6), плоскорезы, ротационный измельчитель и двухрядный кольчато-шпоровый каток (вариант 7), показали одинаковое качество обработки почвы: $D_6 = 0{,}889$ и $D_7 = 0{,}888$. Однако, учитывая более высокую энергоемкость, низкую надежность и производительность агрегатов с активными измельчителями, следует на данном этапе отдать предпочтение агрегату с пассивными рабочими органами.

Таким образом, применение метода обобщенного показателя качества позволяет более объективно проводить сравнение технологических схем агрегатов для обработки почвы по нескольким показателям и выбрать наилучшую из них.

9 ОПТИМИЗАЦИЯ ПАРАМЕТРОВ КОМБИНИРОВАННОГО АГРЕГАТА ДЛЯ ПОДГОТОВКИ ПОЧВЫ К ПОСЕВУ ОЗИМЫХ КОЛОСОВЫХ КУЛЬТУР

Основная цель любой оптимизационной задачи – установление таких параметров, при которых достигается экстремальное значение выбранного критерия.

Выбрать критерий оптимальности при разработке МТА методически трудно, так как эффективность любого агрегата характеризуется показателями, среди которых нет единого и универсального. Неверно принятый критерий оптимизации, как правило, приводит к грубым просчетам, снижает или практически сводит на нет ценность полученных результатов.

Анализ показывает, что в качестве критерия оценки часто принимают различные экономические, технические и эксплуатационные показатели, поэтому оптимальные параметры существенно отличаются друг от друга. Кроме того, в процессе исследования по оптимизации МТА не сравнивают и не соизмеряют эти показатели. Следует также отметить, что существующая методика разработки агротехнических требований, а также их содержание не позволяют связать единым критерием оценки различные машины. Однако можно утверждать, что в основу новой методики целесообразно положить принцип дифференцированной оценки эффективности качественных показателей.

Таким образом, в общем случае в качестве критерия оценки эффективности сравниваемых вариантов должен быть выбран единый обобщенный показатель, обеспечивающий наивыгоднейший компромисс между отдельными показателями.

Для решения поставленной задачи следует выяснить, какая зависимость существует между величиной показателя и его оценкой, как свести воедино отдельные показатели и найти комплексную оценку, а также определить их относительные весомости. Наиболее удачное решение поставленной задачи заключается в преобразовании измеренных значений отдельных показателей в безразмерную шкалу желательности.

Кодированные значения y_i' находят при переводе в безразмерную шкалу значений y' натуральных показателей y_i полинома выражения 8.3.

Обобщенную функцию желательности D – комплексный критерий оценки, определяют как среднее геометрическое значение частных функций желательности d_i.

В дальнейшем при известных функциональных зависимостях $y_i = f_i(x_i)$, которые получают аналитическим способом или на основе многофакторных экспериментов, имеем:

$$D = f\left(x_i\right), \tag{9.1}$$

где x_i – оптимизируемые переменные факторы.

Представленный в таком виде критерий, хотя и служит косвенным показателем, в целом оптимизирует рассматриваемый объект или систему.

Общий порядок оптимизации предусматривает следующие этапы:

- выбор объекта оптимизации и сбор априорной информации;
- выбор отдельных показателей, характеризующих эффективность оптимизируемого объекта;
- обсуждение и выбор оптимизируемых параметров и предполагаемых диапазонов их изменения;
- выбор метода обобщения отдельных показателей (выбор критерия оценки);
- определение весомости (важности) отдельных показателей;
- выбор количественного универсального критерия оценки отдельных показателей (выбор вида функции желательности);

– установление физических связей отдельных показателей и оптимизируемых параметров;
– определение величины переменных оптимизируемых параметров, соответствующих экстремальному значению выбранного критерия оценки.

Предлагаемый метод используется при оптимизации параметров комбинированного почвообрабатывающего агрегата для подготовки почвы за один проход под посев озимых колосовых культур после пропашных предшественников.

Исследования проводились в ОПХ Краснодарского НИИСХ на поле после уборки кукурузы на зерно. Рельеф поля равнинный, микрорельеф – ровный; тип почвы – предкавказский слабовыщелоченный чернозем; абсолютная влажность почвы составила по горизонтам 0–5 см, 5–10 и 10–15 см соответственно 16,2, 21,5, 21,8 %; твердость на глубине 5, 10 и 15 см – 2,04, 2,53, 2,84 МПа.

Технологическая схема агрегата предусматривает послойную обработку почвы и включает два ряда дисковых батарей, плоскорежущие рабочие органы и кольчато-шпоровый каток. Дисковые и плоскорежущие рабочие органы размещены на раме агрегата на разных уровнях по глубине обработки, т. е. весь обрабатываемый слой почвы (8–12 см) делится на три яруса. Второй ряд дисковых батарей смещен по высоте относительно первого на величину h_2, а плоскорезы смещены относительно второго ряда батарей на h_1 и находятся на заданной глубине обработки «a». Такая схема расстановки рабочих органов позволяет получить хорошее крошение почвы.

Чтобы установить физическую связь между выбранными показателями эффективности и переменными факторами (X_1 – рабочая скорость v_p движения агрегата, м/с; X_2 – глубина «a» обработки почвы, см; X_3 и X_4 – углы атаки α_1 и α_2 первого и второго рядов дисковых батарей, рад.; X_5 и X_6 – смещения h_1 и h_2 рабочих органов относительно друг друга, см (таблица 9.1), поставлен многофакторный эксперимент типа 2^{k-1} (k – число факторов).

Таблица 9.1 – Уровни факторов и интервалы их варьирования

	Обозначение	Факторы					
		X_1	X_2	X_3	X_4	X_5	X_6
Основной уровень	0	1,75	11,5	0,175	0,175	3	2
Интервал варьирования:	ΔX_j	0,55	2	0,0875	0,0875	2	2
верхний	+1	2,3	13,5	0,2625	0,2625	5	5
нижний	-1	1,2	9,5	0,0875	0,0875	1	1

В качестве отдельных показателей, характеризующих эффективность оптимизируемого агрегата, приняты приведенные затраты $C_{пр}$, руб. на 1 га, производительность агрегата W, га/ч и качество обработки почвы d_1 – критерий, обобщающий степень крошения почвы y_1 (доля почвенных фракций размером < 50 мм, %), равномерность глубины обработки y_2, см, глыбистость y_3, %, гребнистость y_4, см и удельное тяговое сопротивление y_5, Н/м (таблица 9.2).

Таблица 9.2 – Оценочная шкала качества обработки почвы

Показатель	Желательность, d				
	1,0	0,8	0,63	0,37	0,2
y_1	90 и более	70…89	50…69	30…49	0…29
y_2	0	0,72	1,45	2,1	2,5
y_3	0…15	16…30	31…45	46…60	Более 60
y_4	0…3	4…6	7…9	10…12	Более 12

Приведенные затраты $C_{пр}$ и производительность W рассчитываются по общепринятой методике, на основе которой после несложных преобразований получены следующие выражения:

$$C_{пр} = \frac{(C_з + 0,081G_T + \dfrac{0,655Б_T}{T_{ГТ}} + 0,749B_P)}{W}, \qquad (9.2)$$

$$W = \frac{\left[V_p P_{кр}(0,273 - 4,62 \cdot 10^{-2} T_{TO} - 0,68 \cdot 10^{-2} V_p)\right]}{y_5}, \qquad (9.3)$$

где $C_з$ – зарплата обслуживающего персонала, руб./ч;

G_T – часовой расход горючего при номинальной загрузке трактора на скорости v_p, кг/ч;

$Б_T$ – балансовая стоимость трактора, руб.;

$T_{ГТ}$ – годовая загрузка трактора, ч;

$P_{кр}$ – усилие, развиваемое трактором при номинальной загрузке на скорости v_p, Н;

T_{TO} – время на проведение ежемесячного технического обслуживания трактора, ч.

В результате обработки полного факторного эксперимента получены следующие уравнения связи:

$$y_1 = 116,3 + 3,06X_1 - 5,33X_2 - 139,3X_3 - 37,26X_4 - \\ - 2,11X_5 + 14,3X_2X_3, \qquad (9.4)$$

$$y_2 = 2,52 + 0,31X_1 - 0,69X_3 - 0,12X_5 + 1,03X_3X_5, \qquad (9.5)$$

$$y_3 = 0,84 + 0,51X_2 + 50,73X_3 + 11,46X_4 + \\ +0,48X_6 - 5,16X_3X_4, \qquad (9.6)$$

$$y_4 = 6,9 \qquad (9.7)$$

Регрессионный анализ, широко применяемый при планировании эксперимента, базируется на основе следующих предпосылок:

– серии измерений, состоящие из отдельных наблюдений исследуемого параметра y_1, y_2, y_3,…, y_N при всех возможных сочетаниях значений переменных факторов x_i, служат выборками из генеральных совокупностей, подчиняющихся нормальному закону распределения;

– дисперсии $S^2(y)$ во всех сериях измерений-выборок статистически однородны, т. е. дисперсия $S^2(y)$ воспроизводима от одной серии опытов к другой.

При исследовании тягового сопротивления почвообрабатывающих машин и агрегатов методами математического планирования эксперимента первая предпосылка практически всегда выполняется. Вторая же часто нарушается в связи с изменением характера взаимодействия рабочих органов агрегата с почвой при переходе сочетания переменных факторов x_i с одного уровня их значений на другой. Например, рост скорости движения агрегата или твердости почвы увеличивает дисперсию тягового сопротивления. Кроме того, на величину дисперсии существенное влияние могут оказать тип рабочего органа, его геометрия и расположение.

Обычные приемы для преобразования исследуемого параметра "y" в новый – y' с целью достижения воспроизводимости дисперсий ($y' = \lg y$, $y' = \sqrt{y}$, $y' = 2\sqrt{y}$, $y' = \sqrt{y+1}$, и др.) могут быть рекомендованы лишь в тех случаях, когда существует четко выраженная зависимость $\delta^2(y) = f(\bar{y})$. Это бывает, например, при биномиальном и пуассоновском законах распределения "y". Если же распределение параметра "y" подчиняется нормальному закону, а дисперсии неоднородны, то эти преобразования не приводят к успеху. В конечном итоге воспроизводимость дисперсии требуется для достижения равноточности результатов всех опытов плана эксперимента при одинаковом числе параллельных измерений, что существенно упрощает вычислительные операции. Если имеем достаточно большое число измерений по каждому опыту, то можно получить требуемую точность математического ожидания генеральной совокупности $M(y)$ практически во всех случаях, при любом значении дисперсии $\sigma^2(y)$. Для сохранения же общей процедуры вычислительных операций и для достижения заданной точности рекомендуется все параллельные измерения по каждому опыту разделить на группы, а каждую группу заменить ее средним значением. При этом объем выборки по каждой группе должен обеспечивать требуемую точность измерения средней величины при заданном уровне доверительной вероятности.

Предлагаемый методический подход применен при разработке математической модели удельного тягового сопротивления комбинированного почвообрабатывающего агрегата для подготовки почвы за один проход под посев озимых колосовых культур после пропашных предшественников.

В результате обработки экспериментальных данных после предварительной статистической проверки и исключения резко выделяющихся значений отдельных измерений в каждой выборке определены средние значения $\overline{y}$ и средние квадратические отклонения тягового сопротивления $\sigma_{\overline{y}}$. Проверка по критерию Кохрена G при уровне доверительной вероятности $q_{0,95}$ показывает отсутствие воспроизводимости внутри выборочных дисперсий по опытам при числе измерений в каждой выборке N = 600 ($G_{табл.}$ = 0,0333, $G_{факт.}$ = 0,069).

Корреляционный анализ показывает отсутствие связи между выборочной средней $\overline{y}$ и ее средним квадратическим отклонением. Наиболее вероятным объяснением этого факта может служить наличие зависимости среднего квадратического отклонения $\sigma_{\overline{y}}$ от изменения переменных факторов x_i. Для выяснения характера этой зависимости определено расчетным путем адекватное уравнение регрессии в кодированном выражении:

$$\sigma_{\overline{y}} = 237,9 + 55,3X_1 + 6,5X_2 - 21,3X_3 + 9,4X_5, \qquad (9.8)$$

или в натуральных величинах:

$$\sigma_{\bar{y}} = 492,6 + 986,4V_p + 31,9a - 2395,6a_1 + 46,1h_1. \qquad (9.9)$$

Анализируя уравнение (9.8), можно отметить, что наибольшее влияние на величину $\sigma_{\bar{y}}$ оказывает скорость X_1 (v_p) движения агрегата и угол X_3 (α_1) атаки первого ряда дисковых батарей.

Преобразование каждой выборки в отдельные группы с числом измерений в каждой группе $N \approx 200$ приводит к воспроизводимости межгрупповой дисперсии $\sigma^2_{воспр.}$ = 71,9 (кН/м)^2, которая была использована для расчета математической модели.

В конечном итоге после необходимых расчетов получаем уравнение регрессии тягового сопротивления на единицу ширины захвата комбинированного почвообрабатывающего агрегата в зависимости от переменных факторов:

$$\begin{aligned} y_5 = &-779,9 + 1016,3X_1 + 784,8X_2 + 5961,8X_3 + \\ &+14847,1X_4 + 387,5X_5 - 436,5X_6. \end{aligned} \qquad (9.10)$$

Анализ этого уравнения показывает, что тяговое сопротивление уменьшается при увеличении разности глубины установки первого и второго рядов дисков и уменьшении величины значений всех остальных факторов. Из этого следует, что наименьшее тяговое сопротивление будет достигнуто при установке только одного ряда дисков на глубину хода плоскорезов. Однако наличие второго ряда дисков необходимо для повышения качества обработки почвы, особенно на глубине более 10 см.

Отметим также, что, согласно уравнению, на величину тягового сопротивления максимальное влияние оказывает глубина «a» обработки. Остальные факторы по мере уменьшения их значимости можно расположить в следующей последовательности: α_2, V_p, h_2, h_1, a.

Таким образом, применение метода сгруппированных оценок позволяет получить однородные (воспроизводимые) дисперсии по всем опытам плана эксперимента, что служит необходимым условием составления математической модели при определении тягового сопротивления и его анализа в интервале изменения переменных факторов.

Для показателей качества обработки почвы y_1, y_2, y_3, y_4 преобразование по безразмерной шкале проводится $d_i = \exp\left[-\exp(-y_i')\right]$, а приведенных затрат $C_{пр.}$ и производительности агрегата W – по линейной функции:

$$d_2 = 1,47 - 0,16C_{пр}, \qquad 3 \le C_{пр} \le 7 \tag{9.11}$$

$$d_3 = 0,14 + 0,19W, \qquad 1,2 \le W \le 4,5 \tag{9.12}$$

По отдельным показателям качества обработки почвы d'_1, d'_2, d'_3 и d'_4 находим обобщенный критерий d_1 как среднее геометрическое этих величин:

$$d_1 = \sqrt[n]{\prod_{i=1}^{n} d_i'^{k_i'}}. \tag{9.13}$$

Для характеристик показателей качества обработки почвы используем оценочную шкалу (таблица 9.2), предложенную П.У. Бахтиным.

Весомости таких показателей, как степень крошения почвы k'_1, равномерность глубины обработки k'_2, глыбистость k'_3 и гребнистость k'_4, соответственно равны 0,372, 0,264, 0,272 и 0,092.

Весомость k_1 качества обработки почвы, характеризуемая обобщением показателей y_1, y_2, y_3, y_4, составляет 0,390, приведенных затрат k_2 – 0,265 и производительности k_3 агрегата – 0,345.

Задачу решаем методом итерационных вычислений. Анализ результатов (таблица 9.3.) показывает, что оптимальные параметры МТА в основном определяются принятым критерием. При их оценке по приведенным затратам C_{np} оптимальный вариант агрегата характеризуется следующими параметрами: $B_p = 3{,}82$ м, $v_p = 1{,}90$ м/с, $\alpha_1 = \alpha_2 = 0{,}0875$ рад; $h_1 = 1$, $h_2 = 5$ см. В этом случае комбинированное почвообрабатывающее орудие агрегатируется с трактором Т-4М. При всех остальных критериях оптимальный вариант предусматривает агрегатирование с трактором К-701. При оценке по критерию W: $B_p = 5{,}83$ м, $V_p = 2{,}38$ м/с; по критерию d_1: $B_p = 6{,}33$ м, $V_p = 1{,}18$ м/с; по критерию y_1: $B_p = 2{,}90$ м, $V_p = 3{,}46$ м/с. Углы атаки дисковых батарей α_1 и α_2 при оценке по C_{np}, W и d_1 равны 0,0875 рад, а h_1 и h_2 соответственно равны 1 и 5 см. При оценке по критерию y_1 угол $\alpha_1 = 0{,}42$ и $\alpha_2 = 0{,}0875$ рад, $h_1 = 3$, $h_2 = 5$ см. Это показывает, что параметры регулировки агрегата, соответствующие наилучшим показателям качества обработки почвы, не совпадают с параметрами при оценке по критериям W, C_{np} и d_1, что вызвано разной чувствительностью и направлением изменения этих критериев оценки.

