Министерство сельского хозяйства РФ
ФГБОУ ВО «Иркутский государственный аграрный университет
им. А.А. Ежевского»

Кафедра Математики

ДОКУМЕНТ ПОДПИСАН
ЭЛЕКТРОННОЙ ПОДПИСЬЮ

Сертификат: 0341BE0F0000AEF1994783B067B4B3C28D
Владелец: Дмитриев Николай Николаевич
Действителен: с 16.12.2021 до 16.03.2023

Н.И. Овчинникова

Прикладная математика

Методические указания и контрольные задания

для студентов направления подготовки
21.04.02 - «Землеустройство и кадастры»

Иркутск – 2021

Рекомендовано к изданию научно-методическим советом инженерного факультета Иркутского государственного аграрного университета им. А.А. Ежевского (протокол № 7 от 26 марта 2021 г.).

Составитель: д.т.н., профессор Н.И. Овчинникова

Прикладная математика. Методические указания и контрольные задания для студентов направления подготовки 21.04.02 - «Землеустройство и кадастры» – Иркутск: Изд-во Иркутский ГАУ им. А.А. Ежевского, 2021. - 67 с.

Рецензент: д.т.н., профессор кафедры Информатики и математического моделирования Я.М. Иваньо.

Компьютерный набор и верстка: Н.И. Овчинникова

Данные методические указания включают в себя краткие теоретические сведения дисциплины прикладной математики, из разделов математической статистики: «Описательная статистика», «Элементы корреляционно-регрессионного анализа». Приведенный материал необходим студентам-магистрантам очной и заочной форм обучения для выполнения расчетов и анализа статистических данных при выполнении выпускной квалификационной работы и курсовых контрольных заданий. В указаниях приведены 20 вариантов контрольной работы; решение типового задания из нее; приложения в виде статистических таблиц и библиографический список.

© ФГОУ ВО «Иркутский государственный
аграрный университет им. А.А. Ежевского»

СОДЕРЖАНИЕ

ВВЕДЕНИЕ

Настоящие методические указания предназначены для студентов-магистрантов очной и заочной форм обучения направления подготовки 21.04.02 - «Землеустройство и кадастры», применяющих в своей практике методы математической статистики и интересующихся анализом данных.

Математическая статистика - раздел прикладной математики, посвященный математическим методам систематизации, обработки и использования статистических данных для научных и практических выводов.

В соответствии с действующим учебным планом студенты-магистранты изучают дисциплину «прикладная математика» на 1 курсе, выполняют контрольную работу и сдают зачет.

При выполнении контрольной работы студент должен руководствоваться следующими указаниями:

1. Контрольная работа должна выполняться в отдельной тетради или на листах А-4, на титульном листе которой должны быть ясно написаны: фамилия студента, его инициалы, полный шифр (№ зачетной книжки).

2. Перед решением контрольной задачи надо полностью переписать ее условие.

3. Решение задачи следует излагать с подробными вычислениями, сопроводить чертежом, выполненным аккуратно, с указанием осей координат и единиц масштаба.

5. На каждой странице необходимо оставлять поля шириной 3-4 см. для замечаний преподавателя.

6. Контрольная работа должна выполняться самостоятельно.

7. Студент выполняет тот вариант контрольной работы, который соответствует двум последним цифрам его учебного шифра.

I. ОПИСАТЕЛЬНАЯ СТАТИСТИКА

Описательная статистика предназначена для сбора, представления в удобном виде и описания исходных данных.

1.1. Статистические распределения и их графические представления

Пусть X - одномерный количественный признак и в результате его измерения наблюдалось n его значений (вариант) $x_1, ..., x_n$, среди которых могут быть одинаковые. Пусть среди имеющихся n вариант имеется k различных $x_1, ..., x_k, k \le n$, причем x_1 встречается n_1 раз, x_2 – n_2 раз и т.д. x_k - n_k раз. Количество повторений каждого из значений выборки называется *частотами*, [1]. Сумма всех частот должна быть равна объему выборки

$$\sum_{i=1}^{k} n_i = n. \tag{1}$$

Точечным вариационным рядом (*распределением частот* или *частотным распределением*) называется n различных вариант, записанных в возрастающем порядке вместе с соответствующими частотами.

x_i	x_1	x_2	. . .	x_k
n_i	n_1	n_2	. . .	n_k

(2)

Частотное распределение обычно записывается в одном из видов: в таблице с частотами n_i, (2) и через *относительные частоты* ω_i, заданных в виде доли или в виде процента:

$$\omega_i = \frac{n_i}{n}, \omega_i = \frac{n_i}{n} \cdot 100\%. \tag{3}$$

x_i	x_1	x_2	. . .	x_k
ω_i	ω_1	ω_2	. . .	ω_k

(4)

Если объем выборки n велик ($n \ge 50$) и мы имеем дело с одномерной непрерывной величиной (или с одномерной дискретной, число возможных значений достаточно велико, больше 10), то эти варианты группируют. При этом выбирается определенное число интервалов группировки и, получают, таким образом, *интервальное частотное распределение*, т.е. переходят к так называемым «группированным» выборочным данным. Благодаря группировке данные приобретают систематизированный вид. Алгоритм группировки массива данных $x_1, \dots, x_n$ состоит из следующих шагов:

1) находят минимальную, x_{min} и максимальную, x_{max} варианты;

2) весь диапазон значений признака $[x_{min}, x_{max}]$ разбивают на k интервалов (количество интервалов не должно быть меньше 8-10 и больше 20-25). Для примерной ориентации в выборе k можно пользоваться приближенной формулой Стерджесса:

$$k = 1 + \log_2 n, \ k = 1 + 3{,}332\lg n \text{ или } k = \sqrt{n}. \qquad (6)$$

Полученное число нужно округлить до ближайшего целого (число интервалов дробным быть не может);

3) определяют длину интервала по формуле

$$h = \frac{x_{\max} - x_{\min}}{k}; \qquad (7)$$

4) находят граничные точки каждого из интервалов $(a_i; b_i)$

$$a_1 = x_{\min}, a_2 = x_{\min} + h = b_1 \text{ и т.д.}; \qquad (8)$$

5) подсчитывают число вариант n_i, попавших в интервал (a_i, b_i), причем варианты, попавшие на границы интервалов, относят только к одному из интервалов, результат заносят в таблицу, представляющую собой *интервальный вариационный ряд* (*интервальное частотное распределение*)

$(a_i;b_i)$	$(a_1;b_1)$	$(a_2;b_2)$	$\cdots$	$(a_k;b_k)$
n_i	n_1	n_2	...	n_k

(9)

От интервального ряда (9) можно вновь перейти к точечному ряду вида (2), если в качестве значений случайной величины, соответствующего i-ому интервалу, взять его середину

$$x_i^0 = \frac{a_i + b_i}{2}. \qquad (10)$$

Несмотря на видимую несхожесть, ряды (2), (4) и (9) отражают одно и то же фактическое распределение признака.

Табличное распределение частот дополняют его графическим представлением. *Полигоном частот* (*относительных частот*) называют ломаную, отрезки которой соединяют точки $(x_1, n_1), ..., (x_k, n_k)$ или $(x_1, \omega_1), ..., (x_k, \omega_k)$, (рис. 1), [2]. Полигон относительных частот выглядит точно также, как полигон частот, только меняется масштаб на оси ординат.

Кумулятивная кривая или *кумулята* (кривая сумм) - ломаная, составленная по последовательно суммированным, т.е. накопленным частотам или относительным частотам.

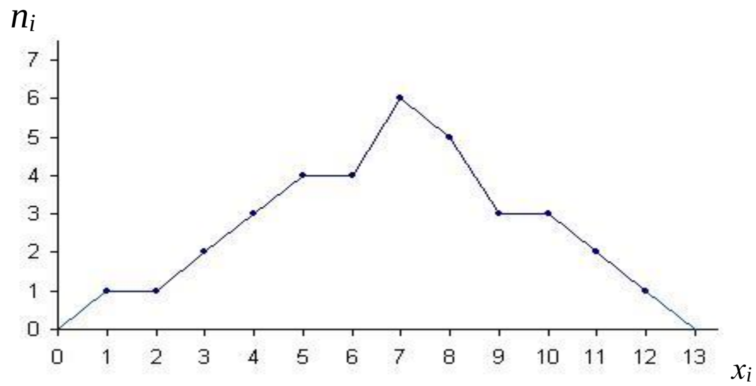


Рисунок 1 - Полигон частот

Накопленная частота n_i^* получается суммированием частот значений, предшествующих данному с частотой n_i, т.е.

$$n_1^* = n_1, n_2^* = n_1 + n_2, n_3^* = n_1 + n_2 + n_3, ..., n_i^* = n_1 + n_2 + ... + n_i. \qquad (11)$$

Отсюда, накопленная частота крайнего правого значения (или максимального элемента выборки) равно объему выборки n. При построении кумулятивной кривой дискретного признака на ось абсцисс наносятся значения признака, а ординатами служат нарастающие итоги частот. Соединением вершин ординат прямыми линиями получают кумуляту, (рис. 2). Кумулятивную кривую называют *полигоном накопленных частот*.

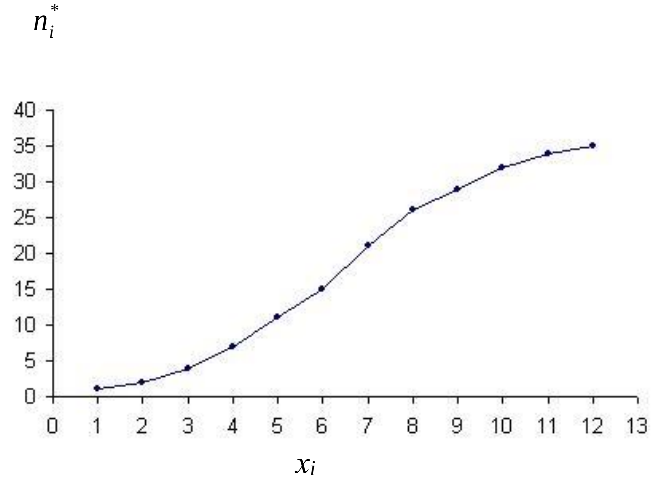


Рисунок 2 - Кумулята

Если на ось ординат нанести значение признака, а накопленные частоты - на ось абсцисс, то получим кривую, называемую *огивой*.

Геометрически интервальный ряд представляется в виде *гистограммы*, которая состоит из последовательности примыкающих друг к другу прямоугольников. Ширина этих прямоугольников равна ширине интервалов группировки h и откладывается по оси абсцисс, высота измеряется по оси ординат и пропорциональна частоте n_i или относительной частоте ω_i. В первом случае имеем *гистограмму частот* с высотами прямоугольников $f = n_i/h$ (рис. 3) и общей площадью, равной объему выборки n. Во втором – *гистограмму относительных частот* с высотами прямоугольников n_i/nh и общей площадью, равной 1.

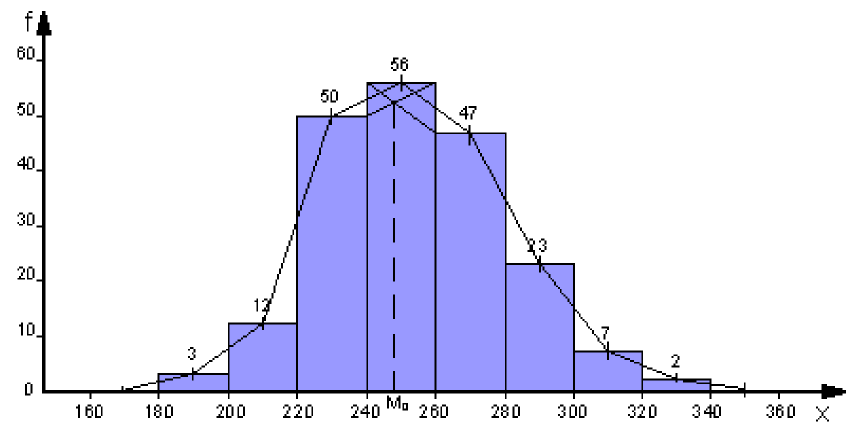


Рисунок 3 – Гистограмма частот

Ступенчатая ломаная, ограничивающая в этом случае сверху постоянную фигуру, является статистической аппроксимацией функции плотности вероятности генеральной совокупности. Если соединить плавной кривой середины верхних оснований прямоугольников, то получим также приближенное представление графика функции плотности распределения исследуемой непрерывной величины. Чем больше опытов проводится, тем ближе построенная гистограмма к теоретической плотности распределения.

1.2. Эмпирическая функция распределения

Статистическая (*эмпирическая*) *функция*, $F^*(x)$ – это функция, которая для каждого значения аргумента равна относительной частоте попадания опытных данных в область, лежащую слева от аргумента, [5]:

$$F^*(x) = \frac{\sum\limits_{x_i < x} n_i}{n} = \frac{n_x}{n} = \sum_{x_i < x} n_i^* \quad . \tag{12}$$

График статистической функции распределения $F^*(x)$ представляет собой ступенчатую фигуру со скачками в точках, определяемых элементами выборки (рис. 4).

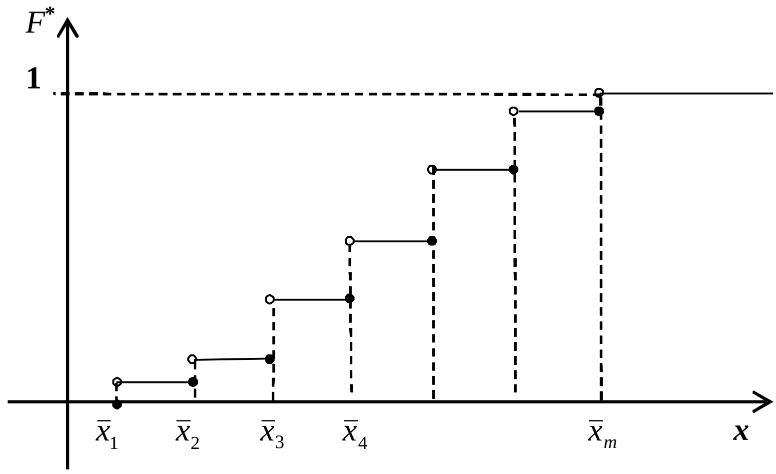


Рисунок 4 – График эмпирической функции распределения

Технология построения статистической функции распределения $F^*(x)$ такая же, как и для теоретической функции распределения в случае дискретных случайных величин: суммируются относительные частоты для всех опытных значений, лежащих слева от аргумента (как раньше суммировались вероятности). На основании закона больших чисел статистическая функция распределения сходится по вероятности к теоретической функции распределения $F(x)$ генеральной совокупности, когда объем выборки неограниченно возрастает.

Эмпирическая функция распределения обладает следующими свойствами:

1) значения функции $F^*(x)$ принадлежат отрезку [0;1];

2) $F^*(x)$ – неубывающая функция;

3) если $x_{\max}, x_{\min}$ - наибольший и наименьший элементы выборки, то справедливо выполнение равенств

$$F^*(x) = \begin{cases} 0, x < x_{\min} \\ 1, x > x_{\max} \end{cases};$$

4) $F^*(x)$ – непрерывна слева.

1.3. Числовые характеристики статистических рядов

1.3.1. Характеристики или меры положения

Для *описания характера расположения распределений* применяют три группы мер: выборочная средняя (арифметическая, геометрическая, степенная и гармоническая), медиана и мода, [5].

Выборочной средней $\overline{x}_в$ называют среднее арифметическое значение признака выборочной совокупности. Формула (13) определяет *невзвешенную* или *простую выборочную среднюю*, если все значения вариант x_i объема *n* различны (для *несгруппированных данных*), а (13') - *взвешенную выборочную среднюю*, если варианты x_1, x_2, …, x_k имеют соответственно частоты n_1, n_2, …, n_k, причем $\sum_{i=1}^{k} n_i = n$ (для *сгруппированных данных*).

$$\bar{x}_в = \frac{\sum_{i=1}^{n} x_i}{n}. \quad (13) \qquad \bar{x}_в = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{\sum_{i=1}^{k} n_i}. \quad (13')$$

Выборочная средняя имеет те же единицы измерения, что и варианты и представляет собой значение, относительно которого может быть «сбалансировано» все эмпирическое распределение, фактически она является абсциссой центра масс гистограммы частот.

Выборочная средняя интервального вариационного ряда вычисляется по формуле (1.13'), где в качестве вариант x_i принимаются середины соответствующих интервалов.

Средняя геометрическая (*невзвешенная* и *взвешенная*) применяется при оценке темпов изменения величин (в частности, при расчете индексов цен) и находится по формулам (14) и (14'):

$$\bar{x}_{геом} = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}. \quad (14) \qquad \bar{x}_{геом} = \sqrt[n]{x_1^{n_1} \cdot x_2^{n_2} \cdot \ldots \cdot x_k^{n_k}}. \quad (14')$$

В практических расчетах удобнее использовать *десятичный* (*натуральный*) *логарифм средней геометрической*

$$lg\bar{x}_{геом} = \frac{1}{n}\sum_{i=1}^{n} lgx_i. \quad (15) \qquad lg\bar{x}_{геом} = \frac{1}{n}\sum_{i=1}^{n} (lgx_i)\cdot n_i. \quad (15')$$

Средняя гармоническая (*невзвешенная и взвешенная*) определяется по формулам:

$$\bar{x}_{гарм.} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{1}{x_i}}. \quad (16) \qquad \bar{x}_{гарм.} = \frac{1}{\frac{1}{n}\sum_{i=1}^{n}\frac{n_i}{x_i}}. \quad (16')$$

Область применения гармонической средней ограничена. В экономике гармонические средние используют при анализе средних норм времени, а также в тех случаях, когда суммируемый признак выражен обратной величиной исследуемого признака.

