

МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Тульский государственный университет»

Институт прикладной математики и компьютерных наук
Кафедра «Вычислительная техника»

Утверждено на заседании кафедры
«Вычислительная техника»
«28 » января 2021 г., протокол № 8

Заведующий кафедрой


_____ А.Н. Ивутин

**МЕТОДИЧЕСКИЕ УКАЗАНИЯ
по выполнению контрольно-курсовой работы
по дисциплине (модулю)
«Математическая статистика»**

**основной профессиональной образовательной программы
высшего образования – программы бакалавриата**

по направлению подготовки
09.03.01 Информатика и вычислительная техника

с направленностью (профилем)
Электронно-вычислительные машины, комплексы, системы и сети

Формы обучения: очная, заочная

Идентификационный номер образовательной программы: 090301-02-21

Тула 2021 год

Разработчик(и) методических указаний

Набродова И.Н., доцент, к.т.н.

(ФИО, должность, ученая степень, ученое звание)



(подпись)

1. Цель и задачи выполнения контрольно-курсовой работы

Контрольно-курсовая работа выполняется с целью закрепления и проверки знаний по курсу «Математическая статистика».

Задачами контрольно-курсовой работы является: изучение практического применения статистических критериев: для решения каких задач они используются, при анализе каких выборок – зависимых или независимых, больших или малых, с известным распределением или неизвестным, при исследовании признаков, выраженных в различных шкалах измерения, изучение вычислений числовых характеристик равномерно распределенной, показательной распределенной и нормально распределенной непрерывной случайной величины: а) математическое ожидание $M(X)$; б) дисперсию $D(X)$; в) среднее квадратическое отклонение $\sigma(X)$; г) вероятность попадания в заданный интервал, получение навыков установления и оценки зависимости изучаемой величины Y от одной или нескольких других величин, изучение определений интервальных оценок с заданным уровнем доверия (доверительные интервалы) для неизвестных параметров распределений с использованием MSExcel (Ooo Calc), изучение аппроксимирования наборов наблюдений линейной функцией с использованием MSExcel (Ooo Calc), изучение проведения проверки статистических гипотез с использованием MSExcel (Ooo Calc).

2. Порядок выполнения контрольно-курсовой работы

Каждый студент получает задание на работу, состоящую из шести заданий. Для выполнения первого задания необходимо изучить практическое применение статистических критериев: для решения каких задач они используются, при анализе каких выборок – зависимых или независимых, больших или малых, с известным распределением или неизвестным, при исследовании признаков, выраженных в различных шкалах измерения (Задания в Приложении 1), для выполнения третьего задания необходимо изучить вычисления числовых характеристик равномерно распределенной, показательной распределенной и нормально распределенной непрерывной случайной величины: а) математическое ожидание $M(X)$; б) дисперсию $D(X)$; в) среднее квадратическое отклонение $\sigma(X)$; г) вероятность попадания в заданный интервал (Задания в Приложении 2), для выполнения четвертого задания необходимо получить навыки установления и оценки зависимости изучаемой величины Y от одной или нескольких других величин (Задания в Приложении 3), для выполнения шестого задания необходимо научиться определять интервальные оценки с заданным уровнем доверия (доверительные интервалы) для неизвестных параметров распределений с использованием MSExcel (Ooo Calc) (Задания в Приложении 4), для выполнения седьмого задания необходимо научиться аппроксимировать набор наблюдений линейной функцией с использованием MSExcel (Ooo Calc) (Задания в Приложении 5), для выполнения восьмого задания необходимо научиться производить проверку статистических гипотез с использованием MSExcel (Ooo Calc) (Задания в Приложении 6).

Теоретические сведения к заданию №1:

Методы математической статистики широко используются при анализе различных процессов и явлений. Если по результатам проведённых экспериментов требуется проверить некоторое предположение относительно генеральной совокупности и сделать обоснованный вывод, то используется статистическая проверка гипотез. Например, если сравниваются различные способы лечения, или разные варианты инвестиций, измерений, технологических процессов, рассматриваются вопросы об эффективности нового метода обучения, управления, о значимости математической модели и т.д. Практической реализации эксперимента предшествует этап, на котором исследователь должен чётко сформулировать предположение, подлежащее проверке.

Предположительное утверждение относительно генеральной совокупности, проверяемое по выборочным данным, называется статистической гипотезой. Далее осуществляется проверка фактического соответствия реальных результатов экспериментов предполагаемой гипотезе. Различают простую и сложную статистические гипотезы.