Оптимальные параметры агрегата при оценке их по обобщенному критерию D были близки к параметрам при оценке по критерию W, изменяющемуся обратно пропорционально тяговому сопротивлению y_5.

Оптимальный вариант при оценке по обобщенному показателю D: $B_p = 5{,}83$ м, $V_p = 2{,}38$ м/с, $\alpha_1 = \alpha_2 = 0{,}0875$ рад., $h_1 = 1$, $h_2 = 5$ см, агрегатирование с трактором К-701. Производительность $W = 3{,}6$ га/ч, $C_{np} = 4{,}29$ руб. на 1 га, степень крошения $y_1 = 65{,}3$ %.

Таблица 9.3 – Оптимальные параметры комбинированного почвообрабатывающего агрегата (α = 10см)

Крите-рий оп-тималь-ности	Марка трактора	Достигнутый уровень критерия оптимизации					Оптимизируемые параметры					
		$C_{пр}$,руб. на 1 га	W,га/ч	d_1	y_1,%	D	B_p,м	V_p,м/с	α_1,рад	α_2,рад	h_1,см	h_2,см
1	2	3	4	5	6	7	8	9	10	11	12	13
$C_{пр}$	ДТ-75М	4,33	1,28	0,7336	63,9	0,8498	2,43	1,96	0,0875	0,0875	1	5
	Т-4М	3,92	1,93	0,7377	63,8	0,8815	3,82	1,90	0,0875	0,0875	1	5
	Т-150К	4,33	1,99	0,6188	67,8	0,8555	2,13	3,22	0,0875	0,0875	1	5
	К-700	5,10	2,33	0,6654	66,4	0,8622	3,27	2,78	0,0875	0,0875	1	5
	К-701	4,29	3,60	0,7015	65,2	0,9139	5,83	2,38	0,0875	0,0875	1	5
W	ДТ-75М	4,44	1,39	0,7592	62,7	0,8570	3,28	1,56	0,0875	0,0875	1	5
	Т-4М	4,14	2,00	0,7586	62,7	0,8840	4,73	1,57	0,0875	0,0875	1	5
	Т-150К	4,45	2,02	0,6740	66,2	0,8651	2,90	2,69	0,0875	0,0875	1	5
	К-700	5,20	2,42	0,7152	64,7	0,8719	4,20	2,21	0,0875	0,0875	1	5
	К-701	4,29	3,60	0,7015	65,2	0,9139	5,83	2,38	0,0875	0,0875	1	5
d_1	ДТ-75М	4,92	1,18	0,7651	54,3	0,8415	2,79	1,56	0,0875	0,0875	5	5
	Т-4М	5,03	1,45	0,7647	54,3	0,8508	3,43	1,57	0,0875	0,0875	5	1
	Т-150К	6,74	1,92	0,7646	54,3	0,8299	4,57	1,57	0,0875	0,0875	5	5
	К-700	4,80	1,74	0,6861	57,7	0,8512	2,50	2,69	0,0875	0,0875	5	5
	К-701	7,88	2,01	0,7815	57,3	0,8099	6,33	1,18	0,0875	0,0875	3	5

Продолжение таблицы 9.3

1	2	3	4	5	6	7	8	9	10	11	12	13
V_p	ДТ-75М	5,63	0,79	0,6006	68,2	0,7820	1,02	2,99	0,42	0,0875	1	5
	Т-4М	4,99	1,15	0,5080	69,3	0,8013	1,35	3,33	0,42	0,0875	1	5
	Т-150К	4,98	1,67	0,5723	68,9	0,8252	2,04	3,22	0,42	0,0875	1	5
	К-700	6,18	1,78	0,5432	69,6	0,7997	2,05	3,44	0,42	0,0875	1	5
	К-701	5,26	2,52	0,5129	69,7	0,8395	2,90	3,46	0,42	0,0875	1	5
D	ДТ-75М	4,44	1,39	0,7592	62,7	0,8570	3,28	1,56	0,0875	0,0875	1	5
	Т-4М	4,14	2,00	0,7586	62,7	0,8840	4,73	1,57	0,0875	0,0875	1	5
	Т-150К	4,45	2,02	0,6740	62,2	0,8651	2,90	2,69	0,0875	0,0875	1	5
	К-700	5,20	2,42	0,7152	64,7	0,8719	4,20	2,21	0,0875	0,0875	1	5
	К-701	4,29	3,60	0,7015	65,2	0,1939	5,83	2,38	0,0875	0,0875	1	5

10 ПОВЫШЕНИЕ ЭФФЕКТИВНОСТИ ПОЧВООБРАБАТЫВАЮЩИХ ДИСКОВЫХ РАБОЧИХ ОРГАНОВ

Для поверхностной обработки почвы в настоящее время созданы и разрабатываются новые комбинированные агрегаты, позволяющие сократить число проходов по полю и повысить качество обработки почвы. Наиболее простыми из них являются комбинированные агрегаты, построенные по принципу рационального сочетания пассивных рабочих органов. Чаще всего в таких агрегатах применяются сферические диски и плоскорезы в сочетании с другими рабочими органами.

Дисковые рабочие органы имеют ряд преимуществ: простота конструкции, относительно небольшой износ рабочих органов и др. Однако эффективность сферических дисков в значительной степени зависит от их расстановки на оси батареи: при недостаточном расстоянии между дисками повышается вероятность заклинивания между ними пласта или отдельных глыб почвы, а увеличение этого расстояния отрицательно сказывается на качестве обработки почвы (высокие гребни дна борозды, неполное подрезание сорняков, появление огрехов).

Эффективность воздействия сферических дисков на обрабатываемый слой почвы можно повысить, если обеспечить оптимальное расстояние между дисками без снижения надежности технологического процесса. Очевидно, что уменьшение междисковых промежутков позволит усилить крошащий эффект рабочих органов, однако для устранения забивания их почвой и растительными остатками необходимо компенсировать «пропускную способность» дисков путем увеличения расстояния между ними в продольном направлении.

Эта гипотеза была реализована в конструкции экспериментальной установки с индивидуальной подвеской сферических дисков, размещенных на жесткой раме (рисунок 10.1).

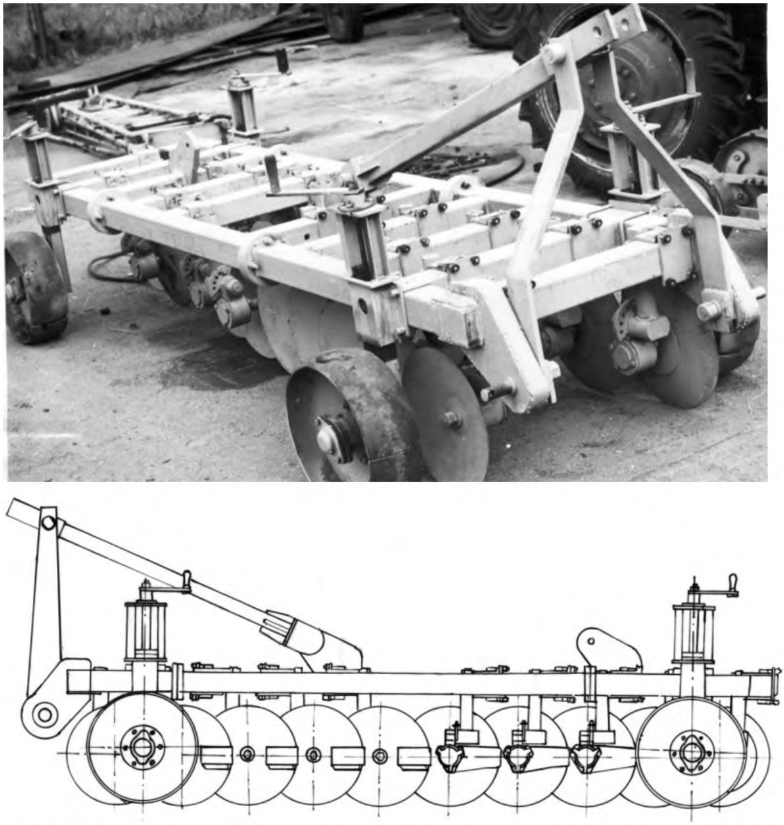


Рисунок 10.1 – Экспериментальная установка с индивидуальной подвеской сферических дисков

Исследования по оценке качества работы установки проводились в ОПХ Краснодарского НИИСХ по стерне кукурузы, убранной на зерно, и по стерне озимой пшеницы.

Рельеф поля – равнинный, микрорельеф – ровный. Тип почвы – предкавказский слабовыщелоченный чернозем.

Следует отметить, что исследования проводились при практически довольно высокой твердости (4,34 и 4,08 МПа) и различной влажности (21,83 и 14,2 %) почвы на глубине хода рабочих органов (до 10 см). Размеры делянок – 10 × 50 м. Установка агрегатировалась с трактором Т-150К. В 1982 г. диски

$\varnothing$450 мм на установке были размещены по схеме «вразвал» с углом атаки α = 15° и расстоянием между ними b = 50 мм. Показатели работы установки сравнивались с показателями бороны БДТ-3,0 за два и три прохода (в трехкратной повторности). В 1983 г. исследовалась работа установки с расположением дисков по схеме «всвал» при углах атаки α = 10, 15 и 20° и одинаковом междисковом расстоянии b = 50 мм. В продольном направлении диски установки были закреплены на расстоянии 340 мм один от другого.

Результаты обработки экспериментальных данных по глубине хода рабочих органов и степени крошения почвы представлены в таблицах 10.1 и 10.2.

Устойчивость хода по глубине. По агротехническим требованиям к машинам поверхностной (мульчирующей) обработки отклонение от средней глубины не должно превышать ± 2 см.

Для статистической оценки процессов изменения глубины обработки почвы применяется выражение, характеризующее вероятность сохранения допуска:

$$\overline{P} = 2\Phi\left(\frac{\Delta}{\sigma}\right), \tag{10.1}$$

где Δ – агротехнический допуск на отклонение глубины обработки от средней величины;

σ – среднеквадратическое отклонение.

Таблица 10.1 – Глубина обработки почвы

Вариант и настроечные параметры (1982)	Скорость км/ч	Глубина обработки, см					Вероятность сохранения допуска (Р)		Вариант и настроечные параметры (1983)	Скорость, км/ч	Глубина обработки, см					Вероятность сохранения допуска (Р)	
		M_{cp}	m	σ	ν	P	Δ=±1 см	Δ=±2 см			M_{cp}	m	σ	ν	P	Δ=±1 см	Δ=±2 см
1.Установка	3,60	7,64	0,42	1,76	23,1	5,40	0,63	0,75	Установка	3,60	7,5	0,27	1,38	18,4	3,6	0,53	0,85
α=15°	7,20	5,35	0,21	0,87	16,1	3,90	0,75	0,98	α=10°	6,55	7,0	0,16	1,16	16,6	2,3	0,61	0,81
b=50мм	8,18	6,06	0,16	0,63	10,3	2,50	0,89	0,99	b=50мм	11,02	5,7	0,17	1,18	20,7	2,9	0,60	0,91
2.БДТ-3																	
α=15°	3,60	4,80	0,28	1,57	32,8	5,90	0,48	0,80	Установка	3,60	10,6	0,24	1,71	16,0	2,3	0,44	0,76
b=220мм	6,43	5,09	0,29	1,67	32,8	5,80	0,45	0,77	α=15°	6,43	9,17	0,15	1,06	11,56	1,64	0,65	0,94
(два прохода)	8,18	6,48	0,36	2,07	32,1	5,57	0,37	0,67	b=50мм	10,59	9,4	0,25	1,83	19,5	2,8	0,41	0,72
2.БДТ-3																	
α=15°	3,60	7,19	0,47	2,41	33,6	6,58	0,33	0,59	Установка	3,60	7,6	0,30	2,20	31,1	4,1	0,35	0,63
b=220мм	6,43	6,17	0,35	1,86	30,2	5,70	0,41	0,72	α=20°	6,48	7,6	0,15	1,10	14,5	2,0	0,64	0,93
(три прохода)	8,18	6,81	0,41	2,34	34,4	6,09	0,33	0,61	b=50мм	10,75	8,2	0,20	1,40	17,0	2,4	0,52	0,84

Таблица 10.2 – Степень крошения почвы

Вариант и настроечные параметры (1982)	Скорость движения, км/ч	Степень крошения (фракции почвы менее 50 мм), %					Вариант и настроечные параметры (1983 г.)	Скорость движения, км/ч	Степень крошения (фракции почвы менее 50 мм), %				
		$M_{ср}$	m	σ	ν	P			$M_{ср}$	m	σ	ν	P
1.Установка α = 15° b = 50мм	3,60 7,20 8,18	64,73 86,84 66,87	9,10 1,45 7,32	15,77 2,52 12,68	24,3 2,9 18,9	14,0 1,6 10,9	Установка α = 10° b = 50мм	3,60 6,55 11,02	99,7 98,6 99,7	0,16 0,83 0,25	0,4 2,03 0,61	0,4 2,0 0,6	0,1 0,8 0,2
2.БДТ-3 α = 15° b = 220 мм (два прохода)	3,60 6,43 8,18	77,48 68,14 73,70	1,75 2,81 3,63	3,03 4,88 6,30	3,9 7,1 8,5	2,9 4,1 4,9	Установка α = 15° b = 50мм	3,60 6,43 10,59	87,98 80,85 93,27	4,06 3,01 3,08	9,94 7,37 7,55	11,2 9,11 8,0	4,6 3,7 3,3
2.БДТ-3 α = 15° b = 220мм (три прохода)	3,60 6,43 8,18	74,40 81,94 86,30	0,83 5,82 4,30	1,44 10,1 7,45	1,8 12,3 8,6	1,0 7,1 4,9	Установка α = 20° b = 50мм	3,60 6,48 10,75	83,7 96,0 97,5	2,82 1,67 1,52	6,92 4,11 3,73	7,8 4,2 3,8	3,1 1,7 1,5

Выполненные расчеты показывают, что при скорости движения в диапазоне V_n =3,6-8,18 км/ч и $\Delta = \pm 1$ см вероятность сохранения этого допуска при полученных значениях среднеквадратических отклонений для рассматриваемых вариантов равна (соответственно скоростям движения): 0,63; 0,75; 0,89 (установка), 0,48; 0,45; 0,37 (БДТ-3,0 за два прохода) и 0,33; 0,41; 0,33 (БДТ-3,0 за три прохода).

При $\Delta = \pm 2$ см вероятность сохранения этого допуска при тех же значениях σ для рассматриваемых вариантов значительно выше и равна (соответственно скоростям): 0,75; 0,98; 0,99 (установка), 0,80; 0,77; 0,67 (БДТ-3,0 за два прохода) и 0,59; 0,72; 0,61 (БДТ-3,0 за три прохода).

Таким образом, хотя допуск $\Delta = \pm 1$ см и является достаточно жестким, он все же выдерживается экспериментальной установкой на 63-89% площади обрабатываемого участка, что значительно выше, чем по сравниваемому варианту (33–48 %).

Допуск $\Delta = \pm 2$ см выдерживается на первой передаче трактора ($V_n = 3{,}6$ км/ч) обоими вариантами на одинаковой площади (75–80 %), а на более высоких скоростях сказывается преимущество установки: 98–99 % против 61–77 % (БДТ-3,0).

Считая, что для различных полей заданный уровень вероятности $\overline{P}_з$ может колебаться в пределах 0,7–0,9, определим допустимое значение $|\sigma|_д$ при заданном уровне вероятности $\overline{P}_з$ сохранении допуска:

1. При $\Delta = \pm 1$ см:

$$|\sigma|_{д_1}^{P_з=0,7} = 0{,}96\,см$$

$$|\sigma|_{д_2}^{P_з=0,9} = 0{,}61\,см$$

2. При $\Delta = \pm 2$ см:

$$\left|\sigma\right|_{\partial_3}^{P_з=0,7} = 1,92\ см$$

$$\left|\sigma\right|_{\partial_4}^{P_з=0,9} = 1,22\ см$$

Таким образом, фактическое значение среднеквадратического отклонения по глубине (таблица 10.1) находится в пределах допустимых значений при $\Delta = \pm 1$ см и $\overline{P}_з = 0,7$ только по установке при скоростях движения V_n = 7,20–8,18 км/ч. При $\Delta = \pm 2$ см и $\overline{P}_з = 0,7$ экспериментальная установка обеспечивает колебания глубины обработки в пределах заданного допуска во всем диапазоне скоростного режима. Если же $\overline{P}_з = 0,9$, то допустимые значения отклонений по глубине достигаются при скоростях V_n = 7,20 и 8,18 км/ч. Для бороны БДТ-3,0 при $\overline{P}_з = 0,9$ все значения $\sigma > \left|\sigma\right|_{\partial}$.