Между выборочной средней, средней геометрической и средней гармонической существует соотношение

$$\overline{x}_{гарм.} \le \overline{x}_{геом.} \le \overline{x}_{в}. \qquad (17)$$

Средняя степенная невзвешенная и взвешенная) определяется по формулам:

$$\overline{x^P} = \frac{\sum_{i=1}^{n} x_i^p}{n}. \qquad (18) \qquad\qquad \overline{x^P} = \frac{\sum_{i=1}^{k} x_i^p \cdot n_i}{\sum_{i=1}^{k} n_i}. \qquad (18')$$

При $p = 2$ имеем *среднюю квадратичную*, при $p = 3$ – *среднюю кубическую* и т.д.

Наряду с приведенными средними величинами в качестве характеристик положения используют *структурные средние*: моду и медиану. *Мода* (модальное значение $x_{мод} = M_o$) – наиболее часто встречающееся в статистической совокупности значение признака. Для дискретного вариационного ряда мода определяется по частотам вариант и соответствует варианте с максимальной частотой. В случае интервального вариационного ряда с равными интервалами *модальный интервал* (интервал, содержащий моду) определяется по наибольшей частоте, а при неравных интервалах - по наибольшей плотности. Вычисление моды производится по следующей формуле:

$$M_0 = x_0 + h\frac{n_{M_0} - n_{M_0-1}}{(n_{M_0} - n_{M_0-1}) + (n_{M_0} - n_{M_0+1})}, \qquad (19)$$

где x_o- левый конец модального интервала; h- ширина модального интервала; n_{Mo}- частота модального интервала; n_{Mo-1} – частота интервала, предшествующего модальному; n_{Mo+1} – частота интервала, следующего за модальным.

Для *графического определения моды* используют 3 соседних столбца гистограммы (самый высокий и 2 прилегающих к нему). Затем правую вершину модального прямоугольника соединяют с правым верхним углом предыдущего прямоугольника. А левую вершину модального прямоугольника – с левым верхним углом последующего прямоугольника. Далее из точки их пересечения опускают перпендикуляр на ось абсцисс (рис. 5).

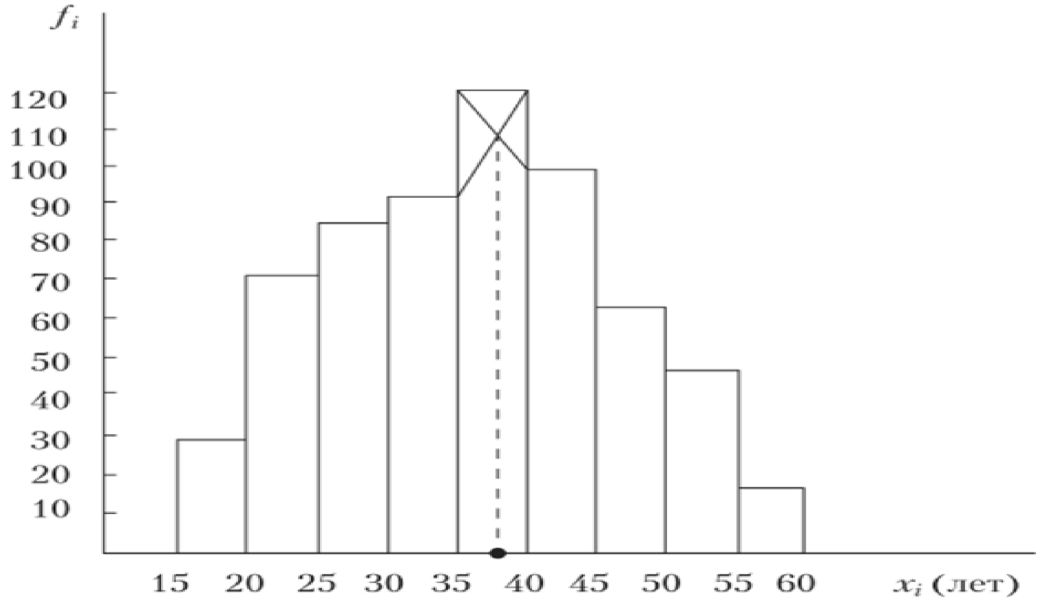


Рисунок 5 – Графическое определение моды

Медиана (медианное значение $x_{мед} = M_e$) - значение изучаемого признака, приходящееся на середину ранжированной совокупности. Для дискретного вариационного ряда медиану определяют по формуле:

$$M_e = \begin{cases} \dfrac{x_l + x_{l+1}}{2}, n = 2l; \\ x_{l+1}, n = 2l + 1. \end{cases} \quad (20)$$

При вычислении медианы интервального вариационного ряда, сначала находят *медианный интервал* $[x_{Me}; x_{Me} + h]$ (интервал, содержащий медиану), где h – длина медианного интервала, путем использования накопленных частот. *Медианному интервалу* соответствует тот, в котором содержится накопленная частота, *превышающая половину объема выборки*. Расчет медианы при постоянной плотности внутри интервала производится по формуле:

$$M_e = x_o + h \cdot \frac{\frac{n}{2} - n^*_{M_e - 1}}{n_{M_e}}, \tag{21}$$

где x_o- левый конец медианного интервала; h- ширина медианного интервала; n^*_{Me-1} – накопленная частота интервала, предшествующего медианному; n_{Me} – частота медианного интервала.

Медиана может быть определена *графически* по *кумуляте*. Для этого последнюю ординату, равную сумме всех частот, т.е. объему выборки n, делят пополам. Из полученной точки восстанавливают перпендикуляр до пересечения с кумулятой. Абсцисса точки пересечения и дает значение медианы (рис.6).

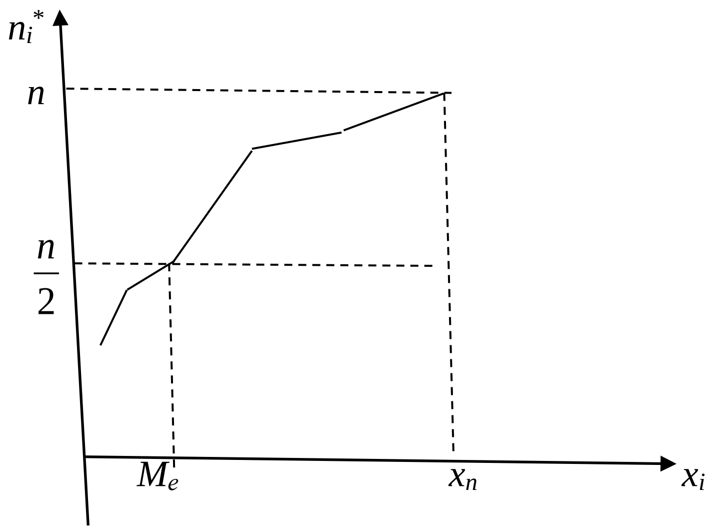


Рисунок 6 – Графическое определение медианы

1.3.2. Характеристики или меры рассеяния

Для измерения вариации применяются следующие характеристики: вариационный размах, среднее линейное отклонение, выборочная дисперсия, выборочное среднее квадратическое отклонение, коэффициент вариации и др., [6].

Вариационный размах (*размах вариации*) R представляет собой разность между наибольшим и наименьшим значением наблюдений:

$$R = x_{max} - x_{min}. \tag{22}$$

Вариационный размах вариации применяется в качестве приблизительной оценки вариации значений признака, широко используется в ряде отраслей промышленности при статистическом изучении качества продукции.

Среднее линейное отклонение (*невзвешенное* и *взвешенное*) определяется по формулам:

$$d = \frac{\sum_{i=1}^{n}\left|x_i - \overline{x_в}\right|}{n}. \qquad (23) \qquad d = \frac{\sum_{i=1}^{k}\left|x_i - \overline{x_в}\right| \cdot n_i}{\sum_{i=1}^{k} n_i}. \qquad (23')$$

Выборочной дисперсией $D_в$ называют среднее арифметическое квадратов отклонения наблюдаемых значений признака от их среднего значения $\overline{x}_в$. Формулы (24) и (24') служат для определения выборочной дисперсии для несгруппированных и сгруппированных данных:

$$D_в = \frac{\sum_{i=1}^{n}\left(x_i - \overline{x_в}\right)^2}{n}. \qquad (24) \qquad D_в = \frac{\sum_{i=1}^{k}\left(x_i - \overline{x_в}\right)^2 \cdot n_i}{\sum_{i=1}^{k} n_i} \qquad (24')$$

На практике для вычисления $D_в$ удобнее использовать формулу:

$$D_в = \overline{x_в^2} - (\overline{x}_в)^2. \qquad (24'')$$

Выборочное среднее квадратическое отклонение $\sigma_в$ представляет собой квадратный корень из выборочной дисперсии:

$$\sigma_в = \sqrt{D_в}\,. \qquad (25)$$

Выборочное среднее квадратическое отклонение выражается в тех же единицах, что и значения исследуемого признака.

Кроме абсолютных величин для характеристики вариации статистической совокупности значений признаков применяются относительные показатели: линейный коэффициент вариации, коэффициент вариации, эмпирический коэффициент детерминации, эмпирическое корреляционное отношение.

Линейным коэффициентом вариации V_d выраженное в процентах отношение среднего линейного отклонения d к к выборочной средней $\overline{x}_в$:

$$V_d = \frac{d}{\overline{x}_в} \cdot 100\% \quad . \qquad (27)$$

Коэффициентом вариации V называют выраженное в процентах отношение выборочного среднего квадратического отклонения $\sigma_в$ к выборочной средней $\overline{x}_в$:

$$V = \frac{\sigma_в}{\overline{x}_в} \cdot 100\% . \qquad (28)$$

Коэффициенты вариации используются для сравнения размеров вариации в вариационных рядах с различными средними, а также для сравнения вариаций разных показателей в одной и той же совокупности. Кроме того, показатели вариации дают характеристику однородности статистических данных. *Статистическая совокупность считается однородной*, если коэффициент вариации не превышает 33%.

Каждому закону распределения соответствует определенный приближенный диапазон значений коэффициента вариации, поэтому он также может служить в качестве критерия при выборе закона распределения. Для нормального закона распределения коэффициент вариации находится в пределах 8%-40%.

1.3.3. Характеристики формы эмпирического распределения

При анализе вариационных рядов смещение от центра и крутизну распределения характеризуют специальные показатели, называемые *характеристиками формы*. Эмпирические распределения, как правило, смещены от центра распределения вправо или влево, асимметричны. Нормальное распределение строго симметрично относительно средней арифметической, что обусловлено четностью функции плотности вероятности. *Асимметрия распределения* возникает вследствие того, что какие-либо факторы действуют в одном направлении сильнее, чем в другом, или процесс развития явления таков, что доминирует какая-то причина. Кроме того, природа некоторых явлений такова, что имеет место асимметричное распределение.

Моментом порядка p распределения вариационного ряда называется величина, определяемая по формуле:

$$\mu_p = \sum_{i=1}^{k} (x_i - a)^p \cdot n_i. \tag{31}$$

В зависимости от значения *a* общая система моментов разбивается на 3 подсистемы. Если $a = 0$, получаем *систему начальных моментов*:

$$m_p = \frac{1}{n} \sum_{i=1}^{k} x_i{}^p \cdot n_i \ . \tag{32}$$

При $a = \overline{x_в}$ получаем *систему центральных моментов*:

$$\mu_p = \frac{1}{n} \sum_{i=1}^{k} (x_i - \overline{x}_в)^p \cdot n_i. \tag{33}$$

Если $a = C = const$ (обычно C близко к середине вариационного ряда, причем, если k – четное, то $C = x_{k\backslash 2}$, если k – нечетное, то $C = x_{(k+1)\backslash 2}$), получаем *систему условных моментов*:

$$M_p^* = \frac{1}{n} \sum_{i=1}^{k} u_i{}^p \cdot n_i = \frac{1}{n} \sum_{i=1}^{k} (x_i - C)^p \cdot n_i, \tag{34}$$

вычисленных для *условных вариант* u_i. Эта система применяется для упрощения расчетов.

Между центральными и начальными моментами имеют место следующие соотношения:

$$\mu_1 = 0\,; \ \mu_2 = m_2 - \left(m_1\right)^2\,; \ \mu_3 = m_3 - 3m_1 \cdot m_2 + 2(m_1)^3\,; \tag{35}$$

$$\mu_4 = m_4 - 4m_1 m_3 + 6m_2 (m_1)^2 - 3(m_1)^4.$$

Имеют место удобные для вычислений формулы, выражающие центральные моменты через условные:

$$\mu_2 = \left[M_2^* - (M_1^*)^2 \right] \cdot h^2 ; \quad \mu_3 = \left[M_3^* - 3M_1^* M_2^* + 2(M_1^*)^3 \right] \cdot h^3 ;$$

$$\mu_4 = \left[M_4^* - 4M_1^* M_3^* + 6M_2^*(M_1^*)^2 - 3(M_1^*)^4 \right] \cdot h^4 . \qquad (36)$$

Центральные моменты 3 и 4 порядков используются для определения коэффициентов асимметрии и эксцесса эмпирических распределений,[8].

Коэффициент асимметрии определяет направление и величину смещения (асимметрию) распределения и рассчитывается по формуле:

$$A_s = \frac{\mu_3}{\sigma_в^3} = \frac{\sum\limits_{i=1}^{k} \left(x_i - \overline{x_в} \right)^3 \cdot n_i}{n \cdot \sigma_в^3} . \qquad (37)$$

Если вершина распределения сдвинута влево и правая часть ветви оказывается длиннее левой, то такая асимметрия является *правосторонней*, в противоположном случае *левосторонней*. При *левосторонней* (*отрицательной*) *асимметрии* коэффициент асимметрии A_s<0, при *правосторонней* (*положительной*) A_s>0. Если $A_s = 0$, то распределение имеет симметричную форму, т.е. варианты, равноудаленные от выборочной средней, имеют одинаковую частоту.

Эксцесс (или *коэффициент крутости*) *эмпирического распределения* находится по формуле:

$$E_k = \frac{\mu_4}{\sigma_в^4} - 3 = \frac{\sum\limits_{i=1}^{k} \left(x_i - \overline{x_в} \right)^4 \cdot n_i}{n \cdot \sigma_в^4} - 3. \qquad (38)$$

Эксцесс у нормальной кривой равен нулю. Кривые, у которых $E_k < 0$, по сравнению с нормальной, менее крутые и имеют более *плоскую вершину*. Кривые, у которых $E_k > 0$ более крутые, чем нормальная кривая и имеют более *острую вершину*.

1.4. Метод произведений

Для упрощения вычислений средних характеристик и показателей рассеяния, а также для уменьшения погрешности округления используют *метод произведений*, [10]. Если выборка задана в виде распределения *равноотстоящих* вариант и соответствующих им частот, то удобно находить выборочную среднюю и дисперсию по формулам:

$$\overline{x}_в = M_1^* \cdot h + C. \qquad (39) \qquad D_в = \left[M_2^* - \left(M_1^* \right)^2 \right] \cdot h^2, \qquad (40)$$

где h – длина интервала группировки, C – ложный ноль (варианта, у которой самая большая частота), M_1^* и M_2^* – условные моменты первого и второго порядка, определяемые по формуле (34).

1.5. Точечное оценивание параметров выборки

Пусть мы располагаем исходными статистическими данными - выборкой (x_1, x_2, … , x_n) объема n из исследуемой генеральной совокупности. Пусть изучаемая случайная величина распределена по закону $P(x;\theta)$, где θ - параметр распределения, значение которого неизвестно до получения выборки. *Задача оценивания* неизвестного параметра θ состоит в построении такой функции $\theta^* = \theta^*(x_1, x_2, \ldots, x_n)$ от имеющихся в нашем распоряжении данных, которая давала бы в определенном смысле наиболее точное приближенное значение истинного (неизвестного нам) параметра θ. Любую функцию выборки называют *статистикой.* Статистика $\theta^* = \theta^*(x_1, x_2, \ldots, x_n)$, используемая для оценки неизвестного параметра θ, называется *статистической оценкой параметра* θ. Оценка, полученная в виде одного числа - точки на числовой прямой, называется *точечной*. Оценка, полученная в виде интервала, называется *интервальной*.

Если по выборке (x_1, x_2, … , x_n) требуется оценить математическое ожидание $M(X) = a$ и дисперсию $D(X) = \sigma^2$ случайной величины X, то в качестве оценки математического ожидания берется выборочная средняя арифметическая, определяемая по формулам (13) или (13'), т.е.

$$a^* = \overline{x}_в = \frac{\sum_{i=1}^{k} x_i}{\sum_{i=1}^{k} n_i}, \quad a^* = \overline{x}_в = \frac{\sum_{i=1}^{k} x_i \cdot n_i}{\sum_{i=1}^{k} n_i},$$

а в качестве оценки дисперсии – используется называемая *исправленная выборочная дисперсия*:

$$S^2 = \frac{n}{n-1}\sigma^2 = \frac{n}{n-1} D_в, \tag{41}$$

Соответственно, оценкой среднего квадратического отклонения служит величина:

$$S_в = \sqrt{S_в^2}. \tag{42}$$

1.6. Интервальные оценки

В практических задачах наряду с точечными оценками необходимо указывать интервалы, которые бы с практической достоверностью, т.е. с вероятностью, близкой к единице, покрывали бы истинное неизвестное значение параметра. Такие интервалы называются *доверительными*, или *интервальными оценками*, а вероятность, с которой доверительный интервал покрывает неизвестное значение параметра, называется *доверительной вероятностью* или *надежностью*, [9]. Доверительные интервалы используются для определения точности оценки θ^* неизвестного параметра θ, а доверительные вероятности – для определения надежности.

Пусть по результатам выборки найдена точечная оценка θ^* неизвестного параметра θ. Очевидно, что чем меньше величина $\left|\theta-\theta^*\right|$, тем точнее оценка. Число δ, для которого $\left|\theta-\theta^*\right|<\delta$, характеризует *точность оценки*. Очевидно, *чем больше* n *- объем выборки, тем точнее будет оценка.* Однако, нельзя точно определить, при каком значении n будет достигнута заданная точность. Можно говорить только о вероятности, с которой данное неравенство будет выполняться.

