Федеральное государственное бюджетное образовательное учреждение высшего образования
«Приморская государственная сельскохозяйственная академия»
Институт землеустройства и агротехнологий

Савельева Е.В.

# Статистические методы обработки результатов. исследований.

Учебное пособие
для обучающихся по направлениям подготовки:
35.03.04 Агрономия; 35.03.03 Агрохимия и агропочвоведение

Уссурийск 2015

УДК 519.2
ББК 22.172
С 781

Рецензент: Лосев А.С., к.ф.-м. наук, доцент Федерального государственного бюджетного учреждения науки Института прикладной математики Дальневосточного отделения РАН

Иванова Е.П., доцент, доцент кафедры Агрохимии, агроэкологии и охраны труда

Статистические методы обработки результатов исследований: учебное пособие для обучающихся по направлениям подготовки: 35.03.04 Агрономия; 35.03.03 Агрохимия и агропочвоведение ФГБОУ ВО Приморская ГСХА / ФГБОУ ВО ПГСХА; сост. Е.В. Савельева.,– Уссурийск, 2015.- 115 с.

Учебное пособие «Статистические методы обработки результатов исследований» представляет собой методические материалы, содержащие достаточный теоретический минимум для освоения дисциплины, задания для самостоятельной работы обучающихся, имеющих прикладной характер.

Основной целью учебного пособия является изучение методов статистической обработки получаемых в ходе эксперимента данных в агрономии, экологии. Учебное пособие состоит их четырех разделов: исследование вариационных рядов; проверка статистических гипотез; элементы дисперсионного анализа; элементы корреляционного и регрессионного анализа.

Для организации самостоятельной работы обучающихся в пособии приведены варианты индивидуальных заданий.

Издается по решению методического совета ФГБОУ ВО Приморская ГСХА

# ВВЕДЕНИЕ

В связи с возрастающей ролью вероятностно-статистических методов в сельском хозяйстве, задания составлены таким образом, чтобы студент овладел практическими навыками по использованию статистических методов обработки и анализа сельскохозяйственных показателей.

В методическом указании дан теоретический минимум, которым должен овладеть студент для выполнения трех индивидуальных задании: вариационные ряды; статистические гипотезы; элементы корреляционного анализа. Надо уметь составить дискретный и интервальный вариационные ряды по данной выборке, и изобразить их геометрически. Вычислить статистические характеристики и по результату оценить их. Найти доверительный интервал генеральной средней с заданной доверительной вероятностью. Все эти вопросы рассматриваются в разделе "Математическая статистика".

Основная задача математической статистики сбор и обработка экспериментальных данных, с целью изучения закономерностей случайных явлений, носящих массовый характер. Например, распределение пахотных земель по урожайности. Обследование всей совокупности заменяют обследованием выборки из генеральной совокупности. Желательно, чтобы результаты обследования выборки отражали характерные, основные черты изучаемого признака; для этого выборка должна быть репрезентативной.

Пособие содержат четыре индивидуальных задания по основным разделам (модулям) дисциплины:

1) исследование вариационных рядов;

2) проверка статистических гипотез;

3) дисперсионный анализ;

4) корреляционный и регрессионный анализ.

Оглавление

# РАЗДЕЛ 1. Выборка и ее представление. Числовые характеристики выборочной совокупности. Доверительная вероятность, доверительный интервал.

### *П.1.Понятие вариационного ряда.*

***Определение 1.1. Генеральной совокупностью*** называется совокупность всех возможных однородных предметов или явлений, над которыми проводятся наблюдения, или совокупность всех возможных наблюдений, проводимых над некоторой случайной величиной в одинаковых условиях.

• На практике сплошное обследование применяется сравнительно редко (если совокупность содержит очень большое число элементов). В таких случаях случайным образом выбирают из всей совокупности ограниченное число объектов и подвергают их изучению.

***Определение 1.2. Выборочной совокупностью (выборкой)*** называется совокупность предметов или явлений, отобранная из соответствующей генеральной совокупности.

***Определение 1.3. Объемом*** совокупности (генеральной или выборочной) называется общее число ее элементов.

- Если $N$ и $n$ – соответственно объемы генеральной и выборочной совокупности, то, очевидно, что $N>n$.
- *Выборочный метод* заключается в том, что из генеральной совокупности берется выборка значительно меньшего объема и определяются характеристики выборки, которые принимаются в качестве приближенных значений соответствующих характеристик генеральной совокупности.

Выборки бывают:

а) ***повторные*** – если отобранный объект (перед отбором следующего) возвращается в генеральную совокупность;

б) ***бесповторные*** – если отобранный объект не возвращается в генеральную совокупность.

Пусть из генеральной совокупности извлечена выборка $X$ объемом $n$. Случайный *выбор элемента* рассматривается как независимое наблюдение над величиной $\xi$, имеющей некоторое распределение вероятностей. Если те значения, которые приняла случайная величина $\xi$ в $n$ наблюдениях, записать не в порядке получения, а в порядке возрастания (то есть ***ранжируя***), то получим упорядоченную выборку $x_1, x_2, \ldots, x_n$ ,, называемую ***вариационным рядом***.

- Выборка и вариационный ряд несут одну и ту же информацию, но с вариационным рядом работать легче в силу его упорядоченности.
- Чтобы по данным выборки можно было достаточно точно судить об исследуемом признаке генеральной совокупности, требуется, чтобы выборка правильно представляла пропорции генеральной совокупности, то есть была ***репрезентативной (представительной).***

***Определение 1.4. Вариантой*** называется значение $x_i$ случайной величины, соответствующее отдельной группе сгруппированного ряда наблюдаемых данных.

***Определение 1.5. Размахом R***вариационного ряда называется разность между его наибольшей $x_{\max}$ и наименьшей $x_{\min}$ вариантами: $R = x_{\max} - x_{\min}$.

- Если изучается дискретная случайная величина, то при достаточно большом объеме выборки в ней могут содержаться повторяющиеся значения. Для каждого такого значения можно подсчитать, сколько раз оно встретилось в ряде наблюдений.

***Определение 1.6 Частотой (весом)*** варианты называется численность отдельной группы сгруппированного ряда наблюдаемых данных.

- Если $i$ – индекс варианты, то $f_i$ – число измеренных значений $i$-ой варианты.

***Определение 1.7. Относительной частотой варианты*** $x_i$ называется отношение $f_i$ числа повторения $x_i$ к объему выборки $n$:

$$\omega_i = \frac{f_i}{n}. \tag{1.1}$$

• Общая сумма частот вариационного ряда равна объему данной совокупности, т. е.

$$\sum_{i=1}^{k} f_i = n,$$

где $k$ – число групп, $n$ – общее число наблюдений, или объем совокупности.

• Общая сумма частностей равна единице

$$\sum \frac{f_i}{n} = 1 \text{ или } \sum \frac{f_i}{n} \cdot 100 = 100\%$$

Данные наблюдений, среди которых много повторяющихся изображают в виде таблицы (таб. 1.1). Такая таблица называется ***таблицей статистического распределения.***

Таблица 1.1

| Значения $x_i$ | $x_1$ | $x_2$ | … | $x_k$ |
|---|---|---|---|---|
| Частоты $f_i$ | $f_1$ | $f_2$ | … | $f_k$ |
| Относительные частоты | $\omega_1 = \frac{f_1}{n}.$ | $\omega_2 = \frac{f_i}{n}.$ | | $\tilde{\omega}_k = \frac{f_i}{n}.$ |

***Замечание***. Если значения признака в выборке не повторяются, в этом случае ряд называется ***несгруппированным***, в этом случае составление таких рядов не имеет смысла, т.к. все частоты равны 1.

| Значения $x_i$ | $x_1$ | $x_2$ | … | $x_k$ |
|---|---|---|---|---|
| Частоты $f_i$ | *1* | *1* | … | *1* |

### ***П.2. Виды вариационных рядов.***

Рассматривают два вида вариационных рядов: ***дискретные, интервальные.***

#### ***1) Составление дискретных рядов.***

***Определение 1.8. Дискретным вариационным рядом*** распределения (***распределением частот*** или ***относительных частот***) называется ранжированная совокупность вариант $x_i$ с соответствующими им частотами или относительными частотами.

*Алгоритм составления дискретного вариационного ряда:*

– подсчитать количество элементов выборки (ее объем);
– найти минимальное ($x_{\min}$) и максимальное ($x_{\max}$) значения выборки и сделать ряд ранжированным;
– в первый столбец таблицы записать варианты значений случайной величины (генеральной совокупности), начиная с $x_{\min}$ и заканчивая $x_{\max}$;
– просмотреть по одному все элементы выборки в протоколе наблюдений, и подсчитать количество значений, соответствующих каждой варианте (то есть $f_i$ ) записать во второй столбец;
– в третий столбец записать значения относительных частот;
– в итоговой строке сделать контроль.

***Пример 1*. (Задание №1 (1))**

При взвешивании новорождённых телят холмогорской породы были получены следующие данные (n=20; min=23; max=45):

42,26,30,33,28,35,40,33,35,28,37,30,40,30,33,33,37,37,33,30.

Составить вариационный дискретный ряд.

***2) Составление интервальных вариационных рядов.***

***Определение 9. Интервальным вариационным рядом (интервальным распределением частот*** или ***относительных частот)*** называется упорядоченная последовательность интервалов варьирования случайной величины с соответствующими частотами или относительными частотами попаданий в каждый из них значений случайной величины.

*Алгоритм составления интервального вариационного ряда:*

– найти минимальное ($x_{\min}$) и максимальное ($x_{\max}$) значения выборки;
– найти объем $n$ выборки;
– определить оптимальное число интервалов по *формуле Стерджеса*:

$$k = 1 + 3{,}322 \cdot \lg n\,. \qquad (1.2)$$

– найти длину интервала по формуле:

$$h = \frac{x_{\max} - x_{\min}}{k}\,. \qquad (1.3)$$

– заполнить первый столбец таблицы интервалами исследуемой совокупности. За начало первого интервала рекомендуется брать величину:

$$x_{нач} = x_{\min}$$

- во втором столбце представить сводку для облегченного подсчета частот в виде конвертов, каждое ребро равно 1 элементу, попавшему в интервал, сводка необходима, если составление происходит вручную;

- просмотреть по одному все элементы выборки в протоколе наблюдений, и подсчитать количество значений, соответствующих каждому интервалу (то есть $f_i$) и заполнить третий столбец;

- в четвертый столбец записать середины интервалов, вычисленные по формуле: $x_i = \dfrac{a_i + b_i}{2}$

- в последний столбец записываем значения накопленных частот

*Определение 7.* ***Накопленной частотой*** $S_{нак}$ в точке $x$ называют суммарную частоту членов статистической совокупности со значениями

***Пример2.*** **(Задание №2 (1) )**

Даны сведения об урожайности пшеницы (ц/га).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 14,3 | 10,1 | 15,2 | 19,9 | 15,9 | 8,4 | 17,7 | 12,6 |
| 19,1 | 15,7 | 12,1 | 16,9 | 21,7 | 10,4 | 16,3 | 14,5 |
| 13,5 | 13,8 | 18,6 | 9,7 | 19,0 | 12,6 | 17,8 | 14,4 |
| 12,2 | 10,3 | 20,6 | 17,3 | 12,3 | 10,2 | 12,5 | 14,8 |
| 21,7 | 15,1 | 13,3 | 13,7 | 18,2 | 9,0 | 16,4 | 12,1 |
| 19,2 | 14,4 | 13,5 | 18,3 | 10,9 | 11,8 | 14,6 | 15,7 |
| 13,0 | 17,4 | 10,4 | 8,2 | 14,0 | 18,4 | 14,8 | 10,5 |
| 17,4 | 13,9 | 12,7 | 11,5 | 13,5 | 15,7 | 12,5 | 20,3 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 11,8 | 18,8 | 15,6 | 8,8 | 16,2 | 14,8 | 10,7 | 9,3 |

Составить интервальный ряд распределения;

### ***Графическое представление статистического ряда.***

Для наглядного представления вариационного ряда (или статистического распределения) пользуются графическим изображением вариационных рядов (полигоном, гистограммой, кумулятой).

*Полигон.*

***Полигон частот*** – это ломаная, отрезки которой соединяют точки $(x_i\ ;\ f_i)$. Для построения полигона частот на оси абсцисс откладывают варианты $x_i$, а на оси ординат – соответствующие им частоты $f_i$. Точки соединяют отрезками прямых и получают полигон частот.

***Пример 3. (задание №1 (2))***

*Гистограмма*

Гистограмма строится только для интервального вариационного ряда (сгруппированной выборки).

***Гистограмма частот*** – это ступенчатая фигура, состоящая из прямоугольников, основаниями которых служат частичные интервалы длиною $h$, а высоты равны частотам $f_i$

- Площадь гистограммы частот равна сумме всех частот, то есть объему выборки.
- Если середины верхних сторон прямоугольников соединить отрезками прямых, а концы этой ломаной еще соединить с серединами соседних интервалов, частоты которых равны 0, а длина равна длине соседнего интервала, то *получим полигон интервального ряда.*

***Кумулята***

***Кумулята*** (кумулятивная кривая) – график накопленных частот, сглаженное графическое изображение эмпирической функции распределения. При построении кумуляты в точке, соответствующей принимаемому значению, для дискретного ряда и в правом конце интервала для интервального ряда строится перпендикуляр, высота которого пропорциональна накопленной частоте, а затем верхние концы перпендикуляров соединяются между собой с помощью отрезков прямых.

***Пример 4. (задание №2 (2))***

### *П. 2. Числовые характеристики выборочной совокупности.*

Для того, чтобы количественно охарактеризовать самые существенные свойства распределения, а также для того, чтобы можно было сравнить разные распределения, вычисляют средние показатели - выборочные числовые характеристики.

### *Характеристики уровня вариационных рядов.*

В статистике используются различные величины в зависимости от того, какие цели при анализе материала ставит исследователь. Понятием средней величины пользуемся в тех случаях, когда требуется определить средний надой по стаду, средний привес, средний прирост стада, средние клинические показатели деятельности сердца, лёгких, среднего состава крови и во многих других случаях.

*1) Выборочная средняя (средняя арифметическая).*

Средняя арифметическая является наиболее распространенной среди средних величин. Ее применяют в тех случаях, когда даны отдельные объекты с индивидуальными значениями признаков, выраженными абсолютными показателями. Среднюю арифметическую определяют как отношение суммы индивидуальных значений признаков к их количеству.

Различают среднюю арифметическую простую и взвешенную.

а) Среднюю арифметическую простую применяют в случае, если индивидуальные значения признака в совокупности встречаются по одному разу (*несгруппированный ряд*

- средняя арифметическая (выборочная средняя) простая:

$$\bar{x}_в = \frac{\sum x_i}{n}, \qquad (1.4)$$

где $\bar{x}_в$ – выборочная средняя; $x_i$ –варианты; $n$–число вариант.

***Пример 5.*** Имеются данные о поголовье бычков, поступивших на мясокомбинат:

19,32,31,25,30,32,27,18,26,30

Необходимо определить среднее число поголовья

Составим расчетную таблицу.

<table>
<tr><td rowspan="2">№ партии</td><td>Поголовье бычков, гол.</td><td rowspan="14">Решение<br>Ряд несгруппированный, значения не повторяются, вычисляем выборочную среднюю:<br><br>Среднее поголовье бычков в партии:<br>$\bar{x}_в = \frac{\sum x_i}{n} = \frac{270}{10} = 27$ гол.</td></tr>
<tr><td>$x$</td></tr>
<tr><td>1</td><td>19</td></tr>
<tr><td>2</td><td>32</td></tr>
<tr><td>3</td><td>31</td></tr>
<tr><td>4</td><td>25</td></tr>
<tr><td>5</td><td>30</td></tr>
<tr><td>6</td><td>32</td></tr>
<tr><td>7</td><td>27</td></tr>
<tr><td>8</td><td>18</td></tr>
<tr><td>9</td><td>26</td></tr>
<tr><td>10</td><td>30</td></tr>
<tr><td>Итого</td><td>$\sum x_i = 270$</td></tr>
</table>

б) Среднюю взвешенную применяют, если индивидуальные значения признака представлены несколькими объектами (*сгруппированный ряд*).

- средняя арифметическая взвешенная:

$$\bar{x}_в = \frac{\sum x_i f_i}{n}, \qquad (1.5)$$

где $f$–частота вариант.

- Если ряд интервальный, то при вычислении среднего арифметического в роли $x_i$ в формулах (4) и (5) представителем каждого интервала берется его середина.
- Если среднее арифметическое находится для генеральной совокупности, то его называют генеральной средней и обозначают $a = \bar{x}_г$.

Рассмотрим методику вычисления простой выборочной средней.

*2) Мода и медиана*

Средние величины, описанные выше, являются обобщающими характеристиками совокупности по тому или иному признаку. Вспомогательными характеристиками являются, так называемые, структурные средние, к которым относятся мода, квартили, децили, медиана и др. Наиболее употребляемыми являются мода и медиана.

***Мода*** – это величина, которая встречается в совокупности наиболее часто, то есть признак с наибольшей частотой. Этот показатель используется в тех случаях, когда требуется охарактеризовать наиболее часто встречающуюся величину признака (наиболее распространенный размер животноводческих ферм на сельскохозяйственных предприятиях, преобладающие цены на сельскохозяйственную продукцию и т. п. ).

а) В дискретном вариационном ряду модой является признак с наибольшей частотой в примере

б) В интервальном вариационном ряду моду находят по формуле:

$$Mo = x_{Mo} + h\,\frac{f_{Mo} - f_{Mo-1}}{2f_{Mo} - f_{Mo-1} - f_{Mo+1}}, \qquad (1.6)$$

где $Mo$ – мода;

$x_{Mo}$–нижняя граница модального интервала;

$h$– длина интервала;

$f_{Mo}$– частота модального интервала;

$f_{Mo-1}$– частота интервала, предшествующего модальному;

$f_{Mo+1}$– частота интервала, следующего за модальным.

*Модальным интервалом* является интервал с наибольшей частотой.

***Медианой*** называется величина, делящая численность упорядоченного вариационного ряда (расположенного в порядке возрастания или убывания признака) на две равные части. Медиана характеризует количественную границу значений изменяющегося признака, которыми обладает половина единиц совокупности. Например, если медианное значение удоя коровы составляет 4735 кг, то это означает, что половина коров имеет удой молока ниже 4735 кг и половина коров выше.

а) Формула расчета медианы для дискретного ряда:

$$M_e = \begin{cases} \frac{1}{2} \cdot \left( x_{\frac{n}{2}} + x_{\frac{n}{2}+1} \right), & \text{если } n - \text{четное}; \\ x_{\frac{n+1}{2}}, & \text{если } n - \text{нечетное}. \end{cases} \qquad (1.7)$$

Формула расчета медианы в интервальном вариационном ряду:

$$Me = x_{Me} + h_{Me} \frac{\frac{n}{2} - s_{Me-1}}{f_{Me}}, \qquad (1.8)$$

где $Me$ – медиана;

$x_{Me}$– нижняя граница медианного интервала;

$h$– длина интервала;

$n$ – обьем выборки;

$s_{Me-1}$ – сумма частот, накопленных в интервалах, предшествующих медианному;

$f_{Me}$– частота медианного интервала.

*Медианным интервалом* является интервал, накопленная частота которого равна или превышает половину суммы частот.

***Пример 6. Задание №1 (3(а))***

***Пример 7. Задание №2 (3(а))***

### ***Показатели колеблемости вариационных рядов.***

Для измерения вариации применяют различные показатели, из которых основными являются размах вариации (лимит), среднее линейное отклонение, дисперсия, среднее квадратическое отклонение, коэффициент вариации.

1) *Размах вариации* определяется как разница между наибольшим и наименьшим значениями признака:

$$R = x_{\max} - x_{\min} ,$$

где $x_{\min}$, $x_{max}$– минимальное и максимальное значение признака. Чем больше разность между $x_{max}$ и $x_{\min}$, тем, в общем, сильнее изменчивость признака.

*2) Среднее линейное отклонение* представляет собой среднюю арифметическую из абсолютных отклонений отдельных вариант от средней арифметической:

- простое $L = \frac{\sum |x - \bar{x}_в|}{n}$;

- взвешенное $L = \frac{\sum |x - \bar{x}_в| f}{n}$, где $L$– среднее линейное отклонение; $\bar{x}_в$ – средняя арифметическая; $x$–варианты;$n$– число вариант;$f$– частоты.

*3) Выборочную дисперсию* $D_в$ рассчитывают как среднюю арифметическую квадратов отклонений вариант от их выборочной средней:

а) простая (для несгруппированного ряда):

$$D_в = \frac{\sum (x_i - \bar{x}_в)^2}{n} \tag{1.9}$$

б) взвешенная (для сгруппированного ряда):

$$D_в = \frac{\sum (x_i - \bar{x}_в)^2 f_i}{n} \tag{1.10}$$

в) упрощенная формула:

$$D_в = \frac{1}{n} \sum_{i=1}^{4} x_i^2 \cdot f_i - \bar{x}_в^2. \tag{1.11}$$

- Для облегчения вычисления дисперсий используют следующие свойства:

1. Дисперсия не изменится, если все значения признака увеличить или уменьшить на постоянное число.
2. При умножений признака на постоянное число дисперсия умножается на квадрат этого числа.

*4) Среднее квадратическое отклонение* $\sigma$ равно корню квадратному из дисперсии:

- простое: $$\sigma = \sqrt{D_в} = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2}{n}} \qquad (1.12)$$

-взвешенное: $$\sigma = \sqrt{D_в} = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2 f_i}{n}} \qquad (1.13)$$

*5) Исправленная дисперсия и среднее квадратическое отклонение. Число степеней свободы.*

Выборочная дисперсия имеет систематическую ошибку, приводящую к уменьшению дисперсий. Чтобы это устранить, вводят поправку, умножая $D_в$ на $\frac{n}{n-1}$. В результате получают исправленную дисперсию: $S^2 = \frac{n}{n-1} D_в$. При малых выборках $S$ ощутимо отличается от $D_в$, например при *n=2* имеем $S^2 = 2D_в$.

Подставим $S^2 = \frac{n}{n-1} D_в$ в (12), (13) получим формулы для вычисления исправленной дисперсий:

$$S^2 = \frac{\sum (x_i - \bar{x}_в)^2}{n-1} \qquad (1.14) \qquad\qquad S^2 = \frac{\sum (x_i - \bar{x}_в)^2 f_i}{n-1} \qquad (1.15)$$

Формулы для вычисления исправленного среднего квадратического отклонения:

$$S = \sqrt{S^2} = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2}{n-1}} \qquad (1.16) \qquad\qquad S = \sqrt{S^2} = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2 f_i}{n-1}} \qquad (1.17)$$

Величина *k=n-1* в знаменателях формул (1.14 – 1.17) называется *числом* *степеней свободы.* Уясним его значение.

Если известен ряд, состоящий из *n* членов, то для него общей характеристикой является средняя арифметическая. Как может быть определено каждое *отдельное значение* ряда? Очевидно, что конкретное значение признака можно узнать, если известны а) средняя арифметическая и б) остальные наблюдения, т.е. *n-1* .

Иначе говоря, определение одного значения в данной совокупности зависит от остальных значений. Так, например, если известно, что 2 кролика в сумме весят 6 кг, а один из них весит 2,5 кг, то вес второго уже точно определен весом первого, т.е. имеется лишь одна степень свободы ( 2-1 =1). Если 3 кролика весят 5 кг, то вес одного всегда точно определяется весом двух других, между которыми уже возможна вариация, т.е. в этом случае имеется 2 степени свободы вариаций (3-1=2).

В общем виде при численности членов совокупности *n* число степеней свободы *k* будет равно *k=n-1*. При большом *n* разница между *n* и (*n-1)* , но при малых выборках (*n<30)* разница будет значительной. Так если *n=6* , а сумма квадратов отклонений равна 60, то средний квадрат отклонений от среднего будет равен не 60:6=10, а 60:5=12.

5) *Коэффициент вариации* представляет собой процентное отношение среднего квадратического отклонения к средней арифметической величине:

$$V = \frac{\sigma_в}{\bar{x}_в}100\%, \qquad (1.18)$$

Коэффициент вариаций является относительным показателем изменчивости выборки Изменчивость принято считать незначительной если $V<10\%$, средней если $10\%<V<20\%$ , значительной если $V>20\%$

Рассмотрим методику расчета показателей вариации для простого дискретного ряда.

***Пример 8.*** Имеются данные о поголовье бычков, поступивших на мясокомбинат

19,32,31,25,30,32,27,18,26,30

Необходимо определить показатели колеблемости поголовья бычков по всей совокупности.

Составим расчетную таблицу.

| № партии | Поголовье бычков, гол. | Поголовье бычков | |
|---|---|---|---|
| | | абсолютные отклонения | квадрат отклонений |
| | $x$ | $\lvert x-\bar{x}_в \rvert$ | $(x-\bar{x}_в)^2$ |
| 1 | 19 | 8 | 64 |
| 2 | 32 | 5 | 25 |
| 3 | 31 | 4 | 16 |
| 4 | 25 | 2 | 4 |
| 5 | 30 | 3 | 9 |
| 6 | 32 | 5 | 25 |
| 7 | 27 | 0 | 0 |
| 8 | 18 | 9 | 81 |
| 9 | 26 | 1 | 1 |
| 10 | 30 | 3 | 9 |
| Итого | $\sum x_i = 270$ | $\sum \lvert x-\bar{x}_в \rvert = 40$ | $\sum (x-\bar{x}_в)^2 = 234$ |

Колеблемость поголовья бычков определяется с помощью средней арифметической простой.

Среднее поголовье бычков в партии:

$$\bar{x}_в = \frac{\sum x_i}{n} = \frac{270}{10} = 27 \text{ гол.}$$

Размах вариации поголовья бычков:

$$R = x_{\max} - x_{\min} = 32 - 18 = 14 \text{ гол.}$$

Среднее линейное отклонение поголовья бычков от средней:

$$L = \frac{\sum \lvert x_i - \bar{x}_в \rvert}{n} = \frac{40}{10} = 4 \text{ гол.}$$

В задаче имеем малую выборку $n = 10 < 30$, следовательно дисперсию и среднее квадратическое отклонение поголовья бычков будет вычислять по исправленным формулам для дисперсий и среднего квадратического отклонения (14), (16) :

$$S^2 = \frac{\sum(x-\bar{x}_в)^2}{n-1} = \frac{234}{9} = 26 \qquad S = \sqrt{26} \approx 5 \text{ гол.}$$

Коэффициент вариации поголовья бычков:

$$V = \frac{S}{\bar{x}_в}100 = \frac{5}{27} \cdot 100\% = 3\%.$$

Среднее квадратическое отклонение показывает, что поголовье бычков по данной совокупности колеблется в пределах ±5 гол.,а коэффициент вариации равен ±3 % по отношению к среднему уровню, что указывает на незначительную изменчивость.

***Пример 6. Задание №1 (3(б))***

***Пример 7. Задание №2 (3(б))***

### *Упрощенный подсчет числовых характеристики (метод условной средней)*

Для упрощенного подсчета числовых характеристик используют построение вариационного ряда и условную среднюю.

Для упрощения счета взвешенного выборочного среднего и выборочной дисперсий применяют следующие формулы:

$$\bar{x}_в = A + \mathrm{h} \cdot \bar{u}, \tag{1.19}$$

$$D = S^2 = \mathrm{h}^2\left(\bar{u^2} - (\bar{u})^2\right) \tag{1.20}$$

где $A$ - условное среднее (близкое к ожидаемому среднему, или варианта с максимальной частотой);

$u_i = \frac{x_i - A}{h}$ - условное отклонение средних значений интервалов от выбранной средней условной (условные варианты);

$\bar{u} = \frac{\sum u_i f_i}{n}$ - среднее взвешенное условных вариант;

$\overline{u^2} = \frac{\sum u_i^2 f_i}{n}$ - среднее взвешенное от квадрата условных вариант.