Простой называют гипотезу, содержащую только одно предположение.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез. Простая гипотеза, в отличие от сложной, полностью определяет теоретическую функцию распределения случайной величины. Например, гипотезы «вероятность появления события в схеме Бернулли равна $1/2$ », «закон распределения случайной величины – нормальный с параметрами $\mu=0$, $\sigma=1$ » – являются простыми, а гипотезы «вероятность появления события в схеме Бернулли заключена между $0,3$ и $0,6$ », «закон распределения не является нормальным» – сложными.

Проверяемую гипотезу обычно называют нулевой или основной и обозначают H_0 . Наряду с нулевой гипотезой H_0 рассматривают альтернативную, или конкурирующую, гипотезу H_1 , являющуюся логическим отрицанием H_0 . Нулевая и альтернативная гипотезы представляют собой две возможности выбора, осуществляемого в задачах проверки статистических гипотез. Например, для простой нулевой гипотезы $H_0: p = 1/2$ сложная альтернативная гипотеза может выглядеть таким образом $H_1: p \neq 1/2$ или так $H_1: p < 1/2$ или $H_1: p > 1/2$.

Правило, по которому принимается решение об отклонении или принятии основной гипотезы H_0 , называется критерием. Суть проверки статистической гипотезы заключается в том, что все выборочное пространство делится на две взаимодополняющие области: критическую область $S_{кр}$ (область неправдоподобно малых и/или неправдоподобно больших значений) и область допустимых значений $S_{кр}$ (область правдоподобных значений). В зависимости от вида альтернативной гипотезы H_1 различают односторонние (критическая область с одной стороны) и двухсторонние критерии (критических области две – «два хвоста распределения»). Затем по выборке $x_1, \dots, x_n$ определяется специально составленная выборочная характеристика – критическая статистика $\theta_{кр}$ ($x_1, x_2, \dots, x_n$), точное или приближенное распределение которой известно. Для этого известного распределения по специальным таблицам находят точки $\theta_{кр.н}$ и $\theta_{кр.в}$ (в случае двухсторонней критической области) или точка $\theta_{кр}$ (в случае

односторонней критической области), разделяющие критическую область $S_{кр}$ и область допустимых значений $S_{кр}$. Для одностороннего критерия область принятия основной гипотезы имеет ограничение только с одной стороны (сверху или снизу), соответственно требуется найти квантиль уровня $1-\alpha$, либо квантиль уровня α . Для двухстороннего критерия область принятия нулевой гипотезы имеет два ограничения – сверху (квантиль уровня $1-\alpha/2$) и снизу (квантиль уровня $\alpha/2$). Рассчитывается эмпирическое значение статистики $\theta_{эмп}$ подстановкой в $\theta_{кр}$ конкретных выборочных значений. Если $\theta_{эмп} \in S_{кр}$, то нулевая гипотеза отклоняется и принимается альтернативная гипотеза. Если же $\theta_{эмп} \notin S_{кр}$, то делается вывод о том, что нет оснований для отклонения нулевой гипотезы.

Так как исследователь работает с выборочными данными, которые попадают из генеральной совокупности случайным образом, то можно совершить ошибки (табл.1).

Таблица 1.

гипотеза H_0	<i>не отвергается</i>	<i>отвергается</i>
верна	правильное решение	ошибка 1-го рода
не верна	ошибка 2-го рода	правильное решение

Если на самом деле верной является нулевая гипотеза, а будет принята альтернативная гипотеза, то такая ошибка называется ошибкой первого рода. Вероятность $P(H_1/H_0)=\alpha$ допустить ошибку 1-го рода называется уровнем значимости критерия.

Ошибка второго рода – это принятие нулевой гипотезы в то время, когда на самом деле верной является альтернативная гипотеза. Вероятность допустить ошибку 2-го рода: $P(H_0/H_1) = \beta$.

При построении процедур проверки гипотез желательно минимизировать значения ошибок обоих родов, но на практике это невозможно: при фиксированном объёме выборки можно минимизировать лишь одну из величин α или β , другая при этом будет увеличиваться. Поэтому поступают таким образом: фиксируют α вероятность ошибки первого рода на определённом уровне (обычно для α используют стандартные значения, например, равные 0,05; 0,01), а вероятность ошибки второго рода – минимизируют.