Вероятность

$$\mathrm{P}\left(\left|\theta-\theta^*\right|<\delta\right)=\gamma \tag{43}$$

называется *доверительной вероятностью* или *надежностью оценки*.

Используя определение модуля, формула (43) может быть записана:

$$\mathrm{P}(\theta^* - \delta < \theta < \theta^* + \delta) = \gamma .$$

Полученная формула читается: вероятность того, что интервал $(\theta^* - \delta; \theta^* + \delta)$, называемый *доверительным интервалом*, заключает в себе (покрывает) неизвестный параметр θ, равна γ. Обычно надежность γ, задается наперед числом, близким к единице (например $\gamma = 0{,}9;\ 0{,}95;\ 0{,}99;\ 0{,}999$). Событие с вероятностью γ считается практически достоверным. *Чем меньше длина доверительного интервала, тем точнее оценка.*

Доверительный интервал для оценки математического ожидания нормального распределения при неизвестной дисперсии определяется по формуле:

$$\bar{x}_в - t_\gamma \frac{S_в}{\sqrt{n}} < a < \bar{x}_в + t_\gamma \frac{S_в}{\sqrt{n}} . \tag{44}$$

Коэффициент t_γ в формуле (44) находится из таблицы 2 распределения Стьюдента по заданной доверительной вероятности и числу степеней свободы k (k = n – 1), зависящему от объема выборки. Величина

$$\delta = t_{\gamma} \frac{S_{в}}{\sqrt{n}} \tag{45}$$

также определяет точность оценки, иногда ее называют *предельной погрешностью оценки*. Из формул (44) и (45) видно, что *увеличение объема выборки* n *приводит к уменьшению длины доверительного интервала*. Очевидно, что *увеличение надежности интервальной оценки влечет за собой увеличение ее точности*. Если задать точность δ и доверительную вероятность γ, то из соотношения (45), можно найти минимальный объем выборки n, который обеспечит заданную точность:

$$n = \frac{t_{\gamma}^{2} \cdot S_{в}^{2}}{\delta^{2}}. \tag{46}$$

Доверительный интервал для оценки среднего квадратического отклонения σ нормально распределенной случайной величины X с надежностью γ находится из двойных неравенств:

$$S_{в}(1-q) < \sigma < S_{в}(1+q) \text{ при } q < 1, \tag{47}$$

$$0 < \sigma < S_{в}(1+q) \text{ при } q > 1,$$

где q находится из таблицы 5 по заданной надежности γ и объему выборки n.

Для оценки дисперсии D нормально распределенной случайной величины X достаточно возвести в квадрат неравенства (47).

1.7. Проверка гипотезы о виде закона распределения. Критерий Пирсона

Напомним, что *статистической гипотезой* называется любое предположение относительно вида или параметров генерального распределения. Гипотеза называется *простой*, если она однозначно определяет распределение случайной величины, в противном случае гипотеза *сложная*. Статистическая гипотеза называется *параметрической*, если она содержит утверждение о значении конечного числа параметров распределения, которое считается неизвестным. *Непараметрическая* гипотеза - это утверждение о виде распределения. Например, простой параметрической гипотезой является предположение о том, что наблюдаемая случайная величина имеет нормальное распределение с параметрами 0 и 1. Та гипотеза, относительно которой ведется проверка, называется *основной* или *нулевой* и обозначается через H_o. Наряду с гипотезой H_o рассматривается *конкурирующая* или *альтернативная* гипотеза H_1, которая должна быть принята в случае отклонения H_o , т.е. H_o и H_1 - две взаимно исключающие гипотезы. В качестве *базисного предположения* принимается утверждение о справедливости одной их этих гипотез.

Критерием проверки гипотез или *статистическим критерием* называется правило, по которому решают, принять или отклонить нулевую гипотезу H_o (соответственно, отклонить или принять альтернативную гипотезу H_1). Решение здесь выносится в зависимости от значения специальным образом случайной величины, называемой *статистикой критерия* или просто *критерием*, распределение которой известно и табулировано.

Критерием согласия называют критерий проверки гипотезы о предполагаемом законе неизвестного распределения, [5]. *Критерием Пирсона* χ^2 («*хи-квадрат*») называют критерий согласия, статистикой которого является случайная величина

$$\chi^2_{\text{расч.}} = \sum_{i=1}^{k} \frac{(n_i - np_i)^2}{np_i} = \sum_{i=1}^{k} \frac{(n_i - m_i)^2}{m_i}, \tag{48}$$

где k – число интервалов разбиения выборки (число степеней свободы), n_i – частота i – го интервала; m_i - теоретические частоты ($m_i = n_i p_i$), n – объем выборки, p_i – теоретическая вероятность попадания значения случайной величины X в i–й интервал.

Критерий согласия Пирсона служит для проверки гипотезы H_o: закон распределения выборки согласуется с гипотезой о нормальном распределении генеральной совокупности с параметрами $M(X) = a$ и $D(X) = \sigma^2$ (критерий аналогично применяется и для других распределений, в этом состоит его преимущество перед другими критериями).

Пусть эмпирическое распределение (выборка объема n) задано в виде равноотстоящих интервалов и соответствующих им частот. Для того, чтобы проверить согласуются ли данные выборки с нулевой гипотезой – гипотезой о нормальном распределении генеральной совокупности, поступают по следующему правилу:

1) выдвигают нулевую гипотезу о нормальном законе распределения случайной величины X и находят оценки его параметров $x_в$ и S;

2) определяют теоретические частоты m_i, соответствующие опытным частотам, если среди опытных частот имеются малочисленные, то их необходимо объединить с соседними:

$$m_i = n \cdot p_i = n \cdot \left[\Phi\left(\frac{x_i - \overline{x_в}}{S_в} \right) - \Phi\left(\frac{x_{i-1} - \overline{x_в}}{S_в} \right) \right], \tag{49}$$

где $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$ - - стандартная функция Лапласа, значения которой находят по статистической таблице 1;

3) по формуле (48) вычисляют статистику критерия;

4) определяют число степеней свободы:

$$s = k - r - 1, \tag{50}$$

где k – число интервалов разбиения выборки; r – число оцениваемых независимых параметров выбранного закона распределения (для нормального закона r = 2, параметры a и σ);

5) задают уровень значимости α и критическую точку критерия определяют по статистической таблице 3 при заданных α и q;

6) если $$\chi^2_{\text{расч.}} < \chi^2_{\text{кр.}}, \tag{51}$$

то нулевую гипотезу принимают, в противном случае нулевую гипотезу отвергают.

II. ЭЛЕМЕНТЫ КОРРЕЛЯЦИОННО-РЕГРЕСИОННОГО АНАЛИЗА

Для целей анализа и планирования хозяйственно-экономической деятельности предприятия широко применяется корреляционно-регрессионный анализ. Корреляционно-регрессионный анализ — классический метод стохастического моделирования хозяйственной деятельности. Он изучает взаимосвязи показателей хозяйственной деятельности, когда зависимость между ними не является строго функциональной и искажена влиянием посторонних, случайных факторов, [3]. При проведении корреляционно-регрессионного анализа строят различные корреляционные и регрессионные модели хозяйственной деятельности. В этих моделях выделяют факторные и результативные показатели (признаки).

Корреляционный анализ ставит задачу измерить тесноту и направление связи между варьирующими переменными и оценить факторы, оказывающие наибольшее влияние на результативный признак.

Регрессионный анализ предназначен для выбора формы связи и типа модели для определения расчетных значений зависимой переменной (результативного признака) .

Методы корреляционного и регрессионного анализа используются в комплексе. Наиболее разработанной в теории и широко применяемой на практике является парная корреляция, когда исследуются соотношения результативного признака и одного факторного признака. Это — однофакторный корреляционный и регрессионный анализ.

2.1. Линейный коэффициент корреляции, оценка его статистической значимости

Для оценки тесноты линейной связи между переменными используют *линейный коэффициент парной корреляции*, [4]:

$$r_{x/y} = a\frac{\sigma_x}{\sigma_y} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y}, \tag{52}$$

где a – коэффициент уравнения линейной парной регрессии при соответствующем факторном признаке ($\hat{y} = ax + b$);

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ – среднее значение фактора *X*;

$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$ – среднее значение результативной переменной *Y*;

$\overline{x \cdot y} = \frac{1}{n}\sum_{i=1}^{n} x_i \cdot y_i$ – среднее значение произведения переменных *X* и *Y*;

$\sigma_x = \sqrt{D_x} = \sqrt{\overline{x^2} - \bar{x}^2}$ – среднеквадратическое отклонение (СКО) переменной *X*;

$\sigma_y = \sqrt{D_y} = \sqrt{\overline{y^2} - \bar{y}^2}$ – среднеквадратическое отклонение (СКО) переменной *Y*;

$\overline{x^2} = \frac{1}{n}\sum_{i=1}^{n} x_i^2$ – среднее значение квадрата переменной *X*;

$\overline{y^2} = \frac{1}{n}\sum_{i=1}^{n} y_i^2$ – среднее значение квадрата результативной переменной *У*.

Можно считать, что:

1) если $r_{x.y} > 0$, то имеется *прямая* линейная связь между переменными *X* и *Y*;

2) если $r_{x.y} < 0$, то имеется *обратная* линейная связь между переменными *X* и *Y*;

3) если $r_{x.y} \approx 0$ ($|r_{x.y}| < 0{,}1$), то линейная связь между переменными *X* и *Y* *отсутствует*.

Качественная оценка тесноты связи величин *X* и *Y* может быть выявлена на основе шкалы Чеддока:

Тестона связи	*Значение коэффициента корреляции*
Слабая	0,1-0,3
Умеренная	0,3-0,5
Заметная	0,5-0,7
Высокая	0,7-0,9
Весьма высокая	0,9-0,99

Графически взаимосвязь двух признаков изображается с помощью *поля корреляции*. В декартовой прямоугольной системе координат (ДПСК) на оси абсцисс откладываются значения факторного признака, а на оси ординат – результативного. Каждая единица статистической совокупности с соответствующими значениями x и y обозначается точкой. При отсутствии тесных связей имеет место беспорядоченное расположение точек на графике. Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определенной линии, выражающей форму связи (рис. 1).

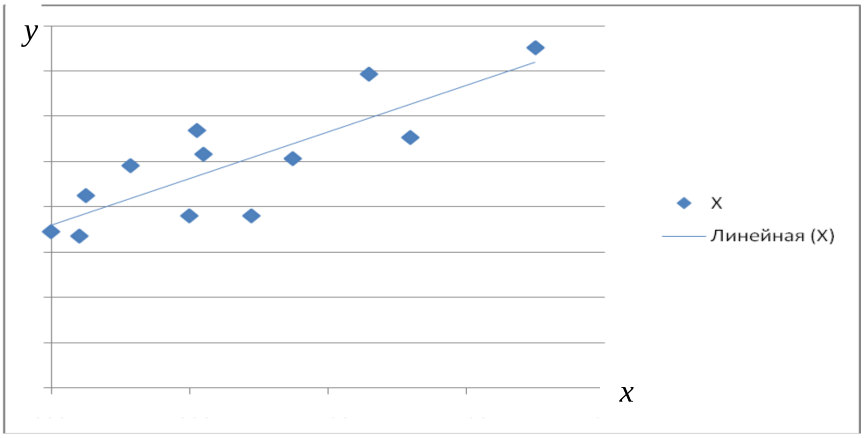


Рисунок 7 – Поле корреляции

Для оценки статистической значимости парного *коэффициента корреляции* при условии линейной формы связи между факторами, можно использовать *t-критерий Стьюдента*, [ ]:

$$t_r = \frac{\left| r_{x/y} \right| \cdot \sqrt{n-2}}{\sqrt{1-r_{x/y}^2}}. \quad (53)$$

Вычисленное значение t_r по формуле (53) сравнивают с критическими ***t***, которое определяют по таблице значений *Стьюдента* при заданном уровне значимости α (0,1; 0,05; 0,02; 0,01; 0,002; и 0,001) и числе степеней свободы $k = n - 2$ (таблица 2). Коэффициент корреляции *признается значимым* (существенным) при условии, если $t_r > t_{табл.}$ В этом случае, практически невероятно, что найденные значения параметров обусловлены только случайными факторами.

2.2. Составление уравнения линейной парной регрессии

Уравнение линейной парной регрессии имеет вид:

$$\hat{y} = ax + b\,. \tag{54}$$

Для определения параметров a, b используется *метод наименьших квадратов* (МНК), в основе которого лежит минимизация суммы квадратов отклонений эмпирических (фактических) значений результативного признака от теоретических, полученных по выбранному уравнению регрессии:

$$S = \sum_{i=1}^{n} \left(y_i - \hat{y}\right)^2 \to min\,.$$

Система нормальных уравнений имеет следующий вид:

$$\begin{cases} nb + a\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \\ b\sum_{i=1}^{n} x_i + a\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i \cdot .y_i \end{cases} \tag{55}$$

Можно воспользоваться готовыми формулами решения системы (55):

$$a = \frac{\operatorname{cov}(x;y)}{\sigma_x^2} = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2}\,,\ b = \bar{y} - a \cdot \bar{x}\,, \tag{56}$$

где $\operatorname{cov}(x;y) = \overline{x \cdot y} - \bar{x} \cdot \bar{y}$ – ковариация переменных X и Y.

Коэффициент регрессии a показывает, на сколько единиц в среднем по совокупности изменится результирующая переменная Y, если факторная переменная X увеличится на одну единицу собственного измерения. Если a >0, то с увеличением факторного признака результативный признак возрастает. Если a<0, то с увеличением факторного признака результативный признак уменьшается.

Коэффициент регрессии b показывает усредненное влияние на результативный признак неучтенных в уравнении факторных признаков.

2.3. Оценка значимости параметров уравнения линейной парной регрессии

При численности объектов анализа до 30 единиц возникает необходимость проверки значимости каждого коэффициента линейной парной регрессии. Для оценки статистической значимости коэффициентов линейной парной регрессии по *t-критерию Стьюдента* выдвигается гипотеза H_0 о случайной природе показателей, т.е. о незначимом их отличии от нуля. Наблюдаемые значения t-критерия рассчитываются по формулам:

$$t_a = \frac{a}{m_a},\ t_b = \frac{b}{m_b}. \qquad (57)$$

где m_a, m_b – случайные ошибки параметров линейной парной регрессии, вычисляемые по формулам:

$$m_a = \frac{S_{ост}}{\sigma_x \sqrt{n-2}},\ m_b = \frac{S_{ост}}{\sqrt{n-2}}, \qquad (58)$$

где $S^2_{ост}$ – остаточная дисперсия на одну степень свободы:

$$S_{ост} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-2}}, \qquad (59)$$

где $\hat{y} = ax + b$ рассчитывается по найденному уравнению линейной парной регрессии для каждого значения x_i. факторного признака X.

Табличное (критическое) значение t-статистики находят по таблице распределения t-Стьюдента (таблица 2). Если $t_{табл} < |t_{расч}|$, то H_0 отклоняется, т.е. коэффициенты регрессии не случайно отличаются от нуля и сформировались под влиянием систематически действующего фактора и статистически значимы.

2.4. Коэффициенты детерминации и эластичности

Для оценки качества уравнения регрессии использую *коэффициент детерминации* R^2, [11]. Коэффициентом детерминации называют отношение *объясненной* части дисперсии результативной переменной (*У)* к *общей дисперсии*:

$$R^2 = \frac{S_{объяс.}}{S_{общая}} = \frac{\sum_{i=1}^{n}(\hat{y} - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \tag{60}$$

Соотношение между объясненной и необъясненной частями общей дисперсии можно представить в альтернативном варианте:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}. \tag{60'}$$

Упрощенный вариант вычисления коэффициента детерминации как квадрата коэффициента корреляции представляет собой:

$$R^2 = r^2_{x/y}. \tag{60''}$$

Коэффициент детерминации показывает, какую часть вариации (изменения) результативной переменной Y объясняет вариация (изменение) фактора X. Коэффициент детерминации принимает значения в диапазоне от нуля до единицы. Значение R^2 является индикатором степени подгонки модели к данным, [7]. Чем ближе R^2 к единице, тем лучше линейное парное регрессионное уравнение и модель объясняет почти всю изменчивость соответствующих переменных. При отсутствии зависимости между результативным и факторным признаками коэффициент детерминации будет близок к нулю. Таким образом, коэффициент детерминации R^2 может применяться для *оценки качества* (*точности*) уравнения регрессии

Коэффициент эластичности дает сравнительную оценку силы связи факторного признака с результативным, который определяется по формуле:

$$Э_{y.x} = a\frac{\bar{x}}{\bar{y}}\ , \tag{61}$$

где $\bar{x}, \bar{y}$ – средние значения факторного X и результативного $У$ признаков соответственно; a – коэффициент уравнения линейной парной регрессии при соответствующем факторном признаке ($\hat{y} = ax + b$).

Коэффициент эластичности показывает на сколько процентов в среднем изменится значение результативного признака при изменении факторного признака на 1%.

2.5. Средняя ошибка аппроксимации

Чтобы иметь *общее суждение* о качестве модели**,** по каждому наблюдению из относительных отклонений определяют *среднюю ошибку аппроксимации*. Средняя ошибка аппроксимации служит для измерения точности построенной модели и определяется по формуле:

$$\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}}{y_i}\right| \cdot 100\% \tag{62}$$

Если $\bar{\varepsilon} \leq$ 15%, то построенное уравнение регрессии достаточно точно выражает закон изменения исследуемого показателя под действием факторного и приемлемо для целей анализа, [8]. Если $\bar{\varepsilon} \leq$ 4%, то данную линейную модель можно использовать для прогнозирования. В остальных случаях ($\bar{\varepsilon}$ >15%) полученное линейное уравнение парной регрессии не может быть использовано для дальнейших исследований.