*Вычислим среднее взвешенное условных вариант и среднее взвешенное от квадрата условных вариант (итоговые суммы 5-ого и 6-ого столбцов разделим на объем выборки =50, ).*

$\overline{u} = \frac{\sum u_i f_i}{n}$ $\overline{u^2} = \frac{\sum u_i^2 f_i}{n}$ -

*Зычислим числовые характеристики по упрощенным :*

Средняя выборочная:

$\overline{x}_в = A + h \cdot \overline{u}$

Выборочная дисперсия:

$S^2 = h^2\left(\overline{u^2} - (\overline{u})^2\right)$

Значения числовых характеристик при первом и втором способе вычисления совпали.

***Пример 8. Задание 2(4)***

### ***Показатели точности выборочной средней.***

*1)* *Ошибка выборочной средней* или ошибка выборки $S_{\bar{x}}$ является мерой отклонения выборочной средней $\overline{x}_в$ от средней всей (генеральной) совокупности $X_{ген}$. Ошибки выборки возникают вследствие неполной представительности выборочной совокупности. Величина ошибок зависит от степени изменчивости изучаемого признака *($S_в$)* и от объема выборки.

Ошибка выборочной средней прямо пропорциональна выборочному отклонению $S_в$ и обратно пропорциональна корню квадратному из объема выборки.

$$S_{\bar{x}} = \frac{S_в}{\sqrt{n}} \qquad (1.21)$$

*2)* Относительная ошибка средней – это ошибка выборки, выраженная в процентах от соответствующей средней

$$S_{\bar{x}\%} = \frac{S_{\bar{x}}}{\overline{x}_в} \cdot 100\% \qquad (1.22)$$

Относительную ошибку средней называют «точностью опыта» Этот показатель указывает на степень отображения свойств генеральной совокупности. Чем меньше ошибки в исследованиях, тем выше точность исследования.

1. если $S_{\bar{x}\%} \le 3\%$, точность опыта высокая;

2. если $S_{\bar{x}\%} = 3\% - 7\%$, точность опыта средняя;

3. если $S_{\bar{x}\%} > 8\%$, точность опыта низкая, исследования считаются ненадежными, опыты рекомендуют браковать.

***Пример 9. Задание 1(4)***

***Пример 10. Задание 2(5)***

### *Показатели распределения*

1) ***Коэффициентом асимметрии ("скошенности")*** *A* случайной величины *X* называется величина

$$A = \frac{\mu_3}{\sigma_X^3}$$

Если A > 0, то кривая распределения более полога справа от $M_0(X)$;

если A < 0, то кривая распределения более полога слева от $M_0(X)$.

*Коэффициент асимметрии* для выборки рассчитывается по формуле:

$$A_s = \frac{n}{(n-1)(n-2)} \cdot \frac{\mu_3}{s^3} = \frac{n}{(n-1)(n-2)} \sum \left( \frac{x - \bar{x}}{s} \right)^3,$$

где $A_s$– коэффициент асимметрии;

$\mu_3 = \dfrac{\sum (X - \bar{X}_в)^3}{n}$ – момент третьего порядка;

$s = \sqrt{\dfrac{\sum (x - \bar{x}_в)^2}{n-1}}$ – выборочное среднее квадратическое отклонение;

$n$ – число вариант.

**2)** ***Коэффициентом эксцесса ("островершинности")*** $\boldsymbol{\varepsilon}$ случайной величины *X* называется величина

$$E = \frac{\mu_4}{\sigma_X^4} - 3$$

Величина Е характеризует островершинность или плосковершинность распределения.

Для нормального закона распределения: *A = 0 и ε = 0;* остальные распределения сравниваются с нормальным:

Если $\boldsymbol{\varepsilon} > 0$ – более островершинные,

а распределения "плосковершинные" имеют $\varepsilon < 0$

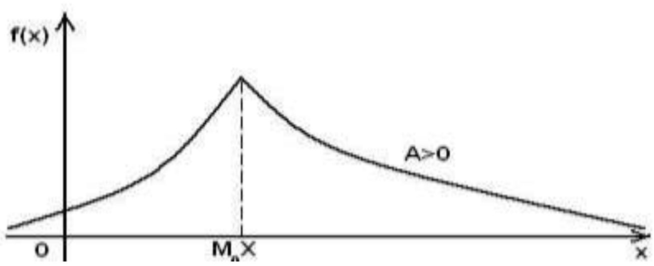
*Показатель эксцесса* для выборки рассчитывается по формуле:

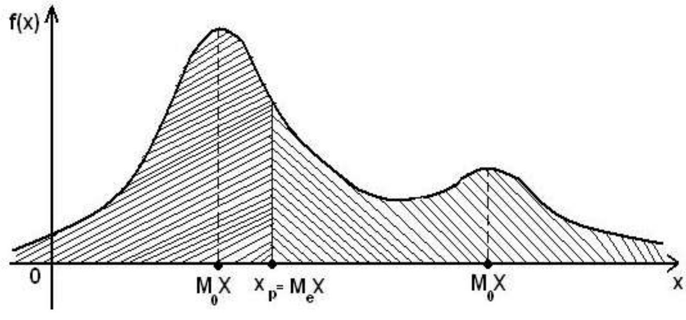
$$\varepsilon_k = \left[\frac{n(n-1)}{(n-1)(n-2)(n-3)} \cdot \frac{\mu_4}{s^4}\right] - \frac{3(n-1)^2}{(n-2)(n-3)} \text{ или}$$

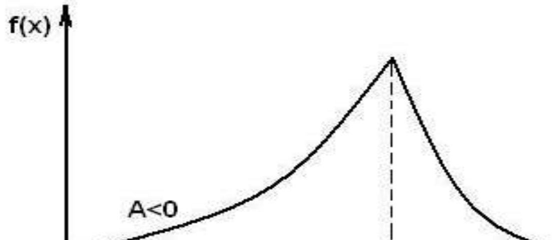
$$\varepsilon_k = \left[\frac{n(n-1)}{(n-1)(n-2)(n-3)} \sum\left(\frac{x-\bar{x}}{s}\right)^4\right] - \frac{3(n-1)^2}{(n-2)(n-3)},$$

где $\varepsilon_k$ – эксцесс;

$\mu_3 = \dfrac{\sum (X - \overline{X})^4}{n}$ – момент четвертого порядка.



*Пример.* Имеются выборочные данные о сервис - периоде по 25 коровам (табл. 1.9).

Т а б л и ц а 1.9.

**Сервис-период коров**

| № п/п | Сервис-период, дн. | Квадрат сервис-периода | Нормированное отклонение от средней | |
|---|---|---|---|---|
| | | | в третьей степени | в четвертой степени |
| | $x$ | $x^2$ | $\left(\frac{x-\bar{x}_в}{s}\right)^3$ | $\left(\frac{x-\bar{x}_в}{s}\right)^4$ |
| 1 | 77 | 5929 | -1,4272 | 1,6069 |
| 2 | 86 | 7396 | -0,0302 | 0,0094 |
| 3 | 99 | 9801 | 0,6477 | 0,5605 |
| 4 | 68 | 4624 | -7,3064 | 14,1777 |
| 5 | 83 | 6889 | -0,1980 | 0,1154 |
| 6 | 83 | 6889 | -0,1980 | 0,1154 |
| 7 | 98 | 9604 | 0,4650 | 0,3602 |
| 8 | 83 | 6889 | -0,1980 | 0,1154 |
| 9 | 89 | 7921 | -0,0001 | 0,0000 |
| 10 | 82 | 6724 | -0,3053 | 0,2056 |
| 11 | 77 | 5929 | -1,4272 | 1,6069 |
| 12 | 91 | 8281 | 0,0028 | 0,0004 |
| 13 | 83 | 6889 | -0,1980 | 0,1154 |
| 14 | 85 | 7225 | -0,0649 | 0,0261 |
| 15 | 92 | 8464 | 0,0124 | 0,0029 |

| | | | | |
|---|---|---|---|---|
| 16 | 106 | 11236 | 3,3668 | 5,0460 |
| 17 | 106 | 11236 | 3,3668 | 5,0460 |
| 18 | 88 | 7744 | -0,0022 | 0,0003 |
| 19 | 83 | 6889 | -0,1980 | 0,1154 |
| 20 | 109 | 11881 | 5,5480 | 9,8216 |
| 21 | 86 | 7396 | -0,0302 | 0,0094 |
| 22 | 107 | 11449 | 4,0143 | 6,3798 |
| 23 | 80 | 6400 | -0,6237 | 0,5328 |
| 24 | 87 | 7569 | -0,0108 | 0,0024 |
| 25 | 108 | 11664 | 4,7399 | 7,9620 |
| Итого | $\sum x = 2236$ | $\sum x^2 = 202918$ | $\sum\left(\frac{x-\bar{x}}{s}\right)^3 = 9,6454$ | $\sum\left(\frac{x-\bar{x}}{s}\right)^4 = 53,934$ |

Требуется рассчитать основные статистические показатели распределения, характеризующие данную выборочную совокупность.

Средняя продолжительность сервис-периода коров:

$$\bar{x}_в = \frac{\sum x}{n} = \frac{2236}{25} = 89,44 \text{ дн.}$$

Выборочная дисперсия исправленная (n<30) сервис-периода коров:

$$s^2 = \frac{\sum (x-\bar{x}_в)^2}{n-1} = \frac{n\sum x^2 - \left(\sum x\right)^2}{n(n-1)} = \frac{25 \cdot 202918 - 2236^2}{25 \cdot (25-1)} = 122,09.$$

Выборочное среднее квадратическое отклонение сервис-периода коров:

$$s = \sqrt{s^2} = \sqrt{122,09} = 11,049 \text{ дн.}$$

Асимметрия сервис-периода коров:

$$A_s = \frac{n}{(n-1)(n-2)} \sum\left(\frac{x-\bar{x}_в}{s}\right)^3 = \frac{25}{(25-1)\cdot(25-2)} \cdot 9,9454 = 0,45.$$

Эксцесс сервис-периода коров:

$$\varepsilon_k = \left[\frac{n(n-1)}{(n-1)(n-2)(n-3)} \sum\left(\frac{x-\bar{x}_в}{s}\right)^4\right] - \frac{3(n-1)^2}{(n-2)(n-3)} =$$

$$= \left[\frac{25\cdot(250-1)}{(25-1)\cdot(25-2)\cdot(25-3)} \cdot 53,934\right] - \frac{3\cdot(25-1)^2}{(25-2)\cdot(25-3)} = -0,529.$$

Полученные значения асимметрии и эксцесса показывают, что данное распределение имеет правостороннюю асимметрию и более высокую вариацию, чем при нормальном распределении.

### *П.3. Точечные и интервальные оценки. Точность и надежность оценки. Доверительная вероятность, доверительный интервал.*

### *Точечные оценки.*

Генеральные совокупности характеризуются некоторыми постоянными числовыми характеристиками распределения. По выборкам можно найти оценки этих характеристик. Из-за случайности выборок значения оценок одной числовой характеристики, вычисленные по разным выборкам из одной и той же генеральной совокупности, бывают, как правило, различными.

Пусть $\theta$ – неизвестный параметр распределения (числовой характеристики генеральной совокупности $X$), а $\theta_n$ – его оценка.

*Определение 1.* ***Статистической оценкой*** $\theta_n$ неизвестного параметра $\theta$ теоретического распределения называют функцию $f(x_1,\ x_2,\ldots x_n)$ от наблюдаемых случайных величин $x_1, x_2, \ldots, x_n$.

• Оценки неизвестного параметра можно находить различными способами. И для того, чтобы установить, какая из оценок лучше, надо знать основные свойства (виды) оценок.

*Определение 2.* Оценка $\theta_n$ неизвестного параметра $\theta$, определяющая одну точку на числовой оси, называется *точечной оценкой*.

*Виды точечных оценок:*

– **несмещенная** – *это такая оценка* $\theta_n$*, среднее значение (математическое ожидание) которой равно оцениваемому параметру* $\theta$ *при любом объеме выборки, т.е.* $M(\theta_n)=\theta$ *( выборочная средняя* $(\bar{x}_в)$ *– это несмещенная оценка генеральной средней (математического ожидания).*

*– если это условие не выполняется, то оценку называют* ***смещенной****, при этом смещение определяется как разность* $(M(\theta_n)-\theta\ ))$ *(выборочная дисперсия при малом объеме выборки* $(D_в)$*– это смещенная оценка генеральной дисперсии.*

### *Интервальные оценки. Надежность оценки.*

Как уже было сказано выше, точечной называют оценку, которая определяется одним числом. Все оценки, рассмотренные выше, - точечные. При выборке малого объёма точечная оценка может значительно отличаться от оцениваемого параметра, т. е. приводить к грубым ошибкам. По этой причине при небольшом объёме выборки следует пользоваться интервальными оценками.

Если на основании имеющихся данных (выборки из генеральной совокупности) находят оценку $\tilde{\theta}(x_1, x_2 ... x_n)$ параметра $\theta$ генеральной совокупности, то при этом понимают, что величина $\tilde{\theta}$ является лишь приближённым значением неизвестного параметра $\theta$ даже в том случае, если эта оценка несмещенная. Полученная оценка называется точечной оценкой. Чтобы получить представление о точности и надёжности оценки $\tilde{\theta}$ неизвестного параметра $\theta$, используют *интервальные оценки*.

*Определение 2.* ***Интервальной*** называют оценку, которая определяется двумя числами – концами интервала. Интервальные оценки позволяют установить точность и надёжность оценок (смысл этих понятий выясняется ниже).

Пусть найденная по данным выборки статистическая характеристика $\tilde{\theta}$ служит оценкой неизвестного параметра $\theta$. Будем считать $\theta$ постоянным числом ($\theta$ может быть и случайной величиной). Ясно, что $\tilde{\theta}$ тем точнее определяет параметр $\theta$, чем меньше абсолютная величина разности $|\tilde{\theta} - \theta|$.

Другими словами, если $\varepsilon > 0$ и $\left|\tilde{\theta} - \theta\right| < \varepsilon$, то чем меньше $\varepsilon$, тем оценка точнее. Таким образом, положительное число $\varepsilon$ характеризует ***точность*** оценки.

Однако статистические методы не позволяют категорически утверждать, что оценка $\tilde{\theta}$ удовлетворяет неравенству $\left|\tilde{\theta} - \theta\right| < \varepsilon$; можно лишь говорить о вероятности $\gamma$, с которой это неравенство осуществляется.

*Определение 3.* ***Надёжностью*** (доверительной вероятностью) оценки по $\tilde{\theta}$ называют вероятность $\gamma$, с которой осуществляется неравенство $\left|\tilde{\theta} - \theta\right| < \varepsilon$.

Пусть вероятность того, что $\left|\tilde{\theta} - \theta\right| < \varepsilon$, равна $\gamma$: $P\left(\left|\tilde{\theta} - \theta\right| < \varepsilon\right) = \gamma$ .

Заменив неравенство $\left|\tilde{\theta} - \theta\right| < \varepsilon$ равносильным ему двойным неравенством

$$\left(\tilde{\theta} - \varepsilon < \theta < \tilde{\theta} + \varepsilon\right) = \gamma \text{ имеем } P\left(\tilde{\theta} - \varepsilon < \theta < \tilde{\theta} + \varepsilon\right) = \gamma. \tag{1.23}$$

Это соотношение следует понимать так: вероятность того, что интервал $\left(\tilde{\theta} - \varepsilon < \theta < \tilde{\theta} + \varepsilon\right)$ заключает в себе (покрывает) неизвестный параметр $\theta$, равна $\gamma$

*Определение 4.* ***Доверительным*** называют интервал $\left(\tilde{\theta} - \varepsilon < \theta < \tilde{\theta} + \varepsilon\right)$, который покрывает неизвестный параметр с заданной надёжностью $\gamma$.

- Надёжность оценки $\gamma$ обычно задают заранее, близкой к единице (обычно $\gamma = 0{,}95$; $0{,}99$; $0{,}999$), тогда зная закон распределения изучаемой величины, находят доверительный интервал.

  Поясним смысл, который имеет заданная надёжность. Надёжность $\gamma$ =0,95 указывает, что если произведено достаточно большое число выборок, то 95% из них определяет такие доверительные интервалы, в которых параметр действительно заключён

*Определение 5.* Число $\alpha = (1 - \gamma)$ называется ***уровнем значимости*** или ***вероятностью ошибки***.

Так, если $\gamma$=0,95, то $\alpha = (1-0{,}95) = 0{,}05$, то лишь в 5 % случаев оцениваемый параметр может выйти за границы доверительного интервала, т.е. в пяти случаев из 100 есть вероятность совершить ошибку.

### *Построение доверительного интервала для генеральной средней*

Предположим, что задана доверительная вероятность $\gamma$, необходимо найти доверительный интервал для генеральной средней

Оценкой генеральной средней $X_{ген.}$ служит выборочная средняя $\overline{x}_в$, тогда доверительный интервал с надежностью $\gamma$ для генеральной средней имеет вид:

$$X_{ген.} \in \left(\overline{x}_в - \varepsilon;\ \overline{x}_в + \varepsilon\right),$$

где $\varepsilon = S_{\overline{x}} \cdot t$ - предельная ошибка выборки, $S_{\overline{x}} = \dfrac{\sigma_в}{\sqrt{n}}$ - ошибка выборочной средней

$$X_{ген.} \in \left(\overline{x}_в - t \cdot S_{\overline{x}};\ \overline{x}_в + t \cdot S_{\overline{x}}\right). \qquad (1.24)$$

или

$$X_{ген.} \in \left(\overline{X}_в - \frac{t\sigma_в}{\sqrt{n}};\ \overline{X}_в + \frac{t\sigma_в}{\sqrt{n}}\right). \qquad (1.25)$$

1) Если выборка ***большая*** *(n>30)*, то выборочная средняя $\overline{x}_в$ приближённо имеет нормальный закон распределения и значение $t$ по заданной надежности $\gamma$ находят с помощью таблицы значений *интегральной функции Лапласа* (*табл*ица 3 *приложения*) из следующего соотношения:

$$2\Phi(t) = \gamma.$$

*Например,* для надежности: $2\Phi(t) = 0{,}95 \Rightarrow \Phi(t) = 0{,}475 \Rightarrow t = 1{,}96.$

Получим: $\boxed{\gamma_1 = 0{,}95 \Rightarrow t_2 = 1{,}96. \quad \gamma_2 = 0{,}99 \Rightarrow t_2 = 2{,}58.}$

2) Если выборка ***малая*** *(n<30)*,то предельная ошибка малой выборки составит:

$$\varepsilon_{м.в.} = \frac{t_{\gamma,n-1} \cdot S_в}{\sqrt{n}}$$

где $S_в$ - исправленное среднее квадаратическое отклонение, величину $t_{\gamma,n-1}$ при заданной доверительной вероятности $\gamma$ и числе степеней свободы $k = n - 1$ находят по таблице *значений критерия Стьюдента* (*табл.* 4 *приложения ).*

Тогда доверительный интервал для генеральной средней (малая выборка $n$<30) имеет вид:

$$\bar{x}_в - \frac{t_{\gamma,n-1} \cdot S_в}{\sqrt{n}} \le X_{ген.} \le \bar{x}_{в.} + \frac{t_{\gamma,n-1} \cdot S_в}{\sqrt{n}}. \tag{1.26}$$

или

$$\bar{x}_в - t_{\gamma,n-1} \cdot S_{\bar{x}} \le X_{ген.} \le \bar{x}_{в.} + t_{\gamma,n-1} \cdot S_{\bar{x}}. \tag{1.27}$$

***Пример 11. Задание 1(5)***
***Пример 12. Задание 2(6)***

| ***Основные числовые характеристики.*** | | |
|---|---|---|
| Наименование характеристик | для простого дискретного (несгруппированного ряда) | для дискретного (сгруппированного ряда) или интервального |
| $\bar{x}_в$<br>Выборочная средняя | $\bar{x}_в = \frac{\sum_{i=1}^{n} x_i}{n}$ | $\bar{x}_в = \frac{\sum_{i=1}^{n} x_i \cdot f_i}{n}$; |
| $D_в = S_в^2$<br>Дисперсия | $D_в = \frac{\sum (x_i - \bar{x}_в)^2}{n}; (n > 30)$<br>$S_в{}^2 = \frac{\sum (x_i - \bar{x}_в)^2}{n-1}; (n < 30)$ | $D_в = \frac{\sum (x_i - \bar{x}_в)^2 \cdot f_i}{n}; (n > 30)$<br>$S_в^2 = \frac{\sum (x_i - \bar{x}_в)^2 \cdot f_i}{n-1}; (n < 30)$ |
| $\sigma_в = S_в$<br>Среднее квадратическое отклонение | $\sigma_в = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2}{n}}; (n > 30)$ | $\sigma_в = \sqrt{\frac{\sum (x_i - \bar{x}_в)^2 \cdot f_i}{n}}$; |

| | | |
|---|---|---|
| | $S_в=\sqrt{\dfrac{\sum(x_i-\bar{x}_в)^2}{n-1}};(n<30)$ | $S_в=\sqrt{\dfrac{\sum(x_i-\bar{x}_в)^2\cdot f_i}{n}};$ |
| Мода | Мода – значение признака, имеющего максимальную частоту | $Mo=x_{Mo}+h\ \dfrac{f_{Mo}-f_{Mo-1}}{2f_{Mo}-f_{Mo-1}-f_{Mo+1}}$ |
| Медиана | $M_e=\begin{cases}\dfrac{1}{2}\cdot\left(x_{\frac{n}{2}}+x_{\frac{n}{2}+1}\right), & если\ n-четное;\\ x_{\frac{n+1}{2}}, & если\ n-нечетное.\end{cases}$ | $Me=x_{Me}+h_{Me}\dfrac{\frac{n}{2}-s_{Me-1}}{f_{Me}}$ |
| ***Показатель изменчивости, однородности выборки.*** | | |
| $V_в$<br>Коэффициент вариации | $V_в=\dfrac{S_в}{\bar{x}_в}\cdot100\%;$<br>1. если *V<10%,* то выборка считается однородной, изменчивость незначительная; 2. если *10% < V < 20%,* то выборка считается средне однородной и средне изменчивость; 3. если *V > 20%,* то выборка считается неоднородной, сильно изменчивой. | |
| ***Показатели точности выборочной средней.*** | | |
| $S_{\bar{x}}$<br>Абсолютная ошибка средней выборочной | $S_{\bar{x}}=\dfrac{S_в}{\sqrt{n}};$<br>Ошибка выборочной средней является мерой отклонения выборочной средней $\bar{x}_в$ от средней генеральной $\bar{x}_г$.Ошибки выражают в тех же единицах измерения, что и признак и приписывают к соответствующим средним со знаком: $\pm$, т.е. $(\bar{x}_в\pm S_{\bar{x}})$ | |
| $S_{\bar{x}\%}$<br>Относительная ошибка выборочной (показатель точности) | $S_{\bar{x}\%}=\dfrac{S_{\bar{x}}}{\bar{x}_в}\cdot100\%;$<br>Этот показатель указывает на степень отображения свойств генеральной совокупности. Чем меньше ошибки в исследованиях, тем выше точность исследования.<br>1. если $S_{\bar{x}\%}\le3\%$, точность опыта высокая; 2. если $S_{\bar{x}\%}=3\%-6\%$, точность опыта средняя; 3. если $S_{\bar{x}\%}>6\%$, точность опыта низкая, исследования считаются ненадежными. | |

## Индивидуальное задание №1 «Исследование вариационных рядов»

***Задание №1. Исследование дискретных вариационных рядов.***

Требуется:

1) Составить вариационный дискретный ряд;
2) Построить полигон;
3) Вычислить числовые характеристики: а) характеристики положения - среднюю выборочную, моду, медиану; б) характеристики изменчивости - выборочную дисперсию; выборочное среднее квадаратическое отклонение; коэффициент вариации;

4) показатели точности выборочной средней;
5) доверительные интервалы для генеральной средней с надежностью 0,99;0,95.

***Задание №2. Исследование интервальных вариационных рядов.***

Требуется:

1) составить интервальный ряд распределения;
2) построить полигон и гистограмму;
3) вычислить числовые характеристики: а) характеристики положения - среднюю выборочную, моду, медиану; б) характеристики изменчивости - выборочную дисперсию; выборочное среднее квадаратическое отклонение; коэффициент вариации.
4) вычислить числовые характеристики упрощенным способом, с помощью метода условного нуля;
5) показатели точности выборочной средней;
6) доверительные интервалы для генеральной средней с надежностью 0,99;0,95.