Исследования работы установки показали, что при скорости движения в диапазоне V_n = 3,60–11,02 км/ч и $\Delta = \pm 1$ см, вероятность сохранения этого допуска находится в пределах 0,35-0,65. В исследованном скоростном режиме наибольшая устойчивость по глубине обработки при $\Delta = \pm 1$ см и углах атаки дисков от 10 до 20° достигается на скорости V_n = 6,43–6,55 км/ч (P = 0,61–0,65). Вероятность соблюдения агротехнического допуска $\Delta = \pm 2$ см при тех же значениях σ для рассматриваемых вариантов значительно выше ($\overline{P}_з = 0,63 - 0,94$), однако закономерность влияния скорости движения установки и угла атаки дисков на величину вероятности остается неизменной: максимальное значение $\overline{P}$ сохраняется при V_n = 6,43–6,55 км/ч и равно 0,81–0,94 при изменении α от 10 до 20°.

Фактические значения среднеквадратического отклонения по глубине находятся в пределах допустимых значений при

$\Delta = \pm 2$ см и $\bar{P} = 0,7$. Исключение составляет только вариант с $\alpha = 20°$ и $V_n = 3,60$ км/ч, когда $\sigma > |\sigma|_{\partial_3}$. При $\bar{P} = 0,9$ установка обеспечивает колебания глубины обработки в допустимых пределах (не более $|\sigma|_{\partial_4}$ = 1,22 см) при скоростях $V_n = 6,43$-6,55 км/ч и $\alpha = 10$–20°, а также при $V_n = 11,02$ км/ч и $\alpha = 10°$.

Значение σ при $\Delta = \pm$ 1см находятся выше допустимых значений при всех настроечных и режимных параметрах работы установки.

Степень крошения почвы. Условия работы объектов исследования по твердости почвы были довольно тяжелыми: 4,15 МПа (1982 г.) и 3,64 МПа (1983 г.) в слое 0–15 см, а на глубине хода рабочих органов (10 см) – 4,34 МПа и 4,08 МПа соответственно. Однако, несмотря на это экспериментальная установка обеспечивает за один проход степень крошения 64,73–86,84 % по стерне кукурузы, убранной на зерно (1982 г.) и 80,85–99,70 % по стерне озимой пшеницы (таблица 10.2).

Анализ полученных данных (1982 г.) показывает, что установка за один проход и борона БДТ-3,0 за два и три прохода обеспечивают практически одинаковую степень крошения при поступательных скоростях V_n = 3,6 и 8,18 км/ч. Только при V_n = 7,2 км/ч установка дает существенно большую степень крошения по сравнению с БДТ-3,0 за два прохода на той же передаче трактора. Следует отметить, что скорость движения бороны в этом случае была меньшей, чем установки на 0,77 км/ч, что, возможно, оказало влияние на качество измельчения почвы в этом опыте.

Оценка качества работы установки по степени крошения почвы при различных углах атаки (1983 г.) показала, что она обеспечивает во всех вариантах степень крошения более 80 %, то есть удовлетворяет агротребованиям по этому показателю. С увеличением скорости до V_n = 10,59–11,02 км/ч крошение существенно увеличивается и имеет наибольшее значение в вариантах опыта с углами атаки 15 и 20° (93,27 и 97,5 %). При

$\alpha = 10°$ получена максимальная степень крошения (98,6–99,7 %), однако этот вариант не сопоставлялся, так как не была обеспечена необходимая глубина обработки почвы.

Увеличение угла атаки дисков с 10 до 20° в исследованном диапазоне поступательных скоростей не оказало существенного влияния на качество крошения почвы.

11 СТАТИСТИЧЕСКИЕ МЕТОДЫ КОНТРОЛЯ КАЧЕСТВА ВЫПОЛНЕНИЯ СЕЛЬСКОХОЗЯЙСТВЕННЫХ РАБОТ

Обеспечение качественного выполнения всех операций единого технологического процесса является одним из основных условий повышения эффективности сельскохозяйственного производства. В этой важной области в последние годы достигнуты большие успехи. Разработаны и внедрены в производство ряд более совершенных сельскохозяйственных агрегатов, способных выполнять весь комплекс работ по возделыванию и уборке большинства культур механизированным способом и с высоким качеством. Повышению качества способствовала также более совершенная система организации производства и материального поощрения за высококачественное и своевременное выполнение работ. Однако имеющиеся возможности по повышению качества выполняемых операций еще не полностью используются. Имеют место случаи нарушения технологической дисциплины, выполнения работ технически неисправными агрегатами, неправильной их регулировки и настройки. Недостаточно внимания уделяется повышению профессионального уровня подготовки непосредственных исполнителей. Во многих случаях неудовлетворительное качество является следствием отсутствия или слабого его контроля. Устранение указанных недостатков требует комплексного подхода к проблеме повышения качества выполнения всех технологических операций, т. е. решения ряда организационных, экономических, технических и методических вопросов.

11.1 Способы отбора данных исследуемых объектов

Основным условием отбора данных исследуемого объекта является случайность, которая и позволяет применять для обработки основные положения теории вероятностей и математической статистики. Даже при **естественном отборе** объектов, что обычно бывает при наблюдениях, приходится организовать случайный отбор данных этих объектов во времени или в случайно отобранных координатах. И, наоборот, отбор данных по заранее заготовленному плану в определенных координатах факторного пространства называется **искусственным отбором**. При искусственном отборе различают два способа: **пристрастный отбор** и **репрезентативный отбор**.

При **пристрастном отборе** для последующего исследования учитываются только элементы с одним или большим количеством общих признаков из множества, из которого отбирается выборка. Например, для сравнения могут быть отобраны автомобили с одинаковым объемом двигателя. Из общего стада с целью исследования молочной продуктивности могут быть отобраны коровы одной породы, не больные и только третьего отела.

Отбор называется **репрезентативным**, если отобранная группа достаточно полно характеризует всю совокупность, из которой был сделан отбор. Однако, как бы тщательно мы не проводили отбор, все же в отобранной выборке всегда будет сохраняться элементы случайности. А не включенные в выборку элементы не отразятся на характеристики группы.

Поэтому понятие «репрезентативность» носит относительный характер. Обычно для получения репрезентативной выборки предварительно отбраковывают нехарактерные элементы из общей совокупности. Обязательным условием получения репрезентативной выборки является **выравнивание условий проведения эксперимента**. Например, при постановке полевых опытов необходимо, чтобы на участке, отведенном под опыт, все варианты располагались после одного и того же предшественника с возможно равномерным плодородием на всей площади участка с выровненной засоренностью и других условий, определяемых конкретным опытом. Все эти вопросы требуют тщательного изучения и широкого предварительного научного обсуждения. Существуют и другие классификации способов отбора данных. В целом следует отметить, что отбор данных для формирования выборки является одним из ответственных решений при планировании эксперимента.

11.2 Выбор числа наблюдений выборки

Необходимое число наблюдений в каждом опыте (повторность) определяется, в основном, изменчивостью исследуемого показателя и его, допустимой экспериментатором, ошибкой. От числа наблюдений n зависит точность определения статистических показателей выборки. Для определения необходимого числа наблюдений достаточно знать стандартное отклонение S и максимально допустимую ошибку средней арифметической по выборке $S_{\bar{X}}$ из прежних исследований или специально проведенных для этой цели испытаний.

$$n = \frac{S^2}{S_{\bar{X}}}, \tag{11.1}$$

Из выражения (11.1) следует заметить, что точность найденных результатов можно повысить увеличивая объем выборки.

11.3 Последовательный анализ

Одним из основных недостатков рекомендуемых методов контроля качества выполнения технологических операций является большая их трудоемкость. Например, контроль таких показателей качества работы, как потери урожая при уборке, повреждение растений на междурядной обработке пропашных культур, равномерность глубины посева, засоренность зерна и другие требуют больших затрат времени. Именно это обстоятельство является одним из главных препятствий широкого применения массовых обследований исследуемых объектов с целью проведения статистических оценок качества протекания технологических процессов. Статистический метод контроля качества протекания технологических процессов представляют собой достаточно сложную процедуру, и требует глубокого понимания всех его аспектов.

В существующих методиках по контролю качества выполнения работ в сельском хозяйстве применяется однократная выборка, т. е. выводы о степени соответствия контролируемого показателя агротехническим допускам применяются только после проведения определенного, заранее фиксированного количества измерений. При этом множество W, составляющее все возможные выборки разбивается на две области: W_O – область принятия гипотезы H_O, т. е. гипотезы, допускающей, что контролируемый показатель соответствует агродопуску и W_1 – область отклонения гипотезы H_O и принятия гипотезы H_1 допускающей, что контролируемый показатель не соответствует агротехническим требованиям. Однако, как бы сильно не отличались измеряемые значения показателей от величины допуска, методика не позволяет уменьшать количество измерений. В связи с этим возникла необходимость разработать такой план контроля, который обеспечил бы при минимальном объеме выборки надежную информацию о соответствии или не соответствии величины контролируемого показателя агротехническим требованиям с целью сокращения количества измерений при контроле качества выполнения сельскохозяйственных операций и повышения надежности выводов о характере протекания технологического процесса. Сокращение количества измерений, т. е. объема выборки, особенно важно при повышенной трудоемкости проведения каждого измерения.

Метод последовательного статистического анализа, основанный на критерии отношения вероятностей, предложенный А. Вальдом, позволяет в среднем на 50 % сократить необходимое количество наблюдений по сравнению с методикой, основанной на однократной выборке. Этот метод предусматривает после каждого измерения принимать одно из следующих трех заключений:

- гипотеза H_O принимается, т. е. контролируемый показатель находится в пределах допуска;
- гипотеза H_O отклоняется и принимается гипотеза H_1 технологический процесс нарушен, контролируемый показатель не соответствует агродопуску;
- измерения надо продолжить.

Если принять, что истинное значение контролируемого показателя μ, а μ_0 и μ_1 – предельные крайние значения, определяемые агротехническим допуском, то после каждого измерения принимается одно из заключений:

- $\mu \leq \mu_0$ – верна гипотеза H_O;
- $\mu \leq \mu_1$ – верна альтернативная гипотеза H_1 и

$\mu_0 \leq \mu \leq \mu_1$ – необходимо продолжить измерения

Последовательный критерий отношения вероятностей предусматривает определять после каждого последующего измерения вероятность $P_{o.n.}$ того, что измеренные случайные значения контролируемого показателя $x_1, x_2, x_3, ..., x_{n-1}, x_n$ взяты из подмножества W_O с параметрами $(x, ...\mu_0)$ и вероятность $P_{1,n}$ того, что они принадлежат подмножеству W_1 с параметрами $(x, ...\mu_1)$. Вероятности $P_{o.n.}$ и $P_{1,n}$ получения выборок, принадлежащих соответственно подмножествам W_O и W_1 выражаются через совместные распределения вероятностей выборки $x_1, x_2, x_3, ..., x_{n-1}, x_n$ и определяются выражениями:

$$P_{o.n.} = f(x_1\mu_0)f(x_2\mu_0)...f(x_{n-1}\mu_0)f(x_n\mu_0), \tag{11.2}$$

если справедлива гипотеза H_O: $\mu \leq \mu_1$, выборка принадлежит подмножеству W_O и :

$$P_{1.n.} = f(x_1\mu_1)f(x_2\mu_1)\ldots f(x_{n-1}\mu_1)f(x_n\mu_1), \tag{11.3}$$

если отклоняется гипотеза H_O и принимается альтернативная гипотеза H_1: $\mu \geq \mu_1$ выборка принадлежит подмножеству W_1.

При каждом испытании отношение вероятностей $P_{o.n.}$ и $P_{1,n}$ и результат сравнивается с заранее установленными соответствующим образом постоянными величинами A и B ($A > B$). Если при этом:

$$\frac{P_{1.n}}{P_{o.n}} \geq A. \tag{11.4}$$

То контроль заканчивается принятием гипотезы H_1. Выборка принадлежит подмножеству W_1 технологический процесс требует подналадки, контролируемый показатель не отвечает агротехническим требованиям. Если:

$$\frac{P_{1.n}}{P_{o.n}} \leq B, \tag{11.5}$$

то процесс заканчивается принятием гипотезы H_O. А это значит, что выборка относится к подмножеству W_O и технологический процесс протекает нормально, а контролируемый показатель находится в пределах допуска. В случае:

$$B < \frac{P_{1.n}}{P_{o.n}} < A, \tag{11.6}$$

необходимо сделать следующее измерение.

Если исследуемая выборка подчиняется нормальному закону распределения, то неравенство 11.5 выразится так:

$$B < \frac{S^n \dfrac{1}{\sqrt{(2\pi)^n}} e^{-\frac{\sum_{i=1}^{n}(x_1-\mu_1)^2}{2S^2}}}{\dfrac{1}{S^n} e^{\frac{\sum_{i=1}^{n}(x_1-\mu_0)^2}{2S^2}}} < A, \qquad (11.7)$$

из которого после ряда несложных преобразований получится неравенство

$$2,3\frac{S^2}{\mu_1-\mu_0}\lg\frac{\beta}{1-\alpha}+n\ddot{\mu} < \sum_{i=1}^{n} x_i < 2,3\frac{S^2}{\mu_1-\mu_0}\lg\frac{\beta}{\alpha}+n\ddot{\mu}, \quad (11.8)$$

где S^2 – дисперсия контролируемой величины.

Значение коэффициентов A и B для практических целей с достаточной точностью можно определить по следующим зависимостям:

$$A = \frac{1-\beta}{\alpha}, \qquad B = \frac{\beta}{1-\alpha}, \qquad (11.9)$$

где α – допустимая вероятность принятия гипотезы H_1 в то время, как истинна гипотеза H_O и называется ошибкой первого рода;

β – допустимая вероятность принятия гипотезы H_O в то время, как истинна гипотеза H_1 и называется ошибкой второго рода.

Для примера рассмотрим контроль качества уборки зерновых колосовых культур, представляющий собой довольно трудоемкую операцию. По существующим методикам фактические потери зерна определяются в пяти-шести местах по диагонали контролируемого участка. Для этого накладывают рамку 0,5×1м^2, в пределах которой подсчитывают потери свободным зерном и в колосьях. Количество повторностей не зависит от условий уборки, характера колебаний потерь и их средней величины. Для подсчета зерен в каждой рамке уходит в среднем даже у самого опытного контролера 10-20 мин. Надежность выводов о степени соответствия контролируемого показателя агротехническим требованиям низкая и зависит от параметров потерь. Для получения необходимых статистических характеристик при проведении контроля на основе последовательной оценки проведены массовые измерения общих потерь зерна за комбайном при различных условиях уборки. Установлено, что между различными выборками имеется разница как по величине средних потерь, так и их дисперсиям (таблица 11.1).

Таблица 11.1 – Статистическая характеристика общих потерь зерноуборочных комбайнов

№ выборки	Число степеней свободы	Величина средних потерь, $\bar{x}, г/м^2$	Дисперсия, $S_0^2, г^2/м^4$	Дисперсия преобразованных выборок
1	14	4,9	228,0	0,119
2	17	1,7	15,8	0,119
3	18	1,9	40,8	0,109
4	24	4,1	67,4	0,069

Статистическая проверка по критерию Бартлетта показала отсутствие однородности дисперсий (при уровне значимости $P = 0,05$ $\chi^2_{факт.} = 56,5 > \chi^2_{табл} = 7,8$).

Для получения нормально распределенной выборки и однородности дисперсий было проведено преобразование при помощи функции:

$$x_i' = \lg x_i. \tag{11.10}$$

Полученная после преобразования однородность дисперсий ($\chi^2_{факт.} = 0,85 < \chi^2_{табл} = 7,8$) дает право на обобщение и получение средневзвешенной дисперсии (таблица 11.1).

$$\overline{S}^2 = \frac{\sum_{i=1}^{n} f_i S_i^2}{\sum_{i=1}^{n} f_i} = 0,1. \tag{11.11}$$

Данная дисперсия характеризует общие потери за комбайном в более широком диапазоне изменения условий работы.