2.6. *F*-критерий Фишера для анализа адекватности уравнения регрессии в целом

Проверить на *адекватность* уравнения регрессии – значит установить, соответствует ли математическая модель, выражающая зависимость между переменными, экспериментальным данным, достаточно ли включенных в уравнение объясняющих переменных для описания зависимой переменной. Анализ адекватности уравнения регрессии в целом осуществляется с помощью *F-критерия Фишера*, [9]. Проверяется гипотеза H_0 о статистической незначимости уравнения регрессии. Для этого рассчитывается фактическое значение критерия по формуле:

$$F_{факт} = \frac{\sum_{i=1}^{n} (\hat{y} - \bar{y})^2}{\sum_{i=1}^{n} (y_i - \hat{y})^2} \cdot (n-2), \qquad (63)$$

где n – число единиц совокупности; y_i – фактические значения результативного признака *У*; $\bar{y}$ – среднее значение результативного признака *У*; $\hat{y}$ - значения результативного признака, рассчитываемые по уравнению линейной модели регрессии. Величина *F*-критерия связана с коэффициентом детерминации R^2 ($r^2_{x/y}$) и ее можно рассчитать по следующей формуле:

$$F_{факт} = \frac{R^2}{1-R^2} \cdot (n-2). \qquad (63')$$

Фактическое значение критерия Фишера сравнивается с табличным– это максимально возможное значение критерия, которое могло сформироваться под влиянием случайных факторов при данных степенях свободы и уровне значимости α (α.=0,01; 0,05). Имеются таблицы критических (табличных) значений *F*-критерия: $F(\alpha;\ k_1;\ k_2)$, где $k_1 = 1$, $k_2 = n-2$ (Таблица 4). Если $F_{табл} < F_{факт}$, то гипотеза H_0 о случайной природе оцениваемых характеристик отклоняется и признается статистическая значимость уравнения регрессии в целом, [12].

При анализе адекватности уравнения регрессии (модели) исследуемому процессу, возможны следующие варианты:

1. Построенная модель на основе F-критерия Фишера в *целом адекватна* и все *коэффициенты* регрессии *значимы.* Такая модель может быть использована для принятия решений и осуществления прогнозов,[13].

2. Модель по F-критерию Фишера *адекватна*, но *часть коэффициентов не значима*. Модель пригодна для принятия некоторых решений, но не для прогнозов.

3. Модель по F-критерию адекватна, но *все коэффициенты регрессии не значимы. М*одель полностью считается *неадекватной,* [15]. На ее основе не принимаются решения и не осуществляются прогнозы.

КОНТРОЛЬНАЯ РАБОТА

Студент выполняет тот вариант контрольной работы, который совпадает с последней цифрой его учебного шифра. При этом, если предпоследняя цифра учебного шифра (№ зачетной книжки) есть число нечетное (1, 3, 5, 7, 9), то номера задач для соответствующего варианта даны в столбце 1, если же предпоследняя цифра учебного шифра есть число четное (2, 4, 6, 8, 0), то номера задач для соответствующего варианта даны в столбце 2.

№ варианта (последняя цифра учебного шифра)	Предпоследняя цифра учебного шифра есть число нечетное	Предпоследняя цифра учебного шифра есть число четное
	1	2
1	1, 21, 41	11, 31, 51
2	2,22, 42	12, 32, 52
3	3, 23, 43	13, 33, 53
4	4, 24, 44	14, 34, 54
5	5, 25, 45	15, 35, 55
6	6, 26,46	16, 36,56
7	7, 27, 47	17, 37, 57
8	8, 28, 48	18, 38, 58
9	9, 29, 49	19, 39, 59
0	10, 30, 50	20, 40,60

ЗАДАНИЕ № 1 (задачи 1 – 20)

Приведена выборка 108 результатов измерений случайной величины X

1. Построить интервальный вариационный ряд (ряд 1) по частотам, относительным частотам и накопленным частотам.
2. От ряда 1 перейти к точечному вариационному ряду (ряд 2).
3. Начертить полигоны частот и относительных частот, кумуляту (по ряду 2) и гистограммы частот и относительных частот (по ряду 1).
4. Записать аналитически и построить графически статистическую функцию распределения.
5. Найти выборочные средние: среднюю арифметическую, среднюю геометрическую, среднюю гармоническую; выборочную дисперсию, выборочное среднеквадратическое отклонение, коэффициенты вариации и асимметрии, эксцесс (по ряду 2). Применить метод произведений.
6. Определить моду и медиану графически и аналитически (по рядам 1 и 2).
7. В предположении, что выборка извлечена из генеральной совокупности, подчиненной нормальному закону, при заданной надежности γ = 0,95 построить доверительные интервалы для неизвестного математического ожидания a, неизвестной дисперсии D и среднего квадратического отклонения σ случайной величины X.

8. На основе анализа гистограммы и статистической функции распределения оценить близость эмпирического распределения к нормальному закону, проверить гипотезу о том, что случайная величина X согласуется с нормальным законом распределения по критерию Пирсона $\left(\chi^2\right)$.

№ задания	Значения случайной величины (признака) X										
1	1,55	1,79	1,64	1,72	1,76	1,82	1,90	1,56	1,60	1,62	1,84	1,68
	1,75	1,78	1,80	1,68	1,78	1,92	1,59	1,75	1,77	1,80	1,78	1,65
	1,66	1,56	1,77	1,74	1,66	1,71	1,67	1,70	1,54	1,69	1,82	1,64
	1,81	1,73	1,54	1,63	1,75	1,68	1,56	1,88	1,84	1,77	1,60	1,72
	1,63	1,69	1,66	1,75	1,69	1,54	1,56	1,79	1,52	1,66	1,70	1,81
	1,69	1,72	1,68	1,56	1,58	1,65	1,73	1,55	1,64	1,56	1,66	1,57
	1,56	1,80	1,74	1,81	1,64	1,67	1,62	1,60	1,55	1,72	1,66	1,80
	1,64	1,68	1,70	1,62	1,59	1,73	1,70	1,65	1,63	1,65	1,58	1,71
	1,71	1,59	1,67	1,58	1,71	1,66	1,59	1,62	1,73	1,69	1,68	1,52
2	11,2	21,0	13,3	21,5	20,6	8,0	19,7	13,4	14,5	18,3	25,1	5,3
	14,2	12,0	17,7	15,9	11,1	18,5	20,0	19,1	9,6	20,5	13,8	21,3
	20,9	7,7	20,1	13,1	14,8	18,0	26,1	5,9	14,6	11,7	18,1	15,6
	11,6	20,2	18,8	8,2	12,0	13,5	22,0	14,4	15,2	15,1	11,0	11,8
	18,1	14,0	12,5	20,8	13,4	21,2	20,9	8,5	17,7	14,2	21,8	7,9
	19,5	13,6	14,3	19,6	25,4	12,3	19,9	19,7	20,2	17,8	6,9	25,8
	8,2	11,7	18,4	18,8	11,3	22,0	14,4	10,2	9,1	21,8	13,7	6,5
	17,9	16,5	16,9	17,3	15,8	14,6	16,8	8,8	12,5	22,3	15,7	16,9
	15,8	14,6	6,2	10,4	7,5	9,4	17,9	16,7	18,9	21,5	24,6	19,9
3	13,6	12,2	13,2	12,8	12,3	13,3	13,0	13,1	14,9	13,8	12,7	11,9
	13,7	13,3	13,0	14,3	13,4	12,8	13,3	13,1	13,2	11,8	12,8	12,3
	13,0	13,9	14,5	12,2	13,0	12,8	13,6	13,2	12,6	12,4	11,7	13,9
	14,1	13,2	14,4	13,8	12,7	13,3	15,0	14,4	13,3	13,4	12,5	14,0
	13,5	12,9	13,8	14,8	15,0	12,6	13,6	13,5	15,0	13,5	13,8	14,2
	13,3	12,6	14,3	14,6	13,4	12,9	14,9	12,7	13,2	11,8	13,9	14,8
	12,6	13,7	14,4	13,8	14,3	12,0	14,4	14,2	14,1	12,8	13,7	12,5
	13,9	13,5	13,9	14,3	12,8	14,6	13,8	12,8	12,5	12,3	14,7	14,9
	12,8	14,6	13,2	14,4	12,5	12,4	12,9	14,7	13,9	14,5	14,6	12,9
4	9,45	9,99	8,74	7,79	8,82	7,43	8,87	9,56	8,57	9,72	7,04	9,85
	8,50	8,68	9,80	8,47	9,89	8,18	8,59	8,24	8,07	8,63	8,48	8,69
	8,86	8,65	8,86	8,12	8,16	8,85	8,68	7,70	8,48	8,09	8,22	7,64
	7,11	8,73	7,15	8,08	8,05	7,18	8,96	7,88	9,84	8,57	9,30	7,58
	7,64	9,42	8,23	9,75	8,02	9,78	9,86	7,07	7,27	9,66	8,70	9,83
	8,69	8,32	9,68	8,06	9,96	8,35	8,42	8,25	9,34	9,46	8,75	8,57
	8,75	8,05	8,54	9,81	8,34	8,27	8,22	7,18	7,55	9,92	8,66	8,23
	8,97	9,67	8,62	8,20	7,59	8,73	7,79	8,65	8,63	7,65	7,78	7,09
	7,78	7,89	9,60	7,58	8,77	7,66	8,50	8,72	8,03	8,97	7,68	8,52

5	170	164	148	199	162	145	157	162	153	135	152	150
	166	155	137	151	147	136	162	163	141	148	189	154
	173	141	153	156	164	181	154	149	146	156	128	152
	134	173	134	166	176	129	147	152	156	154	166	135
	155	148	138	173	136	150	159	142	173	144	120	155
	140	178	168	139	164	154	180	161	154	178	132	160
	126	159	148	139	158	186	167	155	174	181	158	166
	175	134	147	141	166	140	189	160	163	133	154	163
	167	152	168	147	149	125	158	165	176	158	169	193
6	26,9	78,6	58,5	22,1	62,3	36,3	93,0	83,4	44,8	63,8	42,8	61,9
	63,8	43,5	63,0	44,6	63,9	52,8	83,7	43,1	53,2	41,8	24,8	32,3
	33,0	55,9	54,5	52,7	43,0	62,3	53,6	60,2	52,6	52,4	31,7	63,8
	44,1	53,7	44,2	53,8	62,7	73,3	75,5	34,4	73,3	69,4	52,5	54,9
	53,5	62,9	73,3	64,5	55,0	52,6	33,9	43,5	45,0	63,0	63,8	44,2
	83,3	92,5	64,3	54,6	82,9	42,9	54,8	62,7	53,2	51,8	53,9	54,6
	32,6	83,7	54,4	63,8	74,3	82,0	64,4	74,6	64,1	62,8	83,7	92,5
	73,9	53,5	63,7	74,3	52,8	64,6	73,8	51,8	62,5	42,3	34,7	64,1
	62,8	64,6	73,2	44,4	52,4	92,4	72,9	64,7	53,9	54,8	64,6	62,9
7	4,25	2,05	4,02	6,71	4,52	2,82	5,80	2,87	7,35	4,44	6,64	2,68
	6,23	7,18	5,80	8,84	3,75	4,96	7,59	6,73	5,74	3,80	5,78	7,45
	3,66	6,56	2,77	3,74	6,63	8,71	6,67	5,72	6,57	5,69	6,02	5,64
	4,81	5,77	7,54	7,63	5,79	6,48	5,96	3,88	4,89	5,77	6,34	4,72
	2,63	4,69	6,02	4,75	5,05	7,54	6,76	4,79	6,52	6,83	5,70	6,85
	3,99	7,72	5,68	8,56	6,38	6,65	4,73	6,55	5,69	5,96	4,55	7,57
	5,59	5,80	6,73	6,81	5,64	6,07	5,62	5,90	6,55	6,72	2,66	6,80
	6,67	5,68	5,79	5,62	4,59	7,73	6,92	6,65	5,63	4,95	3,58	4,71
	5,91	6,59	7,67	6,58	6,76	5,46	5,09	3,62	6,78	6,69	6,28	6,52
8	260	786	108	585	574	325	165	484	620	535	206	339
	428	660	438	903	347	566	261	786	741	689	439	693
	381	292	543	606	494	261	559	447	396	458	520	452
	639	403	354	537	296	416	348	352	556	354	566	494
	255	418	286	673	536	750	559	542	571	544	350	546
	650	358	568	739	964	554	602	451	654	708	632	567
	246	354	648	639	808	646	567	559	504	441	358	241
	495	544	587	541	569	741	559	706	863	536	854	925
	557	652	468	257	549	835	458	565	646	758	599	543
9	42,3	47,4	58,5	51,2	52,3	43,9	77,0	36,4	43,5	83,9	92,8	71,9
	33,8	46,5	63,0	64,5	63,9	52,8	53,8	43,6	53,3	41,5	84,1	35,4
	53,1	57,9	64,5	62,7	53,0	60,3	43,6	65,2	52,8	62,4	35,7	63,6
	44,9	54,7	60,2	83,8	42,7	73,7	70,5	44,4	73,6	59,3	52,6	58,9
	93,5	62,0	74,3	54,3	55,0	65,8	53,9	63,5	75,0	53,0	63,8	49,2
	73,3	72,5	64,3	54,6	72,9	30,9	44,7	60,7	53,4	58,8	56,9	64,6
	31,6	73,7	84,4	63,8	74,0	80,5	64,4	34,6	54,1	62,1	61,7	92,5
	73,9	53,5	63,7	74,3	52,8	64,6	73,8	51,8	62,5	42,3	34,7	64,1
	52,4	64,8	73,9	41,4	52,4	72,4	62,9	64,1	57,7	50,8	64,0	62,7

10	1,25	2,86	0,32	6,92	4,50	1,82	2,85	8,07	5,31	3,94	4,84	3,67
	4,21	7,32	5,84	3,89	3,76	4,97	7,50	4,23	5,78	4,81	5,78	7,96
	2,66	1,59	4,71	5,04	6,23	8,21	4,67	5,75	4,57	5,62	6,02	5,64
	1,82	4,77	4,50	3,03	4,99	7,43	3,99	4,80	1,82	4,79	5,74	4,99
	2,96	4,25	5,02	4,61	5,15	6,44	7,06	8,39	4,02	5,83	4,97	3,81
	3,59	4,72	5,27	8,56	0,98	4,15	4,53	3,95	5,29	5,16	4,05	3,35
	5,22	5,81	4,73	3,94	5,64	5,07	4,68	5,91	6,05	6,92	2,16	6,88
	6,67	5,68	5,79	5,62	4,59	7,73	6,92	6,65	5,63	4,95	3,58	4,71
	4,31	5,29	7,16	6,58	5,16	4,62	5,19	3,62	1,84	3,89	4,27	3,92
11	41,5	42,3	47,4	51,6	52,3	43,9	49,2	46,6	41,8	57,5	52,3	45,8
	48,0	49,4	57,5	44,4	51,1	49,8	43,8	50,3	49,7	40,9	51,9	50,1
	47,6	44,7	59,6	41,0	61,4	46,7	39,2	49,8	36,6	54,7	40,5	55,9
	60,2	51,8	38,9	43,7	51,7	55,4	45,6	44,9	59,0	42,8	47,6	30,2
	31,4	43,9	42,7	39,2	45,0	43,8	31,1	51,8	58,5	44,9	32,7	44,7
	35,9	40,2	37,3	50,1	62,5	46,1	50,3	62,7	39,9	45,0	44,8	45,3
	50,3	61,6	48,1	45,5	30,2	45,6	54,9	52,5	60,4	61,2	45,1	60,4
	44,9	45,1	62,4	37,4	44,8	50,3	44,9	30,4	44,8	39,6	50,4	43,9
	37,8	39,5	44,0	41,3	52,0	46,2	53,7	54,5	51,2	44,7	52,2	51,6
12	4,97	6,56	4,81	5,68	4,18	2,98	4,59	4,75	5,71	2,87	3,78	1,35
	3,16	2,95	1,70	4,14	5,67	3,77	5,27	3,79	3,56	3,09	3,12	3,94
	4,84	4,78	2,97	3,93	4,31	4,61	2,96	2,88	1,84	2,71	4,63	1,70
	6,03	5,09	1,60	2,85	2,69	4,04	1,06	4,31	2,92	3,66	5,70	3,81
	5,49	2,77	2,88	4,06	3,08	2,95	2,83	2,97	1,64	6,56	3,05	4,57
	2,06	4,01	5,74	1,71	5,65	1,60	3,62	1,60	2,85	2,92	2,26	2,89
	3,95	3,12	4,96	3,62	1,59	5,73	4,75	2,95	3,63	3,15	1,18	3,78
	2,88	6,12	2,05	1,99	2,54	4,24	2,17	3,65	2,15	4,04	4,86	3,14
	3,74	2,98	3,69	2,98	2,91	3,16	1,19	4,62	4,78	3,99	2,98	6,52
13	6,56	8,99	7,94	3,32	3,36	5,82	2,91	3,66	4,60	5,65	7,81	4,98
	4,97	6,56	4,81	5,68	4,18	9,98	4,59	4,75	5,71	8,87	3,78	7,35
	3,16	5,95	7,73	4,14	5,67	3,77	5,27	3,79	3,56	3,09	5,12	6,94
	4,84	4,78	8,97	6,93	4,31	4,61	2,96	2,88	3,84	7,71	4,63	3,70
	6,03	5,09	7,60	2,85	2,69	4,04	5,06	4,31	3,92	5,66	5,70	2,81
	5,49	2,77	2,88	4,06	3,08	2,95	2,83	2,97	1,64	6,56	3,05	4,57
	2,06	4,01	5,74	1,71	5,65	9,60	3,62	4,60	2,85	4,92	2,26	2,89
	3,95	3,12	4,96	5,62	7,59	5,73	4,75	2,95	3,63	3,15	2,18	3,78
	4,08	2,98	3,69	3,98	2,91	3,16	5,19	4,62	4,73	2,99	2,98	5,72
14	650	700	742	769	564	712	681	643	753	538	752	710
	771	629	685	652	691	726	670	731	754	767	719	778
	668	725	663	630	740	691	702	797	711	656	703	821
	679	683	754	773	669	745	811	652	786	674	736	695
	735	691	738	693	642	740	800	684	753	731	695	649
	613	789	846	687	724	758	613	751	754	558	532	863
	686	624	798	819	768	726	711	715	654	761	785	776
	765	840	447	541	666	742	659	560	463	493	654	423
	557	652	468	657	449	535	558	665	646	558	469	843