ВАРИАНТ I

Дана выборочная совокупность урожайности гречихи в ц/га.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 5,7 | 5,1 | 4,6 | 8,1 | 9,6 | 6,6 | 9,8 | 8,9 | 8,0 | 5,5 |
| 10,6 | 5,6 | 4,7 | 5,5 | 7,4 | 9,5 | 5,9 | 6,2 | 4,9 | 4,9 |
| 6,6 | 5,3 | 6,5 | 4,5 | 7,3 | 9,5 | 6,6 | 5,3 | 5,3 | 6,6 |
| 9,4 | 5,5 | 8,4 | 6,6 | 4,4 | 4,9 | 7,3 | 7,8 | 9,8 | 9,0 |
| 4,4 | 5,7 | 6,8 | 8,6 | 6,9 | 7,2 | 6,2 | 5,6 | 6,9 | 5,6 |
| 6,7 | 9,6 | 6,2 | 7,7 | 9,8 | 9,7 | 7,9 | 5,5 | 8,6 | 8,1 |
| 7,4 | 6,6 | 5,6 | 7,8 | 4,7 | 7,5 | 6,1 | 4,8 | 6,5 | 7,8 |
| 9,0 | 7,2 | 8,7 | 5,3 | 6,4 | 5,5 | 8,3 | 8,5 | 7,9 | 8,4 |

ВАРИАНТ 2

Дана выборочная совокупность массы 10 растений сои в граммах на опытном поле.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 12 | 13 | 12,5 | 12 | 14 | 11 | 12 | 13 | 12 |
| 11 | 13,5 | 13 | 11 | 13 | 12,5 | 10 | 12 | 12,5 |
| 13 | 12 | 10 | 10 | 12 | 13 | 13 | 13 | 15 |
| 12 | 11 | 11 | 13 | 13 | 12 | 13,5 | 11 | 13 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 12 | 12 | 12 | 14 | 12 | 10 | 12,5 | 12 | 12 |
| 14 | 13 | 12 | 12 | 12,5 | 11 | 13 | 11,5 | 13,5 |
| 13 | 15 | 13 | 11 | 15 | 12 | 12 | 12,5 | 13 |

ВАРИАНТ 3

Высота растений пшеницы с использованием в качестве удобрений золу Лучегорской ГРЕС.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 30 | 23 | 26 | 28 | 30 | 29 | 25 | 28 |
| 30 | 27 | 24 | 29,5 | 37 | 29 | 24 | 23 |
| 21 | 30 | 31 | 29 | 29 | 23 | 24 | 26 |
| 27 | 28 | 30 | 28 | 24 | 26 | 23 | 26 |
| 22 | 28 | 29 | 25 | 29 | 23 | 22 | 28 |
| 31 | 23 | 28 | 29 | 30 | 27 | 25 | 27 |
| 26 | 21 | 37 | 25 | 23 | 24 | 30 | 31 |

ВАРИАНТ 4

Дана выборочная совокупность густоты всходов кукурузы на 10 $м^2$.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 117 | 10 | 62 | 71 | 108 | 100 | 107 | 74 |
| 88 | 113 | 68 | 103 | 98 | 103 | 98 | 86 |
| 87 | 131 | 74 | 86 | 97 | 120 | 87 | 99 |
| 72 | 99 | 89 | 84 | 82 | 100 | 88 | 88 |
| 110 | 100 | 101 | 90 | 100 | 99 | 76 | 72 |
| 106 | 103 | 98 | 99 | 104 | 79 | 110 | 108 |
| 103 | 108 | 87 | 79 | 110 | 89 | 105 | 103 |
| 89 | 98 | 73 | 92 | 83 | 98 | 106 | 104 |

ВАРИАНТ 5

Дана выборочная урожайность сои по Приморскому краю ц/га.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 10,3 | 8,9 | 10,3 | 11,5 | 8,6 | 9,8 | 10,2 | 9,9 |
| 8,8 | 9,4 | 9,0 | 7,9 | 9,0 | 9,3 | 9,5 | 9,1 |
| 10,1 | 7,9 | 7,7 | 8,2 | 8,5 | 7,8 | 8,2 | 8,3 |
| 11 | 7,6 | 11,2 | 9,6 | 9,7 | 8,8 | 8,0 | 8,7 |
| 9,7 | 7,3 | 9,3 | 9,9 | 9,5 | 11,0 | 10,7 | 10,0 |
| 12,0 | 9,3 | 13,0 | 7,7 | 12 | 10,0 | 10,2 | 8,9 |
| 9,0 | 7,8 | 11,0 | 7,9 | 10,2 | 8,8 | 10,3 | 9,0 |

ВАРИАНТ 6

Дана выборочная совокупность учета всходов ячменя с 1 $м^2$ на опытном поле.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 72 | 63 | 45 | 33 | 49 | 22 | 42 | 48 | 52 | 50 |
| 50 | 59 | 30 | 65 | 50 | 45 | 52 | 53 | 71 | 52 |
| 56 | 50 | 59 | 41 | 57 | 51 | 41 | 62 | 68 | 81 |
| 27 | 64 | 49 | 57 | 50 | 46 | 35 | 80 | 77 | 60 |

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 74 | 65 | 48 | 89 | 64 | 50 | 49 | 64 | 84 | 69 |
| 59 | 51 | 55 | 37 | 46 | 39 | 56 | 55 | 75 | 68 |
| 77 | 59 | 72 | 35 | 35 | 68 | 54 | 71 | 77 | 68 |

ВАРИАНТ 7

Данные по высоте растений кукурузы, начиная с третьего листа на 10 $м^2$ на опытном поле.

| | | | | | | |
|---|---|---|---|---|---|---|
| 64 | 78 | 86 | 81 | 77 | 78 | 80 |
| 70 | 79 | 75 | 80 | 85 | 75 | 77 |
| 73 | 80 | 81 | 76 | 81 | 78 | 83 |
| 80 | 80 | 86 | 78 | 84 | 77 | 81 |
| 86 | 81 | 87 | 73 | 81 | 86 | 85 |
| 78 | 86 | 78 | 81 | 80 | 90 | 80 |
| 76 | 85 | 79 | 85 | 77 | 91 | 70 |
| 80 | 75 | 73 | 80 | 83 | 78 | 81 |

ВАРИАНТ 8

Дана выборочная совокупность урожайности картофеля в ц/га по Приморскому краю.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 92 | 86 | 71 | 89 | 80 | 84 | 75 | 82 |
| 91 | 82 | 60 | 91 | 81 | 70 | 82 | 77 |
| 96 | 93 | 65 | 80 | 80 | 95 | 94 | 77 |
| 100 | 79 | 81 | 89 | 78 | 75 | 98 | 83 |
| 88 | 88 | 90 | 95 | 77 | 90 | 100 | 90 |
| 82 | 81 | 86 | 82 | 74 | 82 | 94 | 75 |
| 81 | 85 | 96 | 70 | 75 | 80 | 75 | 92 |

ВАРИАНТ 9

Дана выборочная совокупность массы 10 растений сои в граммах на опытном поле.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 12 | 10 | 12 | 12 | 10 | 15 | 14,5 | 9 | 15 |
| 13 | 13 | 13 | 13 | 14 | 10 | 10 | 9,5 | 14,5 |
| 12 | 11 | 12 | 10 | 12 | 12 | 15 | 10 | 10 |
| 13 | 13 | 12 | 16 | 17 | 15 | 12 | 8 | 10 |
| 13 | 11 | 12 | 13 | 14 | 12 | 13 | 11 | 15 |
| 13 | 9 | 11 | 13 | 12 | 12 | 12 | 13 | 12 |
| 12 | 13 | 14 | 15 | 12 | 10 | 11 | 16 | 13 |

ВАРИАНТ 10

Даны сведения об урожайности сена многолетних трав в ц/га.

| | | | | |
|---|---|---|---|---|
| 55,9 | 26,0 | 39,3 | 69,3 | 59,8 |
| 32,0 | 32,9 | 57,3 | 28,2 | 54,1 |

| | | | | |
|---|---|---|---|---|
| 29,1 | 58,1 | 17,4 | 45,9 | 71,8 |
| 28,4 | 43,4 | 59,3 | 39,0 | 49,4 |
| 40,1 | 50,5 | 26,8 | 34,3 | 66,0 |
| 17,1 | 47,5 | 25,1 | 38,8 | 30,6 |
| 35,5 | 60,2 | 21,5 | 34,1 | 59,5 |
| 49,9 | 45,0 | 27,3 | 53,8 | 70,5 |

ВАРИАНТ 11

Дана выборочная совокупность урожайности проса в ц/га.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 5,6 | 7,8 | 4,5 | 6,4 | 5,6 | 8,6 | 5,9 | 7,2 | 4,7 |
| 5,9 | 7,5 | 9,9 | 5,1 | 5,5 | 4,3 | 6,6 | 7,0 | 6,4 |
| 6,8 | 5,9 | 8,7 | 5,1 | 6,8 | 5,8 | 4,2 | 7,3 | 5,4 |
| 9,0 | 5,5 | 4,7 | 7,9 | 6,7 | 7,1 | 8,8 | 5,8 | 7,3 |
| 8,7 | 5,6 | 5,5 | 6,3 | 6,2 | 5,0 | 7,1 | 9,8 | 9,2 |
| 8,9 | 7,3 | 7,1 | 7,3 | 8,4 | 7,6 | 6,9 | 7,4 | 6,7 |
| 7,9 | 7,2 | 4,8 | 7,0 | 5,6 | 7,0 | 7,0 | 5,4 | 8,8 |
| 6,8 | 7,4 | 5,5 | 6,8 | 5,9 | 7,1 | 6,9 | 6,6 | 5,1 |

ВАРИАНТ 12

Посевные участки равные по площади дали урожай хлопка - сырца в ц .

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| 20 | 21 | 23 | 18 | 27 | 25 | 25 | 26 | 20 | 22 |
| 22 | 20 | 18 | 27 | 26 | 25 | 24 | 24 | 20 | 21 |
| 23 | 20 | 21 | 22 | 24 | 25 | 20 | 18 | 19 | 20 |
| 21 | 22 | 22 | 23 | 25 | 24 | 24 | 25 | 26 | 28 |
| 20 | 21 | 21 | 20 | 18 | 19 | 21 | 22 | 23 | 28 |

ВАРИАНТ 13

Данные по высоте растений ячменя на делянках с различной обработкой почвы

| | | | | | | |
|---|---|---|---|---|---|---|
| 79 | 83 | 81 | 78 | 81 | 72 | 76 |
| 80 | 80 | 64 | 77 | 76 | 76 | 73 |
| 77 | 79 | 63 | 83 | 83 | 83 | 78 |
| 83 | 84 | 85 | 80 | 89 | 85 | 80 |
| 74 | 73 | 79 | 77 | 90 | 88 | 79 |
| 81 | 79 | 80 | 64 | 72 | 73 | 83 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 76 | 77 | 83 | 90 | 76 | 80 | 78 |

ВАРИАНТ 14

Высота растений пшеницы с использованием в качестве удобрений золу Лучегорской ГРЭС.

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 29 | 26 | 27 | 30 | 37 | 28 | 25 | 27 | 28 |
| 28 | 31 | 24 | 28 | 22 | 27 | 30 | 29 | 24 |
| 26 | 23 | 29 | 24 | 33 | 26 | 34 | 31 | 26 |
| 30 | 31 | 26 | 25 | 25 | 22 | 28 | 32 | 27 |
| 19 | 25 | 29 | 30 | 29 | 28 | 21 | 18 | 26 |
| 26 | 24 | 31 | 28 | 36 | 27 | 31 | 29 | 24 |
| 24 | 25 | 26 | 22 | 31 | 29 | 28 | 27 | 29 |

ВАРИАНТ 15

Дана выборочная совокупность урожайности овса в ц/га.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 15,8 | 14,1 | 14,0 | 10,5 | 15,2 | 9,0 | 16,3 | 13,4 |
| 9,3 | 14 | 11,6 | 13,0 | 17,6 | 13,4 | 18,5 | 13,9 |
| 16,0 | 11,2 | 11,2 | 10,1 | 10,0 | 17,7 | 8,2 | 16,5 |
| 18,4 | 16,3 | 18,5 | 14,5 | 10,5 | 15,2 | 10,3 | 10,2 |
| 16,0 | 13,8 | 10,2 | 14,8 | 12,4 | 10,7 | 14,9 | 15,1 |
| 15,5 | 15,7 | 11,8 | 16,5 | 15,0 | 8,0 | 14,3 | 9,7 |
| 11,9 | 18,3 | 14,7 | 15,0 | 16,2 | 11,1 | 15,2 | 11,3 |
| 15,4 | 14,2 | 14,4 | 15,1 | 18,3 | 17,9 | 12,5 | 18,4 |

ВАРИАНТ 16

Данные по высоте растений ячменя на делянках с различной обработкой почвы.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 90 | 65 | 86 | 82 | 86 | 83 | 83 | 78 |
| 83 | 71 | 87 | 84 | 87 | 82 | 77 | 79 |
| 78 | 78 | 80 | 83 | 87 | 83 | 80 | 83 |
| 79 | 73 | 88 | 77 | 79 | 75 | 81 | 82 |
| 86 | 81 | 90 | 85 | 80 | 80 | 83 | 86 |
| 81 | 83 | 85 | 82 | 75 | 90 | 93 | 88 |
| 83 | 82 | 87 | 79 | 86 | 91 | 84 | 86 |
| 90 | 93 | 82 | 80 | 89 | 77 | 80 | 95 |

ВАРИАНТ 17

Учёт всходов кукурузы в фазе 3-го листа на 10 погонных метрах на опытном поле.

| | | | | | | |
|---|---|---|---|---|---|---|
| 59 | 84 | 98 | 63 | 50 | 77 | 54 |
| 75 | 100 | 83 | 54 | 52 | 66 | 59 |
| 82 | 82 | 59 | 68 | 66 | 104 | 80 |
| 80 | 64 | 70 | 63 | 60 | 87 | 49 |
| 69 | 45 | 75 | 69 | 62 | 70 | 50 |
| 60 | 100 | 83 | 66 | 104 | 54 | 59 |
| 63 | 70 | 54 | 62 | 87 | 80 | 70 |
| 64 | 75 | 63 | 60 | 70 | 59 | 75 |

ВАРИАНТ 18

Даны сведения об урожайности пшеницы (ц/га).

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 14,3 | 10,1 | 15,2 | 19,9 | 15,9 | 8,4 | 17,7 | 12,6 |
| 19,1 | 15,7 | 12,1 | 16,9 | 21,7 | 10,4 | 16,3 | 14,5 |
| 13,5 | 13,8 | 18,6 | 9,7 | 19,0 | 12,6 | 17,8 | 14,4 |
| 12,2 | 10,3 | 20,6 | 17,3 | 12,3 | 10,2 | 12,5 | 14,8 |
| 21,7 | 15,1 | 13,3 | 13,7 | 18,2 | 9,0 | 16,4 | 12,1 |
| 19,2 | 14,4 | 13,5 | 18,3 | 10,9 | 11,8 | 14,6 | 15,7 |
| 13,0 | 17,4 | 10,4 | 8,2 | 14,0 | 18,4 | 14,8 | 10,5 |
| 17,4 | 13,9 | 12,7 | 11,5 | 13,5 | 15,7 | 12,5 | 20,3 |
| 11,8 | 18,8 | 15,6 | 8,8 | 16,2 | 14,8 | 10,7 | 9,3 |

ВАРИАНТ 19

Дана выборочная совокупность урожайности риса в ц/га.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 43,6 | 50,6 | 42,0 | 35,4 | 37,9 | 38,6 | 47,1 | 32,2 |
| 25,2 | 40,0 | 38,9 | 5,5 | 37,8 | 37,9 | 35,8 | 36,2 |
| 46,6 | 42,4 | 38,2 | 36,0 | 28,9 | 36,5 | 38,1 | 35,5 |
| 36,2 | 39,2 | 39,1 | 38,2 | 33,8 | 30,5 | 39,9 | 39,2 |
| 30,8 | 39,1 | 29,8 | 39,0 | 27,8 | 30,4 | 39,4 | 38,9 |
| 41,9 | 39,0 | 30,0 | 39,3 | 33,0 | 43,2 | 48,5 | 38,8 |
| 35,0 | 34,9 | 38,7 | 27,8 | 39,3 | 38,7 | 49,9 | 35,5 |
| 24,0 | 35,3 | 32,8 | 28,9 | 37,2 | 40,6 | 30,7 | 34,2 |

ВАРИАНТ 20

Дана выборочная совокупность урожайности сои по Приморскому краю в ц/га.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 7,4 | 8,8 | 10,2 | 9,8 | 7,4 | 10,3 | 6,9 | 8,6 |
| 7,1 | 7,7 | 9,5 | 9,1 | 7,1 | 9,6 | 7,6 | 9,0 |
| 8,3 | 6,5 | 8,2 | 7,3 | 12,2 | 7,4 | 7,1 | 9,7 |
| 7,2 | 9,3 | 8,0 | 8,6 | 8,5 | 8,8 | 6,8 | 8,5 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 10,5 | 9,0 | 7,8 | 11,5 | 9,2 | 6,3 | 8,9 | 9,4 |
| 12,0 | 8,5 | 9,6 | 9,5 | 7,4 | 7,4 | 8,6 | 12,0 |
| 11,5 | 6,9 | 9,7 | 7,1 | 9,6 | 8,5 | 6,9 | 11,0 |

# РАЗДЕЛ 2. Проверка статистических гипотез.

## П.1. Основные понятия. Виды статистических гипотез. Виды ошибок при проверке статистических гипотез.

Ранее нами было показано, что выборочные характеристики являются оценками генеральных параметров, которые, как правило, остаются неизвестными. Там же описаны точечные и интервальные способы оценки неизвестных параметров по значениям выборочных характеристик.

На разных этапах статистического исследования возникает необходимость в формулировании и экспериментальной проверке различных предположений о характеристиках генеральной совокупности по выборочным данным.

***Определение 2.1. Статистическими гипотезами*** называют предположения или допущения относительно параметров сравниваемых групп, которое выражено в терминах вероятности и может быть проверено по выборочным характеристикам.

В основном рассматривают гипотезы:

1) Гипотеза о виде неизвестного распределения:

   часто необходимо знать закон распределения генеральной совокупности. Если закон распределения неизвестен, но имеются основания предположить, что он имеет определённый вид (назовём его А), выдвигают гипотезу: генеральная совокупность распределена по закону А. Таким образом, в этой гипотезе речь идёт о виде предполагаемого распределения.

2) Гипотеза о параметрах известных распределений:

   возможен случай, когда закон распределения известен, а его параметры неизвестны. Если есть основания предположить, что неизвестный параметр $\Theta$ равен определённому значению $\Theta_0$, выдвигают гипотезу: $\Theta = \Theta_0$. Таким образом, в этой гипотезе речь идёт о предполагаемой величине параметра одного известного распределения.

3) Гипотеза о равенстве параметров двух или нескольких распределений: например, делают *сравнительные оценки* параметров по разности, наблюдаемой между сравниваемыми выборками. Это важно, так как ни одно исследование не обходится без сравнений. Сравнивать приходится данные опыта с контролем, продуктивность одной группы животных с продуктивностью другой и т. д.

Гипотеза «на Марсе есть жизнь» не является статистической, поскольку в ней не идёт речь ни о виде, ни о параметрах распределения.

*Примеры статистических гипотез:*

1. генеральная совокупность распределена по нормальному закону;
2. нормально распределенная случайная величина имеет среднее значение $a = 4$ (среднее квадратическое значение $\sigma$ известно);
3. нормально распределенная случайная величина имеет среднее значение $a \neq 4$ ($\sigma$ известно);
4. дисперсии двух нормальных совокупностей равны между собой.

- В первой из вышеперечисленных гипотез выдвинуто предположение о виде неизвестного распределения; во второй, третьей и четвертой гипотезах – о параметре одного известного распределения; в пятой гипотезе – о параметрах двух известных распределений.

***Определение 2.2. Проверкой статистических гипотез*** называется сопоставление высказанной гипотезы относительно генеральной совокупности с имеющимися выборочными данными, сопровождаемое количественной оценкой степени достоверности получаемого вывода.

*Статистические гипотезы подразделяют на две группы:*

I. *Нулевая (основная)* гипотеза – это выдвинутая гипотеза $H_0$ (т.е. проверяемое утверждение);

II. *Альтернативная (конкурирующая)* гипотеза – это гипотеза $H_1$, отрицающая или исключающая основную гипотезу.

Например, если нулевая гипотеза состоит в предположении, что математическое ожидание $a$ нормального распределения равно 10, то конкурирующая гипотеза, в частности, может состоять в предположении, что $a \neq 10$. Коротко это записывают так: $H_0$: $a$=10; $H_1$: $a \neq 10$.

Различают гипотезы, которые содержат только одно и более одного предположений.

*Простой* называют гипотезу, содержащую только одно предположение. Например, если $\lambda$ - параметр нормального распределения, то гипотеза $H_0 : \lambda = 5$ – простая. Гипотеза $H_0$: математическое ожидание нормального распределения равно 3 ($\sigma$ известно) – простая.

*Сложной* называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез. Например, сложная гипотеза $H$: $\lambda > 5$ состоит из бесчисленного множества простых вида $H_i : \lambda = b_i$, где $b_i$ - любое число, большее 5.

Выдвинутая гипотеза может быть правильной или неправильной, поэтому возникает необходимость её проверки.

Правильное решение может быть принято в двух случаях:

1. гипотеза принимается, причём и в действительности она правильная;
2. гипотеза отвергается, причём и в действительности она неверна.

Гипотезу проверяют на основании выборки, полученной из генеральной совокупности. Из-за случайности выборки в результате проверки могут возникнуть ошибки и приниматься неправильные решения, т. е. могут быть допущены ошибки двух родов.

*Виды ошибок, возникающих при проверке статистических гипотез:*

1. *Ошибка первого рода* – имеет место тогда, когда отвергается правильная гипотеза $H_0$;

2. *Ошибка второго рода* – имеет место тогда, когда принимается неправильная гипотеза $H_0$.

Вероятность совершить ошибку первого рода ($\alpha = P(H_1 / H_0)$) принято обозначать через $\alpha$; её называют *уровнем значимости.* Наиболее часто уровень значимости принимают равным 0,05 или 0,01. Если, например, принят уровень значимости, равный 0,05, то это означает, что в пяти случаях из ста имеется риск допустить ошибку первого рода (отвергнуть правильную гипотезу).

Вероятность ошибки второго рода (надежность оценки) обозначают $\beta$, т.е. $\beta = P(H_0 / H_1)$.

Вероятность принять верную гипотезу равна $P(H_0 / H_0) = 1 - \alpha$. Вероятность отвергнуть неверную гипотезу $H_0$ (*мощность критерия*) равна $P(H_1 / H_1) = 1 - \beta$.

| Виды ошибок | Вероятность ошибки (неправильное решение) | Вероятность правильного решения |
|---|---|---|
| Ошибка первого рода | $\alpha = P(H_1 / H_0)$ | $P(H_0 / H_0) = 1 - \alpha$ |
| Ошибка второго рода | $\beta = P(H_0 / H_1)$ | $P(H_1 / H_1) = 1 - \beta$ |

| Уровень значимости $\alpha$ | Надежность оценки $\beta$ |
|---|---|
| 0,05 | 0,95 |
| 0,01 | 0,99 |

Подчеркнём, что последствия этих ошибок могут оказаться весьма различными. Например, если отвергнуто правильное решение « не продолжать строительство жилого дома», то эта ошибка первого рода повлечёт материальный ущерб: если же принято неправильное решение «продолжать строительство», несмотря на опасность обвала стройки, то эта ошибка второго рода может повлечь гибель людей. Можно привести примеры, когда ошибка первого рода влечёт более тяжёлые последствия, чем ошибка второго рода.

### *П.2. Статистический критерий проверки нулевой гипотезы. Наблюдаемое значение критерия. Критическая область.*

Для проверки нулевой гипотезы используют специально подобранную случайную величину, точное или приближённое распределение которой известно. Обозначим эту величину в целях общности через $K$.

***Определение 2.3. Статистическим критерием*** (или просто критерием) называют случайную величину $K$, которая служит для проверки нулевой гипотезы $H_0$.

Выбор критерия для проверки статистических гипотез может быть выполнен на основании различных принципов. Чаще всего для этого используют *принцип отношения правдоподобия,* который позволяет построить критерий, наиболее мощный среди всех возможных критериев.

***Суть принципа правдоподобия***: выбрать такой критерий (статистику) $K$ с известной функцией плотности $f(k)$ при условии справедливости основной гипотезы $H_0$, чтобы при заданном уровне значимости $\alpha$ можно найти такую критическую точку $k_{кр}$ распределения $f(k)$, которая разделит область значений критерия на две области: *допустимую* (где результаты выборочного наблюдения выглядят наиболее правдоподобно) и *критическую* (где результаты выборочного наблюдения выглядят менее правдоподобно в отношении гипотезы $H_0$).

Проверка каждого типа статистических гипотез выполняется с помощью соответствующего критерия, например:

– проверка гипотезы о *виде закона распределения* случайной величины осуществляется на основании *критерия согласия Пирсона* $\chi^2$*;*

– проверка гипотезы о *равенстве неизвестных значений дисперсии* двух генеральных совокупностей – с помощью *критерия Фишера F*;

– проверка гипотез о *неизвестных значениях параметров* генеральных совокупностей – посредством *критерия Z нормально распределенной случайной величины* и *t –критерия Стьюдента*.

После выбора определённого критерия множество всех его возможных значений разбивают на два непересекающихся подмножества (области): одно из них содержит значения критерия, при которых нулевая гипотеза отвергается, а другая – при которых она принимается.

*Множество значений определенного критерия разделяется на следующие два непересекающиеся подмножества (области):*

1. *Допустимая область* (или *область принятия гипотезы*) – это совокупность значений критерия, при которых гипотезу принимают, т.е. основная гипотеза $H_0$ принимается (не отклоняется).
2. *Критическая область* – это совокупность значений критерия, при которых нулевую гипотезу отвергают, т.е. основная гипотеза $H_0$ отвергается (отклоняется) и принимается альтернативная гипотеза $H_1$.

Поскольку критерий *K* - одномерная случайная величина, все её возможные значения принадлежат некоторому интервалу. Поэтому критическая область и область допустимых значений (область принятия гипотезы) также являются интервалами и, следовательно, существуют точки, которые их разделяют.

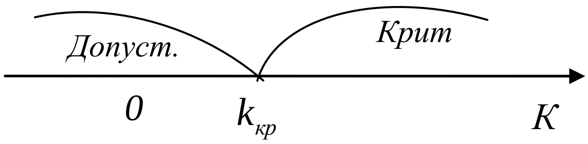
***Определение 2.4.*** Критическими точками (границами) $k_{кр}$ или $k_{теор}$ называют точки, отделяющие критическую область от области принятия гипотезы.

*Виды критических областей*

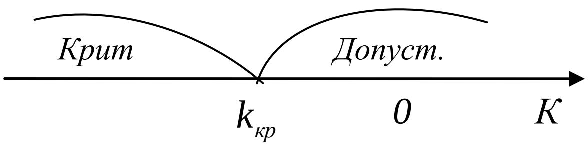
Различают одностороннюю (правостороннюю или левостороннюю) и двустороннюю критические области.

***Односторонние:***

*а) Правосторонней* называют критическую область, определяемую неравенством $K > k_{кр}$, где $k_{кр}$ - положительное число.

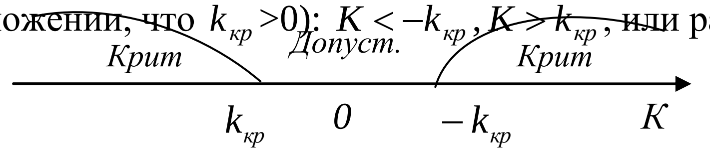



*б) Левосторонней* называют критическую область, определяемую неравенством $K < k_{кр}$, где $k_{кр}$ - отрицательное число.




***Двусторонней*** называют критическую область, определяемую неравенствами $K < k_1, K > k_2$, где $k_2 > k_1$. В частности, если критические точки симметричны относительно нуля, двусторонняя критическая область определяется неравенствами

( в предположении, что $k_{кр}$ >0): $K < -k_{кр}, K > k_{кр}$, или равносильным




неравенством $|K| > k_{кр}$.

Для проверки гипотезы по данным выборок вычисляют частные значения входящих в критерий величин и таким образом получают частное (наблюдаемое) значение критерия.

***Определение 2.5***. ***Наблюдаемым значением*** $k_{набл}$ называют значение критерия, вычисленное по выборкам.

Например, если по двум выборкам найдены исправленные выборочные дисперсии $s_1^2 = 20$ и $s_2^2 = 5$, то наблюдаемое значение критерия $F_{набл} = \frac{s_1^2}{s_2^2} = \frac{20}{5} = 4$.

***Основной принцип проверки статистических гипотез*** можно сформулировать так: *если наблюдаемое значение критерия принадлежит критической области – гипотезу отвергают, если наблюдаемое значение критерия принадлежит области принятия гипотезы (допустимая область) – гипотезу принимают.*

*Алгоритм проверки статистической гипотезы:*

1. Определить основную $H_0$ и альтернативную $H_1$ гипотезы, задать уровень значимости $\alpha$.
2. Выбрать критерий.
3. Найти границы критической области $k_{теор}$ (по альтернативной гипотезе $H_1$, по уровню значимости $\alpha$ – с помощью таблиц).
4. Вычислить наблюдаемое значение $k_{набл}$ по данным выборки.
5. Сравнить значение $k_{набл}$ с критической областью.

Принять решение: а) если значение $k_{набл}$ не принадлежит критической области, то гипотеза $H_0$ не отвергается; б) если значение $k_{набл}$ принадлежит критической области, то гипотеза $H_0$ отвергается, а гипотеза $H_1$ принимается.

### *П.3 Основные виды проверки гипотез*

### *1) Проверка гипотезы о виде закона распределения. Критерий χ2 как критерий согласия*

Критерий $\chi^2$ как критерий согласия используют при проверке принадлежности эмпирического распределения к теоретическому, например, к нормальному, биноминальному, распределению Пуассона и т. п.

В этом случае значение критерия $\chi^2$ определяют, исходя из частот ($f$) эмпирического распределения и частот ($f_o$) теоретического распределения:

$$\chi^2_{набл} = \sum \frac{(f - f_o)^2}{f_o} \qquad (2.1)$$

При этом возможны случаи, когда теоретические частоты заранее известны и когда неизвестны. Во втором случае теоретические частоты определяют на основе теоретического распределения исходя из численности выборки.