Ответственным этапом последовательного контроля является правильный выбор коэффициентов α и β, от значения которых зависит риск допустить некачественную работу и повышенную производительность или высококачественную работу и повышенную производительность или высококачественную работу и большие простои за счет частых остановок для регулировок или устранения других недостатков. Также важное значение имеет и обоснованный выбор агротехнических допусков. Вопросы, связанные с выбором коэффициентов α и β, а также агротехнических допусков в общем плане и для каждого вида работы и даже показателя носит самостоятельный характер и составляет довольно серьезную и сложную задачу. В рассматриваемом случае приемочные и браковочные числа рассчитаны по выражению (11.9) при нескольких значениях α и β и при агротехническом допуске

$\mu_0 = 1{,}4\,\%$ и $\mu_1 = 2{,}2\,\%$, которые затем впоследствии были преобразованы с учетом средней урожайности 40 ц/га по зависимости (11.10). Новые значения μ_0 и μ_1 были получены равными соответственно 0,7482 и 0,9445. Анализ полученных результатов показывает, что по всем четырем выборкам качество работы комбайна по общим потерям зерна удовлетворяет агротехническим требованиям (таблица 11.2). Однако, количество измерений, понадобившееся для принятия окончательного решения, по всем выборкам разное и зависит от величины средних потерь и значений коэффициентов α и β.

Для оценки эффективности метода последовательного статистического контроля для рассматриваемого случая было определено необходимое число измерений при тех же статистических параметрах и для метода, основанного на однократной выборке:

$$\eta = \frac{t^2}{\Delta^2} S_i^2, \tag{11.12}$$

где t – критерий Стьюдента;

Δ – допустимая абсолютная величина отклонения от среднего значения агродопуска.

$$\Delta = \frac{\mu_1 - \mu_0}{2}. \tag{11.13}$$

Расчеты по приведенному примеру показывают, что необходимое количество измерений при применении метода последовательной оценки снизилось до 12 раз при одинаковой для обоих методов надежности получаемых выводов о поведении контролируемого показателя (таблица 11.3).

Таблица11.2 – Таблица контроля качества

№ изме-рения	Приемочное число L_1			Накопленная сумма $\sum_{i=1}^{n} x_i$				Браковочное число L_2		
	α=β= 0,05	α=β=0,1	α=β=0,2	выборка				α=β= 0,05	α=β=0,1	α=β=0,2
				1	2	3	4			
1	- 0.717	-0.3088	0.1175	0.5441	0.1461	0.0000	0.5563	2.3943	2.0014	1.5751
2	0.1346	0.5375	0.9638	1.0492	0.2253	0.1461	1.3824	3.2406	2.8477	2.4214
3	0.9809	1.3838	1.8101	1.9005	0.4806	0.5440	2.0984	4.0869	3.6940	3.2677
4	1.8272	2.2301	2.6564	2.5633	0.7110	0.5652	2.7612	4.9332	4.5403	4.1140
5	2.6735	3.0764	3.5027	3.2876	-	-	3.4855	5.7675	5.3866	4.9603
6	3.5198	3.9227	4.3490	3.9866	-	-	4.1845	6.6138	6.2329	5.8066
7	4.3661	4.7690	5.1953	4.5051	-	-	4.7030	7.4601	7.0792	6.6529
8	5.2124	5.6153	6.0416	5.2211	-	-	5.4190	8.3064	7.9255	7.3992
9	6.0587	6.4616	6.8879	5.5221	-	-	5.7200	9.1527	8.7718	8.3455
10	6.9050	7.3079	7.7347	5.8231	-	-	6.0210	9.9990	9.6181	9.1918

Таблица 11.3 – Количество измерений при применении разных методов контроля

№ выборки	Метод контроля					
	На основе последовательного отношения вероятностей			На основе однократной выборки		
	$\alpha=\beta=0{,}05$	$\alpha=\beta=0{,}1$	$\alpha=\beta=0{,}2$	$\alpha=\beta=0{,}05$	$\alpha=\beta=0{,}1$	$\alpha=\beta=0{,}2$
1	9	7	4	47	33	20
2	3	2	2	47	33	20
3	3	1	1	43	30	18
5	9	7	5	27	19	12

Кроме описанного табличного метода последовательной оценки широкое распространение получает и ее графическое изображение (рисунок 11.1).

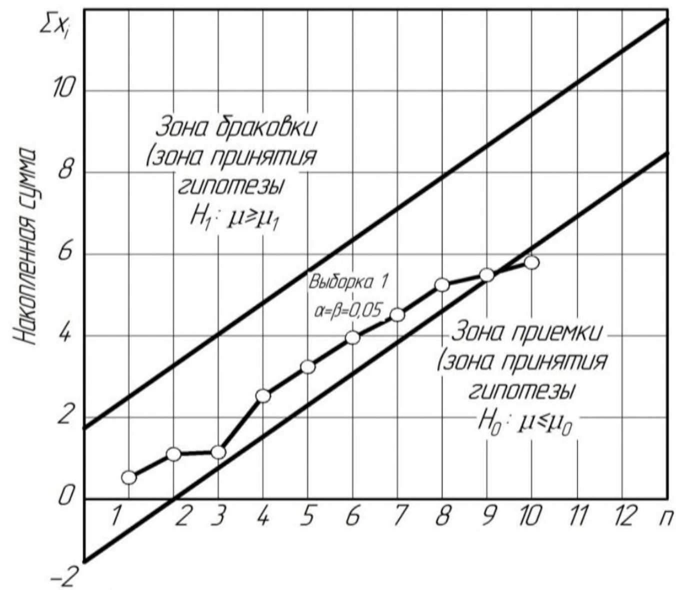


Рисунок 11.1 – Графическое изображение последовательного статистического контроля общих потерь зерна

Для этого строят две параллельные прямые L_1 и L_2, определяемые уравнениями:

$$L_1 = 2{,}3\frac{S^2}{\mu_1 - \mu_0}\lg\frac{\beta}{1-\alpha} + n\bar{\mu}, \tag{11.14}$$

$$L_2 = 2{,}3\frac{S^2}{\mu_1 - \mu_0}\lg\frac{1-\beta}{\alpha} + n\bar{\mu}. \tag{11.15}$$

Рассмотрим другой пример. На хлебоприемном пункте принимают зерно засоренностью не свыше 6 %. Зерно поступает партиями, из которых берутся пробы на анализ. В связи с ошибками анализа и неоднородностью партий возникает средняя квадратичная ошибка результатов, известная по большому числу предыдущих анализов и равна σ=0,24 %. Чтобы гарантировать доброкачественность, требуется браковать зерно, засоренное больше, чем на 6 %, с вероятностью, не меньшей 0,95. Нежелательно впадать и в другую крайность – браковать хорошее зерно. Поэтому ставится еще одно требование – зерно, засоренное не больше чем на 5,8 принимать с вероятностью 0,90.

Покажем, как для контроля качества зерна применять последовательный анализ. В поставленных условиях $\alpha_1 = 5{,}8$, $\alpha_2 = 6$. Вероятности ошибок допускаются:

$$\alpha_1 = 1 - 0{,}9 = 0{,}1;$$
$$\beta = 1 - 0{,}95 = 0{,}05.$$

Непосредственно вычисляем:

$$A_1 = 2{,}3\frac{0{,}24^2}{0{,}2}\lg\frac{0{,}05}{0{,}90} = -8{,}31;$$

$$A_1 = 2,3\frac{0,24^2}{0,2}\lg\frac{0,95}{0,90} = 13,1;$$
$$b = \frac{5,8+6}{2} = 5,9.$$

Область продолжения анализов будет ограничена прямыми, определяемыми функциями:

$$y = -8,31 + 5,9n;$$
$$y = 13,1 + 5,9n.$$

Графическое представление этих прямых представлено на рисунке 11.2.

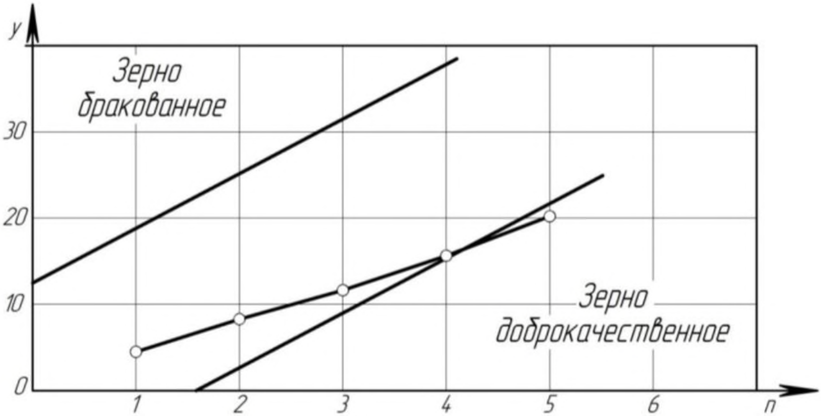


Рисунок 11.2 – Графическое изображение последовательного статистического контроля засоренности зерна

Анализ первой пробы из партии зерна дал результат 4,2 %. Ему соответствует на рисунке 11.2 точка M_1 с абсциссой $n = 1$ и ординатой $y = 4,2$ Точка M_1 находится в «полосе продолжения испытаний», поэтому берем вторую пробу. Ее анализ дал результат 3,9 %, что соответствует точке M_2 с абсциссой $n = 2$ и ординатой $y = 4,2 + 3,9 = 8,1$. Испытания снова нужно продолжать. Третий анализ дал значение 3,6 %; ему соответствует точка M_3 с абсциссой $n=3$ и ординатой $y=8{,}1+3{,}9+3{,}6=11{,}7$. Четвертому анализу с результатом 4,0 % соответствует точка M_4 (n=4, $y=15{,}7$). Наконец, пятая точка M_5, соответствующая еще одному анализу с результатом 4,4 %, получает координаты $n=5$, $y=15{,}7+4{,}4=20{,}1$ и выходит за пределы «полосы продолжения испытаний» вниз. Партию зерна нужно считать доброкачественной.

Может показаться, что доброкачественность зерна видна уже из первого анализа. К сожалению, такой вывод поспешен, так как из-за высокого значения σ мы не можем гарантировать заданные вероятности ошибок α и β.

Для решения поставленной задачи с помощью метода последовательного анализа нам понадобилось пять наблюдений. Если же мы захотим определить число наблюдений n заранее по формулам для фиксированного числа наблюдений, то мы получим:

$$N=(u_{0,90}+u_{0,95})^2\left(\frac{\sigma}{\alpha_2-\alpha_1}\right)^2=(1{,}28+1{,}65)^2\left(\frac{0{,}24}{0{,}2}\right)\approx 13. \quad (11.16)$$

Измерения продолжают до тех пор, пока накопленная сумма $\sum_{i=1}^{n} x_i$ попадает в область, заключенную между прямыми L_1 и L_2 и заканчиваются, если накопленная сумма оказывается в области приемки или браковки. Последовательный анализ нормально распределенной случайной величины наиболее удобно и показательно графическим методом. В качестве еще одного примера можно привести определение зачетного веса и качества зерна при его сдаче на хранение на элеватор по его засоренности и биохимическому составу. Преимущества последовательного анализа неоспоримы, особенно, в тех случаях, когда трудоемкость каждого измерения повышена и требует больших финансовых и материальных затрат.

Выводы

1. Применение последовательного критерия отношения вероятностей при контроле качества выполнения сельскохозяйственных работ позволит по сравнению с общепринятой в настоящее время методикой контроля на основе однократной выборки существенно сократить необходимое количество измерений при одинаковой надежности получаемых выводов о соответствии контролируемого показателя агротехническому допуску.

2. Для получения максимального эффекта от применения метода последовательного статистического анализа рекомендуется его применение для контроля качества продукции или выполнения технологических процессов, характеризующихся высокой трудоемкостью и затратами.

12 ПОСЛЕДОВАТЕЛЬНЫЙ СИМПЛЕКСНЫЙ МЕТОД ОПТИМИЗАЦИИ И НАСТРОЙКИ СЛОЖНЫХ СЕЛЬСКОХОЗЯЙСТВЕННЫХ АГРЕГАТОВ

На технологические, технические и экономические показатели машинно-тракторных агрегатов оказывают влияние большое число факторов. Так, производительность зерноуборочных комбайнов, потери, дробление и засоренность зерна зависят от вида и сорта культуры, влажности зерна и соломы, отношения массы зерна к массе соломы, полеглости засоренности и других факторов, характеризующих условия работы.

Обследование почвообрабатывающих и посевных агрегатов, зерноуборочных комбайнов и других машин показывает, что низкое качество выполнения технологического процесса объясняется несоответствием параметров настройки и регулировки условиям работы. Во многих случаях это несоответствие можно объяснить отсутствием простого и обоснованного направления и изменения настроечных параметров.

Параметры настройки и трудоемкость регулировок при одних и тех же условиях эксплуатации нередко бывают существенно различными. Это приводит в хозяйственных условиях к снижению производительности агрегатов и качества выполняемой работы, а при испытании новой техники – к несравнимости полученных результатов.

Для оптимизации параметров сельскохозяйственных агрегатов и их технологической настройки, как показывает практика, наиболее эффективно использование последовательного симплекс – метода. Последовательный симплекс – метод (ПСМ) является безградиентным методом поиска оптимального сочетания всех параметров. Для этого метода не требуется даже вычисления составляющих градиента ни по одному из всех параметров. Для каждого последующего шага требуется всего лишь один опыт. Основной особенностью ПСМ является совмещение исследования поверхности отклика с передвижением в, определяемом в каждом опыте, направлении к точке оптимума. Симплекс в математическом смысле – это правильный выпуклый многоугольник с $k+1$ вершинами в k-мерном факторном пространстве. При двухфакторном объекте (k = 2) симплексом будет равносторонний треугольник c; при k = 3 –тетраэдр.

Применение правильных симплексов упрощает расчет новых координат симплекса после каждого испытания. Применение симплекс – метода возможно и в случаях, когда отклики оценивают не только по результатам, но и по качественной оценке при внешнем осмотре. Отсюда следует, что симплекс – метод можно применять и при многокритериальной оценке (например, недомолота и дробления при настройке молотильного аппарата). Так как выбор направления движения к оптимуму определяется лишь по результатам сравнения откликов в вершинах симплекса, т. е. основывается на качественной информации, не предъявляется жестких требований к точности оценки значений параметров. Все это указывает на возможность и целесообразность применения этого метода при оптимизации технологических процессов не только в лабораториях, но и в полевых условиях.

Последовательное симплексное планирование необходимо начинать с анализа структуры и свойств факторов, действующих на изучаемый объект, способов регулирования их на заданный уровень, а при наличии дополнительных источников – степени их влияния на выходные показатели объекта. Затем выбирают план симплекса, уровень и шаг варьирования факторов, после чего определяют для каждого опыта координаты вершин симплекса в кодированных и натуральных величинах. Процесс начинается с постановки опытов во всех вершинах начального комплекса. Результаты выходных показателей ранжируются, и выявляется наихудший из них. Вершину симплекса, которой соответствует этот результат, отбрасывают и ставят опыт в новой точке, которая совместно с оставшимися вершинами составляет новый симплекс. При этом направление движения к оптимуму совпадает с направлением от наихудшей вершины через середину противолежащей грани.

Координаты новой вершины в кодированных или натуральных значениях определяют по формуле.

$$x_{i,k+2} = 2k^{-1}\frac{1}{n}\sum_{i=1}^{n} x_{in} - x_i^*, \qquad (12.1)$$

где $x_{i,k+2}$ – координата новой вершины по i-му фактору;

k – число факторов;

x_i^* – координата i-го фактора отбрасываемой вершины;

x_{in} – среднее значение координат всех точек симплекса кроме отбрасываемой точки.

В последующем эту процедуру повторяют. Вращение симплекса вокруг какой-нибудь вершины показывает окончание процесса оптимизации. Параметры, соответствующие этой вершине – оптимальные. Если в процессе эксперимента в двух вершинах симплекса окажется одинаковое минимальное значение выхода, то решение о дальнейшем движении принимается случайным образом. Если значение выхода во вновь определенной вершине снова окажется минимальным, то происходит возврат к исходному симплексу и отбрасывание вершины со следующим по порядку минимальности значением выхода. На любом шаге оптимизации по симплексному методу исследования можно включить в программу новый фактор и образовать (k + 1)-мерный симплекс. После достижения точки оптимума по первому симплексу можно сократить размеры симплекса, а при большой ошибке эксперимента дублировать опыты.

В целях снижения трудоемкости и упрощения процедуры настройки сложного сельскохозяйственного агрегата весьма желательно разделить его на отдельные сборочные единицы с практически независимыми друг от друга параметрами настройки. Если же это не удается, то необходимо рассматривать весь агрегат как единый настроечный с взаимозависимыми параметрами.