15	1,32	1,48	1,49	1,54	1,60	1,27	1,85	1,77	1,25	1,94	1,54	1,37
	1,51	1,47	1,40	1,89	1,61	1,70	1,60	1,63	1,78	1,41	1,48	1,36
	1,64	1,56	1,71	1,64	1,63	1,59	1,47	1,49	1,57	1,62	1,62	1,64
	1,82	1,27	1,59	1,83	1,49	1,46	1,59	1,80	1,42	1,39	1,74	1,69
	1,96	1,66	1,32	1,60	1,54	1,44	1,52	1,37	1,30	1,83	1,87	1,31
	1,55	1,72	1,27	1,56	1,68	1,50	1,38	1,75	1,49	1,46	1,50	1,35
	1,29	1,81	1,74	1,54	1,64	1,70	1,68	1,31	1,58	1,62	1,61	1,88
	1,67	1,68	1,79	1,62	1,59	1,73	1,29	1,65	1,43	1,50	1,58	1,71
	1,31	1,29	1,65	1,58	1,61	1,42	1,37	1,62	1,84	1,80	1,27	1,42
16	31,9	29,8	11,4	10,6	10,3	17,9	39,2	26,0	21,8	37,5	22,3	25,3
	18,0	39,4	17,5	24,4	21,1	40,5	33,8	20,3	30,7	20,9	21,9	10,1
	27,6	14,7	29,6	21,0	21,4	36,7	19,2	19,8	26,1	24,7	30,5	15,9
	20,2	21,8	28,9	23,7	11,7	25,4	25,6	24,9	19,0	12,8	27,6	30,2
	31,4	23,9	22,7	29,2	25,0	23,8	31,1	21,8	28,5	14,9	22,7	24,7
	25,9	20,2	27,3	20,1	22,5	26,1	30,3	32,7	31,9	25,0	24,8	25,3
	20,3	21,6	28,1	25,5	30,2	25,6	24,9	22,5	24,4	21,2	25,1	30,4
	24,9	15,1	32,4	37,4	24,8	20,3	14,8	20,4	19,8	29,6	20,4	23,9
	34,8	36,0	24,0	21,3	22,0	26,2	23,7	24,5	30,2	34,7	22,2	21,6
17	94,5	99,8	71,8	79,6	80,3	77,9	99,2	66,0	71,8	77,5	82,3	85,3
	80,1	69,4	77,5	74,4	81,1	60,5	73,8	80,3	66,7	70,9	61,9	90,1
	87,6	74,7	69,6	71,0	71,4	86,7	79,2	69,8	86,1	82,7	81,5	75,9
	80,2	71,8	68,9	66,7	67,7	75,4	85,6	74,9	90,9	82,8	82,6	80,2
	71,4	73,9	72,7	79,2	75,0	73,8	81,1	71,8	68,5	66,9	82,7	84,7
	85,9	83,2	77,3	80,1	82,5	83,1	70,3	72,7	71,9	85,0	74,8	65,3
	80,3	81,6	78,1	75,5	80,2	85,6	74,9	82,5	84,4	71,2	85,1	80,4
	68,9	75,1	82,4	87,4	74,8	80,3	84,8	80,4	79,8	69,5	80,4	83,9
	74,8	66,5	74,0	81,3	82,0	86,2	83,7	74,5	80,2	84,7	82,2	81,6
18	5,15	4,74	1,78	1,07	2,59	1,21	3,77	9,14	7,71	3,36	4,43	3,77
	2,43	1,20	4,19	3,77	4,41	1,06	1,19	8,67	3,74	2,14	3,26	2,56
	3,17	2,86	1,94	1,23	1,09	3,67	1,95	5,26	1,37	1,39	1,89	1,25
	9,87	7,79	1,82	1,66	8,58	1,75	1,84	2,65	8,99	2,33	2,37	2,59
	4,76	1,64	1,53	5,91	2,02	2,39	6,80	4,87	2,48	2,01	1,69	7,34
	1,83	3,78	2,04	1,82	1,86	1,79	2,05	5,33	1,22	6,17	2,03	1,87
	1,88	1,41	1,86	5,77	1,35	2,11	4,87	1,66	3,05	2,00	1,88	2,11
	4,19	1,85	5,21	2,65	3,69	1,87	4,73	2,18	2,17	7,88	5,51	3,44
	5,56	1,67	4,34	1,79	1,80	5,45	1,84	6,31	1,86	3,95	1,76	5,87
19	1,90	3,18	1,39	0,74	3,20	1,17	2,95	2,77	4,05	1,74	3,24	0,07
	0,81	3,12	1,20	2,69	1,91	2,30	3,20	4,13	1,38	2,41	4,58	2,26
	1,44	0,96	1,61	2,54	3,13	1,51	1,11	2,39	4,37	2,12	0,72	1,24
	1,52	1,82	2,92	0,83	0,98	1,76	4,19	4,30	2,62	0,91	0,74	1,19
	2,60	0,96	0,82	1,20	2,54	3,24	2,92	1,17	3,20	4,53	2,17	3,13
	5,05	1,12	1,97	0,96	3,68	0,99	3,38	2,85	2,59	4,06	4,30	1,15
	2,09	3,81	4,64	3,84	2,34	3,70	2,68	4,31	3,88	2,62	4,61	4,98
	0,97	2,68	3,87	1,62	3,59	2,73	1,19	1,65	1,93	3,50	2,58	2,71
	3,31	2,29	1,65	1,98	2,61	1,42	1,27	1,62	5,04	4,80	3,27	3,42

20	702	471	715	721	724	482	491	740	493	508	702	410
	611	529	685	552	791	726	810	731	614	667	719	738
	608	705	519	630	743	611	702	793	701	656	713	800
	479	483	614	773	569	705	615	602	786	774	736	615
	535	651	538	793	442	540	800	684	753	731	695	649
	613	789	846	687	724	758	613	610	454	558	532	763
	786	424	498	519	468	626	621	615	654	767	781	776
	665	810	447	541	666	742	629	568	463	499	614	423
	539	652	478	617	449	535	758	715	616	558	469	643

ЗАДАНИЕ № 2 (задачи 21 – 40)

При заданной точности δ оценки неизвестного математического ожидания a нормально распределенной случайной величины найти минимальный объем выборки n, при котором с надежностью γ и средним квадратическим отклонением σ обеспечивается заданная точность.

21	γ = 0,975	δ = 0,3	σ = 2	31	γ = 0,788	δ = 0,5	σ = 7
22	γ = 0,925	δ = 0,2	σ = 15	32	γ = 0,904	δ = 1,1	σ = 3
23	γ = 0,955	δ = 2,0	σ = 5	33	γ = 0,896	δ = 2,9	σ = 10
24	γ = 0,976	δ = 1,3	σ = 4	34	γ = 0,909	δ = 0,7	σ = 6
25	γ = 0,982	δ = 1,5	σ = 7	35	γ = 0,972	δ = 1,2	σ = 8
26	γ = 0,990	δ = 0,8	σ = 10	36	γ = 0,821	δ = 0,6	σ = 11
27	γ = 0,922	δ = 2,1	σ = 3	37	γ = 0,918	δ = 1,2	σ = 5
28	γ = 0,967	δ = 1,7	σ = 8	38	γ = 0,793	δ = 2,3	σ = 9
29	γ = 0,938	δ = 0,9	σ = 12	39	γ = 0,884	δ = 0,9	σ = 13
30	γ = 0,981	δ = 1,0	σ = 9	40	γ = 0,945	δ = 0,3	σ = 4

ЗАДАНИЕ № 3 (задачи 41 – 60)

1. Построить поле корреляции и сформулировать гипотезу о форме связи.
2. Рассчитать параметры уравнения линейной регрессии.
3. Оценить тесноту связи с помощью выборочных коэффициентов корреляции и детерминации.
4. Проверить значимость выборочного коэффициента корреляции и параметров линейной модели при уровне значимости α = 0,05.

5. Дать с помощью среднего коэффициента эластичности сравнительную оценку силы связи факторного признака с результативным.

6. Оценить качество уравнения с помощью средней ошибки аппроксимации.

7. Оценить статистическую надежность результатов регрессионного моделирования с помощью *F*-критерия Фишера.

8. Сделать выводы.

Исследуйте зависимость числа поврежденных плодов (*Y*, %) от урожая (*X*, кг.)

41	*X*	15	16	12	26	18	14	8	38	20	34	29	25
	Y	3,2	4,6	3,8	3,0	3,7	3,1	2,5	2,9	4,4	5,2	4,8	4,3
42	*X*	6	7	8	9	10	11	12	13	14	15	16	17
	Y	1,3	1,6	1,5	2,0	1,9	2,1	2,6	2,4	3,0	3,2	3,0	3,5
43	*X*	7	8	10	12	13	14	15	17	18	20	21	22
	Y	2,2	2,3	2,4	2,5	2,6	2,7	2,8	3,0	3,1	3,2	3,3	3,1
44	*X*	72,8	76,4	78,0	70,9	68,6	80,1	82,3	65,7	62,9	90,1	76,3	67,8
	Y	15	18	16	15	13	19	12	14	13	10	12	15
45	*X*	767	819	857	900	642	593	771	626	572	860	426	582
	Y	5,7	5,9	6,2	6,3	86	11,9	13,3	12,9	14,0	13,8	14,5	9,7
46	*X*	52	57	59	62	63	86	119	122	135	144	158	161
	Y	12,1	11,6	11,3	10,2	8,6	9,0	9,4	10,6	11,8	10,9	13,5	14,0
47	*X*	320	300	362	405	410	476	563	548	609	557	612	685
	Y	20,1	24,3	28,7	30,5	31,6	33,7	34,5	37,2	38,6	40,0	41,2	39,4
48	*X*	420	380	350	400	440	380	450	320	487	521	504	312
	Y	5,5	6,0	6,5	6,0	5,0	6,5	4,5	7,3	3,8	6,2	5,9	2,8
49	*X*	25,0	36,2	22,8	15,9	48,1	39,6	42,5	31,8	28,9	44,8	39,6	42,7
	Y	55	62	50	30	75	70	70	55	30	60	43	51
50	*X*	920	880	1400	1170	1100	1090	930	1340	1230	1080	1580	995
	Y	3,35	3,45	5,53	3,80	5,17	5,69	4,25	6,93	5,07	3,81	7,51	4,92
51	*X*	45,2	46,7	47,5	48,3	51,9	59,1	65,8	58,0	61,4	42,5	53,8	64,0
	Y	12	11	29	25	38	16	43	35	18	21	30	37
52	*X*	452	456	451	449	448	447	450	446	454	440	461	445
	Y	32,7	34,3	33,6	31,9	36,3	35,1	37,2	38,0	32,7	30,1	39,2	31,8
53	*X*	12,8	35,2	75,0	21,0	31,5	44,5	60,2	36,4	56,3	68,5	70,0	73,3
	Y	2	3	11	6	7	8	9	4	3	7	8	9
54	*X*	66	61	67	73	51	59	48	47	44	39	36	57
	Y	3,8	3,1	3,6	4,3	4,9	3,3	2,8	2,5	2,0	2,9	3,8	4,0
55	*X*	426	367	515	421	410	440	603	696	528	703	584	356
	Y	3,96	3,89	3,75	3,83	3,10	3,85	3,64	3,95	4,02	4,68	3,81	2,96

56	*X*	21	23	24	25	27	29	30	32	35	41	43	48
	Y	1,82	1,96	2,15	2,27	2,31	2,43	2,50	2,58	3,05	3,14	3,37	3,49
57	*X*	466	676	515	421	470	441	603	696	514	563	538	702
	Y	3,96	3,89	3,75	3,83	4,10	3,85	3,64	3,95	4,30	4,62	4,21	5,13
58	*X*	35,8	40,2	45,6	48,4	49,6	47,8	45,0	41,7	33,1	52,4	31,9	50,3
	Y	4	5	6	7	7	6	8	8	3	9	2	9
59	*X*	501	554	603	529	651	608	523	557	625	486	491	638
	Y	14,0	21,2	17,1	12,2	13,5	23,3	23,1	21,4	24,7	11,8	16,3	20,1
60	*X*	52	46	38	37	45	39	55	59	32	40	64	61
	Y	1,52	1,86	1,25	2,06	1,84	1,47	2,83	3,08	1,17	2,15	3,07	2,88

РЕШЕНИЕ ТИПОВОГО ВАРИАНТА

Задание № 1. Приведена выборка 108 результатов измерений случайной величины X:

150	154	148	149	160	147	158	164	153	135	152	150
168	158	138	151	147	136	160	163	141	148	139	153
171	141	143	156	164	161	159	149	146	156	130	152
139	153	154	136	166	169	147	152	156	154	166	135
155	148	138	173	136	150	159	142	173	144	150	145
150	158	168	139	164	154	150	151	154	158	132	160
146	154	148	139	158	146	167	155	144	141	158	146
155	144	147	141	166	140	159	160	163	133	154	163
157	152	168	157	149	135	158	165	146	158	169	143

Решение. 1. Построить интервальный вариационный ряд (ряд 1) по частотам, относительным частотам и накопленным частотам.

n =108 - объем выборки, тогда количество интервалов находим по формуле (6): $k = \sqrt{108} \approx 10$. $x_{max} = 173, x_{min} = 130$. $h = \frac{x_{max} - x_{min}}{\sqrt{k}} = \frac{173-130}{10} = 4,3$ – шаг.

Тогда $x_1 = 130$, $x_2 = 130 + 4,3 = 134,3$, $x_3 = 134,3 + 4,3 = 138,6$,

$x_4 = 138,6 + 4,3 = 142,9$; $x_5 = 142,9 + 4,3 = 147,2$;, $x_6 = 147,2 + 4,3 = 151,5$;

$x_7 = 151,5 + 4,3 = 155,8$; $x_8 = 155,8 + 4,3 = 160,1$; $x_9 = 160,1 + 4,3 = 164,4$;

$x_{10} = 164,4 + 4,3 = 168,7$; $x_{11} = 168,7 + 4,3 = 173$.

Подсчитаем частоты, относительные частоты и накопленные частоты соответственно по формулам (3) и (11) и внесем в таблицу.

Ряд 1 - Интервальный вариационный ряд

x_i	[130 - 134,3)	[134,3 -138,6)	[138,6 - 142,9)	[142,9 - 147,2)	[147,2 - 151,5)	[151,5 - 155,8)	[155,8 -160,1)	[160,1 - 164,4)	[164,4 - 168,7)	[168,7- 173]
n_i	3	8	10	15	15	17	20	7	8	53
w_i	0,0278	0,0741	0,0926	0,1389	0,1389	0,1574	0,1851	0,0648	0,0741	0,0463
n_i^*	3	11	21	36	51	68	88	95	103	108
$\frac{\boldsymbol{n_i}}{\boldsymbol{h}}$	0,689	1,860	2,326	3,488	3,488	3,953	4,651	1,628	1,860	1,163
$\frac{\boldsymbol{w_i}}{\boldsymbol{h}}$	0,0065	0,4325	0,0215	0,0323	0,0323	0,0366	0,0430	0,0115	0,4325	0,0108

2. От ряда 1 перейти к точечному вариационному ряду (ряд 2).

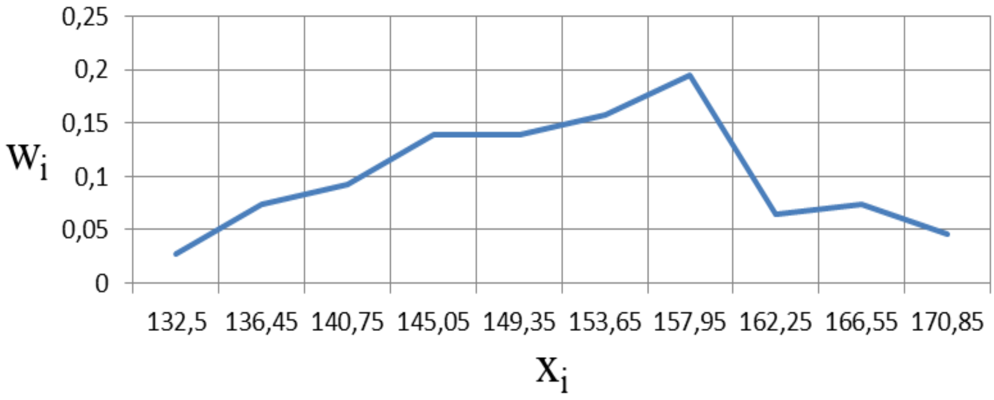
Воспользуемся формулой (10), определив середины интервалов, получим точечный (дискретный) вариационный ряд. $i = \overline{1{,}11}$,

Ряд 2 – Дискретный (точечный) вариационный ряд

x_i^o	132,15	136,45	140,75	145,05	149,35	153,65	157,95	162,25	166,55	170,85
n_i	3	8	10	15	15	17	20	7	8	5
n_i^*	3	11	21	36	51	68	88	95	103	108

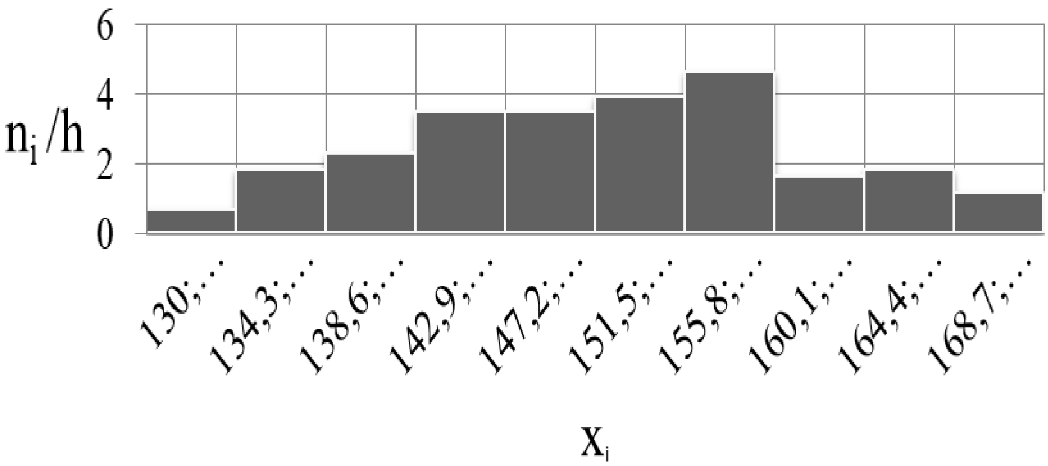
3. Начертить полигоны частот и относительных частот, кумуляту (по ряду 2) и гистограммы частот и относительных частот (по ряду 1). Проиллюстрируем графически ряды 1 и 2.