Мощностью критерия называется вероятность не допустить ошибку 2-го рода $P(H_1/H_1)=1-P(H_0/H_1)=1-\beta$. Оптимальным критерием считается такой, у которого при заданном уровне значимости α достигается максимальное значение функции мощности критерия $1-\beta$ (задача Неймана-Пирсона).

Если использовать терминологию статистического контроля качества продукции, то вероятность α можно интерпретировать как «риск поставщика», т.е. вероятность по результатам выборочного контроля забраковать всю партию, удовлетворяющую стандарту; а вероятность β – «риск потребителя» – вероятность приёмки плохой продукции.

Процедура обоснованного сопоставления сформулированной гипотезы с имеющимися выборочными данными, осуществляемая с помощью статистического критерия, называется статистической проверкой гипотезы.

Разработаны различные статистические критерии проверки гипотез, но последовательность шагов действий укладывается в единую логическую схему:

1. Выдвигается основная гипотеза H_0 (в качестве нулевой гипотезы обычно используют то предположение, которое противоречит наблюдаемым фактам и, скорее всего, будет отклонено). Формулируется альтернативная гипотеза H_1 .

2. Задается уровень значимости критерия α . Логическим обоснованием величины α является вес потерь от ошибочного отклонения гипотезы H_0 (чем больше потери, тем меньшее значение α необходимо выбирать).

3. Определяется некоторая функция результатов наблюдений – критическая статистика – случайная величина, подчиняющаяся определенному закону распределения вероятностей. Вычисляется ее значение для выборочных данных – $\theta_{\text{эмп}}(x_1, x_2, \dots, x_n)$.

4. Из статистических таблиц распределения этой случайной величины находим нижнюю критическую точку для неправдоподобно малых значений случайной величины или верхнюю критическую точку для неправдоподобно больших значений случайной величины. Для одностороннего критерия область принятия основной гипотезы ограничена с одной стороны (и при этом площадь «хвоста» распределения равна α), для двухстороннего критерия область принятия основной гипотезы имеет два ограничения – снизу и сверху (при этом площадь каждого хвоста равна $\alpha/2$).

5. Найденное по таблицам критическое значение $\theta_{\text{кр}}$ сравнивается с расчетным $\theta_{\text{эмп}}(x_1, x_2, \dots, x_n)$. Если расчетное значение принадлежит области правдоподобных значений, то делается вывод «основная гипотеза H_0 не противоречит выборочным данным». В противном случае вывод такой – «основная гипотеза H_0 отклоняется с ошибкой первого рода α ».

По своему прикладному содержанию статистические гипотезы можно подразделить на несколько основных типов:

- о числовых значениях параметров;
- о равенстве числовых характеристик генеральных совокупностей;
- об однородности выборок;
- о согласии эмпирического распределения и выбранной модели;
- о стохастической независимости элементов выборки.

Задача 1. По результатам $n = 9$ замеров установлено, что выборочное среднее время (в секундах) изготовления детали $\bar{x} = 48$. Предполагая, что время изготовления – нормально распределенная случайная величина с дисперсией $\sigma^2 = 9$, рассмотреть гипотезу $H_0: a = 49$ против конкурирующей гипотезы $H_1: a \neq 49$. Доверительная вероятность $\gamma = 95\%$.

Решение:

Здесь рассматривается простая гипотеза $a = 49$ против сложной $a \neq 49$. Тогда $W = (-\infty, -t_{\text{кр}}) \cup (t_{\text{кр}}, +\infty)$ и $\bar{W} = (-t_{\text{кр}}, t_{\text{кр}})$. На уровне $\gamma = 0,95$ находим по таблицам нормального распределения $t_{\text{кр}} = 1,96$. Согласно таблице «Проверка

гипотез о числовых характеристиках» гипотеза принимается при $|\psi| = \left| \frac{(\bar{X}-a)\sqrt{n}}{\sigma} \right|$.
 Так как, $|\psi| = \left| \frac{(48-49)\sqrt{9}}{3} \right| = -1$ и $\psi \in \bar{W}$, то гипотеза H_0 принимается.

Теоретические сведения к заданию №2:

Случайная величина – величина, численное значение которой может меняться в зависимости от результата стохастического эксперимента.

Непрерывной назовём случайную величину, которая может принять любое значение из некоторого промежутка.