При проверке гипотезы о соответствии эмпирического распределения теоретическому сравнивают фактическое значение критерия $\chi^2_{набл}$ с табличным $\chi^2_{теор}$. Если $\chi^2_{набл}$ меньше $\chi^2_{теор}$, следовательно, эмпирическое распределение соответствует теоретическому. В противном случае эмпирическое распределение не соответствует теоретическому, распределение частот в нем носит другой характер.

Рассмотрим алгоритм применения критерия $\chi^2$ как критерия согласия для проверки гипотезы о принадлежности эмпирического распределения к нормальному.

*Алгоритм проверки статистической гипотезы о принадлежности нормальному закону:*

**1.** Определим основную гипотезу:

$H_0$ - генеральная совокупность из которой сделана выборка распределена по нормальному закону;

альтернативная гипотеза $H_1$ - не распределена по нормальному закону. Проверить гипотезу, при уровнях значимости $\alpha = 0{,}05$; $\alpha = 0{,}01$

**2.** Выберем критерий согласия Пирсона $\chi^2$ (хи – квадрат). По таблице $\chi^2$ найдем границу критической области $\chi^2_{теор} = \chi^2_{\alpha,\nu}$,

где $\nu$ – число степеней свободы, вычисляется по формуле: $\nu = (k-1)(l-1)$, где *k*- число наложенных связей (для нормального закона =2), *l*– число интервалов в выборочной совокупности.

**3.** Вычислить наблюдаемое значение $\chi^2_{набл}$ по данным выборки. Для этого выполним следующие действия:

а) находят нормированное отклонение $t$ каждого эмпирического значения от средней выборочной: $t = \frac{x - \bar{x}_в}{\sigma_в}$;

б) по формуле или с помощью таблицы (приложение 3) интеграла вероятностей Лапласа находят значение плотности нормального распределения $\Phi(t)$: $\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-\frac{t^2}{2}} dt$

в) вычисляем вероятность попадания в заданный интервал по формуле

$$p_i(a_i < x < b_i) = \phi\left(\frac{b_i - \bar{x}}{\sigma}\right) - \phi\left(\frac{a_i - \bar{x}}{\sigma}\right)$$

г) вычисляем теоретическую частоту $f_i^0 = p_i \cdot n$, где $p_i$ вероятность попадания в заданный интервал, а $n$ - объем выборки.

д) Подставляем полученные значения частот и вычисляем $\chi^2_{набл} = \sum_{i=1}^{k} \frac{(f_i - f_i^0)^2}{f_i^0}$.

Для упрощенного вычисления $\chi^2_{набл}$ составим таблицу.

| № | Интервалы валового сбора, ц | | частоты | нормированное отклонение | | плотность нормального распределения | | вероятность | теор. частоты | фактическое значение $\chi^2_{факт}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | $a_i$ | $b_i$ | $f$ | $\frac{a_i - x_B}{\sigma_B}$ | $\frac{b_i - x_B}{\sigma_B}$ | $\phi\left(\frac{a_i - x_B}{\sigma_B}\right)$ | $\phi\left(\frac{a_i - x_B}{\sigma_B}\right)$ | $P_i$ | $f_i^0$ | $\sum_{i=1}^{k} \frac{(f_i - f_i^0)^2}{f_i^0}$ |
| 1 | ….. | ….. | … | ….. | …… | …… | ……. | … | … | ….. |
| $\sum$ | | | | | | | | | | $\chi^2_{факт}$ |

**4.** Сравнить значение $\chi^2_{набл}$ с критической областью или $\chi^2_{теор}$.

Принять решение: а) если значение $\chi^2_{набл} < \chi^2_{теор}$ не принадлежит критической области, то гипотеза о принадлежности генеральной совокупности к нормальному закону не отвергается ; б) если значение $\chi^2_{набл}$ принадлежит критической области, то распределение признака не соответствует нормальному закону.

*Пример 1.* Дан интервальный ряд распределения длины волос шерсти у овец, где даны среднее выборочное $\bar{x}_B$=18,4 см; среднее квадратическое отклонение $\sigma_B$= 3,2 см. Требуется проверить гипотезу соответствия распределения «длины шерсти» нормальному закону распределения на уровнях значимости: $\alpha_1 = 0{,}05$; $\alpha_2 = 0{,}01$

*Решение.*

*Нулевая гипотеза $H_0$: Распределение признака «длина волос» соответствует нормальному закону ($\alpha$=0,05).*

Для проверки гипотезы о соответствии нормальному распределению и вычисления фактического значения критерия $\chi^2_{факт} = \sum_{i=1}^{k} \frac{(f_i - f_i^0)^2}{f_i^0}$, где

$f_i$ - частота каждого интервала полученного в опыте;

$f_i^0$ - теоретическая частота, посчитанная по закону распределения, причем $f_i^0 = p_i \cdot n$, где $p_i$ вероятность попадания в заданный интервал, а $n$ - объем выборки, необходимо сделать расчет теоретических частот этого распределения.

Для нормального распределения порядок расчета этих частот следующий:

1) находят нормированное отклонение $t$ каждого эмпирического значения от средней выборочной: $t = \frac{x - \bar{x}_в}{\sigma_в}$;

2) по формуле или с помощью таблицы (приложение 3) интеграла вероятностей Лапласа находят значение плотности нормального распределения $\varphi(t)$: $\varphi(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$

3) вычисляем вероятность попадания в заданный интервал по формуле

$$p_i(a_i < x < b_i) = \phi\left(\frac{b_i - \bar{x}}{\sigma}\right) - \phi\left(\frac{a_i - \bar{x}}{\sigma}\right)$$

4) вычисляем теоретическую частоту $f_i^0 = p_i \cdot n$, где $p_i$ вероятность попадания в заданный интервал, а $n$ - объем выборки.

5) Подставляем полученные значения частот и вычисляем $\chi^2_{факт} = \sum_{i=1}^{k} \frac{(f_i - f_i^0)^2}{f_i^0}$.

Для упрощенного вычисления $\chi^2_{факт}$ составим таблицу 1.

В нашей задаче среднее выборочное $\bar{x}_B$=18,4 см; среднее квадратическое отклонение $\sigma_B$= 3,2 см.

Таблица 2.1.

| № | Интервалы валового сбора, ц | | частоты | нормированное отклонение | | плотность нормального распределения | | вероятность | теор. частоты | фактическое значение $\chi^2_{факт}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| | $a_i$ | $b_i$ | $f$ | $\frac{a_i - x_B}{\sigma_B}$ | $\frac{b_i - x_B}{\sigma_B}$ | $\phi\left(\frac{a_i - x_B}{\sigma_B}\right)$ | $\phi\left(\frac{a_i - x_B}{\sigma_B}\right)$ | $P_i$ | $f_i^0$ | $\sum_{i=1}^{k} \frac{(f_i - f_i^0)^2}{f_i^0}$ |
| 1 | 11,90 | 14,20 | 3 | -2,10 | -1,37 | -0,48 | -0,41 | 0,07 | 1,75 | 0,89 |
| 2 | 14,20 | 16,50 | 4 | -1,37 | -0,64 | -0,41 | -0,24 | 0,17 | 4,25 | 0,01 |
| 3 | 16,50 | 18,80 | 6 | -0,64 | 0,08 | -0,24 | 0,03 | 0,27 | 6,75 | 0,08 |
| 4 | 18,80 | 21,10 | 6 | 0,08 | 0,81 | 0,03 | 0,29 | 0,26 | 6,50 | 0,04 |
| 5 | 21,10 | 23,40 | 5 | 0,81 | 1,54 | 0,29 | 0,44 | 0,15 | 3,75 | 0,42 |
| 6 | 23,40 | 25,70 | 1 | 1,54 | 2,27 | 0,44 | 0,49 | 0,05 | 1,25 | 0,05 |
| ∑ | | | | | | | | | | 1,5 |

По таблице распределения $\chi^2$ (приложение 4) находим $\chi^2_{теорет}$ теоретический критерии на уровне значимости 0,05 с числом степеней свободы: $\nu = l - k = 6 - 3 = 3$ , где $k$ – число наложенных связей (*k=3)* , *l* – число интервалов (*l=6)*. Отсюда $\chi_{теорет}{}^2(0{,}05;3) = 7{,}81$.

Сведем данные в таблицу 2.2, сделаем вывод.

Таблица 2.2.

| | |
|---|---|
| Средняя арифметическая | 18,53 |

| | |
|---|---|
| Среднее квадратичное отклонение | 3,26 |
| Уровень значимости | 0,05 |
| Степени свободы вариации | 5 |
| Фактический уровень значимости | 0,48 |
| Фактическое значение хи-квадрат | 1,5 |
| Табличное значение хи-квадрат | 7,81 |

$\chi^2_{\text{фактическое}}$=1,5

Вывод: Поскольку фактическое значение критерия меньше табличного ($\chi^2_{\text{фактическое}}<\chi^2_{\text{теоретическое}}$), то нулевая гипотеза о соответствии эмпирического распределения теоретическому принимается, т.е. распределение признака «длина волоса шерсти» соответствует нормальному закону ($\alpha$=0,05).

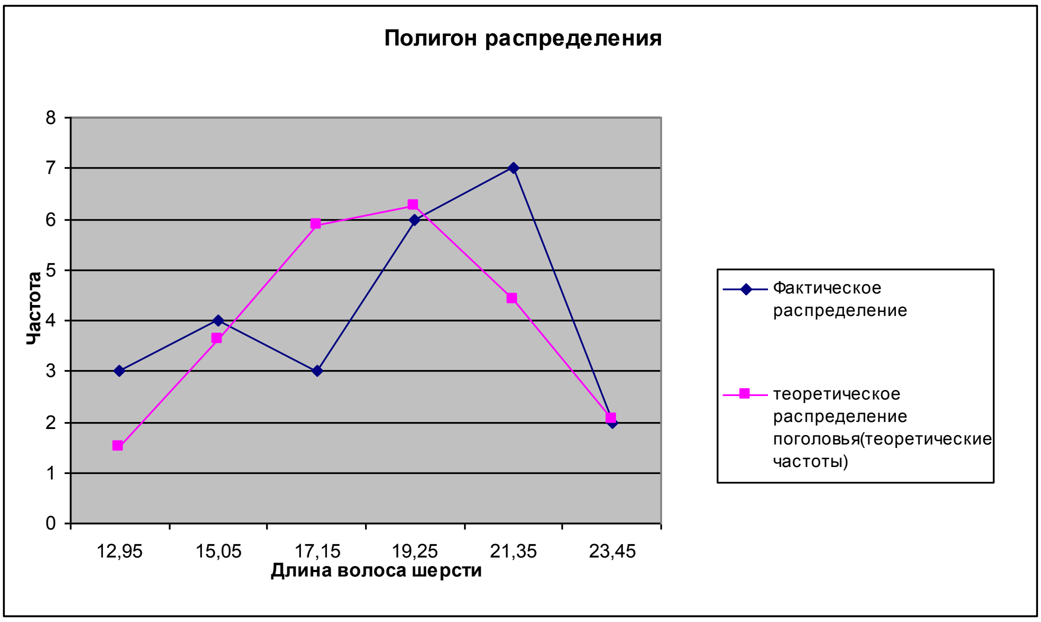



Рисунок 1. Полигон распределения длины волоса шерсти.

***Пример 2.1.*** Дан интервальный ряд распределения длины волоса шерсти у овец, где даны среднее выборочное $\bar{x}_B$=18,4 см; среднее квадратическое отклонение $\sigma_B$= 3,2 см.

| **Номер** | **Группа овец** | **Число овец** | **Середина** | **Накопленная** |
|---|---|---|---|---|

| интервала | нижняя граница | верхняя граница | (абсол. част. интервала) $f_i$ | интервала $X'_i$ | частота $S_{нак}'$ |
|---|---|---|---|---|---|
| 1 | 11,9 | 14,2 | 3 | 13,0 | 3,0 |
| 2 | 14,2 | 16,5 | 4 | 15,3 | 7,0 |
| 3 | 16,5 | 18,8 | 6 | 17,6 | 13,0 |
| 4 | 18,8 | 21,0 | 6 | 19,9 | 19,0 |
| 5 | 21,0 | 23,3 | 5 | 22,2 | 24,0 |
| 6 | 23,3 | 25,6 | 1 | 24,5 | 25,0 |
| итого | - | - | **25** | - | - |

Требуется проверить гипотезу соответствия распределения «длины шерсти» нормальному закону распределения на уровнях значимости: $\alpha_1 = 0{,}05$; $\alpha_2 = 0{,}01$

***Решение.***

***1.Нулевая гипотеза $H_0$:*** распределение признака «длина волос» соответствует нормальному закону, уровни значимости $\alpha_1 = 0{,}05$; $\alpha_2 = 0{,}01$

***2.*** Выберем критерий согласия Пирсона $\chi^2$ (хи – квадрат).

По таблице распределения $\chi^2$ (приложение 2) находим $\chi^2_{теорет}$ теоретический критерии на уровне значимости 0,05 с числом степеней свободы: $\nu = (k-1)(l-1) = (2-1)\cdot(6-1) = 5$, где $k$ – число наложенных связей (*k=2)* , *l* – число интервалов (*l=6).*

Отсюда $\chi^2_{теор} = \chi^2_{0,05;,5} = 11{,}07$, на уровне значимости $\alpha_2 = 0{,}01$ $\chi^2_{0,01;,5} = 12{,}83$

***4.***Вычислить наблюдаемое значение $\chi^2_{набл}$ по данным выборки. Для упрощенного вычисления $\chi^2_{факт}$ составим таблицу 2.3.

Таблица 2.3 .

| № | Интервалы валового сбора, ц | частоты | нормированное отклонение | плотность нормального распределения | вероятность | теор. частоты | фактическое значение $\chi^2_{факт}$ |
|---|---|---|---|---|---|---|---|

| | $a_i$ | $b_i$ | $f$ | $\frac{a_i - x_B}{\sigma_B}$ | $\frac{b_i - x_B}{\sigma_B}$ | $\phi\left(\frac{a_i - x_B}{\sigma_B}\right)$ | $\phi\left(\frac{a_i - x_B}{\sigma_B}\right)$ | $P_i$ | $f_i^0$ | $\sum_{i=1}^{k}\frac{(f_i - f_i^0)^2}{f_i^0}$ |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 11,90 | 14,20 | 3 | -2,10 | -1,37 | -0,48 | -0,41 | 0,07 | 1,75 | 0,89 |
| 2 | 14,20 | 16,50 | 4 | -1,37 | -0,64 | -0,41 | -0,24 | 0,17 | 4,25 | 0,01 |
| 3 | 16,50 | 18,80 | 6 | -0,64 | 0,08 | -0,24 | 0,03 | 0,27 | 6,75 | 0,08 |
| 4 | 18,80 | 21,10 | 6 | 0,08 | 0,81 | 0,03 | 0,29 | 0,26 | 6,50 | 0,04 |
| 5 | 21,10 | 23,40 | 5 | 0,81 | 1,54 | 0,29 | 0,44 | 0,15 | 3,75 | 0,42 |
| 6 | 23,40 | 25,70 | 1 | 1,54 | 2,27 | 0,44 | 0,49 | 0,05 | 1,25 | 0,05 |
| $\sum$ | | | | | | | | | | **1,5** |

$\chi^2$ фактическое=1,5

***5***. Сравним $\chi^2_{набл}$ и $\chi^2_{теор}$ на обоих уровнях значимости:

при $\alpha_1 = 0{,}05$, получим $\chi^2_{набл} = 1{,}5 < \chi^2_{теор1} = 11{,}07$

при $\alpha_2 = 0{,}01$, получим $\chi^2_{набл} = 1{,}5 > \chi^2_{теор2} = 12{,}83$, следовательно на обоих уровнях значимости основная гипотеза принимается поэтому можно считать, что выборка сделана из генеральной совокупности признака «длина волос» подчиняется нормальному закону распределения.

## *2) Проверка гипотез о равенстве параметров двух выборочных совокупностей.*

### *а) Проверка гипотезы о равенстве дисперсий двух выборочных совокупностей.*

Дисперсия является одним из основных показателей вариации (рассеяния). К сравнению дисперсий прибегают всякий раз, когда требуется сопоставить совокупности по степени их изменчивости.

Проверка статистических гипотез относительно дисперсий имеет важное значение и для принятия экономических решений. Например, сравнительная экономическая оценка двух сортов сельскохозяйственной культуры. Основными биологическими признаками сортов, принимаемыми здесь в расчет, являются средняя многолетняя урожайность, качество зерна и вариация урожайности в связи с действием различных факторов. При равенстве средних показателей показатель изменчивости (дисперсия) становится главным в принятии решения об использовании сортов. Для случая различной вариации урожайности возможны такие варианты решения: а) предпочтение отдается сорту, имеющему меньшую колеблемость урожайности по годам, поскольку при этом требуется меньше собственной техники и рабочей силы для уборки урожая без потерь; б) предпочтение отдается сорту, имеющему большую вариации урожайности, поскольку выяснено, что она связана с действием факторов, уровни которых планируется регулировать; в) выбор - на оптимальном сочетанием посевов двух сортов, учитывавшим особенности каждого из них и поставленную экономическую цель.

Методы измерения вариации, оценивания действия факторов опираются на расчет и сравнения дисперсий. Так, сопоставление вариации признака в двух совокупностях достигается проверкой гипотезы о равенстве дисперсий. Для этого применяется $F$ -критерий Фишера. Наблюдаемое значение критерия вычисляется по формуле:

$$k_{набл} = F_{набл} = \frac{S_1^2}{S_2^2}, \tag{2.2}$$

где $S_1^2 \geq S_2^2$. Критическое значение критерия $k_{крит} = F_{крит}$ находят из таблицы $F$ – распределения по заданному уровню значимости и числу степеней свободы $\nu_1$ и $\nu_2$ (соответственно для большой и малой дисперсий, $\nu_1 = n_1 - 1$, $\nu_2 = n_2 - 1$)

Если $F_{набл} > F_{кр}$, то гипотеза о равенстве дисперсий отвергается, в противном случае, т.е. когда $F_{набл} \leq F_{кр}$, гипотеза не отвергается. Использование $F$ -критерия предполагает, что обе совокупности имеют нормальное распределение.

***Пример 2.2***. В результате испытаний двух сортов гороха на трех сортоучастках за 5 лет получены следующие сведения: средняя урожайность сорта А составила 19,8 , сорта Б - 20,6 ц/га; дисперсия урожайности - по сорту А 9,98, по сорту Б - 25,12. Требуется оценить, существенно ли различив в вариации урожайности сортов.

*Решение.*

Поскольку средние значения урожайности примерно равны, для ответа на поставленный вопрос достаточно проверить гипотезу о равенстве дисперсий. По исходным данным $F_{набл} = 25{,}12 / 9{,}98 = 2{,}51$. Критическое значение критерия при $\alpha = 0{,}05$, $\nu_1 = 15 - 1 = 14$ и $\nu_2 = 15 - 1 = 14 \left( n_1 = n_2 = 15 = 15 \cdot 3 \right)$ равно 2,48 (получение данных приложения 2).

Следовательно $F_{набл} > F_{кр}$, и гипотеза о равенстве дисперсий отвергается с рискам ошибиться в 5 случаях из 100. Есть основания полагать, что урожайность сорта Б менее стабильна в связи о различными погодными и почвенными условиями.

Окончательный вывод об экономической целесообразности использования сортов может быть сделан после анализа влияния факторов урожайности. Если влияние будет доказано, а факторы управляемы, то, скорее всего, преимущество будет иметь сорт Б, как более интенсивный. Если сколько-нибудь существенное влияние факторов отсутствует, либо они не управляемы, то предпочтение имеет сорт А, поскольку для получения единицы продукции он требует меньше ресурсов в расчете на своевременную уборку максимального урожая.

***б) Проверка гипотезы о равенстве средних двух выборочных совокупностей -оценка существенности разности средних выборок по t – критерию.***

В экспериментальной и практической работе большое значение имеет определение разности между средними показателями двух сравниваемых групп животных и установление достоверности этой разности.

Например, часто требуется определить, достоверна ли прибавка удоя при введении в рацион коров какого-либо кормового компонента.

В ветеринарных исследованиях очень важно бывает доказать, что примененная доза или новый лекарственный препарат достоверно уменьшает долю заболевших животных по сравнению с долей больных животных в контрольной группе, не подвергавшихся лечению.

В зависимости от содержания исходной информации нулевая гипотеза о равенстве средник значений $H_0 : \left|\tilde{X}_1 - \tilde{X}_2\right| = 0$ двух совокупностей может состоять в двух предположениях:

1) средние значения равны из выборок одной генеральной совокупности, т.е. предполагается, что разность между средними соизмерима с ошибками выборочного наблюдения. В этом случае дисперсии двух выборок равны.
2) средние значения равны из выборок двух генеральных совокупностей, тогда равны генеральные средние. В этом случае имеем неравенство дисперсий двух выборок.

При условии нормального распределения изучаемого признака проверку гипотезы относительно средних проводят с помощью $t$ -критерия Стьюдента при заданном уровне значимости.

*Критическое (теоретическое)* значение критерия находят по заданному уровню значимости $\alpha$ и числу степеней свободы:

$$k_{крит} = t_{крит.} = t(\alpha;\nu)$$

Вычисление степеней свободы $\nu$ проводим с помощью таблицы 2.4.

Таблица 2.4.

| **Гипотеза о равенстве дисперсий принята (выборки из одной генеральной совокупности)** | **Гипотеза о равенстве дисперсий отвергнута (выборки из двух различных генеральных совокупностей)** |
|---|---|
| $\nu = n_1 + n_2 - 2$ | $\nu = (n_1 + n_2 - 2)\cdot\left(0{,}5 + \dfrac{S_x^2 \cdot S_y^2}{S_x^4 + S_y^4}\right)$ |

Наблюдаемое значение $k_{набл} = t_{набл.}$ и $S^2$ - уточненную оценку дисперсии генеральной совокупности, находят по таблице 2.5

Таблица 2.5

| **При различных объемах выборок** $n_1 \neq n_2$ | **При равных объемах выборок** $n_1 = n_2$ |
|---|---|
| | |
| $k_{набл} = t_{набл.} = \frac{\lvert \overline{X_1} - \overline{X_2} \rvert}{\sqrt{S^2}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$ <br> $S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$ | $k_{набл} = t_{набл.} = \frac{\lvert \overline{X_1} - \overline{X_2} \rvert}{\sqrt{S^2}} \cdot \sqrt{n}$ <br> $S^2 = S_1^2 + S_2^2$ |

*Алгоритм проверки гипотезы о равенстве средних двух выборок.*

**1.** Проверим гипотезу о равенстве дисперсий:

а) основная гипотеза: $H_0$ - дисперсий двух совокупностей равны; альтернативная гипотеза $H_1$ - дисперсий не равны. Проверить гипотезу, при уровне значимости $\alpha$ ;

б) Выберем критерий Фишера и вычислим критическое значение критерия $k_{крит} = F_{крит}$ из таблицы $F$ – распределения по заданному уровню значимости и числу степеней свободы $\nu_1$ и $\nu_2$ (соответственно для большой и малой дисперсий, $\nu_1 = n_1 - 1,\ \nu_2 = n_2 - 1$, $\nu_1 \geq \nu_2$)

в) Наблюдаемое значение критерия вычисляется по формуле (2.2): $k_{набл} = F_{набл} = \frac{S_1^2}{S_2^2}$,

где $S_1^2 \geq S_2^2$, $S_1$ и $S_2$ – средние квадратические отклонения двух выборок.

г) Если $F_{набл} > F_{кр}$, то гипотеза о равенстве дисперсий отвергается (выборки сделаны из разных генеральных совокупностей), в противном случае, т.е. когда $F_{набл} \leq F_{кр}$, гипотеза не отвергается (выборки из одной генеральной совокупности).

**2.** Проверим гипотезу о равенстве средних, оценим разность между средними:

а) основная гипотеза: $H_0: \left|\tilde{X}_1 - \tilde{X}_2\right| = 0$ - средние двух совокупностей равны;

альтернативная гипотеза $H_1$ - средние не равны. Проверить гипотезу, при уровне значимости $\alpha$;

б) Выберем $t$ критерий Стьюдента и вычислим критическое значение критерия $k_{крит} = t_{крит}$ из таблицы – $t$ - распределения (приложение 3) по заданному уровню значимости и числа степеней свободы.

Формулу для вычисления степеней свободы выбирают по таблице 1.

- *если принята гипотеза о равенстве дисперсий по формуле:* $\nu = n_1 + n_2 - 2$

- *если гипотеза о равенстве дисперсий отвергнута по формуле:*

$$\nu = (n_1 + n_2 - 2) \cdot \left(0{,}5 + \frac{S_x^2 \cdot S_y^2}{S_x^4 + S_y^4}\right),$$

в) Наблюдаемое значение критерия вычисляется по формуле из таблицы 2:

- *если объемы выборок не равны* $n_1 \neq n_2$ вычисляем по формулам:

$$k_{набл} = t_{набл.} = \frac{\left|\overline{X_1} - \overline{X_2}\right|}{\sqrt{S^2}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

где уточненная оценка генеральной дисперсий $S^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

- *если объемы выборок равны,* $n_1 = n_2 = n$*, то формулам:*

$$t_{набл.} = \frac{\left|\overline{X_1} - \overline{X_2}\right|}{\sqrt{S^2}} \cdot \sqrt{n}$$

где уточненная оценка генеральной дисперсий $S^2 = S_1^2 + S_2^2$

**3.** *Сделаем выводы:*

*1) Если* $t_{набл.}$ *меньше, чем* $t_{кр.}$, то нулевая гипотеза о равенстве средних $H_0: \left|\tilde{X}_1 - \tilde{X}_2\right| = 0$ не отвергается, разность между средними не существенна, изменение фактора не влияет на средние показатели.

В этом случае, если принята гипотеза о равенстве дисперсий (выборки взяты из одной генеральной совокупности), то выбирают любое значение фактора.

В случае неравенства дисперсий (выборки взяты из разных генеральных совокупностей) предпочтение отдается фактору, при котором признак имеет меньшую изменчивость (дисперсию).

*2) Если* $t_{набл.}$ *больше, чем* $t_{кр.}$, то нулевая гипотеза о равенстве средних $H_0 : \left|\tilde{X}_1 - \tilde{X}_2\right| = 0$ отвергается, разность между средними существенна, т.е. изменение фактора существенно влияет на средние показатели

В этом случае, если принята гипотеза о равенстве дисперсий (выборки взяты из одной генеральной совокупности), то, выбирают фактор при котором признак имеет наибольшее среднее.

В случае неравенства дисперсий (выборки взяты из разных генеральных совокупностей) предпочтение отдается фактору, при котором признак имеет наибольшее среднее и меньшую изменчивость (дисперсию).

### *б) Проверка гипотезы о равенстве средних двух выборочных совокупностей -оценка существенности разности средних выборок по t – критерию.*

В экспериментальной и практической работе большое значение имеет определение разности между средними показателями двух сравниваемых сортов культур, видов удобрений установление достоверности этой разности.

Например, часто требуется определить, достоверна ли прибавка удоя при введении в рацион коров какого-либо кормового компонента.

В агрономических исследованиях очень важно бывает доказать, что примененная доза удобрений достоверно увеличивает урожайность по сравнению с контрольной группой без удобрения.

В зависимости от содержания исходной информации нулевая гипотеза о равенстве средник значений $H_0 : \left|\tilde{X}_1 - \tilde{X}_2\right| = 0$ двух совокупностей может состоять в двух предположениях:

3) средние значения равны из выборок одной генеральной совокупности, т.е. предполагается, что разность между средними соизмерима с

ошибками выборочного наблюдения. В этом случае дисперсии двух выборок равны.