Полигон относительных частот по ряду 2

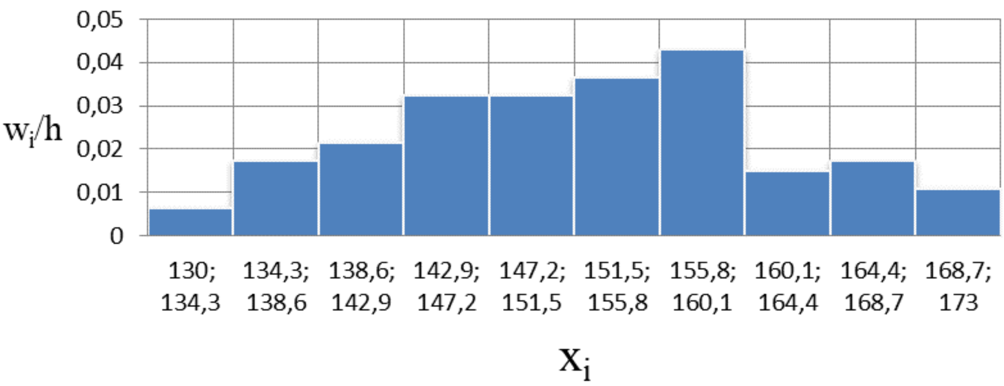


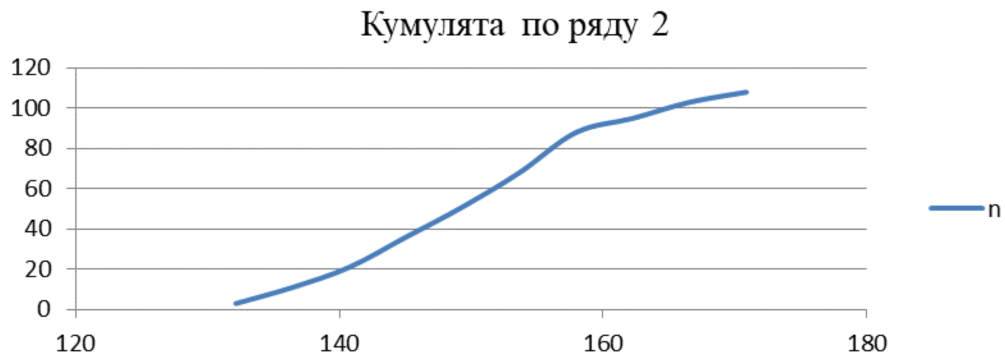
т

Гистограмма частот по ряду 1



Гистограмма относительных частот по ряду 1



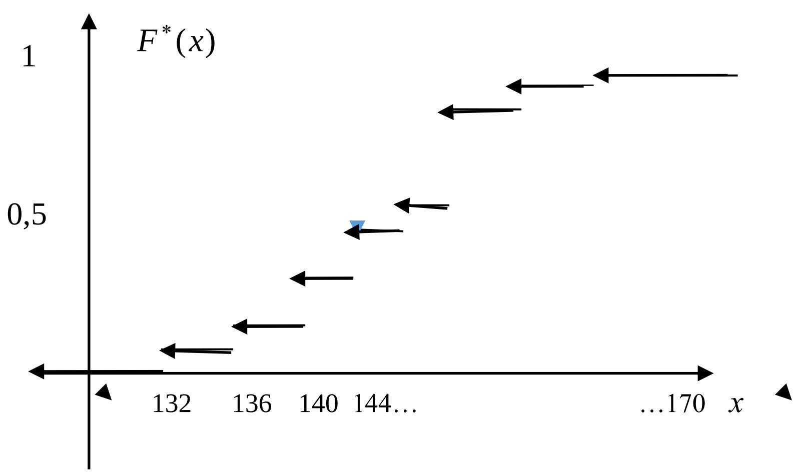


4. Записать аналитически и построить графически статистическую функцию распределения (по ряду 2).

Согласно формуле (12) найдем статистическую функцию распределения:

$$F^*(x)=\begin{cases}0, \ \ x<132{,}15\\ \dfrac{3}{108}, \ 132{,}15\le x\le 136{,}45\\ \dfrac{11}{108}, \ 136{,}45\le x\le 140{,}75\\ \dfrac{21}{108}, \ 140{,}75\le x\le 145{,}05\\ \dfrac{36}{108}, \ 145{,}05\le x\le 149{,}35\\ \dfrac{51}{108}, \ 149{,}35\le x\le 153{,}65\\ \dfrac{68}{108}, \ 153{,}65\le x\le 157{,}95\\ \dfrac{89}{108}, \ 157{,}95\le x\le 162{,}65\\ \dfrac{95}{108}, \ 162{,}65\le x\le 166{,}55\\ \dfrac{103}{108},. \ 166{,}55\le x\le 170{,}85\\ 1, \ \ x\ge 170{,}85\end{cases}$$

График функции распределения



5. Найти выборочные средние: среднюю арифметическую, среднюю геометрическую, среднюю гармоническую; выборочную дисперсию, выборочное среднеквадратическое отклонение, коэффициенты вариации, асимметрии и эксцесс (по ряду 2). Применить метод произведений.

Проведем вычисления выборочных средних по формулам (13') - (16'):

$\bar{x}_в$ = (132,15 · 3 + 136,45 · 8 + 140,75 · 10 + 145,05 · 15 + 149,35 ·15 + 153,65 · 17 + + 157,95 · 20 + 162,25·7 + 166,55 · 8 + 170,85 · 5) /108 = 151,898.

Для вычисления средней геометрической найдем десятичные логарифмы вариант дискретного ряда (2):

lg 132,5 = 2,12; lg 136,45 = 2,13; lg 140,75 = 2,15; lg 145,05 = 2,16;
lg 149,35 = 2,17; lg 153,65 = 2,19; lg 157,95 = 2,20; lg 162,25 = 2,21;
lg 166,55 = 2,22; lg 170,85 = 2,23.

$$lg\bar{x}_{геом.} = \frac{3\cdot 2{,}12+8\cdot 2{,}13+10\cdot 2{,}15+15\cdot 2{,}16+15\cdot 2{,}17+17\cdot 2{,}19+20\cdot 2{,}2+7\cdot 2{,}21+8\cdot 2{,}22+5\cdot 2{,}23}{108} =$$

$$\approx \frac{235{,}46}{108} \approx 2{,}18 \Rightarrow \bar{x}_{геом.} = 10^{2{,}18} = 151{,}36.$$

$$\bar{x}_{гарм.} = \frac{108}{\frac{3}{132{,}5}+\frac{8}{136{,}45}+\frac{10}{140{,}75}+\frac{15}{145{,}05}+\frac{15}{149{,}35}+\frac{17}{153{,}65}+\frac{20}{157{,}95}+\frac{7}{162{,}25}+\frac{8}{166{,}55}+\frac{5}{170{,}85}} \approx 150{,}49.$$

Между выборочной средней, средней геометрической и средней гармонической существует соотношение (17)

$$\overline{x}_{гарм.} \le \overline{x}_{геом.} \le \overline{x}_{в}.$$

Проверим это двойное неравенство:

$$150{,}49 \le 151{,}36 \le 151{,}89,$$

что подтверждает правильность вычислений.

Для подсчета выборочной дисперсии воспользуемся формулой (40), для чего вычислим квадратичную среднюю:

$$\overline{x^2} = (132{,}15^2 \cdot 3 + 136{,}45^2 \cdot 8 + 140{,}75^2 \cdot 10 + 145{,}05^2 \cdot 15 + 149{,}35^2 \cdot 15 + 153{,}65^2 \cdot 17 + 157{,}95^2 \cdot 20 + 162{,}25^2 \cdot 7 + 166{,}55^2 \cdot 8 + 170{,}85^2 \cdot 5)\ /108 =$$

$$= 23164{,}78$$

Тогда $D_в = \overline{x^2} - (\overline{x})^2 = 23164{,}78 - (151{,}898)^2 = 94{,}21$.

А выборочное среднеквадратичное отклонение равно $\sigma_в = \sqrt{94{,}21} = 9{,}71$.

Исправленное среднее квадратичное отклонение определится:

$$S_в^{\ 2} = \frac{n-1}{n} D_в = \frac{108-1}{108} 94{,}21 = 93{,}34; S_в = \sqrt{93{,}34} = 9{,}66.$$

Коэффициент вариации найдем по формуле (28):

$$V = \frac{\sigma_в}{\overline{x}_в} \times 100\% = \frac{9{,}71}{151{,}898} \times 100\% = 6{,}39\%.$$

Вычисления характеристик статистических рядов рациональнее производить методом произведений. Так как значения равностоящие, перейдем к условным вариантам $u_i = \dfrac{x_i - C}{h}$, В качестве ложного нуля выберем варианту $x = C = 157{,}95$ с наибольшей частотой $n_{\max} = 20$, Все вычисления занесем в таблицу. Для контроля вычислений применим формулы:

$$\sum_{i=1}^{k} n_i (u_i + 1)^2 = \sum_{i=1}^{k} n_i u_i^{\ 2} + 2\sum_{i=1}^{k} n_i u_i + n,$$

$$\sum_{i=1}^{k} n_i (u_i + 1)^4 = \sum_{i=1}^{k} n_i u_i^{\ 4} + 4\sum_{i=1}^{k} n_i u_i^{\ 3} + 6\sum_{i=1}^{k} n_i u_i^{\ 2} + 4\sum_{i=1}^{k} n_i u_i + n.$$

x_i	n_i	u_i	n_iu_i	$n_iu_i^2$	$n_i\,(u_i+1)^2$	$n_iu_i^3$	$n_iu_i^4$	$n_i\,(u_i+1)^4$
132,15	3	- 6	- 18	108	75	- 648	3888	1875
136,45	8	- 5	- 40	200	128	- 1000	5000	2048
140,75	10	- 4	- 40	160	90	- 640	2560	810
145,05	15	- 3	- 45	135	60	- 405	1215	240
149,35	15	- 2	- 30	60	15	-120	240	15
153,65	17	- 1	- 17	17	0	- 17	17	0
157,95	20	0	0	0	20	0	0	20
162,25	7	1	7	7	28	7	7	112
166,55	8	2	16	32	72	64	128	648
170,85	5	3	15	45	80	135	405	1280
Σ	108		- 152	764	568	-2624	13460	7048

$$\sum_{i=1}^{10} n_i(u_i+1)^2 = 568, \qquad \sum_{i=1}^{10} n_i u_i^{\,2} + 2\sum_{i=1}^{10} n_i u_i + n = 764 + 2\cdot(-152) + 108 = 568.$$

$$\sum_{i=1}^{10} n_i(u_i+1)^4 = 7048,$$

$$\sum_{i=1}^{10} n_i u_i^{\,4} + 4\sum_{i=1}^{10} n_i u_i^{\,3} + 6\sum_{i=1}^{10} n_i u_i^{\,2} + 4\sum_{i=1}^{10} n_i u_i + n =$$
$$= 13460 + 4\cdot(-2624) + 6\cdot 764 + 4\cdot(-152) + 108 = 7048.$$

Совпадение найденных сумм свидетельствует о правильности вычислений.

Вычислим условные моменты первого, второго, третьего и четвертого порядков по формулам:

$$M_1^* = \frac{\sum_{i=1}^{10} n_i\cdot u_i}{n} = \frac{-152}{108} = -1{,}407; \qquad M_2^* = \frac{\sum_{i=1}^{10} n_i\cdot u_i^2}{n} = \frac{764}{108} = 7{,}074;$$

$$M_3^* = \frac{\sum_{i=1}^{10} n_i\cdot u_i^3}{n} = -\frac{2624}{108} = -24{,}296; \qquad M_4^* = \frac{\sum_{i=1}^{10} n_i\cdot u_i^4}{n} = \frac{13460}{108} = 124{,}629.$$

Шаг h (разность между двумя соседними вариантами) равен 4,3. Выборочная средняя арифметическая и дисперсия, согласно формулам (39) и (40), определятся:

$$\bar{x}_в = -1{,}407\cdot 4{,}3 + 157{,}95 = 151{,}89; \qquad D_в = \left[7{,}074 - (-1{,}407)^2\right]\cdot 4{,}3^2 = 94{,}21.$$

Значения выборочной средней, вычисленной по общей формуле и методу произведений, совпадают. Для нахождения коэффициента асимметрии и эксцесса применим связь между центральными и условными моментами (36):

$$\mu_3 = \left[M_3^* - 3M_1^* M_2^* + 2(M_1^*)^3 \right] \cdot h^3 =$$
$$= \left[-24{,}296 - 3 \cdot (-1{,}407) \cdot 7{,}074) + 2 \cdot (-1{,}407)^3 \right] \cdot 4{,}3^3 = -80{,}143.$$

$$\mu_4 = \left[M_4^* - 4M_1^* M_3^* + 6M_2^* (M_1^*)^2 - 3(M_1^*)^4 \right] \cdot h^4 =$$
$$= \left[124{,}629 - 4 \cdot (-1{,}407) \cdot (-24{,}296) + 6 \cdot 7{,}074 \cdot (-1{,}407)^2 - 3 \cdot (-1{,}407)^4 \right] \cdot 4{,}3^4 =$$
$$= 114062{,}792.$$

$$A_s = \frac{\mu_3}{\sigma_в^3} = -\frac{80{,}143}{9{,}71^3} = -0{,}088.$$

$$E_k = \frac{\mu_4}{\sigma_в^4} - 3 = \frac{114062{,}792}{9{,}71^4} - 3 = 9{,}831,$$

По значению коэффициента асимметрии можно судить о левосторонней асимметрии кривой распределения, а по значению эксцесса, заключаем, что кривая распределения имеет более острую вершину по сравнению с нормальной кривой.

6. Определить моду и медиану графически и аналитически (по рядам 1 и 2).

Нахождение моды и медианы аналитически осуществим по формулам (18) – (21). Модой для дискретного ряда является варианта x_{med} =157,95, которая имеет максимальную частоту $n_{\max} = 20$. Для интервального ряда мода равна

$$M_0 = x_0 + h \frac{n_{M_0} - n_{M_0 - 1}}{(n_{M_0} - n_{M_0 - 1}) + (n_{M_0} - n_{M_0 + 1})} =$$
$$= 155{,}8 + 4{,}3 \cdot \frac{20 - 17}{(20 - 17) + (20 - 7)} = 156{,}6.$$

Для графического определение моды можно использовать гистограмму частот (стр. 43), соединив левый конец прямоугольника модального интервала с правым концом последующего за ним прямоугольника (интервала), а правый конец – с левым предыдущего за ним прямоугольника (интервала) (см. рис.5, стр. 13). Построить самостоятельно.

Медиана для дискретного ряда определится по формуле (20). Так число интервалов – число четное (k = 10), то получим:

$$M_e = \frac{x_5 + x_6}{2} = \frac{149{,}35 + 153{,}65}{2} = 142{,}5.$$

Для нахождения медианы интервального ряда, сначала определим медианный интервал по накопленной частоте, превосходящей половину объема выборки n_i^*= 68. Это будет интервал [151,5 – 155,8). Тогда, согласно формуле (21) получим

$$M_e = 151{,}5 + \frac{4{,}3 \cdot \left(\frac{108}{2} - 68\right)}{15} = 147{,}49.$$

Для графического определение медианы можно использовать кумуляту (стр. 44), разделив последнюю ординату пополам и, опустив перпендикуляр на ось n_i^* , найдем точку пересечения кумуляты с этой прямой, проекция которой на ось Ox и даст значение медианы, полученное графически (см. рис.6, стр. 14). Построить самостоятельно.

7. В предположении, что выборка извлечена из генеральной совокупности, подчиненной нормальному закону, при заданной надежности γ = 0,95 построить доверительные интервалы для неизвестного математического ожидания a, неизвестной дисперсии D и среднего квадратического отклонения σ случайной величины X.