Распределение вероятностей непрерывной случайной величины x можно задавать либо **функцией распределения** $F(x)=P(\xi \leq x)$, либо её производной $f(x)=\frac{dF(x)}{dx}$, называемой **плотностью вероятности**.

Зная $F(x)$, можно найти плотность вероятности по формуле:

$$f(x)=F'(x),$$

а зная $f(x)$, найдём функцию распределения:

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

Для непрерывной случайной величины x вероятность попадания её в промежутки с концами a и b равна:

$$P(a < x < b) = P(a \leq x \leq b) = F(b) - F(a) = \int_a^b f(x)dx.$$

Причём $\int_{-\infty}^{\infty} f(x)dx = F(\infty) = 1$.

Пример. Задана следующая функция распределения:

$$F(x) = \begin{cases} 0, & x < 0 \\ x^2, & 0 \leq x \leq 1 \\ 1, & x > 1 \end{cases}$$

Найти плотность распределения.

Решение:

Зная $F(x)$, можно найти плотность вероятности по формуле:

$$f(x)=F'(x)=\begin{cases} 2x, & 0 \leq x \leq 1 \\ 0, & x \notin [0,1] \end{cases}$$

Равномерное распределение. Случайная величина x называется равномерно распределённой на $[a, b]$, если её плотность распределения $f(x)$ на $[a, b]$ постоянна, а вне $[a, b]$ равна 0:

$$f(x) = \begin{cases} 0, & x \notin [a, b] \\ \frac{1}{b-a}, & x \in [a, b] \end{cases}$$

Пример. Время ожидания автобуса (x) измеряется в минутах и распределено равномерно на отрезке $[0, 30]$. Определить, что ждать придётся не более 10 минут.

Решение:

$$a = 0, b = 30$$

$$f(x) = \frac{1}{b-a} = \frac{1}{30-0} = \frac{1}{30}$$

$$P(x < 10) = \int_0^{10} \frac{1}{30} dx = \frac{1}{30} x \Big|_0^{10} = \frac{1}{30} (10 - 0) = \frac{1}{3}$$

Пример. Задана плотность распределения: $f(x) = \begin{cases} h-2; 1 \leq x \leq 2 \\ 0, x \notin [1,2] \end{cases}$

Найти h .

$$\text{Решение: } f(x) = \begin{cases} 0, x \notin [a, b] \\ \frac{1}{b-a}, x \in [a, b] \end{cases} = \begin{cases} 0, x \notin [1,2] \\ \frac{1}{2-1}, x \in [1,2] \end{cases} = \begin{cases} 0, x \notin [1,2] \\ 1, x \in [1,2] \end{cases}$$

$$h-2=1 \Rightarrow h=3$$

Нормальное распределение. Случайная величина x называется нормально распределённой, если её плотность распределения $f(x)$ имеет вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-a)^2}{2\sigma^2}}, -\infty < x < \infty$$

$$\frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{(x-a)^2}{2\sigma^2}} dx = 1,$$

где a и σ – параметры нормального распределения, $\sigma > 0$.

В этом случае говорят, что x распределено нормально согласно закону $N(a, \sigma)$.

Если $a=0$ и $\sigma=1$, то $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ и эта функция обозначается через $\phi(x)$

и называется плотностью нормированного и центрированного нормального распределения. Функция распределения в этом случае обозначается через

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{y^2}{2}} dy.$$

Значения $\Phi(x)$ затабулированы, $\Phi(+\infty) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{y^2}{2}} dy = 1$.

Пример. Рост мужчины в Москве имеет нормальное распределение. Средний рост мужчины в Москве $a=175$ см, $\sigma=10$ см. Какова вероятность, что рост первого встречного мужчины будет в пределах 160-190 см?

Решение:

$$P(160 < x < 190) = \Phi\left(\frac{\beta-a}{\sigma}\right) - \Phi\left(\frac{\alpha-a}{\sigma}\right) = \Phi\left(\frac{200-175}{10}\right) - \Phi\left(\frac{160-175}{10}\right) = \Phi(2,5) - \Phi(-1,5) = \\ = 0,9938 - 0,0668 = 0,927$$

Правило трёх сигм. Случайная величина x распределена нормально $N(a, \sigma)$.

$$P(|x - a| < 3\sigma) = P(a - 3\sigma < x < a + 3\sigma) = \Phi\left(\frac{a - a + 3\sigma}{\sigma}\right) - \Phi\left(\frac{a - a - 3\sigma}{\sigma}\right) = \Phi(3) - \Phi(-3) = 0,9986 - 0,0014 = 0,9972$$

Пример. Рост мужчины в Москве имеет нормальное распределение. Средний рост мужчины в Москве $a=175$ см, $\sigma=10$ см. Какова вероятность, что рост первого встречного мужчины будет в пределах 145-205 см?