4) средние значения равны из выборок двух генеральных совокупностей, тогда равны генеральные средние. В этом случае имеем неравенство дисперсий двух выборок.

При условии нормального распределения изучаемого признака проверку гипотезы относительно средних проводят с помощью $t$ -критерия Стьюдента при заданном уровне значимости.

*Критическое (теоретическое)* значение критерия находят по заданному уровню значимости $\alpha$ и числу степеней свободы:

$$k_{крит} = t_{крит.} = t(\alpha;\nu)$$

Вычисление степеней свободы $\nu$ проводим с помощью таблицы 2.6

Таблица 2.6.

| **Гипотеза о равенстве дисперсий принята (выборки из одной генеральной совокупности)** | **Гипотеза о равенстве дисперсий отвергнута (выборки из двух различных генеральных совокупностей)** |
|---|---|
| $\nu = n_1 + n_2 - 2$ | $\nu = (n_1 + n_2 - 2)\cdot\left(0{,}5 + \dfrac{S_x^2 \cdot S_y^2}{S_x^4 + S_y^4}\right)$ |

Наблюдаемое значение $k_{набл} = t_{набл.}$ и $S^2$ - уточненную оценку дисперсии генеральной совокупности, находят по таблице 2

Таблица 2

| **При различных объемах выборок** $n_1 \neq n_2$ | **При равных объемах выборок** $n_1 = n_2$ |
|---|---|
| | |

| | |
|---|---|
| $k_{набл}=t_{набл.}=\dfrac{\left\lvert \overline{X_1}-\overline{X_2}\right\rvert}{\sqrt{S^2}}\cdot\sqrt{\dfrac{n_1 n_2}{n_1+n_2}}$<br><br>$S^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ | $k_{набл}=t_{набл.}=\dfrac{\left\lvert \overline{X_1}-\overline{X_2}\right\rvert}{\sqrt{S^2}}\cdot\sqrt{n}$<br><br>$S^2=S_1^2+S_2^2$ |

*Алгоритм проверки гипотезы о равенстве средних двух выборок.*

**4.** Проверим гипотезу о равенстве дисперсий:

а) основная гипотеза: $H_0$ - дисперсий двух совокупностей равны;

альтернативная гипотеза $H_1$ - дисперсий не равны. Проверить гипотезу, при уровне значимости $\alpha$;

б) Выберем критерий Фишера и вычислим критическое значение критерия $k_{крит}=F_{крит}$ из таблицы $F$ – распределения по заданному уровню значимости и числу степеней свободы $\nu_1$ и $\nu_2$ (соответственно для большой и малой дисперсий, $\nu_1=n_1-1,\ \nu_2=n_2-1$, $\nu_1\ge\nu_2$)

в) Наблюдаемое значение критерия вычисляется по формуле (2.5): $k_{набл}=F_{набл}=\dfrac{S_1^2}{S_2^2}$,

где $S_1^2\ge S_2^2$, $S_1$ и $S_2$ – средние квадратические отклонения двух выборок.

г) Если $F_{набл}>F_{кр}$, то гипотеза о равенстве дисперсий отвергается (выборки сделаны из разных генеральных совокупностей), в противном случае, т.е. когда $F_{набл}\le F_{кр}$, гипотеза не отвергается (выборки из одной генеральной совокупности).

**5.** Проверим гипотезу о равенстве средних, оценим разность между средними:

а) основная гипотеза: $H_0:\left\lvert\tilde{X}_1-\tilde{X}_2\right\rvert=0$ - средние двух совокупностей равны;

альтернативная гипотеза $H_1$ - средние не равны. Проверить гипотезу, при уровне значимости $\alpha$;

б) Выберем $t$ критерий Стьюдента и вычислим критическое значение критерия $k_{крит} = t_{крит}$ из таблицы – $t$ - распределения (приложение 3) по заданному уровню значимости и числа степеней свободы.

Формулу для вычисления степеней свободы выбирают по таблице 1.

- *если принята гипотеза о равенстве дисперсий по формуле:* $\nu = n_1 + n_2 - 2$
- *если гипотеза о равенстве дисперсий отвергнута по формуле:*

$$\nu = (n_1 + n_2 - 2) \cdot \left( 0{,}5 + \frac{S_x^2 \cdot S_y^2}{S_x^4 + S_y^4} \right),$$

в) Наблюдаемое значение критерия вычисляется по формуле из таблицы 2:

- *если объемы выборок не равны* $n_1 \neq n_2$ вычисляем по формулам:

$$k_{набл} = t_{набл.} = \frac{\left| \overline{X_1} - \overline{X_2} \right|}{\sqrt{S^2}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

где уточненная оценка генеральной дисперсий $S^2 = \dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}$

- *если объемы выборок равны,* $n_1 = n_2 = n$*, то формулам:*

$$t_{набл.} = \frac{\left| \overline{X_1} - \overline{X_2} \right|}{\sqrt{S^2}} \cdot \sqrt{n}$$

где уточненная оценка генеральной дисперсий $S^2 = S_1^2 + S_2^2$

**6.** *Сделаем выводы:*

*1) Если* $t_{набл.}$ *меньше, чем* $t_{кр.}$, то нулевая гипотеза о равенстве средних $H_0 : \left| \tilde{X}_1 - \tilde{X}_2 \right| = 0$ не отвергается, разность между средними не существенна, изменение фактора не влияет на средние показатели.

В этом случае, если принята гипотеза о равенстве дисперсий (выборки взяты из одной генеральной совокупности), то выбирают любое значение фактора.

В случае неравенства дисперсий (выборки взяты из разных генеральных совокупностей) предпочтение отдается фактору, при котором признак имеет меньшую изменчивость (дисперсию).

*2) Если* $t_{набл.}$ *больше, чем* $t_{кр.}$, то нулевая гипотеза о равенстве средних $H_0 : \left|\tilde{X}_1 - \tilde{X}_2\right| = 0$ отвергается, разность между средними существенна, т.е. изменение фактора существенно влияет на средние показатели

В этом случае, если принята гипотеза о равенстве дисперсий (выборки взяты из одной генеральной совокупности), то, выбирают фактор при котором признак имеет наибольшее среднее.

В случае неравенства дисперсий (выборки взяты из разных генеральных совокупностей) предпочтение отдается фактору, при котором признак имеет наибольшее среднее и меньшую изменчивость (дисперсию).

***Пример 3.*** По приведенным ниже данным сравнить средние удои коров, получавших различные рационы. Для проверки нулевой гипотез о равенстве средних принять уровень значимости $\alpha = 0{,}05$.

*Изменение удоев коров при различных рационах питания.*

| Рацион | Поголовье коров, получавших рацион, гол | Среднесуточный удой молока в пересчете на базисную жирность, кг/гол | Среднеквадратическое отклонение в молочной продуктивности коров, кг/гол |
|---|---|---|---|
| №1 | 10 | 16,2 | 3,8 |
| №2 | 8 | 17,7 | 4,2 |

*Решение.*

**1.** *Проверим гипотезу о равенстве дисперсий:*

а) основная гипотеза: $H_0$ - дисперсий двух совокупностей равны;

альтернативная гипотеза $H_1$ - дисперсий не равны. Проверить гипотезу, при уровне значимости $\alpha = 0{,}05$;

б) Выберем критерий Фишера и вычислим критическое значение критерия $k_{крит} = F_{крит}$ из таблицы $F$ – распределения по заданному уровню значимости

$\alpha = 0{,}05$ и числу степеней свободы $\nu_1$ и $\nu_2$ (соответственно для большой и малой дисперсий, $\nu_1 = n_1 - 1 = 8 - 1 = 7$, $\nu_2 = 10 - 1 = 9$, $\nu_1 \geq \nu_2$), найдем:

$k_{крит} = F_{0,05}(7;8) = 3{,}29$

в) Наблюдаемое значение критерия вычисляется по формуле: $k_{набл} = F_{набл} = \frac{S_1^2}{S_2^2}$,

в числителе подставлена большая дисперсия.

$S_1^2 = 4{,}8 \geq S_2^2 = 3{,}8$; $F_{набл} = 4{,}2^2 / 3{,}8^2 = 1{,}22$,

г) $F_{набл} = 1{,}22 \leq F_{кр} = 4{,}88$, гипотеза о равенстве дисперсий не отвергается (выборки из одной генеральной совокупности). то есть основания предполагать, что выборки со средним $\tilde{X}_1 = 16{,}2$ и $\tilde{X}_2 = 17{,}7$ получены из одной генеральной совокупности, а расхождение между $\tilde{X}_1$ и $\tilde{X}_2$ случайно.

**2.** *Проверим гипотезу о равенстве средних, оценим разность между средними:*

а) основная гипотеза: $H_0 : \left|\tilde{X}_1 - \tilde{X}_2\right| = 0$ - средние двух совокупностей равны;

альтернативная гипотеза $H_1$ - средние не равны. Проверить гипотезу, при уровне значимости $\alpha = 0{,}05$;

б) Выберем $t$ критерий Стьюдента и вычислим критическое значение критерия $k_{крит} = t_{крит}$ из таблицы – $t$ - распределения по заданному уровню значимости и числа степеней свободы (приложение 3).

В пункте (1) принята гипотеза о равенстве дисперсий, следовательно число степеней свободы вычислим по формуле из таблицы 1:

$\nu = n_1 + n_2 - 2$, $\nu = 10 + 8 - 2 = 16$

$t_{кр.}(0{,}05;16) = 2{,}12$

в) Наблюдаемое значение критерия $k_{набл} = t_{набл.}$ вычислим по формуле из таблицы 2 (объемы выборок не равны $n_1 \neq n_2$ ):

$$k_{набл} = t_{набл.} = \frac{\left|\overline{X_1} - \overline{X_2}\right|}{\sqrt{S^2}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}},$$

Вычислим оценку дисперсий :

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} \Rightarrow S^2 = \frac{(10-1)\cdot 3{,}8^2 + (8-1)\cdot 4{,}2^2}{10+8-2} = 15{,}84.$$

Подставим полученное значение $S^2$ и другие величины в формулу: $t_{набл.} = \frac{|16{,}2 - 17{,}7|}{\sqrt{15{,}84}} \cdot \sqrt{\frac{10 \cdot 8}{10+8}} = 0{,}79.$

**3.** *Сделаем вывод:*

$t_{набл.}$ меньше, чем $t_{кр.}$, поэтому нулевая гипотеза не отвергается, разность между средними не существенна, изменение фактора не влияет на средние показатели., При равенстве дисперсий изменение рационов не влияет на средние удои, можно выбрать любой рацион.

***Пример 4.*** Оценить влияние видов минеральных удобрении на средний урожай опытных культур. С этой целью сравнить средние урожаи опыта, применяя $t$- критерий Стьюдента. Уровень значимости принять равным 0,05.

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| 11,8 | 5 | 7,5 | 0,8 |
| 14,3 | 5 | 1,0 | 0,3 |

*Решение.*

**1.** *Проверим гипотезу о равенстве дисперсий:*

а) основная гипотеза: $H_0$ - дисперсий двух совокупностей равны;

альтернативная гипотеза $H_1$ - дисперсий не равны. Проверить гипотезу, при уровне значимости $\alpha = 0{,}05$;

б) Выберем критерий Фишера и вычислим критическое значение критерия $k_{крит} = F_{крит}$ из таблицы $F$ – распределения по заданному уровню значимости

$\alpha = 0{,}05$ и числу степеней свободы $\nu_1$ и $\nu_2$ (соответственно для большой и малой дисперсий, $\nu_1 = n_1 - 1 = 5 - 1 = 4$, $\nu_2 = 5 - 1 = 4$, $\nu_1 \geq \nu_2$), найдем:

$k_{крит} = F_{0,05}(4;4) = 6{,}39$

в) Наблюдаемое значение критерия вычисляется по формуле: $k_{набл} = F_{набл} = \dfrac{S_1^2}{S_2^2}$,

в числителе подставлена большая дисперсия.

$S_1^2 = 0{,}8 \geq S_2^2 = 0{,}3$; $F_{набл} = 0{,}8^2 / 0{,}3^2 = 7{,}11$

г) $F_{набл} = 7{,}11 > F_{кр} = 6{,}39$, гипотеза о равенстве дисперсий отвергается (выборки сделаны из разных генеральных совокупностей).

**2.** *Проверим гипотезу о равенстве средних, оценим разность между средними:*

а) основная гипотеза: $H_0 : \left|\tilde{X}_1 - \tilde{X}_2\right| = 0$ - средние двух совокупностей равны;

альтернативная гипотеза $H_1$ - средние не равны. Проверить гипотезу, при уровне значимости $\alpha = 0{,}05$;

б) Выберем $t$ критерий Стьюдента и вычислим критическое значение критерия $k_{крит} = t_{крит}$ из таблицы – $t$ - распределения по заданному уровню значимости и числа степеней свободы (приложение 3).

В пункте (1) гипотеза о равенстве дисперсий отвергнута, следовательно по таблице 1 , число степеней свободы вычислим по формуле:

$$\nu = (n_1 + n_2 - 2) \cdot \left(0{,}5 + \frac{S_x^2 \cdot S_y^2}{S_x^4 + S_y^4}\right),$$

$$\nu = (5 + 5 - 2)\left(0{,}5 + \frac{0{,}8^2 \cdot 0{,}3^2}{0{,}8^4 + 0{,}3^4}\right) = 5{,}10.$$

$t_{кр.}(0{,}05;5{,}1) = 2{,}12$

в) Наблюдаемое значение критерия по таблице 2 вычислим по формуле (объемы выборок равны, $n_1 = n_2 = 5$):

$t_{набл.} = \dfrac{\left|\overline{X_1} - \overline{X_2}\right|}{\sqrt{S^2}} \cdot \sqrt{n}$, где $S^2 = S_1^2 + S_2^2$ - оценка дисперсий.

$$t_{набл.} = \frac{|7,5-1,0|}{\sqrt{0,8^2+0,3^2}} \cdot \sqrt{5} = 17,0.$$

**3.** *Сделаем вывод:*

$t_{набл.} = 17 > t_{кр.} = 2,12$, поэтому нулевая гипотеза отвергается, разность между средними существенна, изменение фактора влияет на средние показатели. Следовательно, виды внесенных удобрении существенно влияют на урожай следует выбрать вид удобрения с высокой урожайностью 14 кц/га.

## Индивидуальное задание №3. «Проверка статистических гипотез»

По приведенным ниже данным сравнить средние урожаи некоторой культуры, получавших различные виды минеральных удобрений №1 и №2. Для проверки нулевой гипотез о равенстве средних принять уровень значимости $\alpha = 0,05$.

Вариант №1

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 9 | 16,3 | 3,6 |
| №2 | 7 | 17,1 | 4,0 |

Вариант №2

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 9 | 17,3 | 4,6 |
| №2 | 10 | 16,1 | 3,0 |

Вариант №3

| Вид минеральных | Количество | Средний урожай | Среднеквадратическое |
|---|---|---|---|

| | | | |
|---|---|---|---|
| удобрений | опытных образцов культуры | опытных культур ц/га | отклонение урожаев, ц/га |
| №1 | 11 | 12,3 | 3,6 |
| №2 | 8 | 14,1 | 5,0 |

Вариант №4

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 12 | 11,3 | 4,6 |
| №2 | 10 | 15,1 | 6,2 |

Вариант №5

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 14 | 12,3 | 5,6 |
| №2 | 12 | 10,1 | 4,2 |

Вариант №6

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 8 | 18,3 | 2,6 |
| №2 | 12 | 20,1 | 3,2 |

Вариант №7

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 9 | 16,3 | 3,6 |

| | | | |
|---|---|---|---|
| №2 | 11 | 18,1 | 4,2 |

Вариант №8

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 9 | 16,3 | 3,7 |
| №2 | 7 | 18,1 | 5,0 |

Вариант №9

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 15 | 16,3 | 3,6 |
| №2 | 12 | 17,1 | 4,0 |

Вариант №10

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 9 | 16,3 | 3,6 |
| №2 | 7 | 19 | 5,2 |

Вариант №11

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 9 | 16,3 | 3,6 |
| №2 | 7 | 17,1 | 4,0 |

Вариант №12

| Вид минеральных | Количество | Средний урожай | Среднеквадратическое |
|---|---|---|---|

| | | | |
|---|---|---|---|
| удобрений | опытных образцов культуры | опытных культур ц/га | отклонение урожаев, ц/га |
| №1 | 9 | 17,3 | 4,6 |
| №2 | 10 | 16,1 | 3,0 |

Вариант №13

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 11 | 12,3 | 3,6 |
| №2 | 8 | 14,1 | 5,0 |

Вариант №14

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 12 | 11,3 | 4,6 |
| №2 | 10 | 15,1 | 6,2 |

Вариант №15

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 14 | 12,3 | 5,6 |
| №2 | 12 | 10,1 | 4,2 |

Вариант №16

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 8 | 18,3 | 2,6 |
| №2 | 12 | 20,1 | 3,2 |

Вариант №17

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 9 | 16,3 | 3,6 |
| №2 | 11 | 18,1 | 4,2 |

Вариант №18

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 9 | 16,3 | 3,7 |
| №2 | 7 | 18,1 | 5,0 |

Вариант №19

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 15 | 16,3 | 3,6 |
| №2 | 12 | 17,1 | 4,0 |

Вариант №20

| Вид минеральных удобрений | Количество опытных образцов культуры | Средний урожай опытных культур ц/га | Среднеквадратическое отклонение урожаев, ц/га |
|---|---|---|---|
| №1 | 9 | 16,3 | 3,6 |
| №2 | 7 | 19 | 5,2 |

# РАЗДЕЛ 3. Дисперсионный анализ.

## П1.Назначение, сущность и метод дисперсионного анализа

Наряду с методами оценки различия двух выборочных совокупностей путём сравнения их средних значений и стандартных отклонений, а также подбора и оценки моделей распределения статистических совокупностей, особенности эксперимента требуют создания особых способов обработки его результатов. Например, это могут быть группы растений, получивших разные удобрения или уход, когда в опыте ставится цель статистически оценить эффект мероприятия.

Поэтому статистический анализ результатов наблюдений по оценке эффекта мероприятия, зависящего от разных одновременно действующих факторов, выбор этих факторов и оценка «силы» их влияния являются важнейшими предпосылками свертки информации и обоснования математических моделей.

В начале 20-х годов (1925г.) Рональдом Фишером – английским учёным, разработан критерий и метод для такой оценки. Это привело к значительному развитию теории планирования опыта и статистической оценки его эффекта. Этот метод Р. Фишером назван дисперсионным анализом.

Дисперсией называется сумма квадратов отклонений (или средний квадрат отклонений) отдельных вариант (Vi) от их средней арифметической величины (M).

Дисперсия (от латин.dispersus – рассеяние), как и основное отклонение, характеризует степень разнообразия (варьирования) отдельных вариантов ряда распределения вокруг среднего значения.

Задачей дисперсионного анализа является исследования тех или иных факторов на изменчивость средних значений изучаемых случайных величин.

Дисперсионный анализ состоит в выделении и оценке отдельных факторов, вызывающих изменчивость. С этой целью производится <u>разложение дисперсии</u> частичной совокупности на составляющие. Каждая из этих составляющих, даёт оценку дисперсии в общей совокупности.

Одна часть отражает изменчивость, вызываемую действием учитываемого в опыте фактора, другая является следствием совокупности случайных, не учитываемых причин. Эти две составляющие образуют общую меру изменчивости. Если варьирование характеризовать дисперсиями, то будем иметь:

$$D_o = D\text{ф} + Dc$$

Где $D_o$ - общее варьирование или общая дисперсия,

$D\text{ф}$– факториальная дисперсия

$Dc$– случайная дисперсия

Значения этих дисперсий будут выражаться:

$D_o = \sum ( Vi - M_o)^2$ - сумма квадратов отклонений отдельных вариант *(Vi)* всего комплекса, умноженная на число вариантов в группах *(n);*

$Dc = \sum [\sum(Vi - Mr)^2]$ – сумма из сумм квадратов отклонений отдельных вариант от их групповых средних *(Vi–Mr);*

Следовательно, будем иметь равенство:

$\sum ( Vi - M_o)^2 = n \sum(Mr - M_o)^2 + \sum [\sum(Vi - Mr)^2]$.

Отношением сумм квадратов отклонений к числам степеней свободы *(К)* соответственно общей **(*D*o),** факториальной *(Dф)* и случайной *(Dc)* дисперсий, получаются выборочные дисперсии, т.е.:

$S_o^2 = \frac{\text{До}}{\text{Ко}}$

$S\text{ф}^2 = \frac{\text{Дф}}{\text{Кф}}$

$Sc^2 = \frac{\text{Дс}}{\text{Кс}}$

Которые служат оценками соответствующих генеральных параметров:

$S_o^2$ - общей дисперсии комплекса $(\sigma_o^2)$;

$S\text{ф}^2$ - межгрупповой дисперсии $(\sigma\text{ф}^2)$;

$Sc^2$ - внутригрупповой дисперсии $(\sigma c^2)$.

Отношение дисперсии межгрупповой, или факториальной к дисперсии внутригрупповой, или остаточной, т.е.:

$$F = \frac{S^2\text{ф}}{S^2 c}$$

Служит критерием оценки влияния на признак регулируемых в опыте факторов.

Если $F \geq Fst$ (табличного), то нулевая гипотеза (предположение) опровергается, т.е. различия, наблюдаемые между групповыми средними комплекса, признается существенным, влияние фактора существенное.

Если же $F<Fst$ (табличного), то различие, наблюдаемое между групповыми средними, признается несущественным - случайным .

Заключительный этап дисперсионного анализа – оценка силы влияния отдельных факторов или их совместного действия на результативный признак.

Дисперсионный анализ характеризуется строгой логичностью и последовательностью вычислительных операций. Ценность этого метода заключается в том, что он позволяет выявить суммарное действие факторов, действие каждого регулируемого в опыте фактора, а также действие различных сочетаний факторов друг с другом на результативный признак.

Дадим конкретные понятия результативного признака и факторов:

*Признаки,* изменяющиеся под воздействием тех или иных причин, называются *результативными;* а причины, вызывающие изменение величины результативного признака или признаков, называются *факторами.*

Например, масса или линейные размеры, по которым мы судим об организме, его физическое развитие, успеваемость студентов, урожайность той или иной культуры и т.п.- все это признаки, на которые оказывают влияние различные факторы: элементы или режим питания, физические или умственные упражнения, дозы лекарственных или токсических веществ и т.п.

*Результативными признаками* в агрономий могут быть:

- высота, диаметр и объем растения;
- количество подроста в лесу;
- урожайность и др.

*Действующий фактор* – это любое влияние (воздействие), которое отражается на результативном признаке.

*Факторами* могут быть: различная интенсивность изреживания древостоя, удобрения, температура, воздух, возраст, деревьев или древостоев, тип возрастной структуры и др. Факторы обозначаются заглавными буквами латинского алфавита А, В, С**…,** а учитываемые признаки через *X, Y, Z*….

Факторов, воздействующих на один и тот же признак много. В опыте же регулируются (учитываются) лишь некоторые из них: они называются *регулируемыми* или *организованными факторами* в отличие от тех**,** которые регулированию не подвергаются. Хотя и оказывают воздействие на величину результативного признака. Обычно каждый регулируемый фактор испытывается серийно, т.е. в виде нескольких обособленных др. от др. групп, называемых *градациями.* Их принято обозначать теми же буквами, которыми обозначаются факторы.

Например, градации фактора *А* обозначаются через *Aı A2 A3* и т.д., а фактора *В* - *Bı B2 B3* и т.д. Числа градаций того или иного фактора определяются условиями опыта, например**,** испытываемыми дозами удобрений. Количеством сортов при изучении их урожайности и т.д.

Результативные признаки тоже могут подразделяться на отдельные градации, на которых испытывается действие регулируемых факторов.

При образовании любого статистического комплекса, а также при планировании эксперимента с целью дисперсионного анализа необходимо соблюдать по крайней мере *два основных условия*:

1) Во-первых, учитываемые факторы должны быть независимы друг от друга;
2) Во-вторых, выборка, группируемая в статистический комплекс, должна формироваться по принципу рандомизации, т.е. методом случайного отбора вариант из генеральной совокупности, распределяющейся по нормальному закону.

Структура комплекса определяется числом регулируемого фактора или факторов, а также числом подразделений или групп, образуемых по результативному признаку.

Форма дисперсионного комплекса задается таблицей, в которой *число строк* равно числу градаций результативного признака, а *число столбцов* соответствует числу градаций регулируемого фактора или нескольких факторов с их градациями.

Дисперсионный комплекс называется *однофакторным,* если испытывается действие на признак одного регулируемого фактора, а если одновременно испытывается действие на признак двух, трёх и большего числа регулируемых факторов комплекс будет соответственно, двух-, трёх- и многофакторным.

Результаты массовых испытаний, выраженные в виде числовых выражений результативного признака, т.е. масса вариант, могут распределяться по градациям комплекса равномерно и неравномерно, пропорционально и непропорционально. Отсюда дисперсионные комплексы называются *равномерными, пропорциональными и неравномерными.*

### П.2. Дисперсионный анализ однофакторного комплекса

Разумеется, однофакторные дисперсионные комплексы, как отмечено выше могут быть равномерными и неравномерными. Независимо от этого техника дисперсионного анализа однофакторных комплексов сводится главным образом к расчету показателей варьирования, которыми в области дисперсионного анализа служат средние квадраты отклонений, или дисперсии, а также и к расчету групповых, или частных средних ($M_\iota$) и общей средней арифметической для всего комплекса в целом ($M_0$).

Дисперсионный анализ проводится обычно по той или иной схеме. Для однофакторных равномерных комплексов такой схемой может служить следующий порядок операций:

1. Исходные данные группируются в виде комбинированной таблицы таким образом, чтобы градации регулируемого признака **(А)** располагались по горизонтали в верхней части таблицы, образуя ее

графы или столбцы, а значения результативного признака *(Vi),*т.е. варианты или даты, группировались соответственно по градациям фактора *А*.

2. Сгруппировав выборку, как указано в пункте 1, переходят к расчету вспомогательных величин, нужных для определению сумм квадратов отклонений:

$Д_o = \sum (Vi – M0)^2;$

$Dф = n\sum (Mr – M0)^2;$

$Dc = \sum [\sum (Vi – Mr)^2]$, *или*

$Dc = Д0 – Дф.$

3. Закончив расчёт, определяют числа степеней свободы **(К):**

   $K0 = N – 1$**-** для общего варьирования;

   $Kф = a – 1$-для вариации межгрупповой (факториальной);

   $Kc = N – a$, или $Kc = (N – 1) – (a – 1)$ –для вариации внутригрупповой, или остаточной.

Здесь *а* – число градаций фактора *А.* Следует иметь в виду, что числа степеней свободы в соответствии с равенством $D0 = Dф + Dc$ находятся между собой в определённых количественных соотношениях, т.е.

$K0 = Kф + Kc$

По этому равенству можно контролировать правильность расчета этих величин.

4. Определение средних квадратов отклонений или дисперсий по отношениям сумм квадратов отклонений к соответствующим числам свободы, т.е. общая дисперсия для всего комплекса равна:

   $S_o^2 = \frac{Д_o}{N-1}$

   $Sф^2 = \frac{Дф}{a-1}$ – межгрупповая дисперсия;

   $Sc^2 = \frac{Дc}{N-a}$ – внутригрупповая (остаточная) дисперсия.

5. Наконец, определяется эффективность действия фактора *A* на результативный признак. Для этого служит дисперсионное отношение, или критерий Р. Фишера *(Fф),* который вычисляется по формуле:

$Fф = \frac{Sф^2}{Sc^2}$*(при* $Sф^2 \geq Sc^2$*).*

Результаты дисперсионного анализа сводятся в табл. 3.1.