Для интервальной оценки неизвестного математического ожидания a воспользуемся формулой (44)

$$\overline{x}_в - t_\gamma \frac{S_в}{\sqrt{n}} < a < \overline{x}_в + t_\gamma \frac{S_в}{\sqrt{n}},$$

где $\gamma = 0{,}95$, $\overline{x_в} = 151{,}89; n = 108; S_в = 9{,}66.$ Параметр t_γ определим из статистической таблицы 2 по заданным числу степеней свободы $r = n - 1$ (или объему выборки) и уровнем значимости $\alpha = 1 - \gamma$, причем $t_\gamma \frac{S_в}{\sqrt{n}}$ также определяет точность оценки, иногда ее называют предельной погрешностью оценки. Для нашего примера имеем t_γ (120; 0,05) = 1,98.

$$151{,}89-1{,}98\frac{9{,}66}{\sqrt{108}}<a<151{,}89+1{,}98\frac{9{,}66}{\sqrt{108}}\Rightarrow 150{,}05<a<153{,}73.$$

Доверительный интервал для оценки среднего квадратического отклонения σ нормально распределенной случайной величины X с надежностью γ находится из двойных неравенств (47):

$$S_в(1-q)<\sigma<S_в(1+q) \text{ при } q<1,$$
$$0<\sigma<S_в(1+q) \text{ при } q>1,$$

где q находится из статистической таблицы 5 по заданной надежности γ и объему выборки. Так как $q(0{,}95;100)=0{,}143<1$, то имеем

$$9{,}66\cdot(1-0{,}143)<\sigma<9{,}66\cdot(1+0{,}143)\Rightarrow 8{,}28<\sigma<11{,}04.$$

Для интервальной оценки неизвестной дисперсии D возведем полученное неравенство в квадрат

$$68{,}56<D<121{,}88.$$

9.На основе анализа гистограммы и статистической функции распределения оценить близость эмпирического распределения к нормальному закону, проверить гипотезу о том, что случайная величина X согласуется с нормальным законом распределения по критерию Пирсона $\left(\chi^2\right)$.

Анализ гистограммы и графика статистической функции распределения показывает, что эмпирическое распределение не близко к нормальному распределению случайной величины X.

Если предположить, что выборка извлечена из генеральной совокупности, подчиненной нормальному закону, то теоретическое распределение с параметрами $a=\overline{x}_в=151{,}89$ и $\sigma=S_в=9{,}66$ будет иметь вид

$$f(x)=\frac{1}{9{,}66\sqrt{2\pi}}e^{-\frac{(x-151{,}89)^2}{2\cdot 9{,}66^2}}=\frac{1}{9{,}66\sqrt{2\pi}}e^{-\frac{(x-151{,}89)^2}{186{,}63}}.$$

Оценим близость эмпирического распределения к нормальному закону по критерию Пирсона, статистикой которого является случайная величина

$$\chi^2_{\text{расч.}} = \sum_{i=1}^{k} \frac{\left(n_i - np_i\right)^2}{np_i} = \sum_{i=1}^{k} \frac{\left(n_i - m_i\right)^2}{m_i},$$

где k – число интервалов разбиения выборки (k =10); n_i – частота i – го интервала; n – объем выборки; p_i – теоретическая вероятность попадания значения случайной величины X в i–й интервал; m_i - теоретические частоты ($m_i = n \cdot p_i$).

Вычислим $z_i = \frac{x_i - \bar{x}_в}{S_в}$ и найдем значения функций Лапласа от полученных переменных, используя таблицу 1, причем полагают наименьшее значение $z_{\min} = -\infty$, а наибольшее $z_{\max} = \infty$, используя свойства функции Лапласа $\Phi(-\infty) = -0{,}5$; $\Phi(\infty) = 0{,}5$. Например, $z_2 = \frac{134{,}3 - 151{,}89}{9{,}66} = -1{,}82$; Ф(-1,82) = -0,46582 (значение находится на пересечении строки с x = 1,8 и столбца № 2, с применением свойства нечетности функции Лапласа). Все вычисления занесем в следующие две таблицы.

$[x_i, x_{i+1})$	n_i	$[z_i, z_{i+1})$	$\Phi(z_i)$	$\Phi(z_{i+1})$	p_i	m_i
[130;134,3)	3	[-∞;-1,82)	-0,5	-0,46582	0,03418	4
[134,3;138,6)	8	[-1,82; -1,38)	-0,46582	-0,41621	0,04961	5
[138,6;142,9)	10	[-1,38;-0,93)	-0,41621	-0,32381	0,0924	10
[142,9;147,2)	15	[-0,93;-0,49)	-0,32381	-0,18439	0,13942	15
[147,2;151,5)	15	[-0,49;-0,39)	-0,18439	-0,15173	0,03266	4
[151,5;155,8)	17	[-0,39;0,40)	-0,15173	0,15542	0,30715	33
[155,8;160,1)	20	[0,40;0,85)	0,15542	0,30234	0,14692	16
[160,1; 164,4)	7	[0,85;1,29)	0,30234	0,40147	0,09913	11
[164,4; 168,7)	8	[1,29;1,74)	0,40147	0,45907	0,0576	6
[168,7; 173]	5	[1,74; ∞)	0,45907	0,5	0,04093	4
$\sum$	108				1	108

Вероятность попадания в i-й интервал $[z_i, z_{i+1})$ для нормального распределения находится по известной формуле $p_i = \Phi(z_{i+1}) - \Phi(z_i)$, где $\Phi(z)$ - функция Лапласа. Далее, для нахождения $\chi^2_{набл}$, составим следующую таблицу:

n_i	m_i	n_i - m_i	$(n_i - m_i)^2$	$\frac{(n_i - m_i)^2}{m_i}$
3	4	-1	1	0,25
8	5	3	9	1,80
10	10	0	0	0,00
15	15	0	0	0,00
15	4	11	121	30,25
17	33	-16	256	7,56
20	16	4	16	1
7	11	-4	16	1,45
8	6	2	4	0,67
5	4	1	1	0,25
$\sum$	108			$\chi^2_{набл} = 43,23$

Число степеней свободы равно s = 10 – 2 – 1 = 7. Уровень значимости равен α = 1 – 0,95 = 0,05. По таблице 3 распределения χ^2 по данным s, α найдем $\chi_{кр}{}^2(7;0,05) = 14,1$. Так как $\chi^2_{набл} > \chi^2_{кр}$, то гипотезу H_o о нормальном распределении выборки отвергаем. Случайная величина не согласуется с нормальным законом распределения. *Эмпирические и теоретические частоты различаются значимо.*

Задание № 2. При заданной точности δ = 0,6 оценки неизвестного математического ожидания a нормально распределенной случайной величины найти минимальный объем выборки n, при котором с надежностью γ = 0,821 и средним квадратическим отклонением σ = 11 обеспечивается заданная точность.

Решение. Для определения минимального объема выборки n воспользуемся следующей формулой:

$$n = \frac{t^2 \cdot \sigma^2}{\delta^2}.$$

Из формулы $2\Phi(t) = \gamma = 0,821,\ \Phi(t) = 0,4105,$ найдем из таблицы 1 (справа налево) t = 1,34. Тогда

$$n = \frac{(1,34)^2 \cdot 11^2}{(0,6)^2} = 603,52.$$

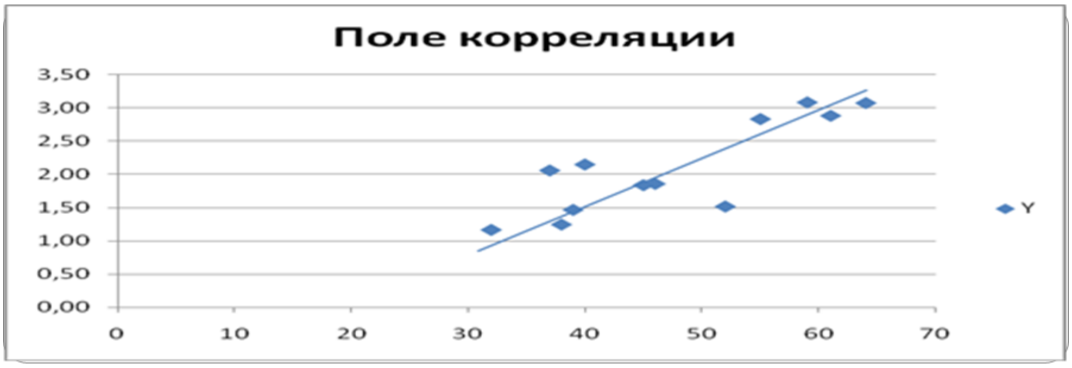
Округляя получаем, что минимальный объем выборки должен быть $n = 604$.

Задание 3. Зависимость числа поврежденных плодов (Y, %) от урожая (X, кг) представлена 12 данными

X	52	46	38	37	45	39	55	59	32	40	64	61
Y	1,52	1,86	1,25	2,06	1,84	1,47	2,83	3,08	1,17	2,15	3,07	2,88

Решение. 1. Построить поле корреляции и сформулировать гипотезу о форме связи.

Наносим точки (x;y) в системе координат и получаем поле корреляции. Проведем прямую линию, около которой сгруппированы построенные точки. Будем считать, что связь между x и y *линейная* $\hat{y} = ax + b$.



2. Рассчитать параметры уравнения линейной регрессии.

Для определения коэффициентов уравнения линейной парной регрессии воспользуемся формулой (55). Составим расчетную таблицу.

	x	y	x^2	y^2	$x \cdot y$	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
	52	1,52	2704	2,3104	79,04	0,5783	0,3344
	46	1,86	2116	3,4596	85,56	0,2383	0,0568
	38	1,25	1444	1,5625	47,5	0,8483	0,7196
	37	2,06	1369	4,2436	76,22	0,0383	0,0015
	45	1,84	2025	3,3856	82,8	0,2583	0,0667
	39	1,47	1521	2,1609	57,33	0,6283	0,3948
	55	2,83	3025	8,0089	155,65	0,7317	0,5354
	59	3,08	3481	9,4864	181,72	0,9817	0,9637
	32	1,17	1024	1,3689	37,44	0,9283	0,8617
	40	2,15	1600	4,6225	86	0,0517	0,0027
	64	3,07	4096	9,4249	196,48	0,9717	0,9442
	61	2,88	3721	8,2944	175,68	0,7817	0,6111
Сумма	568,00	**25,18**	28126	58,3286	1261,42		5,4926
Среднее	47,3333	2,0983	2343,8333	4,8607	105,1183		

Система нормальных уравнений примет вид:

$$\begin{cases} 12b + 568a = 25{,}18 \\ 568b + 28126a = 1261{,}42 \end{cases}.$$

Решение системы дает следующие результаты:

$$a = \frac{\overline{y \cdot x} - \bar{y} \cdot \bar{x}}{\overline{x^2} - \bar{x}^2} = \frac{105{,}1183 - 2{,}0983 \cdot 47{,}3333}{2343{,}8333 - (47{,}3333)^2} = \frac{5{,}7988}{103{,}392} = 0{,}0561.$$

$$b = \bar{y} - a \cdot \bar{x} = 2{,}0983 - 0{,}0561 \cdot 47{,}3333 = -0{,}5571.$$

Уравнение линейной парной регрессии запишется:

$$\hat{y} = 0{,}0561x - 0{,}5571. \qquad (*)$$

По полученному уравнению можно сделать вывод, что при увеличении урожая на один килограмм увеличится в среднем число поврежденных плодов на 0,0561%.

3. Оценить тесноту связи с помощью выборочных коэффициентов корреляции и детерминации.

Для вычисления линейного коэффициента корреляции воспользуемся формулой (52), предварительно вычислив средние квадратические отклонения результативного и факторного признаков:

$$\sigma_x = \sqrt{D_x} = \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{2343{,}8333 - (47{,}3333)^2} = \sqrt{103{,}39201} = 10{,}1682$$

$$\sigma_y = \sqrt{D_y} = \sqrt{\overline{y^2} - \bar{y}^2} = \sqrt{4{,}8607 - (2{,}0983)^2} = \sqrt{0{,}4578} = 0{,}6766.$$

$$r_{x/y} = a\frac{\sigma_x}{\sigma_y} = 0{,}0561 \cdot \frac{10{,}1682}{0{,}6766} = 0{,}8431 \text{ или}$$

$$r_{x/y} = \frac{\overline{x \cdot y} - \bar{x} \cdot \bar{y}}{\sigma_x \cdot \sigma_y} = \frac{105{,}1183 - 47{,}3333 \cdot 2{,}0983}{10{,}1682 \cdot 0{,}6766} = 0{,}8431 = 84{,}31\%.$$

По значению коэффициента корреляции согласно шкале Чеддока делаем вывод о том, что связь между числом поврежденных плодов Y и урожаем X – высокая. Коэффициент детерминации R^2 определим по формуле (60):

$$R^2=\frac{S_{объяс.}}{S_{общая}}=\frac{\sum_{i=1}^{n}(\hat{y}-\bar{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=\frac{3{,}9045}{5{,}4926}=0{,}7108=71{,}08\%;$$

$$R^2=1-\frac{\sum_{i=1}^{n}(y_i-\hat{y})^2}{\sum_{i=1}^{n}(y_i-\bar{y})^2}=1-\frac{1{,}5917}{5{,}4926}=0{,}7102;$$

$$R^2 = (r_{x/y})^2 = 0{,}8431^2 = 0{,}7108=71{,}08\%$$

Незначительное расхождение в вычислениях коэффициента детерминации наблюдается по формуле, что объясняется погрешностью округлений.

4. Проверить значимость выборочного коэффициента корреляции и параметров линейной модели при уровне значимости α = 0,05.

Для оценки значимости линейного коэффициента корреляции применим формулу (53) и вычислим расчетное значение t – критерия Стьюдента

$$t_r=\frac{|r_{x/y}|\cdot\sqrt{n-2}}{\sqrt{1-r_{x/y}^2}}=\frac{0{,}8431\cdot\sqrt{12-2}}{\sqrt{1-(0{,}8431)^2}}=4{,}958.$$

Сравним это значение с табличным (критическим) (таблица 2):

$$t_{крит.} = t(\alpha;\ r) = t\,(0{,}05;\ 10) = 2{,}23,$$

где число степеней свободы r определялось: $r = n - 2 = 12 - 2 = 10$.

Так как $t_r > t_{крит.}$, т.е. 4,958 > 2,23, то найденный *линейный коэффициент корреляции* является *статистически значимым*.

Проведем оценку значимости параметров a и b уравнения линейной парной регрессии, используя формулы (57) – (59). Для подсчета случайных ошибок параметров линейной парной модели и остаточной дисперсии составим расчетную таблицу, расширив ее столбцами для вычисления средней ошибки аппроксимации.

x	y	$\hat{y}$ = 0,0561x – - 0,5571.	$y_i - \hat{y}$	$\left\lvert \frac{y_i - \hat{y}}{y_i} \right\rvert$	$(y_i - \hat{y})^2$	$\hat{y} - \bar{y}$	$(\hat{y} - \bar{y})^2$
52	1,52	2,3601	-0,8401	0,5527	0,7058	0,2618	0,0685
46	1,86	2,0235	-0,1635	0,0879	0,0267	-0,0748	0,0056
38	1,25	1,5747	-0,3247	0,2598	0,1054	-0,5236	0,2742
37	2,06	1,5186	-0,5414	0,2628	0,2931	-0,5797	0,3361
45	1,84	1,9674	-0,1274	0,0692	0,0162	-0,1309	0,0171
39	1,47	1,6308	-0,1608	0,1094	0,0259	-0,4675	0,2186
55	2,83	2,5284	0,3016	0,1066	0,0909	0,4301	0,1849
59	3,08	2,7528	0,3272	0,1062	0,1071	0,6545	0,4284
32	1,17	1,2381	-0,0681	0,0582	0,0046	-0,8602	0,7399
40	2,15	1,6869	0,4631	0,2154	0,2145	-0,4114	0,1692
64	3,07	3,0333	0,0367	0,0119	0,0013	0,935	0,8742
61	2,88	2,8650	0,015	0,0052	0,0002	0,7667	0,5878
Сумма 568,00	25,18	**25,1796**		1,8451	1,5917		3,9045
Среднее **47,3333**	2,0983			0,1538			

$$S_{ост} = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-2}} = \sqrt{\frac{1,5917}{10}} = 0,399,$$

$$m_a = \frac{S_{ост}}{\sigma_x\sqrt{n-2}} = \frac{0,399}{10,1682 \cdot \sqrt{10}} = 0,0124, \quad t_a = \frac{a}{m_a} = \frac{0,0561}{0,0124} = 4,524,$$

$$m_b = \frac{S_{ост}}{\sqrt{n-2}} = \frac{0,399}{\sqrt{10}} = 0,1262, \quad t_b = \frac{b}{m_b} = \frac{-0,5571}{0,1262} = -4,4144.$$

Сравним эти расчетные значения t – критерия с табличным (критическим) (таблица 2):

$$t_{крит.} = t(\alpha; r) = t(0,05; 10) = 2,23,$$

$$t_{табл} < |t_a|, \text{ т.е } 2,23 < 4,524 \text{ и } t_{табл} < |t_b|, \text{ т.е } 2,23 < |-4,4144|.$$

Выполнение неравенств указывает на статистическую значимость параметров линейной парной регрессии.

5. Дать с помощью среднего коэффициента эластичности сравнительную оценку силы связи факторного признака с результативным.

Вычислим коэффициент эластичности по формуле (61):

$$Э_{y.x} = a\frac{\bar{x}}{\bar{y}} = 0{,}0561 \cdot \frac{47{,}3333}{2{,}0983} = 1{,}265.$$

По полученному значению коэффициента эластичности можно сделать следующий вывод: при увеличении урожая на 1 % число поврежденных плодов возрастет на 1, 265%.

6. Оценить качество уравнения с помощью средней ошибки аппроксимации.

Среднюю ошибку аппроксимации вычислим по формуле (62), используя вторую расчетную таблицу:

$$\bar{\varepsilon} = \frac{1}{n}\sum_{i=1}^{n}\left|\frac{y_i - \hat{y}}{y_i}\right| \cdot 100\% = \frac{1}{12} \cdot 1{,}8451 = 0{,}1538 = 15{,}38\%.$$

Средняя ошибка аппроксимации незначительно превосходит 15%. Это говорит о том, что полученную линейную парную регрессионную модель нецелесообразно использовать в дальнейших исследованиях.

7. Оценить статистическую надежность результатов регрессионного моделирования с помощью F-критерия Фишера.