Решение:

$$P(145 < x < 205) = 0,9972$$

Правило двух сигм. Случайная величина x распределена нормально $N(a, \sigma)$.

$$P(|x - a| < 2\sigma) = P(a - 2\sigma < x < a + 2\sigma) = \Phi\left(\frac{a - a + 2\sigma}{\sigma}\right) - \Phi\left(\frac{a - a - 2\sigma}{\sigma}\right) = \Phi(2) - \Phi(-2) = 0,9772 - 0,0228 = 0,9544$$

Правило одной сигмы. Случайная величина x распределена нормально $N(a, \sigma)$.

$$P(|x - a| < \sigma) = P(a - \sigma < x < a + \sigma) = \Phi\left(\frac{a - a + \sigma}{\sigma}\right) - \Phi\left(\frac{a - a - \sigma}{\sigma}\right) = \Phi(1) - \Phi(-1) = 0,8413 - 0,1587 = 0,6826$$

Математическое ожидание. Математическим ожиданием дискретной случайной величины X , принимающей конечное число значений x_i с вероятностями p_i , называется сумма:

$$M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + \dots + x_n \cdot p_n. \quad (1)$$

Математическим ожиданием непрерывной случайной величины X называется интеграл от произведения ее значений x на плотность распределения вероятностей $f(x)$:

$$M(X) = \int_{-\infty}^{\infty} x f(x) dx. \quad (2)$$

Несобственный интеграл (2) предполагается абсолютно сходящимся (в противном случае говорят, что математическое ожидание $M(X)$ не существует). Математическое ожидание характеризует *среднее значение* случайной величины X . Его размерность совпадает с размерностью случайной величины.

Свойства математического ожидания:

$$1) M(c \cdot X) = c \cdot M(X), \quad c \in R,$$

$$2) M(X + Y) = M(X) + M(Y), \quad X, Y \in E,$$

$$3) M(X \cdot Y) = M(X) \cdot M(Y) \text{ для независимых случайных величин } X \text{ и } Y.$$

Дисперсия. Дисперсией случайной величины X называется число:

$$D(X) = M\{[X - M(X)]^2\} = M(X^2) - [M(X)]^2. \quad (3)$$

Дисперсия является *характеристикой рассеяния* значений случайной величины X относительно ее среднего значения $M(X)$. Размерность дисперсии равна размерности случайной величины в квадрате. Исходя из определений дисперсии и математического ожидания для дискретной случайной величины и для непрерывной случайной величины получим аналогичные выражения для дисперсии:

$$D(X) = M(X - m)^2 = \begin{cases} \sum_{i=1}^n (x_i - m)^2 p_i & \text{для дискретных случайных величин,} \\ \int_{-\infty}^{\infty} (x - m)^2 f(x) dx & \text{для непрерывных случайных величин.} \end{cases}$$

Здесь $m = M(X)$.

Свойства дисперсии:

$$1) D(c \cdot X) = c^2 \cdot D(X), \quad c \in R,$$

$$2) D(X + Y) = D(X) + D(Y) \text{ для независимых случайных величин } X \text{ и } Y.$$

Среднее квадратичное отклонение:

$$\sigma = \sqrt{D(X)}. \quad (4)$$

Так как размерность среднего квадратичного отклонения та же, что и у случайной величины, оно чаще, чем дисперсия, используется как мера рассеяния.