Таблица 3.1.

| *Вариация* | *Числа степеней свободы (К)* | *Сумма квадратов (D)* | *Средние квадраты* $(S^2)$ | *Дисперсионное отношение (Fф)* |
|---|---|---|---|---|
| *Межгрупповая* | $Кф = a - 1$ | *Dф* | $Sф^2 = \frac{Дф}{Кф}$ | $Fф = \frac{Sф^2}{Sc^2}$ |
| *Внутригрупповая (Остаточная)* | $Кс = N - a$ | *Dc* | $Sc^2 = \frac{Дс}{Кс}$ | |
| *Общая* | $K_o = N - 1$ | $D_o$ | $S_o^2 = \frac{Д_о}{К_о}$ | |

Так как дисперсионное отношение $Fф = \frac{Sф^2}{Sc^2}$ -Величина случайная, ее необходимо сравнивать с табличным (стандартным) значением **(Fst)**критерия Р. Фишера для принятого уровня значимости ( а = 0,01; 0,05) и чисел степеней свободы *(Кф и Кс).* При этом, как указывалось выше, число степеней свободы для большей дисперсии находятся по горизонтали, а для меньшей дисперсии – в первом столбце (значения критерия Фишера *Fst),* которая приводится в приложении 1.

Нулевая гипотеза отвергается и эффективность действия фактора *A* на результативный признак признается статистически достоверной, если $Fф \geq Fst$**.** В противном случае нулевая гипотеза (предположение) сохранится, т.е. действие фактора *A* на результативный признак несущественное, случайное.

Вычисляется также показатель силы влияния $(\eta^2)$ как отношение факториальной дисперсии **(***Dф)* к общей *($D_o$),* т.е.

$$\eta^2 = \frac{Дф}{Д_о}$$

И его ошибкой по формуле

$$m\eta = \pm (1 - \eta^2)\, \frac{a-1}{N-a}$$

Затем делается вывод относительно силы влияния действующего фактора (возраста) на увеличение результативного признака (высота деревьев).

Технические приемы дисперсионного анализа разберем на конкретном примере (однофакторном и двухфакторном комплексах).

**Пример 3.1.** Провести дисперсионный анализ данных однофакторного вегетационного опыта с водными культурами по изучению действия соотношения натрия, фосфора и калия $N:P_2O_5:K_2O$ при питании рассады томатов на урожай плодов, установить, значимо ли различие в действии соотношений удобрений при питании на урожай плодов. Нулевая гипотеза равна нулю, т.е. все разности между средними по вариантам статистически не существенны.

| | | **Ранний урожай плодов** | | | |
|---|---|---|---|---|---|
| № п/п | Варианты $N:P_2O_5:K_2O$ | Урожай, Х | | | |
| | | I | II | III | IV |
| 1 | (st) 1:1:1 | 454 | 470 | 430 | 500 |
| 2 | 1:1:1 | 502 | 550 | 490 | 507 |
| 3 | 1:2:1 | 601 | 670 | 550 | 607 |
| 4 | 1:2:2 | 407 | 412 | 475 | 402 |
| 5 | 2:1:1 | 418 | 470 | 460 | 412 |

***Решение.***

**1.** *Составляем расчетную таблицу, располагая в ней исходные данные по рядам и столбцам, определяем суммы и средние по вариантам, общую сумму и среднее значение результативного признака по опыту (табл. 3.2).*

Таблица 3.2.

| Ранний урожай плодов | | | | | | | |
|---|---|---|---|---|---|---|---|
| Варианты $N:P_2O_5:K_2O$ | Урожай, X | | | | Число наблюдений n | Суммы V | Средние |
| | I | II | III | IV | | | |
| 1(st) | 454 | 470 | 430 | 500 | 4 | 1854 | 463,5 |
| 2 | 502 | 550 | 490 | 507 | 4 | 2049 | 512,3 |
| 3 | 601 | 670 | 550 | 607 | 4 | 2428 | 607,0 |
| 4 | 407 | 412 | 475 | 402 | 4 | 1696 | 424,0 |
| 5 | 418 | 470 | 460 | 412 | 4 | 1760 | 440,0 |
| Общая сумма | - | - | - | | 20 | 9787 | 489,4 |

**2.** *Вычисляем суммы квадратов отклонений.*

Для вычисления суммы квадратов отклонений исходные даты целесообразно преобразовать по соотношению $X_1 = X - A$, приняв за условную среднюю А число 500, близкое к среднему урожаю по опыту $\bar{x} = 489{,}4$ (табл. 11 )

Таблица 3.3

| Таблица преобразованных дат | | | | | | |
|---|---|---|---|---|---|---|
| Варианты | Повторности, $X_i$=X-A | | | | Суммы V | $V^2$ |
| | I | II | III | IV | | |
| 1 | -46 | -30 | -70 | 0 | -146 | 21316 |
| 2 | 2 | 50 | -10 | 7 | 49 | 2401 |
| 3 | 101 | 170 | 50 | 107 | 428 | 183184 |
| 4 | -93 | -88 | -25 | -98 | -304 | 92416 |
| 5 | -82 | -30 | -40 | -88 | -240 | 57600 |
| Общая сумма | - | - | - | - | $\sum x_1$ =-213 | 45369 |

Вычисления суммы квадратов отклонений проведем в такой последовательности:

**Общее число наблюдений: (**5 вариантов) * (4 повторений) =20

$$N = \sum n = 20$$

**Корректирующий фактор**

$$C = \left(\sum X_1\right)^2 : N = (213)^2 : 20 = 2268$$

Зависимость между этими источниками варьирования выразится равенством: $\boldsymbol{C_y = C_v + C_z}$, где: $\boldsymbol{C_y}$ - общая сумма квадратов;

$\boldsymbol{C_v}$ - варьирование между выборками (вариантами);

$\boldsymbol{C_z}$ – варьирование внутри выборок.

Здесь вариация между выборками (вариантами) представляет ту часть общей дисперсии, которая обусловлена действием изучаемых факторов, а дисперсия внутри выборок характеризует случайное варьирование изучаемого признака, т.е. ошибку эксперимента.

**Общая сумма квадратов отклонений:**

$$C_y = \sum X_1^2 - C = (46^2 + 30^2 + \dots + 88^2) - 2268 = 104941$$

**Сумма квадратов для вариантов:**

$$C_v = \sum V^2 : n - C$$

$$C_v = (146^2 + 49^2 + \dots + 240^2) : 4 - 2268 = 86961$$

**Остаточную сумму квадратов**, соответствующую случайному варьированию вычисляем по разности:

$$C_z = C_y - C_v = 104941 - 86961 = 1798$$

*3. Определяем фактическое и теоретическое значение критерия Фишера $F_{ф}$ и $F_{05}$, составим таблицу 12 дисперсионного анализа.*

**Общее число степеней свободы** $(\boldsymbol{N-1})$ также расчленяется на две части – **степени свободы для вариантов** $(\boldsymbol{l-1})$ и для **случайного варьирования** – $(\boldsymbol{N-l})$ остатка:

$$\boldsymbol{N-1 = (l-1) + (N-l),}$$

где: $\boldsymbol{N}$ - общее число наблюдений; $\boldsymbol{l}$ – число вариантов.

Отнесением сумм квадратов отклонений к числам степеней свободы получают дисперсии: $\boldsymbol{S_y^2 = \frac{C_y}{N-1}}$, $\boldsymbol{S_v^2 = \frac{C_v}{l-1}}$, $\boldsymbol{S_z^2 = \frac{C_z}{N-l}}$, которые служат оценками соответствующих генеральных параметров;

$\boldsymbol{S_y^2}$ **- общей дисперсии комплекса;**

$\boldsymbol{S_v^2}$ **- межгрупповой дисперсии;**

$\boldsymbol{S_z^2}$ **– дисперсии внутригрупповой или остаточной.**

Отношение дисперсии межгрупповой или факториальной к дисперсии внутригрупповой или остаточной:

$\frac{S_v^2}{S_z^2} = F_ф$ – служит **критерием оценки влияния на признак регулируемых в опыте факторов**.

Таблица 3.4

| **Результаты дисперсионного анализа** | | | | | |
|---|---|---|---|---|---|
| Дисперсия | Сумма квадратов | Степени свободы | Средний квадрат | $F_ф$ | $F_{0,5}$ |
| Общая *Cy* | $\sum X^2 - C$ | *N-1* | | | |
| Вариантов *Cv* | $\sum \frac{V^2}{n} - C$ | $l\text{-}1=f_1$ | $S_v^2$ | $\frac{S_v^2}{S_z^2}$ | По таблице |
| Остаток (ошибки) *Cz* | $C_y - C_v$ | $N\text{-}l= f_2$ | $S_z^2$ | | |

В нашей задаче получим таблицу дисперсионного анализа.

| **Результаты дисперсионного анализа** | | | | | |
|---|---|---|---|---|---|
| Дисперсия | Сумма квадратов | Степени свободы | Средний квадрат | $F_ф$ | $F_{05}$ |
| Общая *Cy* | 104941 | 19 | - | - | - |
| Вариантов *Cv* | 86961 | $4= f_1$ | 21740 | 18,14 | 3,06 |
| Остаток (ошибки) *Cz* | 17980 | $15= f_2$ | 1199 | - | - |

**Фактическое значение:** $\frac{S_v^2}{S_z^2} = F_ф$ , $= \frac{21740}{1199} \approx 18{,}14$

Теоретическое значение $F_{05}$ находят из приложения 4, на 5%-ном уровне значимости по $f_1$=4 степеням свободы для дисперсии вариантов и $f_2$=15 степеням для остатка, получим: $\mathbf{F_{05} = 3{,}06}$.

*При* $F_ф > F_{05}$ *в опыте есть существенные различия по вариантам и нулевая гипотеза* $H_0: d = 0$ *отвергается, следовательно фактор действия соотношения удобрений при питании рассады томатов на урожай плодов значимо .*

*Критерий F* устанавливает только факт наличия существенных различий между средними, но не указывает, между какими средними имеются эти различия. В практике опытной работы используется несколько методов для оценки существенности разности между средними, один из таких методов применим для нашей задачи.

***4.*** *Определяем ошибку опыта и существенность частных различий:*

а) *ошибка опыта*

$$S_{\bar{x}} = \sqrt{\frac{S_z^2}{n}} = \sqrt{\frac{1199}{4}} = 17{,}3$$

б) *ошибка разности средних*

$$S_d = \sqrt{\frac{2S_z^2}{n}} = \sqrt{\frac{2 * 1199}{4}} = 24{,}5$$

в) *наименьшую существенную разность* для 5%-ного уровня значимости в абсолютных и относительных показателях

$$HCP_{05} = t_{05} * S_d = 2{,}13 * 24{,}5 = 52{,}2;$$

$$HCP_{05} = \frac{t_{05} * S_d}{\bar{x}} * 100 = \frac{52{,}2}{489{,}4} * 100 = 10{,}7\%.$$

Значение критерия $t_{05}$ возьмем из приложения 2 для 15 степеней свободы дисперсии остатка (ошибки).

*Критерий наименьшей существенной разности (НСР) = t*Sd указывает предельную ошибку для разности двух выборочных средних. Если фактическая разность d≤НСР, то она существенна, значима, а если d<НСР – несущественна, незначима.*

Итоги результатов опыта и статистической обработки данных записываем в таблицу 3.5.

Таблица 3.5

| **Итоговая таблица дисперсионного анализа** | | | |
|---|---|---|---|
| Варианты | Урожай, г/сосуд | Разность со стандартом | |
| | | г | % |
| 1(st) | 463,5 | - | - |
| 2 | 512,3 | 48,8 | 10,0 |

| | | | |
|---|---|---|---|
| 3 | 607,0 | 143,5 | 29,3 |
| 4 | 424,0 | -39,5 | -8,1 |
| 5 | 440,0 | -23,5 | -4,8 |
| НСР, г-сосуд | - | 52,2 | 10,7 |

**Вывод.** Увеличенное питание рассады фосфором и калием обеспечивает получение более высоких урожаев (соотношение 1:2:2); при усилении азотного питания имеет место тенденция к снижению урожая, однако статистически оно несуществен

## Индивидуальное задание №4 «Дисперсионный анализ»

*Вариант 1-10 (задание таблицы 3.6,3.6а)*

Провести дисперсионный анализ данных однофакторного вегетационного опыта с культурами по изучению влияние азотных удобрений на урожайность лука, г/сосуд, установить, значимо ли различие видов азотных удобрении на урожай лука. Нулевая гипотеза равна нулю, т.е. все разности между средними по вариантам статистически не существенны.

Влияние азотных удобрений на урожай лука.

Таблица 3.6.

| № п/п | Варианты, свет л/к | Повторность | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | Без азота (st) | 15,8 | 16,1 | 16,0 | 17,2 | 14,9 | 15,5 | 16,1 |
| 2 | Сульфат аммония | 29,4 | 30,1 | 27,3 | 28,1 | 25,4 | 27,6 | 26,1 |
| 3 | Аммиачная селитра | 46,1 | 13,9 | 44,2 | 44,1 | 46,0 | 44,4 | 48,6 |
| 4 | Мочевина | 38,4 | 39,9 | 40,2 | 38,5 | 39,6 | 40,1 | 39,3 |

Данные урожайности по столбцам:

Таблица 3.6а.

| № варианта | Повторность | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| 1 | 1 | 2 | 3 | 4 |
| 2 | 2 | 3 | 4 | 5 |
| 3 | 3 | 4 | 5 | 6 |
| 4 | 1 | 2 | 3 | 5 |
| 5 | 4 | 5 | 6 | 7 |
| 6 | 1 | 2 | 3 | 6 |
| 7 | 2 | 3 | 5 | 6 |
| 8 | 1 | 2 | 3 | 7 |
| 9 | 2 | 3 | 5 | 7 |

| | | | | |
|---|---|---|---|---|
| 10 | 1 | 2 | 4 | 7 |

*Вариант 11-20 (задание таблицы 3.7, 3.7а)*

Провести дисперсионный анализ данных однофакторного вегетационного опыта с культурами по изучению влияние удобрений на урожайность томатов, г/сосуд, установить, значимо ли различие видов удобрении на урожай томатов. Нулевая гипотеза равна нулю, т.е. все разности между средними по вариантам статистически не существенны.

Влияние удобрений на урожайность томатов, г/сосуд.

Таблица 3.6.

| № п/п | Варианты опыта | Повторность | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
| 1 | Без удобрений | 95 | 101 | 85 | 100 | 91 | 87 | 108 |
| 2 | *NPK (фон)* | 185 | 200 | 190 | 195 | 180 | 191 | 201 |
| 3 | *NPK+Cu* | 215 | 236 | 234 | 225 | 198 | 215 | 245 |
| 4 | *NPK+B* | 370 | 354 | 361 | 370 | 351 | 385 | 370 |
| 5 | *NPK+Mn* | 214 | 200 | 198 | 213 | 198 | 186 | 195 |

Данные урожайности по столбцам:

Таблица 3.6а

| № варианта | Повторность | | | |
|---|---|---|---|---|
| | I | II | III | IV |
| 11 | 1 | 2 | 3 | 4 |
| 12 | 2 | 3 | 4 | 5 |
| 13 | 3 | 4 | 5 | 6 |
| 14 | 1 | 2 | 3 | 5 |
| 15 | 4 | 5 | 6 | 7 |
| 16 | 1 | 2 | 3 | 6 |
| 17 | 2 | 3 | 5 | 6 |
| 18 | 1 | 2 | 3 | 7 |
| 19 | 2 | 3 | 5 | 7 |
| 20 | 1 | 2 | 4 | 7 |

# РАЗДЕЛ 4. Элементы корреляционного и регрессионного анализа.

## П.1.Функциональная, статистическая, корреляционная зависимости

*Функциональная зависимость* между двумя переменными характеризуется тем, что каждому значению одной переменной соответствует вполне определённое значение другой переменной.

- *Примеры функциональных зависимостей:* путь и время; количество купленного товара и его стоимость; количество потреблённой энергии и плата за неё.

***Определение 4.1.*** Зависимость между случайными величинами *X* и *Y*, состоящая в том, что каждому значению одной величины соответствует распределение другой величины, называется ***статистической зависимостью***.

- *Примеры статистических зависимостей*: урожайность и количество внесённых удобрений, рост человека и его масса, объём дневной продукции и её себестоимость по различным предприятиям.

На практике часто используют связь между изменениями одной случайной величины *X* и изменениями средней другой величины *Y*.

***Определение 4.2.*** Зависимость между двумя переменными величинами состоящая в том, что при изменении одной величины изменяется ***условное среднее*** значение другой величины, называется ***корреляционной зависимостью***.

***Условным средним*** $\bar{y}_x$ называется среднее арифметическое наблюдаемых значений *Y*, соответствующих $X=x$.

*Например*, если при $X=3$, величина *Y* принимает значения $y_1=2$, $y_2=7$, $y_3=6$, то $\bar{y}_x = \frac{2+7+6}{3} = 5$.

*Примеры корреляционных зависимостей:* живая масса коров первотёлок и их средний удой ; яйценоскость кур и средний размер яиц и т.д.

Условные средние являются функциями соответственно от $x$ и $y$.

***Выборочным уравнением регрессии Y на X*** называется зависимость вида:

$$\bar{y}_x = f^*(x) \qquad (4.1)$$

***Выборочным уравнением регрессии X на Y*** называется зависимость вида

$$\bar{x}_y = \varphi^*(y). \qquad (4.2)$$

Графики соответствующих функций $f^*(x)$, $\varphi^*(y)$ называются ***выборочными линиями регрессии.***

Статистическую связь между переменными можно изучать методами корреляционного и регрессионного анализа.

***Основной задачей регрессионного анализа является установление формы зависимости между переменными и изучение этой зависимости.***

***Основная задача корреляционного анализа – выявление связи между переменными и оценка её тесноты.***

Для решения этих задач используют соответствующий математический аппарат. Рассмотрим задачу в общем виде.

Пусть имеется два ряда наблюдаемых зависимых между собой величин $X$ и $Y$. Если $x_i$ и $y_i$ встречаются по одному разу, то их значения записывают в виде таблицы 4.1.

Таблица 4.1

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $X_i$ | $x_1$ | $x_2$ | $x_3$ | … | $x_k$ | … | $x_n$ |
| $Y_i$ | $y_1$ | $y_2$ | $y_3$ | … | $y_k$ | … | $y_n$ |

Графически, если взаимосвязь двух признаков задана таблицей 1, то в системе координат строят точки $(x_i;y_i)$ и по расположению точек судят о форме зависимости (линейная, показательная, параболическая и т.д.). Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определённой линии, выражающей форму связи (см. рис. 4.1).


Если каждому значению $y_i$ отвечает несколько значений $x$, а каждому $x_j$ – несколько значений $y$, то эти данные записывают в виде таблицы 4.2 .

Таблица 4.2

| Y / X | $x_1$ | $x_2$ | $\cdots$ | $x_k$ | $\sum\limits_{j=1}^{k}$ |
|---|---|---|---|---|---|
| $y_1$ | $n_{11}$ | $n_{12}$ | $\cdots$ | $n_{1k}$ | $\sum n_{1j}=l_1$ |
| $y_2$ | $n_{21}$ | $n_{22}$ | $\cdots$ | $n_{2k}$ | $\sum n_{2j}=l_2$ |
| $\dots$ | $\dots$ | $\dots$ | $\dots$ | $\dots$ | $\dots$ |
| $y_m$ | $n_{m1}$ | $n_{m2}$ | $\cdots$ | $n_{mk}$ | $\sum n_{mj}=l_m$ |
| $\sum\limits_{i=1}^{m}$ | $\sum n_{i1}=n_1$ | $\sum n_{i2}=n_2$ | $\cdots$ | $\sum n_{ik}=n_k$ | $n=\sum\limits_{i=1}^{m} l_i=\sum\limits_{j=1}^{k} n_j$ |

- В таблице 5.2. $n_{ij}$ – частота, показывающая число повторений парные значений $x_i$ и $y_j$ . Если парные значения не появляются ни разу, то в соответствующей клетке ставят прочерк.

Таблица 5.2, в которой результаты наблюдений записаны в порядке возрастания с указанием частоты $n_{ij}$, называется ***корреляционной таблицей.***

Корреляционная таблица может быть составлена как для дискретных так и для непрерывных признаков. В последнем случае строят интервальный ряд для значений $X$, $Y$ затем переходят к дискретному ряду, заменяя соответствующий интервал $(x_{i-1};x_i)$ его серединой $x_i*=\dfrac{x_{i-1}+x_i}{2}$ . Также поступают со случайной величиной $Y$.

Графически, если взаимосвязь двух признаков задана таблицей 5.1, то в системе координат строят точки $(x_i;y_i)$ и по расположению точек судят о форме зависимости (линейная, показательная, параболическая и т.д.). Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определённой линии, выражающей форму связи (см. рис. 5.1).

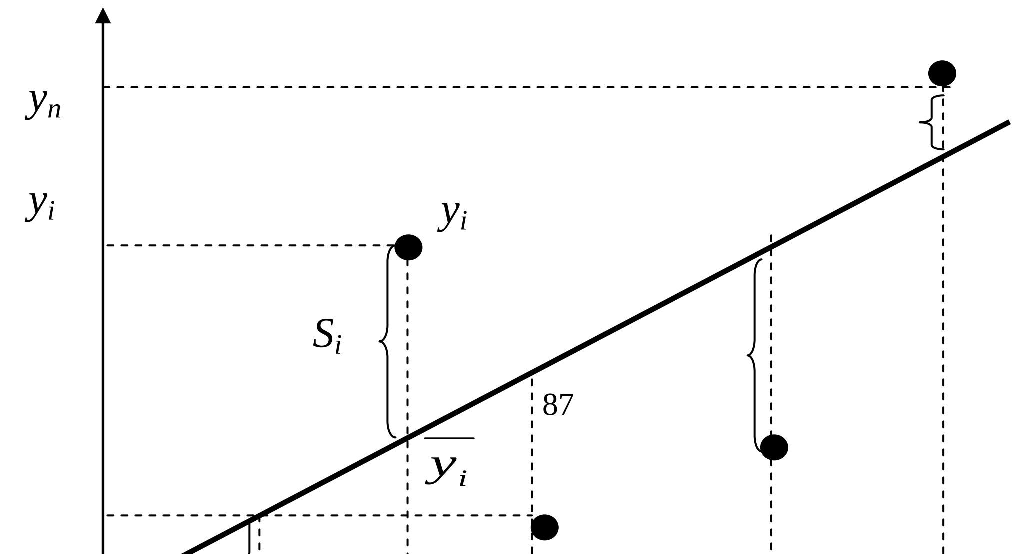
- Данные корреляционной таблицы также можно изобразить графически. Для этого каждую пару значений $(x_i;y_j)$ изображают точкой в системе координат; частоту $n_{ij}$, с которой эта пара встречается в таблице, изображают в виде числа, помещённого возле соответствующей точки.
- Графическое изображение корреляционной таблицы называется ***полем корреляции***. На таком графике наглядно можно проследить общую тенденцию в изменении $Y$ с изменением $X$ (см. рис. 5.2).

## П.2. Отыскание параметров выборочного уравнения *линейной регрессии*

Пусть известны результаты опыта, целью которого является исследование зависимости между двумя величинами $X$ и $Y$ (например, длина волос шерсти (см) у овец и настриг шерсти (кг) ). Предположим, что различные значения $x$ признака $X$ и соответствующие значения $y$ признака $Y$ наблюдаются по одному разу (см. табл. 4.1).Строим точки $(x_i;y_i)$ на координатной плоскости.

Предположим, что точки расположены приближённо вдоль некоторой прямой, уравнение которой ищем в виде:

$$y=a+bx\ . \tag{4.3}$$



Среди всех прямых линий $y=a+bx$ ищут наиболее близкую к данной системе точек, причём близость будем измерять суммой квадратов отклонений:

$$S_i = \sum_{i=1}^{n} \left[ y_i - \overline{y_i} \right]^2$$

$$S_i = \sum_{i=1}^{n} \left[ y_i - (a + bx_i) \right]^2 .$$

Из всех прямых выбирают ту, для которой сумма квадратов отклонений $S$ минимальна.

Так как минимизируется сумма квадратов отклонений экспериментальных и теоретических значений, то предложенный метод называют *методом наименьших квадратов*. А прямую построенную по этому методу, называют *МНК-прямой*. Для определения параметров уравнения прямой линии $y$ на $x$, получают систему *нормальных уравнений*.

***Система нормальных уравнений*** для отыскания неизвестных параметров прямой линии регрессии имеет вид:

$$\begin{cases} \left( \sum_{i=1}^{n} x_i^{\,2} \right) \cdot b + \left( \sum_{i=1}^{n} x_i \right) \cdot a = \sum_{i=1}^{n} x_i \cdot y_i \\ \left( \sum_{i=1}^{n} x_i \right) \cdot b + na = \sum_{i=1}^{n} y_i \end{cases} . \qquad (4.4)$$

Решая систему (4) получим формулы для определения неизвестных параметров регрессии $a$ и $b$ (или *МНК– оценок*):

$$b=\frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2} \quad ; \qquad a=\bar{y}-b\bar{x} \ . \tag{4.5}$$

где $\bar{x}=\frac{\sum x}{n}$; $\bar{y}=\frac{\sum y}{n}$ - средние арифметические

Обозначив отклонения выборочных данных от средних значений:

$\Delta x=(x_i-\bar{x})$; $\Delta y=(y_i-\bar{y})$, получим формулы из формулы (4.5):

$$b=\frac{\sum_{i=1}^{n}\Delta x\cdot\Delta y}{\sum_{i=1}^{n}\Delta x^2} \quad ; \qquad a=\bar{y}-b\bar{x} \ . \tag{4.5a}$$

Уравнение регрессии *Y* по *X* можно также записать в виде:

$$y=b(x-\bar{x})+\bar{y} \ . \tag{4.6}$$

*Угловой коэффициент* $b$ в уравнении регрессии называется *выборочным* ***коэффициентом регрессии Y по X .*** Обозначим его $b_{yx}$ ,тогда уравнение регрессии можно записать в виде:

$$y=b_{yx}(x-\bar{x})+\bar{y} \quad \text{или} \quad y-\bar{y}=b_{yx}(x-\bar{x}). \tag{4.7}$$

**Коэффициент регрессии *Y* по *X* показывает на сколько единиц в *среднем* изменяется переменная *Y* при увеличении переменной *X* на одну единицу.** Например, если b=0,2 - это значит при увеличений длины волос на 1 см, настриг шерсти возрастет на 0,2 кг.

### П.3. Эмпирический коэффициент корреляции и его свойства

Предположим, что наблюдения проводят над системой двух случайных величин *X* и *Y*, связанных линейной зависимостью. Для оценки тесноты связи между ними по результатам выборки вычисляют статистику, называемую *коэффициентом корреляции*.

***Коэффициент корреляции*** вычисляют по формулам:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}; \tag{4.7}$$

### *Свойства коэффициента корреляции:*

1.Значения коэффициента корреляции изменяются на отрезке [-1;1]: $-1 \le r \le 1$ .

2. Чем больше $|r|$ , тем теснее связь между изучаемыми признаками.

- В зависимости от того насколько $|r|$ близок к 1, различают связь *слабую, умеренную, заметную, достаточно тесную, тесную и весьма тесную.*

3. Если $|r| = 1$, то корреляционная связь становится функциональной.

4. Если $r = 0$, то между изучаемыми признаками нет линейной корреляционной зависимости, но не исключается существование какого либо другого вида корреляционной зависимости (параболической, гиперболической, показательной и т.д.).

5. Если $r > 0$, то связь между признаками *X, Y* прямая; если $r < 0$, то эта связь обратная.