При существующем методе настройки молотильного аппарата зерноуборочного комбайна в зависимости от убираемой культуры и ее состояния устанавливают возможный диапазон изменения частоты вращения барабана, а также в определенном соотношении зазоры между барабаном и подбарабаньем на входе и выходе. Затем проводят корректировку этих параметров последовательным сравнением качественных показателей (недомолота и дробления зерна) при соответствующих значениях регулировки и выбирают наилучший из них. Однако такой метод корректировки недостаточно логичен, трудоемкость его в большой степени зависит от опыта наладчика, а окончательные параметры настройки не всегда действительно оптимальные. Для настройки молотильного аппарата однобарабанного зерноуборочного комбайна СК-5 «Нива» методом последовательного симплексного планирования был принят план (рисунок12.1.), включающий два фактора: $\tilde{X}_1$ – зазор между барабаном и подбарабаньем на входе (при этом зазор на выходе был установлен 20 мм), $\tilde{X}_2$ – частота вращения барабана.

С учетом технологических соображений, чувствительности молотильного аппарата и средств измерения выходных показателей выбираются нулевые уровни регулируемых параметров и шаг их варьирования.

$$\tilde{X}_1 = 3\,\text{мм};\ \Delta\tilde{X}_1 = 1\,\text{мм};\ \tilde{X}_2 = 989{,}4\,\text{мин}^{-1};\ \Delta\tilde{X}_2 = 14{,}5\,\text{мин}^{-1}$$

Исходя из условий работы комбайна, принят исходный симплекс с вершинами 5, 11, 12. Дальнейшая схема движения симплекса и величины выходных показателей показаны на рисунке 12.1 и в таблице 12.1. При комплексной оценке качества работы предпочтение отдавали степени вымолота. При этом руководствовались агротехническими допусками по недомолоту – 1 % (3-6 зерен в 100 колосьях, взятых из обмолоченной соломы) и по дроблению – 2 % (3-4 половинки в пробе из 100 зерен).

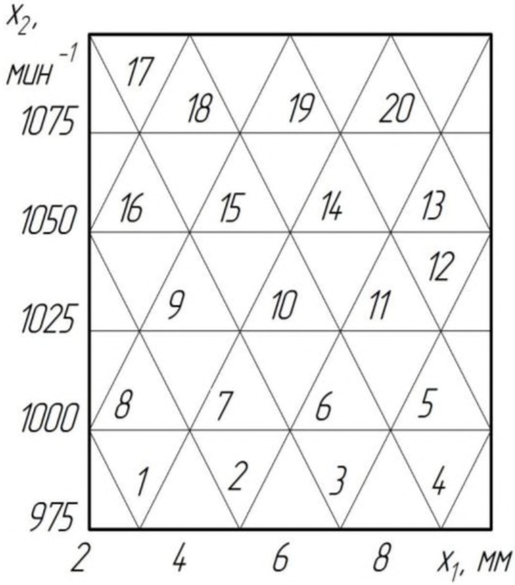


Рисунок 12.1 – Схема движения симплекса

По результатам выполненных исследований оптимальные параметры настройки молотильного аппарата в конкретных условиях составляют: зазоры на входе – 20 мм, на выходе – 4 мм, частота вращения барабана – 1050 мин$^{-1}$.

Существующие методы технологической настройки сельскохозяйственных агрегатов базируются на довольно сложных логических основах при выборе направления и величины изменения настроечных параметров для достижения их оптимальных значений и требуют высокой квалификации наладчика.

Таблица 12.1 – План настройки молотильного аппарата

Симплекс	Вершина симплекса	$\tilde{X}_1$, мм	$\tilde{X}_2$, мин$^{-1}$	Потери зерна от недомолота в 100 колосьях		Дробление в пробе	
				шт.	%	шт	%
1	5	8	1000	50	4.5	0	0,00
	11	7	1025	24	2.2	1	0,45
	12	9	1025	46	4.1	0	0,00
2	11	7	1025	24	2.2	1	0,45
	12	9	1025	46	4.1	0	0,00
	13	8	1050	34	3.1	0	0,00
3	11	7	1025	24	2.2	1	0,45
	13	8	1050	34	3.1	0	0,00
	14	6	1050	16	1.4	2	0,90
4	10	5	1025	10	0.9	2	0,90
	11	7	1025	24	2.2	1	0,45
	14	6	1050	16	1.4	2	0,90
5	10	5	1025	10	0.9	2	0,90
	14	6	1050	16	1.4	2	0,90
	15	4	1050	4	0.4	3	1,40
6	9	3	1025	2	0.2	6	2,80
	10	5	1025	10	0.9	2	0,90
	15	4	1050	4	0.4	3	1,40
7	9	3	1025	2	0.2	6	2,80
	15	4	1050	4	0.4	3	1,40
	16	2	1050	0	0.0	8	3,60
8	15	4	1050	4	0.4	3	1,40
	16	2	1050	0	0.0	8	3,60
	17	6	1025	0	0.0	7	3,30
9	15	4	1050	4	0.4	3	1,40
	17	3	1075	0	0.0	7	3,30
	18	5	1075	8	0.7	6	2,80
10	14	6	1050	16	1.4	2	0,90
	15	4	1050	4	0.4	3	1,40
	18	5	1075	8	0.7	6	2,80

Применение последовательного симплекс метода при технологической настройке сельскохозяйственных агрегатов повышает уверенность в оптимальности выбранных параметров, сокращает трудоемкость процесса настройки, не требует высокой квалификации от исполнителей. Простота процедуры последовательного симплекс – метода позволяет использовать его не только при испытаниях сложных по технологической настройке сельскохозяйственных агрегатов инженерами – испытателями, но и в производственных полевых условиях механизаторами и контролерами.

Выводы

1. Существующие методы технологической настройки сельскохозяйственных агрегатов базируются на довольно сложных логических основах при выборе направления и величины изменения настроечных параметров для достижения их оптимальных значений и требуют весьма высокой квалификации наладчика.

2. Применение последовательного симплекс – метода при технологической настройке сельскохозяйственных агрегатов повышает уверенность в оптимальности выбранных параметров, сокращает трудоемкость процесса настройки, не требует высокой квалификации от исполнителей.

3. Простота процедуры последовательного симплекс – метода позволяет использовать его не только при испытаниях сложных по технологической настройке сельскохозяйственных агрегатов инженерами – испытателями, но и в производственных условиях механизаторами и контролерами.

Контрольные вопросы

1. Что такое репрезентативный отбор.
2. Сущность последовательного анализа.
3. Понятие симплексного метода оптимизации.

СПИСОК ЛИТЕРАТУРЫ

1. Адлер Ю. П. Планирование эксперимента при поиске оптимальных условий [Текст] / Ю. П. Адлер, Е. И. Маркова, Ю. В. Грановский. – М. : Наука, 1976. – 280 с.
2. Болдин А. П. Основы научных исследований [Текст] / А. П. Болдин, В. А. Максимов. – М. : Академия, 2012. – 336 с.
3. Валге А. М. Обработка экспериментальных данных и моделирование динамических систем при проведении исследований по механизации сельскохозяйственного производства [Текст] / А. М. Валге. – СПб. : СЗНИИМЭСХ, 2002. – 176 с.
4. Вентцель Е. С. Теория вероятностей [Текст] / Е. С. Вентцель. – М. : Изд-во Физико-математической литературы, 1968. – 464 с.
5. Вирченко Н. А. Графики функций [Текст] : справочник / Н. А. Вирченко, И. И. Ляшко, К. И. Швецов. – Киев : Наукова думка, 1979. – 320 с.
6. Вознесенский В. А. Принятие решений по статистическим моделям [Текст] / В. А. Вознесенский, А. Ф. Ковальчук. – М. : Статистика, 1978. – 192 с.
7. Доспехов Б. А. Методика полевого опыта [Текст] / Б. А. Доспехов. – М. : Колос,1979. – 416 с.
8. Зедгинидзе И. Г. Планирование эксперимента для исследования многокомпонентных систем [Текст] / И. Г. Зедгинидзе. – М. : Наука, 1976. – 383 с.
9. Кириченко А. К. Оценка равномерности глубины обработки почвы комбинированным агрегатом [Текст] / А. К. Кириченко, К. А. Сохт // Механизация и электрификация социалистического сельского хозяйства. – 1983. – № 7. – С. 55.
10. Кириченко А. К. Оценка качества обработки почвы сферическими дисками с индивидуальной подвеской [Текст] / А. К. Кириченко, К. А. Сохт // Механизация работ в производстве зерна и селекционном процессе : сб. науч. тр. КНИИСХ. – Краснодар, 1985. – № 18. – С. 18–24.

11. Комаров Б. А. Оценка качества работ в зерновом хозяйстве [Текст] / Б. А. Комаров, А. Г. Рыбалко. – М. : Россельхозиздат, 1977. – 64 с.
12. Основы научных исследований [Текст] : сборник заданий / В. С. Кравченко, Е. И. Трубилин, В. С. Курасов, В. В. Куцеев, Е. В. Труфляк. – Краснодар : КубГАУ, 2011. – 105 с.
13. Основы научных исследований [Текст] : учебное пособие. / В. С. Кравченко, Е. И. Трубилин, В. С. Курасов, В. В. Куцеев, Е. В. Труфляк. – Краснодар : КубГАУ, 2005. – 136 с.
14. Круг Г. К. Теоретические основы планирования экспериментальных исследований [Текст] / Г. К. Круг. – М. : Изд-во МЭИ, 1973. – 180 с.
15. Литтл Т. Сельскохозяйственное опытное дело. Планирование и анализ [Текст] / Т. Литтл, Ф. Хиллз. – М. : Колос, 1981. – 320 с.
16. Маркова Е. В. Рандомизация и статистический вывод [Текст] / Е. В. Маркова, А. А. Маслак. – М. : Финансы и статистика, 1986. – 208 с.
17. Митков А. Л. Статистические методы в сельхозмашиностроении [Текст] / А. Л. Митков, С. В. Карда–шевский. – М. : Машиностроение, 1978. – 360 с.
18. Михайлов В. И. Планирование экспериментов в судостроении [Текст] / В. И. Михайлов, К. М. Федосов. – Л. : Судостроение, 1978. – 160 с.
19. Налимов В. В. Статистические методы планирования экстремальных экспериментов [Текст] / В. В. Налимов, Н. А. Чернова. – М. : Наука, 1965. – 310 с.
20. Орманджи К. С. Операционная технология уборки колосовых культур [Текст] / К. С. Ормаджи, Е. А. Мащков, Н. Г. Тарасов. – М. : Россельхозиздат, 1977. – 102 с.
21. Орманджи К. С. Оценка качества механизированных работ в полеводстве [Текст] / К. С. Ормаджи. – М. : Россельхозиздат, 1977. – 74 с.

22. Перегудов В. Н. Планирование многофакторных полевых опытов с удобрениями и математическая обработка их результатов [Текст] / В. Н. Перегудов. – М. : Колос, 1978. – 184 с.
23. Пирятин В. Д. Обработка результатов экспериментальных измерений по способу наименьших квадратов [Текст] / В. Д. Пирятин. – Харьков : ХГУ, 1962. – 216 с.
24. Пугачёв А. Н. Контроль качества уборки зерновых культур. М. : Колос, 1980. – 254 с.
25. Пустыльник Е. И. Статистические методы анализа и обработки наблюдений [Текст] / Е. И. Пустыльник. – М. : Наука, 1968. – 288 с.
26. Смирнов Н. В. Курс теории вероятностей и математической статистик [Текст] / Н. В. Смирнов, И. В. Дунин-Барковский. – М. : Наука, 1969. – 512 с.
27. Снедекор Дж. У. Статистические методы в применении к исследованиям в сельском хозяйстве и биологии [Текст] / Дж. У. Снедекор. – М. : Изд-во. Сельскохозяйственной литературы, журналов и плакатов, 1962. – 504 с.
28. Сохт К. А. Машинные технологии возделывания зерновых культур [Текст] : дис. на соискание ученой степени доктора технических наук: 05.20.01 / Сохт Казбек Аюбович. – Краснодар : Изд-во КубГАУ, 2002. – 60 с.
29. Сохт К. А. Метод последовательного статистического контроля качества уборки зерновых культур [Текст] / К. А. Сохт, В. Д. Карпенко // Сборник научных трудов КНИИСХ. – 1979. – № 18. – С. 18–22.
30. Сохт К. А. Применение метода обобщенного показателя качества при выборе технологической схемы сельскохозяйственных машин [Текст] / К. А. Сохт, А. К. Кириченко // Сб. науч. тр. КНИИСХ. Механизация производства зерна в Краснодарском крае: – 1979. – № 18. – С. 108–113.

31. Сохт К. А. Метод последовательного статистического контроля качества выполнения сельскохозяйственных работ [Текст] / К. А. Сохт, В. Д. Карпенко // Механизация и электрификация социалистического сельского хозяйства. – 1979. – № 8. – С. 4–5.
32. Сохт К. А. Прогнозирование оптимальных параметров машинно – тракторных агрегатов [Текст] / К. А. Сохт, А. К. Кириченко // Механизация и электрификация социалистического сельского хозяйства. – 1980. – № 9. – С. 31–33.
33. Сохт К. А. Технологическая настройка агрегатов [Текст] / К. А. Сохт, В. Д. Карпенко // Механизация и электрификация социалистического сельского хозяйства. – 1981. – № 5. – С. 49–50.
34. Сохт К. А. Воспроизводимость дисперсии тягового сопротивления почвообрабатывающих агрегатов [Текст] / К. А. Сохт, А. К. Кириченко // Механизация и электрификация социалистического сельского хозяйства. – 1984. – № 2. – С. 61–62.
35. Файффер П. Повышение производительности, снижение потерь и сохранение качества при комбайновой уборке [Текст] / П. Файффер. – Кведлинбург / Лейпциг, 1978. – 114 с.
36. Юдин М. И. Планирование эксперимента и обработка его результатов [Текст] / М. И. Юдин. – Краснодар : Печатный двор Кубани, 2004. – 240 с.

Приложения

Приложение А

Таблица А1 – Значения функции Лапласа $\Phi_0(t)=\frac{1}{\sqrt{2\pi}}\int_0^t e^{-\frac{t^2}{2}}dt$

t	***Ф₀(t)***	***t***	***Ф₀(t)***	***t***	***Ф₀(t)***	***t***	***Ф₀(t)***
0,00	0,0000	**0,29**	0,1141	**0,58**	0,2190	**0,87**	0,3078
0,01	0,0040	**0,30**	0,1179	**0,59**	0,2224	**0,88**	0,3106
0,02	0,0080	**0,31**	0,1217	**0,60**	0,2257	**0,89**	0,3133
0,03	0,0120	**0,32**	0,1255	**0,61**	0,2291	**0,90**	0,3159
0,04	0,0160	**0,33**	0,1293	**0,62**	0,2324	**0,91**	0,3186
0,05	0,0199	**0,34**	0,1331	**0,63**	0,2357	**0,92**	0,3212
0,06	0,0239	**0,35**	0,1368	**0,64**	0,2389	**0,93**	0,3238
0,07	0,0279	**0,36**	0 1406	**0,65**	0,2422	**0,94**	0,3264
0,08	0,0319	**0,37**	0,1443	**0,66**	0,2454	**0,95**	0,3289
0,09	0,0359	**0,38**	0,1480	**0,67**	0,2486	**0,96**	0,3315
0,10	0,0398	**0,39**	0,1517	**0,68**	0,2517	**0,97**	0,3340
0,11	0,0438	**0,40**	0,1554	**0,69**	0,2549	**0,98**	0,3365
0,12	0,0478	**0,41**	0,1591	**0,70**	0,2580	**0,99**	0,3389
0,13	0,0517	**0,42**	0,1628	**0,71**	0,2611	**1,00**	0,3413
0,14	0,0557	**0,43**	0,1664	**0,72**	0,2642	**1,01**	0,3438
0,15	0,0596	**0,44**	0,1700	**0,73**	0,2673	**1,02**	0,3461
0,16	0,0636	**0,45**	0,1736	**0,74**	0,2703	**1,03**	0,3485
0,17	0,0675	**0,46**	0,1772	**0,75**	0,2734	**1,04**	0,3508
0,18	0,0714	**0,47**	0,1808	**0,76**	0,2764	**1,05**	0,3531
0,19	0,0753	**0,48**	0,1844	**0,77**	0,2794	**1,06**	0,3554
0,20	0,0793	**0,49**	0,1879	**0,78**	0,2823	**1,07**	0,3577
0,21	0,0832	**0,50**	0,1915	**0,79**	0,2852	**1,08**	0,3599
0,22	0,0871	**0,51**	0,1950	**0,80**	0,2881	**1,09**	0,3621
0,23	0,0910	**0,52**	0,1985	**0,81**	0,2910	**1,10**	0,3643
0,24	0,0948	**0,53**	0,2019	**0,82**	0,2939	**1,11**	0,3665
0,25	0,0987	**0,54**	0,2054	**0,83**	0,2967	**1,12**	0,3686
0,26	0,1026	**0,55**	0,2088	**0,84**	0,2995	**1,13**	0,3708
0,27	0,1064	**0,56**	0,2123	**0,85**	0,3023	**1,14**	0,3729
0,28	0,1103	**0,57**	0,2157	**0,86**	0,3051	**1,15**	0,3749