Моменты распределения. Понятия математического ожидания и дисперсии являются частными случаями более общего понятия для числовых характеристик случайных величин – *моментов распределения*. Моменты распределения случайной величины вводятся как математические ожидания некоторых простейших функций от случайной величины. Так, моментом порядка k относительно точки x_0 называется математическое ожидание $M(X - x_0)^k$. Моменты относительно начала координат $x = 0$ называются *начальными моментами* и обозначаются:

$$\alpha_k = M(X^k). \quad (5)$$

Начальный момент первого порядка есть центр распределения рассматриваемой случайной величины:

$$\alpha_1 = M(X) = m. \quad (6)$$

Моменты относительно центра распределения $x = m$ называются *центрными моментами* и обозначаются:

$$\mu_k = M(X - m)^k. \quad (7)$$

Из (7) следует, что центральный момент первого порядка всегда равен нулю:

$$\mu_1 = M(X - m) = M(X) - M(m) = m - m = 0. \quad (8)$$

Центральные моменты не зависят от начала отсчета значений случайной величины, так как при сдвиге на постоянное значение C ее центр распределения сдвигается на то же значение C , а отклонение от центра не меняется: $X - m = (X - C) - (m - C)$.

Теперь очевидно, что *дисперсия* – это *центральный момент второго порядка*:

$$D(X) = M(X - m)^2 = \mu_2. \quad (9)$$

Асимметрия. Центральный момент третьего порядка:

$$\mu_3 = M(X - m)^3 \quad (10)$$

служит для оценки *асимметрии распределения*. Если распределение симметрично относительно точки $x=m$, то центральный момент третьего порядка будет равен нулю (как и все центральные моменты нечетных порядков). Поэтому, если центральный момент третьего порядка отличен от нуля, то распределение не может быть симметричным. Величину асимметрии оценивают с помощью безразмерного *коэффициента асимметрии*:

$$C_s = \frac{\mu_3}{(\mu_2)^{3/2}} = \frac{M(X - m)^3}{\sqrt{[D(X)]^3}} = \frac{M(X - m)^3}{[\sigma(X)]^3}. \quad (11)$$

Знак коэффициента асимметрии (11) указывает на правостороннюю или левостороннюю асимметрию (рис. 1).

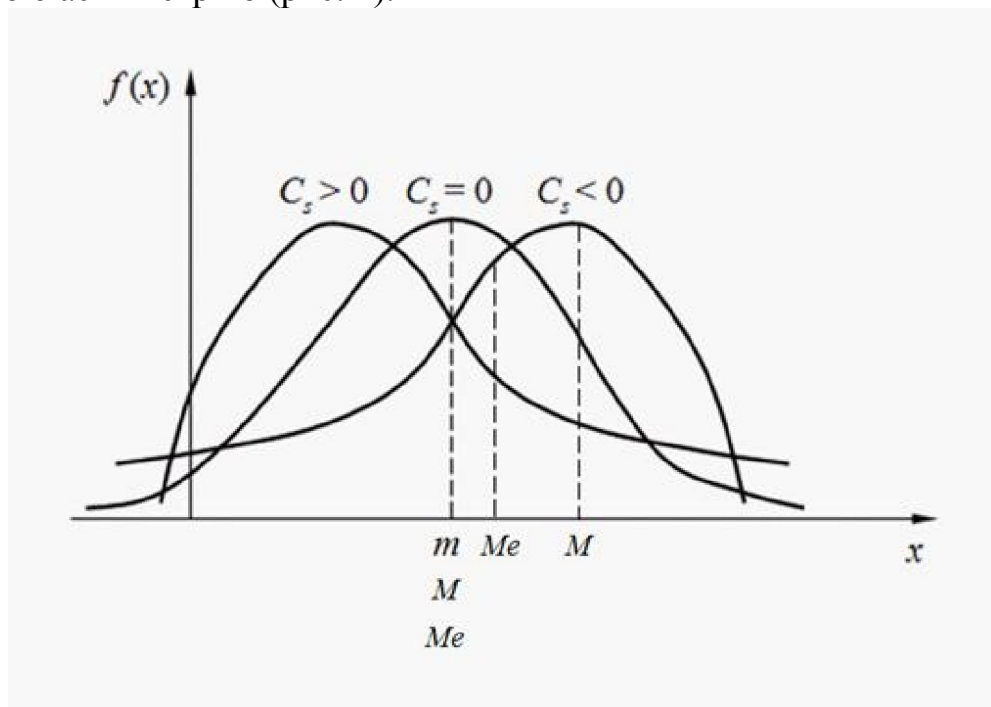


Рисунок 1 – Виды асимметрии распределений

Экссесс. Центральный момент четвертого порядка:

$$\mu_4 = M(X - m)^4 \quad (12)$$

служит для оценки так называемого *экссесса*, определяющего степень крутости (островершинности) кривой распределения вблизи центра распределения по отношению к кривой нормального распределения. Так как для нормального распределения $\mu_4 = 3\sigma^4$, то в качестве экссесса принимается величина:

$$E = \frac{\mu_4}{[\sigma(X)]^4} - 3. \quad (13)$$

На рис. 2 приведены примеры кривых распределения с различными значениями эксцесса. Для нормального распределения $E=0$. Кривые, более островершинные, чем нормальная, имеют положительный эксцесс, более плосковершинные – отрицательный.