Чтобы оценить значимость уравнения регрессии в целом, найдем расчетное значение F-критерия Фишера по формулам:

$$F_{факт} = \frac{\sum_{i=1}^{n}(\hat{y} - \bar{y})^2}{\sum_{i=1}^{n}(y_i - \hat{y})^2} \cdot (n-2) = \frac{3{,}9045}{1{,}5917} \cdot 10 = 24{,}53;$$

$$F_{факт} = \frac{R^2}{1 - R^2} \cdot (n-2) = \frac{0{,}7108}{1 - 0{,}7108} \cdot 10 = 24{,}57.$$

При уровне значимости $\alpha = 0{,}05$ для парной линейной регрессии количестве степеней свободы равно $k_1 = 1$, $k_2 = n - 2 = 12 - 2 = 10$. Тогда F-критерий Фишера табличный равен $F_{крит}(0{,}05;\ 1;\ 10) = 4{,}96$ (таблица 4).

$$F_{факт.} > F_{табл}$$

Значит уравнение значимо в целом.

8. Сделать выводы.

1. На основании поля корреляции выдвинута гипотеза о линейной форме связи между числом поврежденных плодов и урожаем.

2. По методу наименьших квадратов получена линейная парная модель регрессии $\hat{y}$ = 0,0561x – 0,5571, по которому видно, что при увеличении урожая на один килограмм увеличится в среднем число поврежденных плодов на 0,0561%.

3. Значения коэффициентов корреляции и детерминации соответственно равны $r_{x/y}$ = 0,8431, R^2 = 0,7108, что свидетельствует о, высокой связи между числом поврежденных плодов Y и урожаем X.

4. Проведя оценку статистической значимости линейного коэффициента корреляции и параметров линейной парной регрессии по критерию Стьюдента, выявлено, что полученные коэффициенты статистически значимы.

5. Коэффициент эластичности равен: $Э_{y.x} = 1{,}265$, значение которого можно интерпретировать следующим образом: при увеличении урожая на 1 % число поврежденных плодов возрастет на 1, 265%.

6. Средняя ошибка аппроксимации определилась значением $\bar{\varepsilon} = 0{,}1538 = 15{,}38\%$. Это говорит о существенности среднего отклонения расчетных данных от фактических, по которым построена линейная парная регрессионная модель.

7. Анализ статистической надежности результатов регрессионного моделирования с помощью F-критерия Фишера показал, что в целом построенное линейное парное уравнение регрессии адекватно и может быть использовано для принятия решений и осуществления прогнозов.

СТАТИСТИЧЕСКИЕ ТАБЛИЦЫ

Таблица 1 - Значения функции $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_0^x e^{-\frac{t^2}{2}} dt$

x	0	1	2	3	4
0,0	0,00000	0,00399	0,00798	0,01197	0,01595
0,1	0,03983	0,04338	0,04776	0,05172	0,05567
0,2	0,07926	0,08317	0,08706	0,09095	0,09483
0,3	0,11791	0,12172	0,12552	0,12930	0,13307
0,4	0,15542	0,15910	0,16276	0,16640	0,17003
0,5	0,19146	0,19497	0,19847	0,20194	0,20540
0,6	0,22575	0,22907	0,23237	0,23565	0,23891
0,7	0,25804	0,26115	0,26424	0,26730	0,27035
0,8	0,28814	0,29103	0,29389	0,29673	0,29955
0,9	0,31594	0,31859	0,32121	0,32381	0,32639
1,0	0,34134	0,34375	0,34614	0,34850	0,35083
1,1	0,36433	0,36650	0,36864	0,37076	0,37286
1,2	0,38493	0,38686	0,38877	0,39065	0,39251
1,3	0,40320	0,40490	0,40658	0,40824	0,40988
1,4	0,41924	0,42073	0,42220	0,42364	0,42507
1,5	0,43319	0,43448	0,43574	0,43699	0,43822
1,6	0,44520	0,44630	0,44738	0,44845	0,44950
1,7	0,45543	0,45637	0,45728	0,45818	0,45907
1,8	0,46407	0,46485	0,46582	0,46638	0,46712
1,9	0,47128	0,47193	0,47257	0,47320	0,47381
2,0	0,47725	0,47778	0,47831	0,47882	0,47932
2,1	0,48214	0,48257	0,48300	0,48341	0,48382
2,2	0,48610	0,48645	0,48679	0,48713	0,48745
2,3	0,48928	0,48956	0,48983	0,49010	0,49036
2,4	0,49180	0,49202	0,48224	0,49245	0,49266
2,5	0,49379	0,49396	0,49413	0,49430	0,49446
2,6	0,49534	0,49547	0,49560	0,49573	0,49585
2,7	0,49653	0,49664	0,49674	0,49683	0,49693
2,8	0,49744	0,49752	0,49760	0,49767	0,49774
2,9	0,49813	0,49819	0,49825	0,49831	0,49836
3,0	0,49865		3,10000	0,49903	3,20000
3,5	0,49977		3,60000	0,49984	3,70000
4,0	0,49996				

x	5	6	7	8	9
0,0	0,01994	0,02392	0,02790	0,03188	0,03586
0,1	0,05962	0,06356	0,06749	0,07142	0,07535
0,2	0,09871	0,10257	0,10642	0,11026	0,11409
0,3	0,13683	0,14058	0,14431	0,14803	0,15173
0,4	0,17364	0,17724	0,18082	0,18439	0,18793
0,5	0,20884	0,21226	0,21566	0,21904	0,22240
0,6	0,24215	0,24537	0,24857	0,25175	0,25490
0,7	0,27337	0,27637	0,27935	0,28230	0,28524
0,8	0,30234	0,30511	0,30785	0,31057	0,31327
0,9	0,32894	0,33147	0,33398	0,33646	0,33891
1,0	0,35314	0,35543	0,35769	0,35993	0,36214
1,1	0,37493	0,37698	0,37900	0,38100	0,38298
1,2	0,39435	0,39617	0,39796	0,39973	0,40147
1,3	0,41149	0,41309	0,41466	0,41621	0,41774
1,4	0,42647	0,42786	0,42922	0,43056	0,43189
1,5	0,43943	0,44062	0,441 79	0,44295	0,44408
1,6	0,45053	0,45154	0,45254	0,45352	0,45449
1,7	0,45994	0,46080	0,46164	0,46246	0,46327
1,8	0,46784	0,46856	0,46926	0,46995	0,47062
1,9	0,47441	0,47500	0,47558	0,47615	0,47670
2,0	0,47982	0,48030	0,48077	0,48124	0,48169
2,1	0,48422	0,48461	0,48500	0,48537	0,48574
2,2	0,48778	0,48809	0,48840	0,48870	0,48899
2,3	0,49061	0,49086	0,49111	0,49134	0,49158
2,4	0,49286	0,49305	0,49324	0,49343	0,49361
2,5	0,49461	0,49477	0,49492	0,49506	0,49520
2,6	0,495098	0,49609	0,49621	0,49632	0,49643
2,7	0,49702	0,49711	0,49720	0,49728	0,49736
2,8	0,49781	0,49788	0,49795	0,49801	0,49807
2,9	0,49841	0,49846	0,49851	0,49856	0,49861
3,0	0,49931	3,30000	0,49952	3,40000	0,49986
3,5	0,49989	3,80000	0,49993	3,90000	0,49995
4,0					

Таблица 2 - Критические точки распределения Стьюдента (t-критерия)

Число степеней свободы k	Уровень значимости α (двусторонняя критическая область)					
	0.10	0.05	0.02	0.01	0.002	0.001
1	6.313	12.706	31.82	63.656	318.31	636.62
2	2.920	4.303	6.964	9.924	22.327	31.599
3	2.353	3.182	4.540	5.840	10.214	12.924
4	2.132	2.776	3.746	4.604	7.173	8.610
5	2.015	2.571	3.649	4.032	5.893	6.863
6	1943	2.447'	3.142	3.707	5.207	5.958
7	1.895	2.365	2.998	3.499	4.785	5.408
8	1.860	2.306	2.896	3.355	4.501	5.041
9	1.833	2.262	2.821	3.249	4.297	4.780
10	1.812	2.228	2.764	3.169	4.144	4.587
11	1.796	2.201	2.718	3.105	4.024	4.437
12	1.782	2.179	2.680	3.085	3.929	4.321
13	1.771	2.160	2.650	3.112	3.852	4.220
14	1.761	2.145	2.625	2.976	3.787	4.140
15	1.753	2.131	2.603	2.947	3.732	4.072
16	1.746	2.120	2.583	2.920	3.686	4.015
17	1.740	2.110	2.567	2.898	3.646	3.965
18	1.734	2.101	2.551	2.878	3.611	3.922
19	1.729	2.093	2.539	2.861	3.579	3.883
20	1.725	2.086	2.528	2.845	3.552	3.849
21	1.721	2.080	2.521	2.837	3.534	3.824
22	1.717	2.074	2.514	2.824	3.513	3.790
23	1.714	2.069	2.506	2.815	3.501	3.775
24	1.711	2.064	2.498	2.802	3.475	3.742
25	1.708	2.060	2.490	2.792	3.452	3.723
26	1.706	2.056	2.485	2.784	3.446	3.718
27	1.703	2.052	2.471	2.771	3.422	3.697
28	1.701	2.048	2.469	2.769	3.407	3.669
29	1.699	2.045	2.467	2.758	3.401	3.658
30	1.697	2.042	2.462	2.754	3.395	3.645
40	1.684	2.021	2.427	2.708	3.312	3.575
60	1.671	2.000	2.395	2.665	3.234	3.460
120	1.658	1.980	2.362	2.627	3.172	3.378
∞	1.645	1.960	2.338	2.580	3.091	3.290
	0.05	0.025	0.01	0.005	0.001	0.0005
	Уровень значимости α (односторонняя критическая область)					

Таблица 3 - Критические точки распределения χ^2

Число степеней свободы, s	Уровень значимости α					
	0.01	0.025	0.05	0.95	0.975	0.99
1	6.63	5.02	3.84	0.039	0.0982	0.0157
2	9.21	7.38	5.99	0.103	0.0506	0.0201
3	11.3	9.35	7.81	0.352	0.216	0.115
4	13.3	11.1	9.49	0.711	0.484	0.297
5	15.1	12.8	11.1	1.15	0.831	0.554
6	16.8	14.4	12.6	1.64	1.24	0.872
7	18.5	16.0	14. 1	2.17	1.69	1.24
8	20.1	17.5	15.5	2.73	2.18	1.65
9	21.7	19.0	16.9	3.33	2.70	2.09
10	23.2	20.5	18.3	3.94	3.25	2.56
11	24.7	21.9	19.7	4.57	3.82	3.05
12	26.2	23.3	21.0	5.23	4.40	3.57
13	27.7	24.7	22.4	5.89	5.01	4.11
14	29.1	26.1	23.7	6.57	5.63	4.66
15	30.6	27.5	25.0	7.26	6.26	5.23
16	32.0	28.8	26.3	7.96	6.91	5.81
17	33.4	30.2	27.6	8.67	7.56	6.41
18	34.8	31.5	28.9	9.39	8.23	7.01
19	36.2	32.9	30.1	10.1	8.91	7.63
20	37.6	34.2	31.4	10.9	9.59	8.26
21	38.9	35.5	32.7	11.6	10.3	8.90
22	40.3	36.8	33.9	12.3	11.0	9.54
23	41.6	38.1	35.2	13.1	11.7	10.2
24	43.0	39.4	36.4	13.8	12.4	10.9
25	44.3	40.6	37.7	14.6	13.1	11.5
26	45.6	41.9	38.9	15.4	13.8	12.2
27	47.0	43.2	40.1	16.2	14.6	12.9
28	48.3	44.5	41.3	16.9	15.3	13.6
29	49.6	45.7	42.6	17.7	16.0	14.3
30	50.9	47.0	43.8	18.5	16.8	15.0

Таблица 4 - Критические точки распределения Фишера

(k_1 – число степеней свободы большей дисперсии, k_2 – число степеней свободы меньшей дисперсии)

Уровень значимости α = 0.01												
k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	11	12
1	4052	4999	5403	5625	5764	5889	5928	5981	6022	6056	6082	6106
2	98.49	99.01	90.17	99.25	99.33	99.30	99.34	99.36	99.36	99.40	99.41	99.42
3	34.12	30.81	29.46	28.71	28.24	27.91	27.67	27.49	27.34	27.23	27.13	27.05
4	21.20	18.00	16.69	15.98	15.52	15.21	14.98	14.80	14.66	14.54	14.45	14.37
5	16.26	13.27	12.06	11.39	10.97	10.67	10.45	10.27	10.15	10.05	9.96	9.893
6	13.74	10.92	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72
7	12.25	9.55	8.45	7.85	7.46	7.19	7.00	6.84	6.71	6.62	6.54	6.473
8	11.26	8.65	7.59	7.01	6.63	6.37	6.19	6.03	5.91	5.82	5.74	5.67
9	10.56	8.02	6.99	6.42	6.06	5.80	5.62	5.47	5.35	5.26	5.18	5.11
10	10.04	7.56	6.55	5.99	5.64	5.39	5.21	5.06	4.95	4.85	4.78	4.71
11	9.86	7.20	6.22	5.67	5.32	5.07	4.88	4.74	4.63	4.54	4.46	4.40
12	9.33	6.93	5.95	5.41	5.06	4.82	4.65	4.50	4.39	4.30	4.22	4.16
13	9,07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96
14	8.86	6.51	5.56	5.03	4.69	4.46	4.28	4.14	4.03	3.94	3.86	3.80
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.61	3.55
17	8.40	6.11	5.18	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.45

Продолжение таблицы 4

Уровень значимости α = 0.05												
$k_2 \backslash k_1$	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	18.5	19.0	19.16	19.25	19:30	19.33	19.36	19.37	19.38	1939	1940	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	694	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	9.99	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

Таблица 5 – Значения параметра $q = q(\gamma, n)$

γ / n	0,95	0,99	0,999
5	1,37	2,67	5,64
6	1,09	2,01	3,88
7	0,92	1,62	2,98
8	0,80	1,38	2,42
9	0,71	1,20	2,06
10	0,65	1,08	1,80
11	0,59	0,98	1,60
12	0,55	0,90	1,45
13	0,52	0,83	1,33
14	0,48	0,78	1,23
15	0,46	0,73	1,15
16	0,44	0,70	1,07
17	0,42	0,66	1,01
18	0,40	0,63	0,96
19	0,39	0,60	0,92
20	0,37	0,58	0,88
25	0,32	0,49	0,73
30	0,28	0,43	0,63
35	0,26	0,38	0,56
40	0,24	0,35	0,50
45	0,22	0,32	0,46
50	0,21	0,30	0,43
60	0,188	0,269	0,38
70	0,174	0,245	0,34
80	0,161	0,226	0,31
90	0,151	0,211	0,29
100	0,143	0,198	0,27
150	0,115	0,160	0,211
200	0,099	0,136	0,185
250	0,089	0,120	0,162

Библиографический список

1. Алибеков И. Ю. Теория вероятностей и математическая статистика в среде MATLAB. Учебное пособие. М.: Лань, 2019. - 184 с.

2. Гмурман В. Е. Теория вероятностей и математическая статистика. Учебник. М.: Юрайт, 2018. - 480 с.

3. Дрейпер, Н. Прикладной регрессионный анализ/ Н. Дрейпер, Г. Смит. – М.: Вильямс, 2016. - 912 с.

4. Есаулов, И.Г. Регрессионный анализ данных в пакете Mathcad: Учебное пособие/И.Г.Есаулов. – СПб.:Лань П, 2016. - 224 с.

5. Ивченко Г. И., Медведев Ю. И. Математическая статистика. Учебник. М.: Либроком, 2020. - 352 с.

6. Кобзарь, А.И. Прикладная математическая статистика. Для инженеров и научных работников / А.И. Кобзарь. - М.: ФИЗМАТЛИТ, 2012. - 816 с.

7. Кочетков Е.С. Теория вероятностей и математическая статистика: Учебник / Е.С. Кочетков, С.О. Смерчинская, В.В. Соколов. - 2-е изд., испр. и перераб. - М.: Форум: НИЦ ИНФРА-М, 2014. - 240 с

8. Сидняев, Н. И. Теория планирования эксперимента и анализ статистических данных : учеб. пособие для магистров / Н. И. Сидняев. – М. : Юрайт, 2012. – 399 с.

9. Соколов, Г.А. Введение в регрессионный анализ и планирование регрессионных экспериментов в экономике: Учебное пособие/Г.А. Соколов, Р,В, Сагитов. – М.:ИНФРА - М, 2012. – 202 с.

10. Чашкин, Ю.Р. Математическая статистика. Анализ и обработка данных: Учебное пособие / Ю.Р. Чашкин; Под ред. С.Н. Смоленский. - Рн/Д: Феникс, 2010. - 236 с.

11. Хамидуллин Р. Я. Теория вероятностей и математическая статистика. М.: Издательский дом Университета "Синергия", 2020. 276 с.

Интернет-ресурсы

12. http://www.znanium.com/bookread.php?book=447828, www.math.uah.edu/stat - Виртуальная лаборатория теории вероятностей и статистики

13. www.mathcs.carleton.edu/probweb/probweb.html - Коллекция Интернет-ресурсов по теории вероятностей и статистике, предназначенная для научных работников и преподавателей.

14. www.mathguide.de - Каталог математических ресурсов, упорядоченных по типу и тематике.

15. http://stattrek.com/Tables/ChiSquare.aspx - Обучающий сайт по статобработке.

Н.И. Овчинникова

Прикладная математика

Методические указания и контрольные задания

Прикладная математика. Методические указания и контрольные задания для студентов направления подготовки 21.04.02 - «Землеустройство и кадастры» – Иркутск: Изд-во Иркутский ГАУ им. А.А. Ежевского, 2021. - 67 с.

Компьютерный набор и верстка Овчинниковой Н.И.

Редактор Тесля В.И.

Лицензия ЛР № 070444 от 11.03.98 г.
Подписано к печати
Формат 60×84. Печ. 4,7 л. Тираж 50 экз.

664038, Иркутская обл., Иркутский р-он, п. Молодежный,
Иркутский ГАУ им. А.А. Ежевского