Продолжение таблицы А1

t	**$\Phi_0(t)$**	**t**	**$\Phi_0(t)$**	**t**	**$\Phi_0(t)$**	**t**	**$\Phi_0(t)$**
1,16	0,3770	**1,48**	0,4306	**1,80**	0,4641	**2,24**	0,4875
1,17	0,3790	**1,49**	0,4319	**1,81**	0,4649	**2,26**	0,4881
1,18	0,3810	**1,50**	0,4332	**1,82**	0,4656	**2,28**	0,4887
1,19	0,3830	**1,51**	0,4345	**1,83**	0,4664	**2,30**	0,4893
1,20	0,3849	**1,52**	0,4357	**1,84**	0,4671	**2,32**	0,4898
1,21	0,3869	**1,53**	0,4370	**1,85**	0,4678	**2,34**	0,4904
1,22	0,3883	**1,54**	0,4382	**1,86**	0,4686	**2,36**	0,4909
1,23	0,3907	**1,55**	0,4394	**1,87**	0,4693	**2,38**	0,4913
1,24	0,3925	**1,56**	0,4406	**1,88**	0,4699	**2,40**	0,4918
1,25	0,3944	**1,57**	0,4418	**1,89**	0,4706	**2,42**	0,4922
1,26	0,3962	**1,58**	0,4429	**1,90**	0,4713	**2,44**	0,4927
1,27	0,3980	**1,59**	0,4441	**1,91**	0,4719	**2,46**	0,4931
1,28	0,3997	**1,60**	0,4452	**1,92**	0,4726	**2,48**	0,4934
1,29	0,4015	**1,61**	0,4463	**1,93**	0,4732	**2,50**	0,4938
1,30	0,4032	**1,62**	0,4474	**1,94**	0,4738	**2,52**	0,4941
1,31	0,4049	**1,63**	0,4484	**1,95**	0,4744	**2,54**	0,4945
1,32	0,4066	**1,64**	0,4495	**1,96**	0,4750	**2,56**	0,4948
1,33	0,4082	**1,65**	0,4505	**1,97**	0,4756	**2,58**	0,4951
1,34	0,4099	**1,66**	0,4515	**1,98**	0,4761	**2,60**	0,4953
1,35	0,4115	**1,67**	0,4525	**1,99**	0,4767	**2,62**	0,4956
1,36	0,4131	**1,68**	0,4535	**2,00**	0,4772	**2,64**	0,4959
1,37	0,4147	**1,69**	0,4545	**2,02**	0,4783	**2,66**	0,4961
1,38	0,4162	**1,70**	0,4554	**2,04**	0,4793	**2,68**	0,4963
1,39	0,4177	**1,71**	0,4564	**2,06**	0,4803	**2,70**	0,4965
1,40	0,4192	**1,72**	0,4573	**2,08**	0,4812	**2,72**	0,4967
1,41	0,4207	**1,73**	0,4582	**2,10**	0,4821	**2,74**	0,4969
1,42	0,4222	**1,74**	0,4591	**2,12**	0,4830	**2,76**	0,4971
1,43	0,4236	**1,75**	0,4599	**2,14**	0,4838	**2,78**	0,4973
1,44	0,4251	**1,76**	0,4608	**2,16**	0,4846	**2,80**	0,4974
1,45	0,4265	**1,77**	0,4616	**2,18**	0,4854	**2,82**	0,4976
1,46	0,4279	**1,78**	0,4625	**2,20**	0,4661	**2,84**	0,4977

Продолжение таблицы А1

1,47	0,4292	**1,79**	0,4633	**2,22**	0,4868	**2,86**	0,4979
2,88	0,4980	**3,3**	0,49952	**3,9**	0,49995	**4,5**	0,5
2,90	0,4981	**3,35**	0,4996	**3,95**	0,49996	**4,55**	0,5
2,92	0,4982	**3,4**	0,49966	**4**	0,49997	**4,6**	0,5
2,94	0,4984	**3,45**	0,49972	**4,05**	0,49997	**4,65**	0,5
2,96	0,49846	**3,5**	0,49977	**4,1**	0,49998	**4,7**	0,5
2,98	0,49856	**3,55**	0,49981	**4,15**	0,49998	**4,75**	0,5
3,00	0,49865	**3,6**	0,49984	**4,2**	0,49999	**4,8**	0,5
3,05	0,49886	**3,65**	0,49987	**4,25**	0,49999	**4,85**	0,5
3,1	0,49903	**3,7**	0,49989	**4,3**	0,49999	**4,9**	0,5
3,15	0,49918	**3,75**	0,49991	**4,35**	0,49999	**4,95**	0,5
3,2	0,49931	**3,8**	0,49993	**4,4**	0,49999	**5**	0,5
3,25	0,49942	**3,85**	0,49994	**4,45**	0,5	-	-

Примичание – в случае отсутсвия необходимого аргумента в приведенных таблицах необходимо применить ***интерполяцию*** – нахождение приблеженного значения функции для заданного аргумента. Линейная интерполяция определяется по выражению:

$$y = y_1 + \frac{(x - x_1)(y_2 - y_1)}{x_2 - x_1},$$

где y, x – соответсвенно значения искомой функции и заданного аргумента;

y_1, x_1, y_2, x_2 – ближайшие к искомой функции соотвественно значения функции и аргумента (по данным таблицы).

Приложение Б

Таблица Б1 – Квантили распределения Стьюдента (*t*–критерий)

Число степеней свободы, f	α=0,9 P=0,1 (10%)	α=0,8 P=0,2 (20%)	α=0,7 P=0,3 (30%)	α=0,6 P=0,4 (40%)	α=0,5 P=0,5 (50%)	α=0,4 P=0,6 (60%)	α=0,3 P=0,7 (70%)	α=0,2 P=0,8 (80%)	α=0,1 P=0,9 (90%)	α=0,05 P=0,95 (95%)	α=0,02 P=0,98 (98%)	α=0,01 P=0,99 (99%)	α=0,005 P=0,995 (99,5%)
1	0.1584	0.3249	0.5095	0.7265	1.0000	1.3764	1.9626	3.0777	6.3138	12.7062	31.8205	63.657	127.32
2	0.1421	0.2887	0.4447	0.6172	0.8165	1.0607	1.3862	1.8856	2.9200	4.3027	6.9646	9.9248	14.089
3	0.1366	0.2767	0.4242	0.5844	0.7649	0.9785	1.2498	1.6377	2.3534	3.1824	4.5407	5.8409	7.4533
4	0.1338	0.2707	0.4142	0.5686	0.7407	0.9410	1.1896	1.5332	2.1318	2.7764	3.7469	4.6041	5.5976
5	0.1322	0.2672	0.4082	0.5594	0.7267	0.9195	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321	4.7733
6	0.1311	0.2648	0.4043	0.5534	0.7176	0.9057	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074	4.3168
7	0.1303	0.2632	0.4015	0.5491	0.7111	0.8960	1.1192	1.4149	1.8946	2.3646	2.9980	3.4995	4.0293
8	0.1297	0.2619	0.3995	0.5459	0.7064	0.8889	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554	3.8325
9	0.1293	0.2610	0.3979	0.5435	0.7027	0.8834	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498	3.6897
10	0.1289	0.2602	0.3966	0.5415	0.6998	0.8791	1.0931	1.3722	1.8125	2.2281	2.7638	3.1693	3.5814
11	0.1286	0.2596	0.3956	0.5399	0.6974	0.8755	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058	3.4966
12	0.1283	0.2590	0.3947	0.5386	0.6955	0.8726	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545	3.4284
13	0.1281	0.2586	0.3940	0.5375	0.6938	0.8702	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123	3.3725
14	0.1280	0.2582	0.3933	0.5366	0.6924	0.8681	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768	3.3257
15	0.1278	0.2579	0.3928	0.5357	0.6912	0.8662	1.0735	1.3406	1.7531	2.1314	2.6025	2.9467	3.286
16	0.1277	0.2576	0.3923	0.5350	0.6901	0.8647	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208	3.252
17	0.1276	0.2573	0.3919	0.5344	0.6892	0.8633	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982	3.2224

Продолжение таблицы Б1

Число степеней свободы, f	α=0,9 P=0,1 (10%)	α=0,8 P=0,2 (20%)	α=0,7 P=0,3 (30%)	α=0,6 P=0,4 (40%)	α=0,5 P=0,5 (50%)	α=0,4 P=0,6 (60%)	α=0,3 P=0,7 (70%)	α=0,2 P=0,8 (80%)	α=0,1 P=0,9 (90%)	α=0,05 P=0,95 (95%)	α=0,02 P=0,98 (98%)	α=0,01 P=0,99 (99%)	α=0,005 P=0,995 (99,5%)
18	0.1274	0.2571	0.3915	0.5338	0.6884	0.8620	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784	3.1966
19	0.1274	0.2569	0.3912	0.5333	0.6876	0.8610	1.0655	1.3277	1.7291	2.0930	2.5395	2.8609	3.1737
20	0.1273	0.2567	0.3909	0.5329	0.6870	0.8600	1.0640	1.3253	1.7247	2.0860	2.5280	2.8453	3.1534
21	0.1272	0.2566	0.3906	0.5325	0.6864	0.8591	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314	3.1352
22	0.1271	0.2564	0.3904	0.5321	0.6858	0.8583	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188	3.1188
23	0.1271	0.2563	0.3902	0.5317	0.6853	0.8575	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073	3.104
24	0.1270	0.2562	0.3900	0.5314	0.6848	0.8569	1.0593	1.3178	1.7109	2.0639	2.4922	2.7969	3.0905
25	0.1269	0.2561	0.3898	0.5312	0.6844	0.8562	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874	3.0782
26	0.1269	0.2560	0.3896	0.5309	0.6840	0.8557	1.0575	1.3150	1.7056	2.0555	2.4786	2.7787	3.0669
27	0.1268	0.2559	0.3894	0.5306	0.6837	0.8551	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707	3.0565
28	0.1268	0.2558	0.3893	0.5304	0.6834	0.8546	1.0560	1.3125	1.7011	2.0484	2.4671	2.7633	3.0469
29	0.1268	0.2557	0.3892	0.5302	0.6830	0.8542	1.0553	1.3114	1.6991	2.0452	2.4620	2.7564	3.038
30	0.1267	0.2556	0.3890	0.5300	0.6828	0.8538	1.0547	1.3104	1.6973	2.0423	2.4573	2.75	3.0298
31	0.1267	0.2555	0.3889	0.5298	0.6825	0.8534	1.0541	1.3095	1.6955	2.0395	2.4528	2.744	3.0221
32	0.1267	0.2555	0.3888	0.5297	0.6822	0.8530	1.0535	1.3086	1.6939	2.0369	2.4487	2.7385	3.0149
33	0.1266	0.2554	0.3887	0.5295	0.6820	0.8526	1.0530	1.3077	1.6924	2.0345	2.4448	2.7333	3.0082
34	0.1266	0.2553	0.3886	0.5294	0.6818	0.8523	1.0525	1.3070	1.6909	2.0322	2.4411	2.7284	3.002
35	0.1266	0.2553	0.3885	0.5292	0.6816	0.8520	1.0520	1.3062	1.6896	2.0301	2.4377	2.7238	2.996
36	0.1266	0.2552	0.3884	0.5291	0.6814	0.8517	1.0516	1.3055	1.6883	2.0281	2.4345	2.7195	2.9905

Продолжение таблицы Б1

Число степеней свободы, f	α=0,9 P=0,1 (10%)	α=0,8 P=0,2 (20%)	α=0,7 P=0,3 (30%)	α=0,6 P=0,4 (40%)	α=0,5 P=0,5 (50%)	α=0,4 P=0,6 (60%)	α=0,3 P=0,7 (70%)	α=0,2 P=0,8 (80%)	α=0,1 P=0,9 (90%)	α=0,05 P=0,95 (95%)	α=0,02 P=0,98 (98%)	α=0,01 P=0,99 (99%)	α=0,005 P=0,995 (99,5%)
37	0.1265	0.2552	0.3883	0.5289	0.6812	0.8514	1.0512	1.3049	1.6871	2.0262	2.4314	2.7154	2.9852
38	0.1265	0.2551	0.3882	0.5288	0.6810	0.8512	1.0508	1.3042	1.6860	2.0244	2.4286	2.7116	2.9803
39	0.1265	0.2551	0.3882	0.5287	0.6808	0.8509	1.0504	1.3036	1.6849	2.0227	2.4258	2.7079	2.9756
40	0.1265	0.2550	0.3881	0.5286	0.6807	0.8507	1.0500	1.3031	1.6839	2.0211	2.4233	2.7045	2.9712
41	0.1264	0.2550	0.3880	0.5285	0.6805	0.8505	1.0497	1.3025	1.6829	2.0195	2.4208	2.7012	2.967
42	0.1264	0.2550	0.3880	0.5284	0.6804	0.8503	1.0494	1.3020	1.6820	2.0181	2.4185	2.6981	2.963
43	0.1264	0.2549	0.3879	0.5283	0.6802	0.8501	1.0491	1.3016	1.6811	2.0167	2.4163	2.6951	2.9592
44	0.1264	0.2549	0.3878	0.5282	0.6801	0.8499	1.0488	1.3011	1.6802	2.0154	2.4141	2.6923	2.9555
45	0.1264	0.2549	0.3878	0.5281	0.6800	0.8497	1.0485	1.3006	1.6794	2.0141	2.4121	2.6896	2.9521
46	0.1264	0.2548	0.3877	0.5281	0.6799	0.8495	1.0483	1.3002	1.6787	2.0129	2.4102	2.687	2.9488
47	0.1263	0.2548	0.3877	0.5280	0.6797	0.8493	1.0480	1.2998	1.6779	2.0117	2.4083	2.6846	2.9456
48	0.1263	0.2548	0.3876	0.5279	0.6796	0.8492	1.0478	1.2994	1.6772	2.0106	2.4066	2.6822	2.9426
49	0.1263	0.2547	0.3876	0.5278	0.6795	0.8490	1.0475	1.2991	1.6766	2.0096	2.4049	2.68	2.9397
50	0.1263	0.2547	0.3875	0.5278	0.6794	0.8489	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778	2.937
100	0.1260	0.2540	0.3864	0.5261	0.6770	0.8452	1.0418	1.2901	1.6602	1.9840	2.3642	2.6259	2.8707
1000	0.1257	0.2534	0.3854	0.5246	0.6747	0.8420	1.0370	1.2824	1.6464	1.9623	2.3301	2.5808	2.8133
∞	0.1257	0.2534	0.3853	0.5244	0.6745	0.8416	1.0364	1.2816	1.6449	1.9600	2.3263	2.5758	2.8070

Приложение В

Таблица В1 – Квантили распределения Фишера (F-критерий F_{1-P}) $\left(F = \dfrac{S_1^2}{S_2^2}, (S_1^2 > S_2^2)\right)$

Степени свободы для мень-шей дис-персии (знам-ля)	Степени свободы для большей дисперсии (числителя)												
	α=0,1 P=0,9 (90%)												
	1	2	3	4	5	6	7	8	9	10	12	24	∞
1	39,864	49,500	53,593	55,833	57,241	58,204	58,906	59,439	59,858	60,195	60,705	62,002	63,328
2	8,5263	9,0000	9,1618	9,2434	9,2926	9,3255	9,3491	9,3668	9,3805	9,3916	9,4081	9,4496	9,4913
3	5,5383	5,4624	5,3908	5,3427	5,3092	5,2847	5,2662	5,2517	5,2400	5,2304	5,2156	5,1764	5,1337
4	4,5448	4,3246	4,1908	4,1073	4,0506	4,0098	3,9790	3,9549	3,9357	3,9199	3,8955	3,8310	3,7607
5	4,0604	3,7797	3,6195	3,5202	3,4530	3,4045	3,3679	3,3393	3,3163	3,2974	3,2682	3,1905	3,1050
6	3,7760	3,4633	3,2888	3,1808	3,1075	3,0546	3,0145	2,9830	2,9577	2,9369	2,9047	2,8183	2,7222
7	3,5894	3,2574	3,0741	2,9605	2,8833	2,8274	2,7849	2,7516	2,7247	2,7025	2,6681	2,5753	2,4708
8	3,4579	3,1131	2,9238	2,8064	2,7265	2,6683	2,6241	2,5893	2,5612	2,5380	2,5020	2,4041	2,2926
9	3,3603	3,0065	2,8129	2,6927	2,6106	2,5509	2,5053	2,4694	2,4403	2,4163	2,3789	2,2768	2,1592
10	3,2850	2,9245	2,7277	2,6053	2,5216	2,4606	2,4140	2,3772	2,3473	2,3226	2,2841	2,1784	2,0554
11	3,2252	2,8595	2,6602	2,5362	2,4512	2,3891	2,3416	2,3040	2,2735	2,2482	2,2087	2,1000	1,9721
12	3,1765	2,8068	2,6055	2,4801	2,3940	2,3310	2,2828	2,2446	2,2135	2,1878	2,1474	2,0360	1,9036
13	3,1362	2,7632	2,5603	2,4337	2,3467	2,2830	2,2341	2,1953	2,1638	2,1376	2,0966	1,9827	1,8462
14	3,1022	2,7265	2,5222	2,3947	2,3069	2,2426	2,1931	2,1539	2,1220	2,0954	2,0537	1,9377	1,7973
15	3,0732	2,6952	2,4898	2,3614	2,2730	2,2081	2,1582	2,1185	2,0862	2,0593	2,0171	1,8990	1,7551