Коэффициент корреляции, возведенный в квадрат, называется ***коэффициентом детерминации*** $r^2$. Он показывает долю изменений, которые вызваны факторным признаком. Коэффициент детерминации $r^2$ является прямым способом выражения зависимости одного признака от другого. Если известно, что У находится в причинной связи с Х, то $r^2$ – это доля вариации У, обусловленная влиянием Х.

***Замечание.*** Корреляционные зависимости наблюдаются между очень многими признаками организмов – морфологическими, физиологическими, а также между различными биологическими процессами. Различают положительную и отрицательную корреляции. При положительной корреляции с увеличением одного признака увеличивается и другой. Например, с увеличением живой массы коров первотёлок возрастает и удой; чем выше процент жира в молоке, тем выше и процент белка в нём. При отрицательной корреляции с увеличением удоя у коров снижается жирность молока; куры с высокой яйценоскостью имеют более мелкие яйца.

***Пример 4.1.*** С помощью корреляционного анализа определите влияние длины волоса шерсти на настриг шерсти. Для этого постройте линейное уравнение регрессии, рассчитайте коэффициенты корреляции и детерминаций. Сделайте выводы.

*Первичные данные по настригу (кг) и длине волоса шерсти (см) овец*

| **№** | **Длина волоса шерсти, см (X)** | **Настриг шерсти, кг (Y)** |
|---|---|---|
| **1** | 11,9 | 4,2 |
| **2** | 20,4 | 5,1 |
| **3** | 17,1 | 4,8 |
| . | ---------------------- | ------------------------------- |
| . | ------------------------ | ---------------------------- |
| . | -------------------------- | ------------------------------- |
| **22** | 15,0 | 4,5 |
| **23** | 22,3 | 4,9 |
| **24** | 16,0 | 4,3 |
| **25** | 13,9 | 4,2 |

***Решение.***

Строим точки ($x_i$;$y_i$) на координатной плоскости (рисунок 4.1)

***1) Найдем параметры линейной регрессий.***

Предположим, что точки расположены приближённо вдоль некоторой прямой (рис.2), уравнение которой ищем в виде:

$$y=a+bx\ .$$

Уравнение является уравнением линейной регрессий

Предположим, что точки расположены приближённо вдоль некоторой прямой, уравнение которой ищем в виде:

$$y=a+bx\ .$$

Уравнение является уравнением линейной регрессий

Для нахождения параметров линейной регрессий воспользуемся формулами:

$$b=\frac{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)\left(y_i-\bar{y}\right)}{\sum_{i=1}^{n}\left(x_i-\bar{x}\right)^2}\quad;\qquad a=\bar{y}-b\bar{x}\ .$$

где $\bar{x}=\frac{\sum x}{n};\quad \bar{y}=\frac{\sum y}{n};\quad .$

Обозначим: $\Delta x_i=x_i-\bar{x};\quad \Delta y_i=y_i-\bar{y}$, тогда имеем:

$$b=\frac{\sum_{i=1}^{n}\Delta x_i\,\Delta y_i}{\sum_{i=1}^{n}\Delta x_i{}^2}.$$

Угловой коэффициент $b$ в уравнении регрессии называется *выборочным коэффициентом регрессии Y по X* . Обозначим его $b_{yx}$, тогда уравнение регрессии можно записать в виде:

$$y=b_{yx}\left(x-\bar{x}\right)+\bar{y}\quad \text{или}\quad y-\bar{y}=b_{yx}\left(x-\bar{x}\right).$$

Коэффициент регрессии *Y* по *X* показывает на сколько единиц в *среднем* изменяется переменная *Y* при увеличении переменной *X* на одну единицу.

Для вычислений удобно составить расчетную таблицу 4.1:

| N | $x_i$ | $y_i$ | $\Delta x_i$ | $\Delta y_i$ | $\Delta x_i^2$ | $\Delta y_i^2$ | $\Delta x_i \Delta y_i$ | $x_i^2$ | $x_i y_i$ | $\bar{y}_i$ |
|---|---|---|---|---|---|---|---|---|---|---|
| **1** | **2** | **3** | **4** | **5** | **6** | **7** | **8** | **9** | **10** | **11** |
| 1 | 11,9 | 4,2 | -6,6 | -0,5 | 44,0 | 0,3 | 3,3 | 141,6 | 50,0 | 4,0 |
| 2 | 20,4 | 5,1 | 1,9 | 0,4 | 3,5 | 0,2 | 0,7 | 416,2 | 104,0 | 4,9 |
| - | - | - | - | - | - | - | - | - | - | - |
| 24 | 16 | 4,3 | -2,5 | -0,4 | 6,4 | 0,2 | 1,0 | 256,0 | 68,8 | 4,4 |
| 25 | 13,9 | 4,2 | -4,6 | -0,5 | 21,5 | 0,3 | 2,3 | 193,2 | 58,4 | 4,2 |
| $\sum$ | **463,3** | **117,6** | **0,0** | **0,0** | **256,0** | **3,7** | **27,2** | **8841,9** | **2206,6** | **117,6** |

Найдём среднюю выборочную по $X$ и по $Y$. Для этого используем суммы, записанные в последней строке первого и второго столбцов:

$$\bar{x} = \frac{\sum x}{n} = \frac{463{,}3}{25} = 18{,}5; \quad \bar{y} = \frac{\sum y}{n} = \frac{117{,}6}{25} = 4{,}7 .$$

Заполняем таблицу с четвёртого по восьмой столбец. В столбцах 6,7,8 подсчитываем суммы. Суммы, полученные по столбцам 6 и 8, подставляем в формулу для нахождения $b$ :

$$b = \frac{27{,}2}{256} = 0{,}11; \quad a = 4{,}7 - 0{,}11 \cdot 18{,}5 \approx 2{,}7$$

Тогда искомое уравнение регрессии*:* **<u>*y=2,7 +0,11x*</u>**, где

$b_{yx} = 0{,}11$ – коэффициент регрессии $Y$ по $X$.

***Вывод:*** При увеличении длины волоса шерсти на 1 см., настриг шерсти в среднем увеличивается на 110 г.

Теперь по уравнению регрессии рассчитаем теоретические значения $\hat{y}_i$ и заполним столбец 9. Например: $y_1 = 0{,}1 + 0{,}2 \cdot 11{,}9 = 3$.

***2) Вычислим коэффициенты корреляции и детерминаций***

**Для оценки тесноты связи между ними по результатам выборки вычисляют статистику, называемую *коэффициентом корреляции*.**

R (или r) – коэффициент корреляции. Устанавливает, есть ли связь между признаками, и насколько она тесная *-1≤R≤1*

Если же модуль коэффициента корреляции ~1,то связь близка к линейной (прямий).

***Коэффициент корреляции*** вычисляют по формулам:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} = \frac{\sum_{i=1}^{n}\Delta x_i \cdot \Delta y_i}{\sqrt{\sum_{i=1}^{n}\Delta x_i^{\ 2} \cdot \sum_{i=1}^{n}\Delta y_i^{\ 2}}}.$$

Итак, получаем: $r = \frac{340{,}2}{\sqrt{1377{,}3 \cdot 86{,}4}} = 0{,}878$.

***Вывод:*** Т.к. коэффициент корреляций близок к единице, то связь между признаками весьма тесная, прямая, т.е. чем больше длина шерсти, тем больше настриг.

Найдем коэффициент ***детерминации*** $r^2 = 0{,}76$ который показывает, что вариация настрига шерсти обусловлена на 76% влиянием длины волоса шерсти. Остальные 24% вариации настрига обусловлены неучтенными факторами.

***3) Построим найденную прямую регрессии в программе Microsoft Excel (рис.3).***

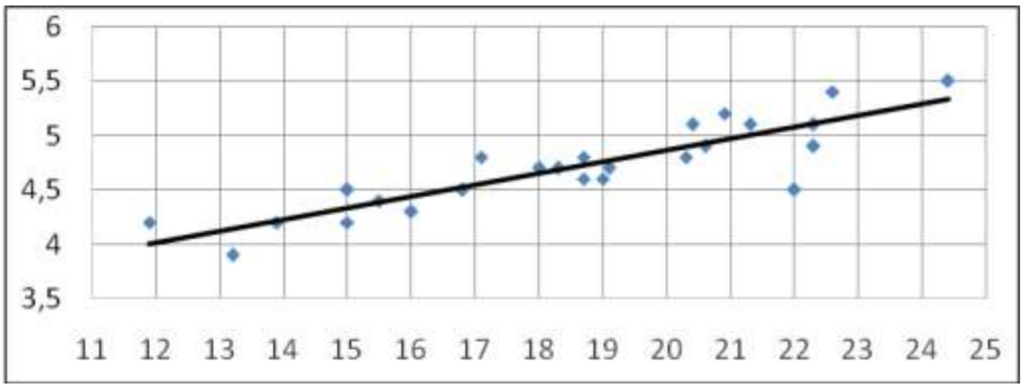



Рисунок 4.3

## Индивидуальное задание № 4. «Элементы корреляционного и регрессионного анализа»

Найти линейную регрессию $Y$ на $X$ и выборочный коэффициент корреляции, если заданы эмпирические зависимости.

Вариант 1.

Зависимость между толщиной стебля (мм) и высотой растения (см) раннеспелого сорта сои Искра 226:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Толщина стебля (мм) | 3,6 | 3,3 | 3 | 3,1 | 3,3 | 3,4 | 3,7 | 3,4 | 3,2 | 3,6 |
| Высота растения (см) | 54 | 49,5 | 40,7 | 39,5 | 31,6 | 40,4 | 40 | 43,3 | 35,3 | 45,9 |

Вариант 2.

Зависимость между толщиной стебля (мм) и высотой растения (см) раннеспелого сорта сои Венера высокая:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Толщина стебля (мм) | 3,9 | 3,1 | 3,6 | 3,1 | 2,8 | 3,4 | 4 | 3 | 3,5 | 3,5 |
| Высота растения см) | 40 | 47,4 | 30,5 | 36 | 44,1 | 31,6 | 42,6 | 40,2 | 41,4 | 39,3 |

Вариант 3.

Зависимость между числом бобов на растении (шт) и урожайностью (ц/га) раннеспелого сорта сои Тимирязевская 144:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Число бобов (шт) | 32 | 18 | 20 | 18 | 19 | 22 | 16 | 23 | 19 | 24 |
| Урожайность (ц/га) | 21 | 17,6 | 16,5 | 12,8 | 17,9 | 19 | 15,4 | 17,9 | 18,2 | 24 |

Вариант 4.

Зависимость между числом бобов на растении (шт) и урожайностью (ц/га) среднеспелого сорта сои Днепропетровская:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Число бобов (шт) | 23 | 20 | 16 | 18 | 17 | 29 | 28 | 20 | 14 | 18 |
| Урожайность (ц/га) | 23,2 | 21,2 | 16,2 | 17,6 | 14,8 | 30,8 | 20,1 | 19,8 | 13,1 | 15,6 |

Вариант 5.

Зависимость между вегетационным периодом (дн) и урожайностью семян (ц/га) сорта сои Приморская 529:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Вегетационный период (дн) | 97 | 98 | 99 | 95 | 94 | 94 | 98 | 93 | 94 | 96 |
| Урожайность семян (ц/га) | 21 | 17,6 | 16,5 | 12,8 | 17,9 | 19 | 17,9 | 18,2 | 15,4 | 24 |

Вариант 6.

Зависимость между содержанием масла (%) и белка (%) в семенах сои сорта Янтарный:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Содержание масла (%) | 18,7 | 19 | 18,7 | 18 | 17,6 | 18,6 | 18,9 | 18,5 | 19,8 | 18,3 |
| Содержание белка (%) | 39,7 | 40,6 | 39,7 | 40,2 | 42,2 | 39,2 | 40,8 | 40,3 | 39,3 | 40,3 |

Вариант 7.
Зависимость между высотой стебля (см) и полеганием гречихи (балл) скороспелой северной группы:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Высота стебля (см) | 5,4 | 5,4 | 5,2 | 3,9 | 5,5 | 4,8 | 5,8 | 5,9 | 4,2 | 4,1 |
| Полегание (балл) | 3 | 3,3 | 3,6 | 3,3 | 3,3 | 3,3 | 3,3 | 3,6 | 3,6 | 3 |

Вариант 8.
Зависимость между числом ветвей (шт) и количеством соцветий (шт) гречихи произрастающей в Приморском крае:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Число ветвей (шт) | 3,2 | 3,5 | 3,6 | 4,1 | 3,6 | 3,5 | 3,6 | 3,3 | 4,2 | 3,6 |
| Количество соцветий (шт) | 28,4 | 19,1 | 23,2 | 19,3 | 26,4 | 18,6 | 25,3 | 26,7 | 22,4 | 18,5 |

Вариант 9.
Зависимость между урожайностью образцов гречихи с $1м^2$ от вегетационного периода (дн):

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Вегетационный период (дн) | 65 | 66 | 69 | 67 | 71 | 73 | 69 | 67 | 66 | 68 |
| Урожайность | 40 | 43,3 | 97,1 | 98,6 | 139,8 | 62 | 76,2 | 80,7 | 64,4 | 103,9 |

Вариант 10.
Зависимость между количеством белка (%) от вегетационного периода (дн):

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Вегетационный период (дн) | 65 | 65 | 71 | 67 | 67 | 66 | 66 | 67 | 68 | 69 |
| Белок (%) | 15,9 | 17,5 | 17,9 | 17,4 | 18 | 17,1 | 17,5 | 16,8 | 17,2 | 15,3 |

Вариант 11.
Зависимость между количеством белка (%) от вегетационного периода (дн):

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Вегетационный период (дн) | 65 | 65 | 71 | 67 | 67 | 66 | 66 | 67 | 68 | 69 |
| Белок (%) | 15,9 | 17,5 | 17,9 | 17,4 | 18 | 17,1 | 17,5 | 16,8 | 17,2 | 15,3 |

Вариант 12.
Зависимость между количеством жира (%) в зерне пшеницы от вегетационного периода (дн):

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| Вегетационный период (дн) | 69 | 70 | 68 | 67 | 68 | 67 | 69 | 61 | 68 |
| Жирность (%) | 3,2 | 2,9 | 3,9 | 2,2 | 2,6 | 2,9 | 3,8 | 3,5 | 2,6 |

Вариант 13.
Зависимость между массой 1000 зерен (г) пшеницы и энергии прорастания (%):

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Масса 1000 зерен (г) | 26,15 | 26,1 | 25,34 | 25,23 | 25,04 | 25,06 | 24,65 | 24,57 |
| Энергия прорастания (%) | 93 | 90 | 88 | 86 | 83 | 86 | 80 | 79 |

Вариант 14.

Зависимость между массой 1000 зерен (г) и полегания (балл) образцов гречихи, произрастающей в Приморском крае:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Масса 1000 зерен (г) | 19 | 18,1 | 20,4 | 21,3 | 19,4 | 18,5 | 22,1 | 20 | 23,6 | 24,6 |
| Энергия прорастания (%) | 3,6 | 2,6 | 3 | 3,3 | 2,6 | 3 | 3 | 3 | 3,3 | 3,6 |

Вариант 15.

Зависимость между массой 1000 зерен (г) и плёнчатости (%) образцов гречихи среднеспелой Южной группы:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Масса 1000 зерен (г) | 21,6 | 24,8 | 22,4 | 22,5 | 21 | 21,8 | 18,3 | 24,6 | 24,4 | 23,2 |
| Пленчатости (%) | 22,6 | 24,1 | 21 | 18,1 | 21 | 23,2 | 21,4 | 25,3 | 24,3 | 21,9 |

Вариант 16.

Зависимость между жаростойкостью (балл) и массой 1000 зерен (г) овса скороспелой Северной группы:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Жаростойкость (балл) | 2 | 2,3 | 4 | 2 | 2,3 | 2 | 3,3 | 2 | 1,6 | 3 |
| Масса 1000 зерен (г) | 25,4 | 26,7 | 23,2 | 24,9 | 25,4 | 24,4 | 24,6 | 22,3 | 23,1 | 23 |

Вариант 17.

Зависимость между толщиной (мм) и высотой (см) стебля гречихи:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Толщина стебля (мм) | 0,32 | 0,37 | 0,4 | 0,38 | 0,4 | 0,43 | 0,43 | 0,4 | 0,55 | 0,43 |
| Высота растения (см) | 3,9 | 5,4 | 5,4 | 5,2 | 3,9 | 5,2 | 4,8 | 5,8 | 5,9 | 4,2 |

Вариант 18.

Зависимость между толщиной стебля (мм) и высотой растения (см) раннеспелого сорта сои Искра 226:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Толщина стебля (мм) | 3 | 3,4 | 3,5 | 2,9 | 3,2 | 3,6 | 2,9 | 3,6 | 3,6 | 3,4 |
| Высота растения (см) | 34,5 | 40,6 | 31,4 | 30,7 | 37,9 | 37,8 | 29,9 | 38,8 | 29,7 | 32,8 |

Вариант 19.

Зависимость между толщиной стебля (мм) и высотой растения (см) среднераннего сорта сои Венера высокая:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Толщина стебля (мм) | 3,4 | 3,6 | 3,8 | 3,8 | 3,6 | 3,6 | 4,1 | 4,1 | 3,5 | 3,4 |
| Высота растения (см) | 43,1 | 43,2 | 44,6 | 34,6 | 53,4 | 50,6 | 43,7 | 39,6 | 44,9 | 48,8 |

Вариант 20.

Зависимость между числом бобов на растении (шт) и урожайностью (ц/га) среднераннего сорта сои Днепропетровская:

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Число бобов (шт) | 22 | 26 | 20 | 16 | 15 | 17 | 22 | 17 | 18 | 27 |

| Урожайность (ц/га) | 15,9 | 27,4 | 17,3 | 15,6 | 15,1 | 14,8 | 21,5 | 17 | 17,9 | 24 |
|---|---|---|---|---|---|---|---|---|---|---|

## Задание 2.

Вариант №1

В таблице приведены данные о зависимости X массы 1000 зерен (г) пшеницы и Y энергии прорастания (%).

| x/Y | 80 | 83 | 86 | 90 | 93 | $n_i$ |
|---|---|---|---|---|---|---|
| 24 | 3 | | | 1 | | 4 |
| 25 | 1 | 5 | 1 | | 3 | 10 |
| 26 | | 2 | 4 | 5 | 7 | 18 |
| 27 | | 1 | 10 | 12 | 15 | 38 |
| $n_j$ | 4 | 8 | 15 | 18 | 25 | 70 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить величину энергии прорастания (%), если масса 1000 зерен пшеницы равна 28 г.

4)Найти какую массу 1000 зерен пшеницы нужно взять, чтобы получить энергию прорастания 88 %.

Вариант №2

В таблице приведены данные о зависимости между X урожайностью образцов гречихи с 1 м$^2$ от Y вегетационного периода (дн).

| x/Y | 16 | 17 | 18 | 19 | 20 | $n_i$ |
|---|---|---|---|---|---|---|
| 65 | 2 | | 1 | 1 | | 4 |
| 67 | 3 | 5 | 6 | | | 14 |
| 69 | 1 | | 2 | 3 | 9 | 15 |
| 73 | | 1 | 5 | 7 | 14 | 27 |
| $n_j$ | 6 | 6 | 14 | 11 | 23 | 60 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить величину урожая гречихи, если вегетационный период равен 15 дней.

4)Найти с каким вегетационным периодом нужно взять гречиху, чтобы ее урожайность составляла 68 ц/га.

Вариант №3

В таблице приведены данные о зависимости между X толщиной стебля (мм) от Y высотой растения (см) среднераннего сорта сои.

| x/Y | 34 | 43 | 45 | 50 | 53 | $n_i$ |
|---|---|---|---|---|---|---|
| 3 | 11 | 2 | | 1 | | 14 |
| 4 | 4 | | 4 | 6 | 12 | 26 |
| 5 | | 3 | 2 | 3 | 10 | 18 |
| $n_j$ | 15 | 5 | 6 | 10 | 22 | 58 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить высоту растения, если толщина стебля равна 6 мм.

4)Найти какой толщины должен быть стебель сои, чтобы его высота составляла 47 (см).

Вариант №4

В таблице приведены данные о зависимости между X числом бобов на растении (шт) от Y урожайностью сои (ц/га).

| x/Y | 10 | 12 | 16 | 18 | 20 | $n_i$ |
|---|---|---|---|---|---|---|
| 12 | 11 | 2 | | | 1 | 14 |
| 16 | 4 | | 4 | 6 | 12 | 26 |
| 18 | | 4 | 2 | 3 | | 9 |
| 22 | 2 | | 5 | 7 | 17 | 31 |
| $n_j$ | 17 | 6 | 11 | 16 | 30 | 80 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить урожайность сои, если число бобов на растении равна 20 (шт)

4)Найти число бобов на растении при урожайности сои 15 ц/га.

Вариант №5

В таблице приведены данные о зависимости между X (%) содержанием масла от Y (%) белка в семенах сои.

| x/Y | 38 | 39 | 40 | 42 | 44 | $n_i$ |
|---|---|---|---|---|---|---|
| 16 | | | 3 | 7 | 10 | 20 |
| 17 | | 1 | 4 | 6 | 12 | 23 |
| 18 | 7 | 8 | | 3 | | 18 |
| 20 | 6 | 10 | 1 | | 2 | 19 |
| $n_j$ | 13 | 19 | 8 | 16 | 24 | 80 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить содержание белка в семенах сои (%), если содержание масла составляет 19 (%).

4)Найти содержание масла (%) если содержание белка 43 (%).

Вариант №6

В таблице приведены данные о зависимости между X (балл) жаростойкостью от Y (г) массы 1000 зерен овса.

| x/Y | 22 | 23 | 24 | 25 | 26 | $n_i$ |
|---|---|---|---|---|---|---|
| 1 | | | 3 | 7 | 10 | 20 |
| 2 | | 1 | 4 | 6 | 12 | 23 |
| 3 | 7 | 8 | | 3 | | 18 |
| 4 | 11 | 10 | 1 | | 2 | 24 |
| $n_j$ | 18 | 19 | 8 | 16 | 24 | 85 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2) Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3) Определить массу 1000 зерен овса при балле жаростойкости 5.

4)Найти жаростойкость (балл) ,если масса 1000 зерен овса составляет 21 г.

Вариант №7

В таблице приведены данные о зависимости X массы 1000 зерен (г) пшеницы и Y энергии прорастания (%).

| x/Y | 80 | 83 | 86 | 90 | 92 | $n_i$ |
|---|---|---|---|---|---|---|
| 24 | 3 | | | 1 | | 4 |
| 25 | 1 | 5 | 1 | | 3 | 10 |
| 26 | | 2 | 4 | 10 | 7 | 23 |
| 27 | | 1 | 10 | 12 | 15 | 38 |
| $n_j$ | 4 | 8 | 15 | 23 | 25 | 75 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить величину энергии прорастания (%), если масса 1000 зерен пшеницы равна 28 г.

4)Найти какую массу 1000 зерен пшеницы нужно взять, чтобы получить энергию прорастания 88 %.

Вариант №8

В таблице приведены данные о зависимости между X урожайностью образцов гречихи с 1 $м^2$ от Y вегетационного периода (дн).

| x/Y | 16 | 17 | 18 | 19 | 20 | $n_i$ |
|---|---|---|---|---|---|---|
| 65 | 2 | | 1 | 1 | | 4 |
| 67 | 3 | 5 | 6 | | | 14 |
| 69 | 1 | | 2 | 3 | 9 | 15 |
| 71 | | 6 | 5 | 7 | 14 | 32 |
| $n_j$ | 6 | 11 | 14 | 11 | 23 | 65 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить величину урожая гречихи, если вегетационный период равен 21 день.

4)Найти с каким вегетационным периодом нужно взять гречиху, чтобы ее урожайность составляла 70 ц/га.

Вариант №9

В таблице приведены данные о зависимости между X толщиной стебля (мм) от Y высотой растения (см) среднераннего сорта сои.

| x/Y | 34 | 43 | 45 | 50 | 53 | $n_i$ |
|---|---|---|---|---|---|---|
| 3 | 11 | 2 | | 1 | | 14 |
| 5 | 4 | | 4 | 6 | 12 | 26 |
| 6 | | 3 | 2 | 5 | 10 | 20 |
| $n_j$ | 15 | 5 | 6 | 12 | 22 | 60 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить высоту растения, если толщина стебля равна 4 мм.

4)Найти какой толщины должен быть стебель сои, чтобы его высота составляла 52 (см).

Вариант №10

В таблице приведены данные о зависимости между X числом бобов на растении (шт) от Y урожайностью сои (ц/га).

| x/Y | 10 | 12 | 16 | 18 | 20 | $n_i$ |
|---|---|---|---|---|---|---|
| 12 | 11 | 2 | | | 1 | 14 |
| 16 | 4 | | 4 | 6 | 12 | 26 |

| | | | | | | |
|---|---|---|---|---|---|---|
| 18 | | 5 | 2 | 3 | | 10 |
| 22 | 2 | | 9 | 7 | 17 | 35 |
| $n_j$ | 17 | 7 | 15 | 16 | 30 | 85 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить урожайность сои, если число бобов на растении равна 14 (шт)

4)Найти число бобов на растении при урожайности сои 14 ц/га.

Вариант №11

В таблице приведены данные о зависимости между X (%) содержанием масла от Y (%) белка в семенах сои.

| x/Y | 38 | 39 | 40 | 42 | 44 | $n_i$ |
|---|---|---|---|---|---|---|
| 16 | | | 3 | 7 | 10 | 20 |
| 17 | | 1 | 4 | 6 | 12 | 23 |
| 18 | 7 | 8 | | 3 | | 18 |
| 20 | 6 | 10 | 1 | | 7 | 24 |
| $n_j$ | 13 | 19 | 8 | 16 | 29 | 85 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить содержание белка в семенах сои (%), если содержание масла составляет 15 (%).

4)Найти содержание масла (%) если содержание белка 41 (%).

Вариант № 12

В таблице приведены данные о зависимости между X (балл) жаростойкостью от Y (г) массы 1000 зерен овса.

| x/Y | 21 | 23 | 24 | 25 | 26 | $n_i$ |
|---|---|---|---|---|---|---|
| 1 | | | 3 | 7 | 10 | 20 |
| 2 | | 1 | 4 | 6 | 12 | 23 |
| 3 | 7 | 8 | | 3 | | 18 |
| 4 | 11 | 10 | 1 | | 2 | 24 |
| $n_j$ | 18 | 19 | 8 | 16 | 24 | 85 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить массу 1000 зерен овса при балле жаростойкости 5.

4)Найти  жаростойкость (балл) ,если масса 1000 зерен овса составляет 21 г.

Вариант №13

В таблице приведены данные о зависимости X массы 1000 зерен (г) пшеницы и Y энергии прорастания (%).

| x/Y | 81 | 83 | 86 | 90 | 93 | $n_i$ |
|---|---|---|---|---|---|---|
| 24 | 3 | | | 1 | | 4 |
| 25 | 1 | 5 | 1 | | 3 | 10 |
| 26 | | 2 | 4 | 5 | 7 | 18 |
| 28 | | 1 | 10 | 12 | 15 | 38 |
| $n_j$ | 4 | 8 | 15 | 18 | 25 | 70 |

Требуется:

1)  Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить величину энергии прорастания (%), если масса 1000 зерен пшеницы равна 23 г.

4)Найти какую массу 1000 зерен пшеницы нужно взять, чтобы получить энергию прорастания 87 %.