Продолжение таблицы В1

Степени свободы для меньшей дисперсии (знам-ля)	Степени свободы для большей дисперсии (числителя)												
	α=0,1 P=0,9 (90%)												
	1	2	3	4	5	6	7	8	9	10	12	24	∞
16	3,0481	2,6682	2,4618	2,3327	2,2438	2,1783	2,1280	2,0880	2,0553	2,0281	1,9854	1,8656	1,7182
17	3,0262	2,6446	2,4374	2,3077	2,2183	2,1524	2,1017	2,0613	2,0284	2,0009	1,9577	1,8362	1,6856
18	3,0070	2,6239	2,4160	2,2858	2,1958	2,1296	2,0785	2,0379	2,0047	1,9770	1,9333	1,8103	1,6567
19	2,9899	2,6056	2,3970	2,2663	2,1760	2,1094	2,0580	2,0171	1,9836	1,9557	1,9117	1,7873	1,6308
20	2,9747	2,5893	2,3801	2,2489	2,1582	2,0913	2,0397	1,9985	1,9649	1,9367	1,8924	1,7667	1,6074
21	2,9609	2,5746	2,3649	2,2333	2,1423	2,0751	2,0232	1,9819	1,9480	1,9197	1,8750	1,7481	1,5862
22	2,9486	2,5613	2,3512	2,2193	2,1279	2,0605	2,0084	1,9668	1,9327	1,9043	1,8593	1,7312	1,5668
23	2,9374	2,5493	2,3387	2,2065	2,1149	2,0472	1,9949	1,9531	1,9189	1,8903	1,8450	1,7159	1,5490
24	2,9271	2,5383	2,3274	2,1949	2,1030	2,0351	1,9826	1,9407	1,9063	1,3775	1,8319	1,7019	1,5327
25	2,9177	2,5283	2,3170	2,1843	2,0922	2,0241	1,9714	1,9292	1,8947	1,8658	1,8200	1,6890	1,5176
26	2,9091	2,5191	2,3075	2,1745	2,0822	2,0139	1,9610	1,9188	1,8841	1,8550	1,8090	1,6771	1,5036
27	2,9012	2,5106	2,2987	2,1655	2,0730	2,0045	1,9515	1,9091	1,8743	1,8451	1,7989	1,6662	1,4906
28	2,8939	2,5028	2,2906	2,1571	2,0645	1,9959	1,9427	1,9001	1,8652	1,8359	1,7895	1,6560	1,4784
29	2,8871	2,4955	2,2831	2,1494	2,0566	1,9878	1,9345	1,8918	1,8568	1,8274	1,7808	1,6465	1,4670
30	2,8807	2,4887	2,2761	2,1422	2,0492	1,9803	1,9269	1,8841	1,8490	1,8195	1,7727	1,6377	1,4564
40	2,8354	2,4404	2,2261	2,0909	1,9968	1,9269	1,8725	1,8289	1,7929	1,7627	1,7146	1,5741	1,3769
60	2,7914	2,3933	2,1774	2,0410	1,9457	1,8747	1,8194	1,7748	1,7380	1,7070	1,6574	1,5107	1,2915
120	2,7478	2,3473	2,1300	1,9923	1,8959	1,8238	1,7675	1,7220	1,6843	1,6524	1,6012	1,4472	1,1926
∞	2,7055	2,3026	2,0838	1,9449	1,8473	1,7741	1,7167	1,6702	1,6315	1,5987	1,5458	1,3832	1,0000

Таблица В2 – Квантили распределения Фишера (F-критерий F_{1-P})

Степени свободы для меньшей дисперсии (знам-ля)	Степени свободы для большей дисперсии (числителя) α=0,05 P=0,95 (95%)												
	1	2	3	4	5	6	7	8	9	10	12	20	∞
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88	161,45	249,05	254,32
2	18,513	19,000	19,164	19,247	19,296	19,330	19,353	19,371	19,385	19,396	18,513	19,454	19,496
3	10,128	9,5521	9,2766	9,1172	9,0135	8,9406	8,8868	8,8452	8,8123	8,7855	10,128	8,6385	8,5265
4	7,7086	6,9443	6,5914	6,3883	6,2560	6,1631	6,0942	6,0410	5,9988	5,9644	7,7086	5,7744	5,6281
5	6,6079	5,7861	5,4095	5,1922	5,0503	4,9503	4,8759	4,8183	4,7725	4,7351	6,6079	4,5272	4,3650
6	5,9874	5,1433	4,7571	4,5337	4,3874	4,2839	4,2066	4,1468	4,0990	4,0600	5,9874	3,8415	3,6688
7	5,5914	4,7374	4,3468	4,1203	3,9715	3,8660	3,7870	3,7257	3,6767	3,6365	5,5914	3,4105	3,2298
8	5,3177	4,4590	4,0662	3,8378	3,6875	3,5806	3,5005	3,4381	3,3881	3,3472	5,3177	3,1152	2,9276
9	5,1174	4,2565	3,8626	3,6331	3,4817	3,3738	3,2927	3,2296	3,1789	3,1373	5,1174	2,9005	2,7067
10	4,9646	4,1028	3,7083	3,4780	3,3258	3,2172	3,1355	3,0717	3,0204	2,9782	4,9646	2,7372	2,5379
11	4,8443	3,9823	3,5874	3,3567	3,2039	3,0946	3,0123	2,9480	2,8962	2,8536	4,8443	2,6090	2,4045
12	4,7472	3,8853	3,4903	3,2592	3,1059	2,9961	2,9134	2,8486	2,7964	2,7534	4,7472	2,5055	2,2962
13	4,6672	3,8056	3,4105	3,1791	3,0254	2,9153	2,8321	2,7669	2,7144	2,6710	4,6672	2,4202	2,2064
14	4,6001	3,7389	3,3439	3,1122	2,9582	2,8477	2,7642	2,6987	2,6458	2,6021	4,6001	2,3487	2,1307
15	4,5431	3,6823	3,2874	3,0556	2,9013	2,7905	2,7066	2,6408	2,5876	2,5437	4,5431	2,2878	2,0658
16	4,4940	3,6337	3,2389	3,0069	2,8524	2,7413	2,6572	2,5911	2,5377	2,4935	4,4940	2,2354	2,0096
17	4,4513	3,5915	3,1968	2,9647	2,8100	2,6987	2,6143	2,5480	2,4943	2,4499	4,4513	2,1898	1,9604
18	4,4139	3,5546	3,1599	2,9277	2,7729	2,6613	2,5767	2,5102	2,4563	2,4117	4,4139	2,1497	1,9168

Продолжение таблицы В2

Степени свободы для меньшей дисперсии (знаменателя)	Степени свободы для большей дисперсии (числителя)												
	α=0,05 P=0,95 (95%)												
	1	2	3	4	5	6	7	8	9	10	12	20	∞
19	4,3808	3,5219	3,1274	2,8951	2,7401	2,6283	2,5435	2,4768	2,4227	2,3779	4,3808	2,1141	1,8780
20	4,3513	3,4928	3,0984	2,8661	2,7109	2,5990	2,5140	2,4471	2,3928	2,3479	4,3513	2,0825	1,8432
21	4,3248	3,4668	3,0725	2,8401	2,6848	2,5727	2,4876	2,4205	2,3661	2,3210	2,1757	2,0540	1,8117
22	4,3009	3,4434	3,0491	2,8167	2,6613	2,5491	2,4638	2,3965	2,3419	2,2967	2,1508	2,0283	1,7831
23	4,2793	3,4221	3,0280	2,7955	2,6400	2,5277	2,4422	2,3748	2,3201	2,2747	2,1282	2,0050	1,7570
24	4,2597	3,4028	3,0088	2,7763	2,6207	2,5082	2,4226	2,3551	2,3002	2,2547	2,1077	1,9838	1,7331
25	4,2417	3,3852	2,9912	2,7587	2,6030	2,4904	2,4047	2,3371	2,2821	2,2365	2,0889	1,9643	1,7110
26	4,2252	3,3690	2,9751	2,7426	2,5868	2,4741	2,3883	2,3205	2,2655	2,2197	2,0716	1,9464	1,6906
27	4,2100	3,3541	2,9604	2,7278	2,5719	2,4591	2,3732	2,3053	2,2501	2,2043	2,0558	1,9299	1,6717
28	4,1960	3,3404	2,9467	2,7141	2,5581	2,4453	2,3593	2,2913	2,2360	2,1900	2,0411	1,9147	1,6541
29	4,1830	3,3277	2,9340	2,7014	2,5454	2,4324	2,3463	2,2782	2,2229	2,1768	2,0275	1,9005	1,6377
30	4,1709	3,3158	2,9223	2,6896	2,5336	2,4205	2,3343	2,2662	2,2107	2,1646	2,0148	1,8874	1,6223
40	4,0848	3,2317	2,8387	2,6060	2,4459	2,3359	2,2400	2,1802	2,1240	2,0772	1,9245	1,7929	1,5089
60	4,0012	3,1504	2,7581	2,5252	2,3683	2,2540	2,1665	2,0970	2,0401	1,9926	1,8364	1,7001	1,3893
120	3,9201	3,0718	2,6802	2,4472	2,2900	2,1750	2,0867	2,0164	1,9588	1,9105	1,7505	1,6084	1,2539
∞	3,8415	2,9957	2,6049	2,3719	2,2141	2,0986	2,0096	1,9384	1,8799	1,8307	1,6664	1,5173	1,0000

Таблица В3 – Квантили распределения Фишера (F-критерий F_{1-P})

Степени свободы для меньшей дисперсии (знам-ля)	Степени свободы для большей дисперсии (числителя)												
	α=0,01 P=0,99 (99%)												
	1	2	3	4	5	6	7	8	9	10	12	20	∞
1	4052,2	4999,5	5403,3	5624,6	5763,7	5859,0	5928,3	5981,1	6022,5	6055,8	6157,3	6234,6	6366,0
2	98,503	99,000	99,166	99,249	99,299	99,332	99,356	99,374	99,388	99,399	99,432	99,458	99,499
3	34,116	30,817	29,457	28,710	28,237	27,911	27,672	27,489	27,345	27,229	26,872	26,598	26,125
4	21,198	18,000	16,694	15,977	15,522	15,207	14,976	14,799	14,659	14,546	14,198	13,929	13,463
5	16,258	13,274	12,060	11,392	10,967	10,672	10,456	10,289	10,158	10,051	9,7222	9,4665	9,0204
6	13,745	10,925	9,7795	9,1483	8,7459	8,4661	8,2600	8,1016	7,9761	7,8741	7,5590	7,3127	6,8861
7	12,246	9,5466	8,4513	7,8467	7,4604	7,1914	6,9928	6,8401	6,7188	6,6201	6,3143	6,0743	5,6495
8	11,259	8,6491	7,5910	7,0060	6,6318	6,3707	6,1776	6,0289	5,9106	5,8143	5,5151	5,2793	4,8588
9	10,561	8,0215	6,9919	6,4221	6,0569	5,8018	5,6129	5,4671	5,3511	5,2565	4,9621	4,7290	4,3105
10	10,044	7,5594	6,5523	5,9943	5,6363	5,3858	5,2001	5,0567	4,9424	4,8492	4,5582	4,3269	3,9090
11	9,6460	7,2057	6,2167	5,6683	5,3160	5,0692	4,8861	4,7445	4,6315	4,5393	4,2509	4,0209	3,6025
12	9,3302	6,9266	5,9526	5,4119	5,0643	4,8206	4,6395	4,4994	4,3875	4,2961	4,0096	3,7805	3,3608
13	9,0738	6,7010	5,7394	5,2053	4,8616	4,6204	4,4410	4,3021	4,1911	4,1003	3,8154	3,5868	3,1654
14	8,8616	6,5149	5,5639	5,0354	4,6950	4,4558	4,2779	4,1399	4,0297	3,9394	3,6557	3,4274	3,0040
15	8,6831	6,3589	5,4170	4,8932	4,5556	4,3183	4,1415	4,0045	3,8948	3,8049	3,5222	3,2940	2,8684
16	8,5310	6,2262	5,2922	4,7726	4,4374	4,2016	4,0259	3,8896	3,7804	3,6909	3,4089	3,1808	2,7528
17	8,3997	6,1121	5,1850	4,6690	4,3359	4,1015	3,9267	3,7910	3,6822	3,5931	3,3117	3,0835	2,6530
18	8,2854	6,0129	5,0919	4,5790	4,2479	4,0146	3,8406	3,7054	3,5971	3,5082	3,2273	2,9990	2,5660

Продолжение таблицы В3

Степени свободы для меньшей дисперсии (знаменателя)	Степени свободы для большей дисперсии (числителя)												
	α=0,01 P=0,99 (99%)												
	1	2	3	4	5	6	7	8	9	10	12	20	∞
19	8,1850	5,9259	5,0103	4,5003	4,1708	3,9386	3,7653	3,6305	3,5225	3,4338	3,1533	2,9249	2,4893
20	8,0960	5,8489	4,9382	4,4307	4,1027	3,8714	3,6987	3,5644	3,4567	3,3682	3,08 80	2,8594	2,4212
21	8,0166	5,7804	4,8740	4,3688	4,0421	3,8117	3,6396	3,5056	3,3981	3,3098	3,0299	2,8011	2,3603
22	7,9454	5,7190	4,8166	4,3134	3,9880	3,7583	3,5867	3,4530	3,3458	3,2576	2,9780	2,7488	2,3055
23	7,8811	5,6637	4,7649	4,2635	3,9392	3,7102	3,5390	3,4057	3,2986	3,2106	2,9311	2,7017	2,2559
24	7,8229	5,6136	4,7181	4,2184	3,8951	3,6667	3,4959	3,3629	3,2560	3,1681	2,8887	2,6591	2,2107
25	7,7698	5,5680	4,6755	4,1774	3,8550	3,6272	3,4568	3,3239	3,2172	3,1294	2,8502	2,6203	2,1694
26	7,7213	5,5263	4,6366	4,1400	3,8183	3,5911	3,4210	3,2884	3,1818	3,0941	2,8150	2,5848	2,1315
27	7,6767	5,4881	4,6009	4,1056	3,7848	3,5580	3,3882	3,2558	3,1494	3,0618	2,7827	2,5522	2,0965
28	7,6356	5,4529	4,5681	4,0740	3,7539	3,527(	3,3581	3,2259	3,1195	3,0320	2,7530	2,5223	2,0642
29	7,5976	5,4205	4,5378	4,0449	3,7254	3,4993	3,3302	3,1982	3,0920	3,0045	2,7256	2,4946	2,0342
30	7,5625	5,3903	4,5097	4,0179	3,6990	3,4735	3,3045	3,1726	3,0665	2,9791	2,7002	2,4689	2,0062
40	7,3141	5,1785	4,3126	3,8283	3,5138	3,2910	3,1238	2,9930	2,8876	2,8005	2,5216	2,2880	1,8047
60	7,0771	4,9774	4,1259	3,6491	3,3389	3,1187	2,9530	2,8233	2,7185	2,6318	2,3523	2,1154	1,6006
120	6,8510	4,7865	3,9491	3,4796	3,1735	2,9559	2,7918	2,6629	2,5586	2,4721	2,1915	1,9500	1,3805
∞	6,6349	4,6052	3,7816	3,3192	3,0173	2,8020	2,6393	2,5113	2,4073	2,3209	2,0385	1,7908	1,0000

Приложение Г

Таблица Г1 – Квантили распределения Пирсона (χ^2_{1-p} – критерий)

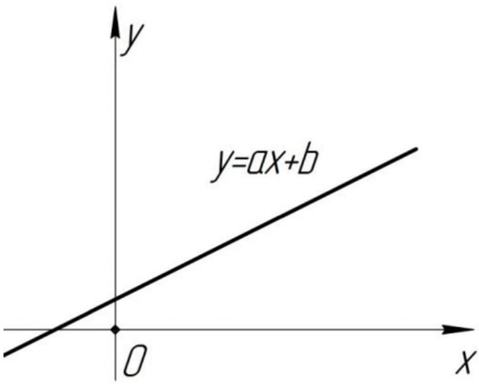
Число степеней свободы	Уровень значимости										
	0,99	0,98	0,95	0,90	0,80	0,20	0,10	0,05	0,02	0,01	0,001
1	0,00016	0,00063	0,00393	0,0158	0,0642	1,642	2,706	3,841	5,412	6,635	10,827
2	0,0201	0,0404	0,103	0,211	0,446	3,219	4,605	5,991	7,824	9,210	13,815
3	0,115	0,185	0,352	0,584	1,005	4,642	6,251	7,815	9,837	11,341	16,268
4	0,297	0,429	0,711	1,064	1,649	5,989	7,779	9,488	11,668	13,277	18,465
5	0,554	0,752	1,145	1,610	2,343	7,289	9,236	11,070	13,388	15,086	20,517
6	0,872	1,134	1,635	2,204	3,070	8,558	10,645	12,592	15,033	16,812	22,457
7	1,239	1,564	2,167	2,833	3,822	9,803	12,017	14,067	16,622	18,475	24,322
8	1,646	2,032	2,733	3,490	4,594	11,030	13,362	15,507	18,679	20,090	26,125
9	2,088	2,532	3,325	4,168	5,380	12,242	14,684	16,919	19,679	21,666	27,877
10	2,588	3,059	3,940	4,865	6,179	13,442	15,987	18,307	21,161	23,209	29,588
11	3,053	3,609	4,575	5,578	6,989	14,631	17,275	19,675	22,618	24,725	31,264
12	3,571	4,178	5,226	6,304	7,807	15,812	18,549	21,026	24,054	26,217	32,909
13	4,107	4,765	5,892	7,042	8,634	16,985	19,812	22,362	25,472	27,688	34,528
14	4,660	5,368	6,571	7,790	9,467	18,151	21,064	23,685	26,873	29,141	36,123
15	5,229	5,985	7,262	8,547	10,307	19,311	22,307	24,996	28,259	30,578	37,697
16	5,812	6,614	7,962	9,312	11,152	20,465	23,542	26,296	29,633	32,000	39,252
17	6,408	7,255	8,672	10,085	12,002	21,615	24,769	27,587	30,995	33,409	40,790
18	7,015	7,906	9,390	10,865	12,857	22,760	25,989	28,869	32,346	34,805	42,312
19	7,633	8,567	10,117	11,651	13,716	23,900	27,204	30,144	33,687	36,191	43,820

Продолжение таблицы Г1

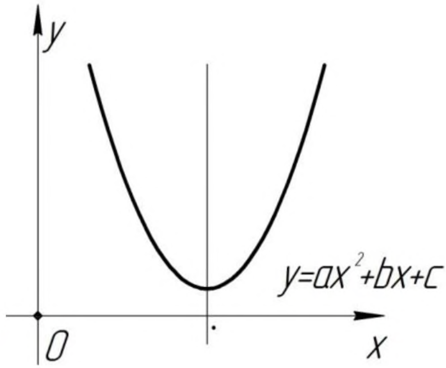
Число степеней свободы	Уровень значимости										
	0,99	0,98	0,95	0,90	0,80	0,20	0,10	0,05	0,02	0,01	0,001
20	8,260	9,237	10,851	12,443	14,578	25,038	28,412	31,410	35,020	37,566	45,315
21	8,897	9,915	11,591	13,240	15,445	26,171	29,615	32,671	36,343	38,932	46,797
22	9,542	10,600	12,338	14,041	16,314	27,301	30,813	33,924	37,659	40,289	48,268
23	10,196	11,298	13,091	14,848	17,187	28,429	32,007	35,172	38,968	41,638	49,728
24	10,856	11,992	13,848	15,659	18,062	29,553	33,196	36,415	40,270	42,980	51,179
25	11,542	12,697	14,611	16,473	18,940	30,675	34,382	37,652	41,566	44,314	52,620
26	12,198	13,409	15,379	17,292	19,820	31,795	35,563	38,885	42,856	45,642	54,052
27	12,879	14,125	16,151	18,114	20,703	32,912	86,741	40,113	44,140	46,963	55,476
28	13,565	14,847	16,928	18,939	21,588	34,027	37,916	41,337	45,419	48,278	56,893
29	14,256	15,574	17,708	19,768	22,475	35,139	39,087	42,557	46,693	49,588	58,302
30	14,953	16,306	18,493	20,599	23,364	36,250	40,256	43,773	47,962	50,892	59,703

Приложение Д

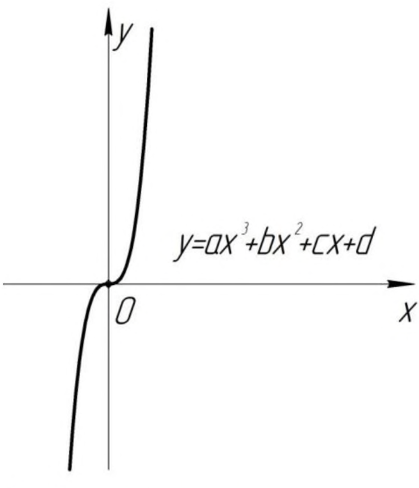
Графики основных функциональных зависимостей



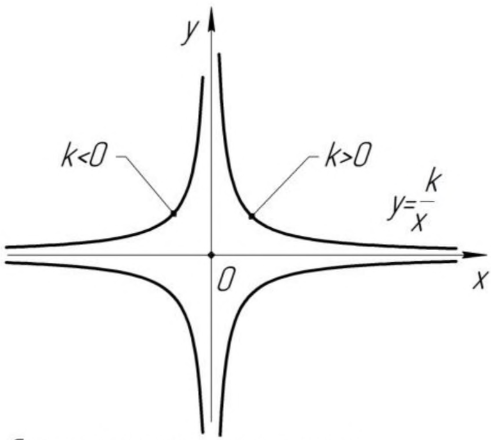
Линейная функция



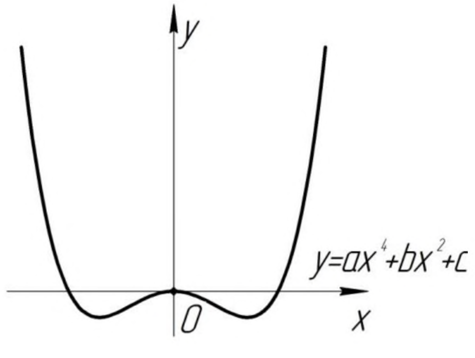
Квадратная функция



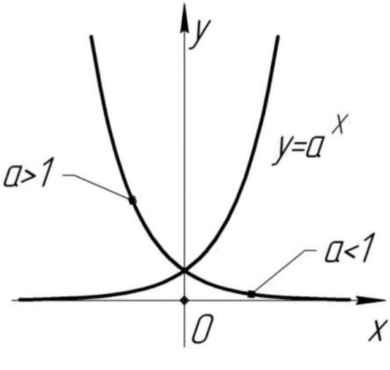
Кубическая функция



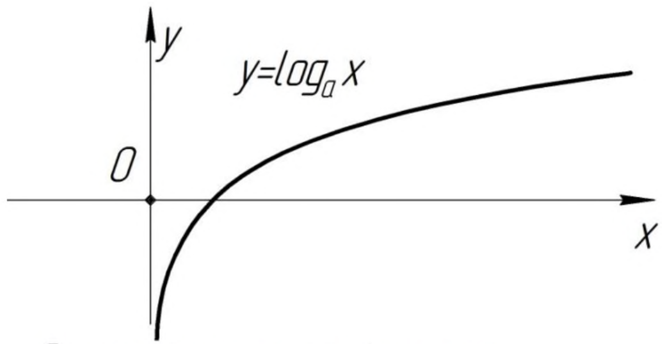
Степенная функция с целым отрицательным показателем



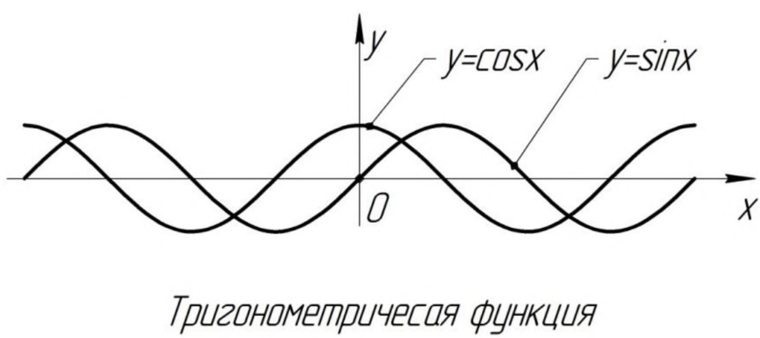
Биквадратная функция



Показательная функция



Логарифмическая функция



Тригонометричесая функция

Приложение Е

Таблица Е1 – Случайные числа

1	2	3	4	5	6	7	8	9	10
5489	5583	3156	835	1988	3912	938	7460	869	4420
3522	935	7877	5665	7020	9555	7379	7124	7878	5544
7555	7579	2550	2487	9477	864	2349	1012	8250	2633
5759	3554	5080	9074	7001	6249	3224	6368	9102	2672
6303	6895	3371	3196	7231	2918	7380	438	7547	2644
7351	5634	5323	2623	7803	8374	2191	464	696	9529
7068	7803	8832	5119	6350	120	5026	3684	5657	304
3613	1428	1796	8447	503	5654	3254	7336	9536	1944
5143	4534	2105	368	7890	2473	4240	8652	9435	1422
9815	5144	7649	8638	6137	8070	5345	4865	2456	5708
5780	1277	6316	1013	2867	9938	3930	3203	5696	1769
1187	951	5991	5245	5700	5564	7352	891	6249	6568
4184	2179	4554	9083	2254	2435	2965	5154	1209	7069
2916	2972	9885	275	144	8034	8122	3213	7666	230
5524	1341	9860	6565	6981	9842	171	2284	2707	3008
146	5291	2354	5694	377	5336	6460	9585	3415	2358
4920	2826	5238	5402	7937	1993	4332	2327	6875	5230
7978	1947	6380	3425	7267	7285	1130	7722	164	8573
7453	653	3645	7497	5969	8682	4191	2976	361	9334
1473	6938	4899	5348	1641	3652	852	5296	4538	4456
8162	8797	8000	4707	1880	9660	8446	1883	9768	881

Приложение Ж

Таблица Ж1 – Квантили распределения Кохрена G_{1-p} (G – критерий, $G = S_{\max}^2 \Big/ \sum_{1}^{N} S_i^2$)

Степени свободы k	Степени свободы f													
	P=0,05 (95%)													
	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	9985	9750	9392	9057	8772	8534	8332	8159	8010	7880	7341	6602	5813	5000
3	9669	8709	7977	7457	7071	6771	6530	6333	6167	6025	5466	4748	4031	3333
4	9065	7679	6841	6287	5895	5598	5365	5175	5017	4884	4366	3720	3093	2500
5	8412	6838	5981	5441	5065	4783	4564	4387	4241	4118	3645	3066	2513	2000
6	7808	6161	5321	4803	4447	4184	3980	3817	3682	3568	3135	2612	2119	1667
7	7271	5612	4800	4307	3974	3726	3535	3384	3259	3154	2756	2278	1833	1429
8	6798	5157	4377	3910	3595	3362	3185	3043	2926	2829	2462	2022	1616	1250
9	6385	4775	4027	3584	3286	3067	2901	2768	2659	2568	2226	1820	1446	1111
10	6020	4450	3733	3311	3029	2823	2666	2541	2439	2353	2032	1655	1308	1000
12	5410	3924	3264	2880	2624	2439	2299	2187	2098	2020	1737	1403	1100	0833
15	4709	3346	2758	2419	2195	2034	1911	1815	1736	1671	1429	1144	0889	0667
20	3894	2705	2205	1921	1735	1602	1501	1422	1357	1303	1108	0879	0675	0500
24	3434	2354	1907	1656	1493	1374	1286	1216	1160	1113	0942	0743	0567	0417
30	2929	1980	1593	1377	1237	1137	1061	1002	0958	0921	0771	0604	0457	0333
40	2370	1576	1259	1082	0968	0887	0827	0780	0745	0713	0595	0462	0347	0250
60	1737	1131	0895	0765	0682	0623	0583	0552	0520	0497	0411	0316	0234	0167
120	0998	0632	0495	0419	0371	0337	0312	0292	0279	0266	0218	0165	0120	0083
∞	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

Продолжение таблица Ж1

Степени свободы k	Степени свободы f													
	P=0,01 (99%)													
	1	2	3	4	5	6	7	8	9	10	16	36	144	∞
2	9999	9950	9794	9586	9373	9172	8998	8823	8674	8539	7949	7067	6062	5000
3	9933	9423	8831	8335	7933	7606	7335	7107	6912	6743	6059	5153	4230	3333
4	9676	8643	7814	7212	6761	6410	6129	5897	5702	5536	4884	4057	3251	2500
5	9279	7885	6957	6329	5875	5531	5259	5037	4854	4697	4094	3351	2644	2000
6	8828	7218	6258	5635	5195	4866	4608	4401	4229	4084	3529	2858	2229	1667
7	8376	6644	5685	5080	4659	4347	4105	3911	3751	3616	3105	2494	1929	1429
8	7945	6152	5209	4627	4226	3932	3704	3522	3a73	3248	2779	2214	1700	1250
9	7544	5727	4810	4251	3870	3592	3378	3207	3067	2950	2514	1992	1521	1111
10	7175	5358	4469	3934	3572	3308	3106	2945	2813	2704	2297	1811	1376	1000
12	6528	4751	3919	3428	3099	2861	2680	2535	2419	2320	1961	1535	1157	0833
15	5747	4069	3317	2882	2593	2386	2228	2104	2002	1918	1612	1251	0934	0667
20	4799	3297	2654	2288	2048	1877	1748	1646	1567	1501	1248	0960	0709	0500
24	4247	2871	2295	1970	1759	1608	1495	1406	1338	1283	1060	0810	0595	0417
30	3632	2412	1913	1635	1454	1327	1232	1157	1100	1054	0867	0658	0480	0333
40	2940	1915	1508	1281	1135	1033	0957	0898	0853	0816	0668	0503	0363	0250
60	2151	1371	1069	0902	0796	0722	0668	0625	0594	0567	0461	0344	0245	0167
120	1225	0759	0585	0489	0429	0387	0357	0334	0316	0302	0242	0178	0125	0083
∞	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000	0000

Все квантили G_{1-p} меньше единицы, поэтому в таблице приведены лишь десятичные знаки, следующие после запятой, перед которой при пользовании таблицей нужно ставить нуль.

Приложение И

Таблица И1 – Матрица полного факторного эксперимента от 2^2 до 2^4

План	Номер опыта	x_1	x_2	x_3	x_4	y_i
	1	+	+	+	+	y_1
2^2	2	-	+	+	+	y_2
	3	+	-	+	+	y_3
	4	-	-	+	+	y_4
2^3	5	+	+	-	+	y_5
	6	-	+	-	+	y_6
	7	+	-	-	+	y_7
	8	-	-	-	+	y_8
	9	+	+	+	-	y_9
	10	-	+	+	-	y_{10}
2^4	11	+	-	+	-	y_{11}
	12	-	-	+	-	y_{12}
	13	+	+	-	-	y_{13}
	14	-	+	-	-	y_{14}
	15	+	-	-	-	y_{15}
	16	-	-	-	-	y_{16}

ОГЛАВЛЕНИЕ

У ч е б н о е и з д а н и е

Сохт Казбек Аюбович
Трубилин Евгений Иванович
Коновалов Владимир Иванович

СТАТИСТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЙ ПРОЦЕССОВ И МАШИН В АГРОБИЗНЕСЕ

Учебное пособие

В авторской редакции

Подписано в печать 22.01.2016. Формат $60 \times 84\ ^1/_{16}$.
Усл. печ. л. – 12,6. Уч.-изд. л. – 9,9.
Тираж 100 экз. Заказ № 32

Типография
Кубанского государственного аграрного университета.
350044, г. Краснодар, ул. Калинина, 13