Вариант № 14

В таблице приведены данные о зависимости между X урожайностью образцов гречихи с 1 м$^2$ от Y вегетационного периода (дн).

| x/Y | 15 | 17 | 18 | 19 | 20 | $n_i$ |
|---|---|---|---|---|---|---|
| 65 | 2 | | 1 | 1 | | 4 |
| 67 | 3 | 5 | 6 | | | 14 |
| 69 | 1 | | 2 | 3 | 9 | 15 |
| 74 | | 1 | 5 | 7 | 14 | 27 |
| $n_j$ | 6 | 6 | 14 | 11 | 23 | 60 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить величину урожая гречихи, если вегетационный период равен 16 дней.

4)Найти с каким вегетационным периодом нужно взять гречиху, чтобы ее урожайность составляла 68 ц/га.

Вариант № 15

В таблице приведены данные о зависимости между X толщиной стебля (мм) от Y высотой растения (см) среднераннего сорта сои.

| x/Y | 38 | 43 | 45 | 50 | 53 | $n_i$ |
|---|---|---|---|---|---|---|
| 2 | 11 | 2 |  | 1 |  | 14 |
| 4 | 4 |  | 4 | 6 | 9 | 23 |
| 5 |  | 3 | 2 | 3 | 10 | 18 |
| $n_j$ | 15 | 5 | 6 | 10 | 19 | 55 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить высоту растения, если толщина стебля равна 3 мм.

4)Найти какой толщины должен быть стебель сои, чтобы его высота составляла 44 (см).

Вариант №16

В таблице приведены данные о зависимости X массы 1000 зерен (г) пшеницы и Y энергии прорастания (%).

| x/Y | 80 | 83 | 86 | 90 | 93 | $n_i$ |
|---|---|---|---|---|---|---|
| 24 | 3 |  |  | 1 |  | 4 |
| 25 | 1 | 5 | 1 |  | 3 | 10 |
| 26 |  | 2 | 4 | 5 | 7 | 18 |
| 27 |  | 1 | 10 | 12 | 15 | 38 |
| $n_j$ | 4 | 8 | 15 | 18 | 25 | 70 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить величину энергии прорастания (%), если масса 1000 зерен пшеницы равна 28 г.

4)Найти какую массу 1000 зерен пшеницы нужно взять, чтобы получить энергию прорастания 88 %.

Вариант №17

В таблице приведены данные о зависимости между X урожайностью образцов гречихи с 1 $м^2$ от Y вегетационного периода (дн).

| x/Y | 16 | 17 | 18 | 19 | 20 | $n_i$ |
|---|---|---|---|---|---|---|
| 65 | 2 |  | 1 | 1 |  | 4 |
| 67 | 3 | 5 | 6 |  |  | 14 |
| 69 | 1 |  | 2 | 3 | 9 | 15 |
| 73 |  | 1 | 5 | 7 | 14 | 27 |

| | | | | | | |
|---|---|---|---|---|---|---|
| $n_j$ | 6 | 6 | 14 | 11 | 23 | 60 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить величину урожая гречихи, если вегетационный период равен 15 дней.

4)Найти с каким вегетационным периодом нужно взять гречиху, чтобы ее урожайность составляла 68 ц/га.

Вариант №18

В таблице приведены данные о зависимости между X толщиной стебля (мм) от Y высотой растения (см) среднераннего сорта сои.

| x/Y | 34 | 43 | 45 | 50 | 53 | $n_i$ |
|---|---|---|---|---|---|---|
| 3 | 11 | 2 | | 1 | | 14 |
| 4 | 4 | | 4 | 6 | 12 | 26 |
| 5 | | 3 | 2 | 3 | 10 | 18 |
| $n_j$ | 15 | 5 | 6 | 10 | 22 | 58 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить высоту растения, если толщина стебля равна 6 мм.

4)Найти какой толщины должен быть стебель сои, чтобы его высота составляла 47 (см).

Вариант №19

В таблице приведены данные о зависимости между X числом бобов на растении (шт) от Y урожайностью сои (ц/га).

| x/Y | 10 | 12 | 16 | 18 | 20 | $n_i$ |
|---|---|---|---|---|---|---|
| 12 | 11 | 2 | | | 1 | 14 |
| 16 | 4 | | 4 | 6 | 12 | 26 |
| 18 | | 4 | 2 | 3 | | 9 |
| 22 | 2 | | 5 | 7 | 17 | 31 |
| $n_j$ | 17 | 6 | 11 | 16 | 30 | 80 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить урожайность сои, если число бобов на растении равна 20 (шт)

4)Найти число бобов на растении при урожайности сои 15 ц/га.

Вариант №20

В таблице приведены данные о зависимости между X (%) содержанием масла от Y (%) белка в семенах сои.

| x/Y | 38 | 39 | 40 | 42 | 44 | $n_i$ |
|---|---|---|---|---|---|---|
| 16 | | | 3 | 7 | 10 | 20 |
| 17 | | 1 | 4 | 6 | 12 | 23 |
| 18 | 7 | 8 | | 3 | | 18 |
| 20 | 6 | 10 | 1 | | 2 | 19 |
| $n_j$ | 13 | 19 | 8 | 16 | 24 | 80 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить содержание белка в семенах сои (%), если содержание масла составляет 19 (%).

4)Найти содержание масла (%) если содержание белка 43 (%).

***Решение типового примера.***

В таблице приведены данные о зависимости между X толщиной стебля (мм) от Y высотой растения (см) среднераннего сорта сои.

| x/Y | 34 | 43 | 45 | 50 | 53 | $n_i$ |
|---|---|---|---|---|---|---|
| 3 | 11 | 2 | | 1 | | 14 |
| 4 | 4 | | 4 | 6 | 12 | 26 |
| 5 | | 3 | 2 | 3 | 10 | 18 |
| $n_j$ | 15 | 5 | 6 | 10 | 22 | 58 |

Требуется:

1)Найти коэффициент корреляции, сделать вывод о тесноте связи.

2)Составить уравнения прямых регрессии Y на X и X на Y, построить графики.

3)Определить высоту растения, если толщина стебля равна 6 мм.

4)Найти какой толщины должен быть стебель сои, чтобы его высота составляла 47 (см).

***Решение.***

1. Для нахождения коэффициента корреляции используем формулу:

$$r=\frac{\overline{xy}-\overline{x}\cdot\overline{y}}{S_x\cdot S_y} \qquad \text{или} \qquad r=\frac{\mu}{S_x\cdot S_y}.$$

Найдём средние выборочные значения и средние квадратические отклонения переменных $X$ и $Y$:

$$\overline{x}=\frac{1}{58}(3\cdot 14+4\cdot 26+5\cdot 18)=4{,}07; \quad \overline{y}=\frac{1}{58}(34\cdot 15+43\cdot 5+45\cdot 6+50\cdot 10+53\cdot 22)=45{,}88$$

$$S_x{}^2=\overline{x^2}-\left(\overline{x}\right)^2; \quad S_x=\sqrt{S_x{}^2}; \quad S_y{}^2=\overline{y^2}-\left(\overline{y}\right)^2; \quad S_y=\sqrt{S_y{}^2}.$$

$$S_x^2=\frac{1}{58}(3^2\cdot 14+4^2\cdot 26+5^2\cdot 18)-4{,}07^2=0{,}53854828$$

$$S_y^2=\frac{1}{58}(34^2\cdot 15+43^2\cdot 5+45^2\cdot 6+50^2\cdot 10+53^2\cdot 22)-45{,}88^2=59{,}38766897$$

Таким образом: $S_x=\sqrt{0{,}53854828}\approx 0{,}734$

$S_y=\sqrt{59{,}38766897}\approx 7{,}706$.

Чтобы найти ковариацию $\mu$ вычислим величину $\overline{xy}$.

$$\overline{xy}=\frac{1}{58}(34\cdot 3\cdot 11+34\cdot 4\cdot 4+43\cdot 3\cdot 2+43\cdot 5\cdot 3+45\cdot 4\cdot 4+45\cdot 5\cdot 2+50\cdot 3\cdot 1+50\cdot 4\cdot 6+$$
$$+50\cdot 5\cdot 3+53\cdot 4\cdot 12+53\cdot 5\cdot 10)=190{,}224$$

Тогда получаем, что

$$\mu=\overline{xy}-\overline{x}\cdot\overline{y}=190{,}224-4{,}07\cdot 45{,}88=3{,}4924$$

Следовательно, коэффициент вариации $r=\dfrac{3{,}4924}{0{,}734\cdot 7{,}706}\approx 0{,}617$

***Вывод: связь между признаками умеренная, прямая.***

2. Найдём коэффициенты регрессии, используя формулы:

$$b_{yx}=\frac{\mu}{S_x^2}=\frac{3{,}4924}{0{,}53854828}\approx 6{,}485$$

$$b_{xy}=\frac{\mu}{S_y^2}=\frac{3{,}4924}{59{,}38766897}\approx 0{,}059$$

Составим теперь уравнения регрессий:

а) уравнение регрессии $Y$ на $X$:

$$y-\overline{y}=b_{yx}\left(x-\overline{x}\right)\Rightarrow y-45{,}88=6{,}485(x-4{,}07)\Rightarrow y=6{,}485x+19{,}486$$

б) уравнение регрессии $X$ на $Y$:

$$x-\bar{x}=b_{xy}\left(y-\bar{y}\right)\Rightarrow x-4{,}07=0{,}059(x-45{,}88)\Rightarrow x=0{,}059x+1{,}363$$

Графики найденных регрессий приведены на рис. 3.

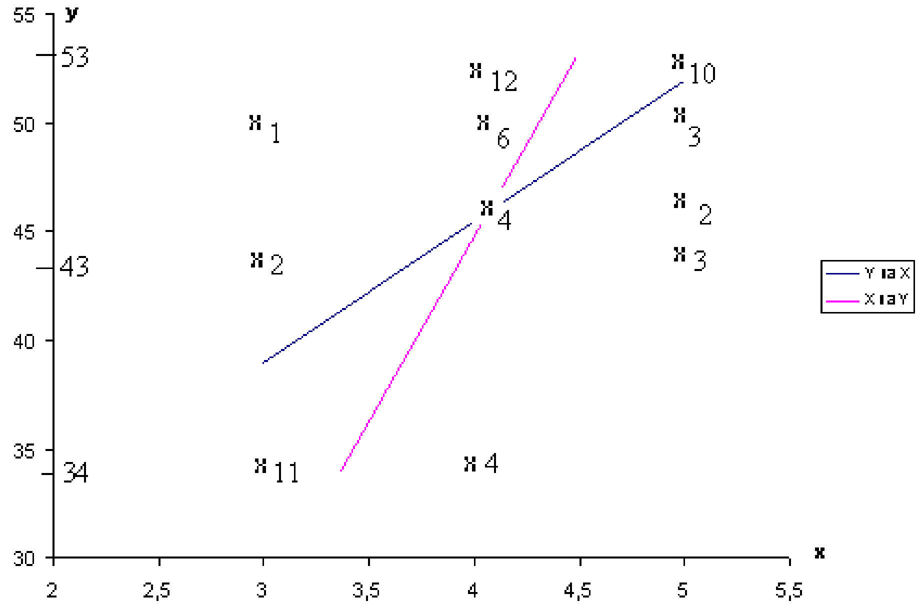



Рисунок 4.4

**3.** Чтобы найти высоту растения, если толщина стебля равна 6 мм, используя уравнение регрессии *Y* на *X*, найдём значение *y* при *x*=6:

$$y=6{,}485\cdot 6+19{,}486\approx 58{,}396\ \text{(см)}$$

Значит, если будет толщина стебля 6 мм, то получим высоту растения примерно 58,396 см.

**4**. Если высота составляла 47 см, то используя уравнение регрессии *X* на *Y* найдем толщину стебля: $x=0{,}059\cdot 47+1{,}363\approx 4{,}136$ (мм).

Значит, если высота растения примерно 47 см, то толщина стебля примерно 4,136 мм.

Литература.

*1. Основная литература.*

1. Доспехов Б.А. Методика полевого опыта (с основами статистической обработки результатов исследований). – 6 – е изд., стереотип. – М.: ИД Альянс, 2011. – 352 с., ил.

2. Глуховцев В.В., Кириченко В.Г., Зудилин С.Н., Практикум по основам научных исследований в агрономий. – М.: Колос, 2006. – 240 с.

*2. Дополнительная литература:*

1. Зайцев И.А. Высшая математика: учеб. для вузов/ И.А. Зайцев. – 4-е изд., стереотип. – М.: Дрофа, 2005.- 398с.

2. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высш.шк., 2001.

3. Гмурман В.Е. Теория вероятностей и математическая статистика. – М.: Высш.шк., 2001.

4.Практикум по статистике: учеб. пособие / под ред. А.П. Зинченко. – М.: КолосС, 2004. – 393с. МСХ.

2.Яковлев В.Б. Статистика. Расчеты в Microsoft Excel: учеб. пособие / В.Б. Яковлев. – М.: Колос, 2005. – 352с.МСХ.

**Таблица 1** Значения функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \cdot \int_0^x e^{-\frac{t^2}{2}} dt$

| x | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,0 | 0,0000 | 0040 | 0080 | 0120 | 0160 | 0199 | 0239 | 0279 | 0319 | 0359 |
| 0,1 | 0398 | 0438 | 0478 | 0517 | 0557 | 0596 | 0636 | 0675 | 0714 | 0753 |
| 0,2 | 0793 | 0832 | 0871 | 0910 | 0948 | 0987 | 1026 | 1064 | 1103 | 1141 |
| 0,3 | 1179 | 1217 | 1255 | 1293 | 1331 | 1368 | 1406 | 1443 | 1480 | 1517 |
| 0,4 | 1554 | 1591 | 1628 | 1664 | 1700 | 1736 | 1772 | 1808 | 1844 | 1879 |
| 0,5 | 1915 | 1950 | 1985 | 2019 | 2054 | 2088 | 2123 | 2157 | 2190 | 2224 |
| 0,6 | 2257 | 2291 | 2324 | 2357 | 2389 | 2422 | 2454 | 2486 | 2517 | 2549 |
| 0,7 | 2580 | 2611 | 2642 | 2673 | 2708 | 2734 | 2764 | 2794 | 2823 | 2852 |
| 0,8 | 2881 | 2910 | 2939 | 2967 | 2995 | 3023 | 3051 | 3078 | 3106 | 3133 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| 0,9 | 3159 | 3186 | 3212 | 3238 | 3264 | 3289 | 3315 | 3340 | 3365 | 3389 |
| 1,0 | 3413 | 3438 | 3461 | 3485 | 3508 | 3531 | 3554 | 3577 | 3599 | 3621 |
| 1,1 | 3643 | 3665 | 3696 | 3708 | 3729 | 3749 | 3770 | 3790 | 3810 | 3830 |
| 1,2 | 3849 | 3869 | 3883 | 3907 | 3925 | 3944 | 3962 | 3980 | 3997 | 4015 |
| 1,3 | 4032 | 4049 | 4066 | 4082 | 4099 | 4115 | 4131 | 4147 | 4162 | 4177 |
| 1,4 | 4192 | 4207 | 4222 | 4236 | 4251 | 4265 | 4279 | 4292 | 4306 | 4319 |
| 1,5 | 4332 | 4345 | 4357 | 4370 | 4382 | 4394 | 4406 | 4418 | 4429 | 4441 |
| 1,6 | 4452 | 4463 | 4474 | 4484 | 4495 | 4505 | 4515 | 4525 | 4535 | 4545 |
| 1,7 | 4554 | 4564 | 4573 | 4582 | 4591 | 4599 | 4608 | 4616 | 4625 | 4633 |
| 1,8 | 4641 | 4649 | 4656 | 4664 | 4671 | 4678 | 4686 | 4693 | 4699 | 4706 |
| 1,9 | 4713 | 4719 | 4726 | 4732 | 4738 | 4744 | 4750 | 4756 | 4761 | 4767 |
| 2,0 | 4772 | 4778 | 4783 | 4788 | 4793 | 4798 | 4803 | 4808 | 4812 | 4817 |
| 2,1 | 4821 | 4826 | 4830 | 4834 | 4838 | 4842 | 4846 | 4850 | 4854 | 4857 |
| 2,2 | 4861 | 4864 | 4868 | 4871 | 4875 | 4878 | 4881 | 4884 | 4887 | 4890 |
| 2,3 | 4893 | 4896 | 4898 | 4901 | 4904 | 4906 | 4909 | 4911 | 4913 | 4916 |
| 2,4 | 4918 | 4920 | 4922 | 4925 | 4927 | 4929 | 4931 | 4932 | 4034 | 4936 |
| 2,5 | 4938 | 4940 | 4941 | 4943 | 4945 | 4946 | 4948 | 4949 | 4951 | 4951 |
| 2,6 | 4953 | 4955 | 4956 | 4067 | 4959 | 4960 | 4961 | 4962 | 4963 | 4964 |
| 2,7 | 4965 | 4966 | 4967 | 4968 | 4969 | 4970 | 4971 | 4972 | 4973 | 4974 |
| 2,8 | 4974 | 4975 | 4976 | 4977 | 4977 | 4978 | 4979 | 4979 | 4980 | 4981 |
| 2,9 | 4981 | 4982 | 4982 | 4983 | 4984 | 4984 | 4985 | 4985 | 4986 | 4986 |

| x | | x | | x | | x | |
|---|---|---|---|---|---|---|---|
| 3,0 | 0,49865 | 3,5 | 0,49977 | 4,0 | 0,499968 | 4,5 | 0,4999966 |
| 3,1 | 0,49903 | 3,6 | 0,49984 | 4,1 | 0,499979 | 4,6 | 0,4999979 |
| 3,2 | 0,49931 | 3,7 | 0,49989 | 4,2 | 0,499987 | 4,7 | 0,4999987 |
| 3,3 | 0,49952 | 3,8 | 0,49993 | 4,3 | 0,499991 | 4,8 | 0,4999992 |
| 3,4 | 0,49966 | 3,9 | 0,49995 | 4,4 | 0,499995 | 4,9 | 0,4999995 |

**Таблица 2**

Теоретические значения $x^2$ теор при разных степенях свободы (V) и критических уровнях значимости

| Число степеней свободы, V | Уровень значимости | | | Число степеней свободы, V | Уровень значимости | | |
|---|---|---|---|---|---|---|---|
| | P=0,05 | P=0,01 | P=0,001 | | P=0,05 | P=0,01 | P=0,001 |
| **1** | 3,84 | 6,63 | 10,8 | **18** | 28,87 | 34,81 | 42,3 |
| **2** | 5,99 | 9,21 | 13,8 | **19** | 30,14 | 36,19 | 43,8 |
| **3** | 7,81 | 11,34 | 16,3 | **20** | 31,41 | 37,57 | 45,3 |
| **4** | 9,49 | 13,28 | 18,5 | **21** | 32,67 | 38,93 | |
| **5** | 11,07 | 12,83 | 20,5 | **22** | 33,92 | 40,29 | |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| **6** | 12,59 | 16,81 | 22,5 | **23** | 35,17 | 41,64 | |
| **7** | 14,07 | 18,48 | 24,3 | **24** | 36,42 | 42,98 | |
| **8** | 15,51 | 20,09 | 26,1 | **25** | 37,65 | 44,31 | |
| **9** | 16,92 | 21,67 | 27,9 | **26** | 38,89 | 45,64 | |
| **10** | 18,31 | 23,21 | 29,6 | **27** | 40,11 | 46,96 | |
| **11** | 19,68 | 24,72 | 31,3 | **28** | 41,34 | 48,28 | |
| **12** | 21,03 | 26,22 | 32,9 | **29** | 42,56 | 49,59 | |
| **13** | 22,36 | 27,69 | 34,5 | **30** | 43,77 | 50,89 | |
| **14** | 23,68 | 29,14 | 36,1 | | | | |
| **15** | 25,00 | 30,58 | 37,7 | **50** | 67,50 | 76,15 | |
| **16** | 26,30 | 32,00 | 39,3 | **80** | 101,88 | 112,3 | |
| **17** | 27,59 | 33,41 | 40,8 | **100** | 124,34 | 135,8 | |

**Таблица 3.** Таблица F – критерия Фишера.

**Значение *F*-критерия Фишера при уровне значимости 0,05**

| $V_2$ | $V_1$ - степени свободы (для большого среднего квадрата) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | 161 | 200 | 216 | 225 | 230 | 234 | 237 | 239 | 241 | 242 | 243 | 244 |
| 2 | 18,51 | 19,00 | 19,16 | 19,25 | 19,30 | 19,33 | 19,36 | 19,37 | 19,38 | 19,39 | 19,40 | 19,41 |
| 3 | 10,13 | 9,55 | 9,28 | 9,12 | 9,01 | 8,94 | 8,88 | 8,84 | 8,81 | 8,78 | 8,76 | 8,74 |
| 4 | 7,71 | 6,94 | 6,59 | 6,39 | 6,26 | 6,16 | 6,09 | 6,04 | 6,00 | 5,96 | 5,93 | 5,91 |

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 5 | 6,61 | 5,79 | 5,41 | 5,19 | 5,05 | 4,95 | 4,88 | 4,82 | 4,78 | 4,74 | 4,70 | 4,68 |
| 6 | 5,99 | 5,14 | 4,76 | 4,53 | 4,39 | 4,28 | 4,21 | 4,15 | 4,10 | 4,06 | 4,03 | 4,00 |
| 7 | 5,59 | 4,74 | 4,35 | 4,12 | 3,97 | 3,87 | 3,79 | 3,73 | 3,68 | 3,63 | 3,60 | 3,57 |
| 8 | 5,32 | 4,46 | 4,07 | 3,84 | 3,69 | 3,58 | 3,50 | 3,44 | 3,39 | 3,34 | 3,31 | 3,28 |
| 9 | 5,12 | 4,26 | 3,86 | 3,63 | 3,48 | 3,37 | 3,29 | 3,23 | 3,18 | 3,13 | 3,10 | 3,07 |
| 10 | 4,96 | 4,10 | 3,71 | 3,48 | 3,33 | 3,22 | 3,14 | 3,07 | 3,02 | 2,97 | 2,94 | 2,91 |
| 11 | 4,84 | 3,98 | 3,59 | 3,36 | 3,20 | 3,09 | 3,01 | 2,95 | 2,90 | 2,86 | 2,82 | 2,79 |
| 12 | 4,75 | 3,88 | 3,49 | 3,26 | 3,11 | 3,00 | 2,92 | 2,85 | 2,80 | 2,76 | 2,72 | 2,69 |
| 13 | 4,67 | 3,80 | 3,41 | 3,18 | 3,02 | 2,92 | 2,84 | 2,77 | 2,72 | 2,67 | 2,63 | 2,60 |
| 14 | 4,60 | 3,74 | 3,34 | 3,11 | 2,96 | 2,85 | 2,77 | 2,70 | 2,65 | 2,60 | 2,56 | 2,53 |
| 15 | 4,54 | 3,68 | 3,29 | 3,06 | 2,90 | 2,79 | 2,70 | 2,64 | 2,59 | 2,55 | 2,51 | 2,48 |
| 16 | 4,49 | 3,63 | 3,24 | 3,01 | 2,85 | 2,74 | 2,66 | 2,59 | 2,54 | 2,49 | 2,45 | 2,42 |
| 17 | 4,45 | 3,59 | 3,20 | 2,96 | 2,81 | 2,70 | 2,62 | 2,55 | 2,50 | 2,45 | 2,41 | 2,38 |
| 18 | 4,41 | 3,55 | 3,16 | 2,93 | 2,77 | 2,66 | 2,58 | 2,51 | 2,46 | 2,41 | 2,37 | 2,34 |
| 19 | 4,38 | 3,52 | 3,13 | 2,90 | 2,74 | 2,63 | 2,55 | 2,48 | 2,43 | 2,38 | 2,34 | 2,31 |
| 20 | 4,35 | 3,49 | 3,10 | 2,87 | 2,71 | 2,60 | 2,52 | 2,45 | 2,40 | 2,35 | 2,31 | 2,28 |
| 21 | 4,32 | 3,47 | 3,07 | 2,84 | 2,68 | 2,57 | 2,49 | 2,42 | 2,37 | 2,32 | 2,28 | 2,25 |
| 22 | 4,30 | 3,44 | 3,05 | 2,82 | 2,66 | 2,55 | 2,47 | 2,40 | 2,36 | 2,30 | 2,26 | 2,23 |
| 23 | 4,28 | 3,42 | 3,03 | 2,80 | 2,64 | 2,53 | 2,45 | 2,38 | 2,32 | 2,28 | 2,24 | 2,20 |
| 24 | 4,26 | 3,40 | 3,01 | 2,78 | 2,62 | 2,51 | 2,43 | 2,36 | 2,30 | 2,26 | 2,22 | 2,18 |
| 25 | 4,24 | 3,38 | 2,99 | 2,76 | 2,60 | 2,49 | 2,41 | 2,34 | 2,28 | 2,24 | 2,20 | 2,16 |
| 26 | 4,22 | 3,37 | 2,98 | 2,74 | 2,59 | 2,47 | 2,39 | 2,32 | 2,27 | 2,22 | 2,18 | 2,15 |
| 27 | 4,21 | 3,35 | 2,96 | 2,73 | 2,57 | 2,46 | 2,37 | 2,30 | 2,25 | 2,20 | 2,16 | 2,13 |

**Таблица 4**

Критерий Стьюдента t -критерий

| Число степеней Свободы, *V* | У зовни значимости (*P*) | | |
|---|---|---|---|
| | 0,05 | 0,01 | 0,001 |
| 1 | 12,7 | 63,66 | - |
| 2 | 4,30 | 9,93 | 31,60 |
| 3 | 3,18 | 5,84 | 12,94 |
| 4 | 2,78 | 4,60 | 8,61 |
| 5 | 2,57 | 4,03 | 6,86 |
| 6 | 2,45 | 3,71 | 5,96 |
| 7 | 2,37 | 3,50 | 5,41 |
| 8 | 2,31 | 3,36 | 5,04 |
| 9 | 2,26 | 3,25 | 4,78 |
| 10 | 2,23 | 3,17 | 4,59 |
| 11 | 2,20 | 3,11 | 4,44 |
| 12 | 2,18 | 3,06 | 4,32 |
| 13 | 2,16 | 3,01 | 4,22 |
| 14 | 2,15 | 2,98 | 4,14 |
| 15 | 2,13 | 2,95 | 4,07 |
| 16 | 2,12 | 2,92 | 4,02 |
| 17 | 2,11 | 2,90 | 3,97 |
| 18 | 2,10 | 2,88 | 3,92 |
| 19 | 2,09 | 2,86 | 3,88 |
| 20 | 2,09 | 2,85 | 3,85 |
| 21 | 2,08 | 2,83 | 3,82 |
| 22 | 2,07 | 2,82 | 3,79 |
| 23 | 2,07 | 2,81 | 3,77 |
| 24 | 2,06 | 2,80 | 3,75 |
| 25 | 2,06 | 2,79 | 3,73 |
| 26 | 2,06 | 2,78 | 3,71 |
| 27 | 2,05 | 2,77 | 3,69 |
| 28 | 2,05 | 2,76 | 3,67 |
| 29 | 2,05 | 2,76 | 3,66 |
| 30 | 2,04 | 2,75 | 3,65 |
| ТО | 1,96 | 2,58 | 3,29 |

Савельева Екатерина Владимировна

Статистические методы обработки результатов исследований: учебное пособие для обучающихся по направлениям подготовки: 35.03.04 Агрономия; 35.03.03 Агрохимия и агропочвоведение

Подписано в печать _____2015 г. Формат 60х90 1/16. Бумага писчая.

Печать офсетная. Уч.-изд.л. __. Тираж __ экз. Заказ _________

ФГБОУ ВО Приморская ГСХА

Адрес: 692510, г. Уссурийск, пр-т. Блюхера, 44

Участок оперативной полиграфии ФГБОУ ВПО Приморская ГСХА

692500, г. Уссурийск, ул. Раздольная, 8