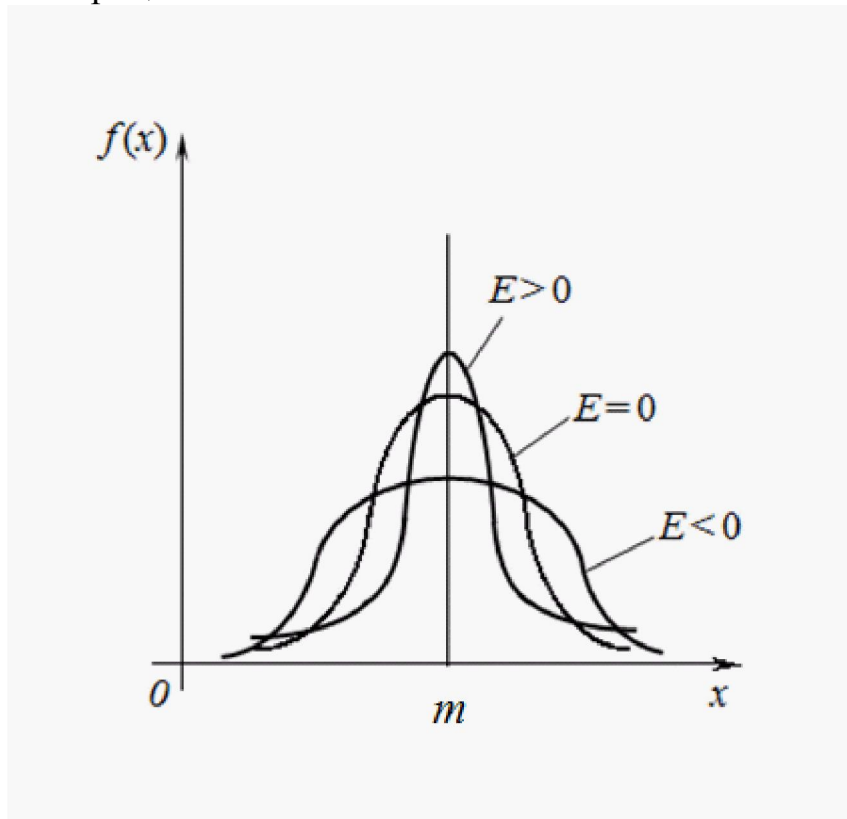


Рисунок 2 – Кривые распределения с различной степенью крутости (эксцессом)

Моменты более высоких порядков в инженерных приложениях математической статистики обычно не применяются.

Мода дискретной случайной величины – это ее наиболее вероятное значение. *Модой непрерывной* случайной величины называется ее значение, при котором плотность вероятности максимальна (рис.1). Если кривая распределения имеет один максимум, то распределение называется *унимодальным*. Если кривая распределения имеет более одного максимума, то распределение называется *полимодальным*. Иногда встречаются распределения, кривые которых имеют не максимум, а минимум. Такие распределения называются *антимодальными*. В общем случае мода и математическое ожидание случайной величины не совпадают. В частном случае, для *модального*, т.е. имеющего моду, симметричного распределения и при условии, что существует математическое ожидание, последнее совпадает с модой и центром симметрии распределения.

Медиана случайной величины X – это ее значение Me , для которого имеет место равенство: $P\{X < Me\} = P\{X > Me\}$, т.е. равновероятно, что случайная величина X окажется меньше или больше Me . Геометрически *медиана* – это абсцисса точки, в которой площадь под кривой распределения делится пополам (рис.1). В случае симметричного модального распределения медиана, мода и математическое ожидание совпадают.

Пример. Случайная величина X задана функцией распределения $F(x)$:

$$F(x) = \begin{cases} 0, & x \leq 0 \\ \frac{x}{4}, & 0 < x < 4 \\ 1, & x \geq 4 \end{cases}$$

Решение: Найдем плотность распределения $f(x)$, как производную от функции распределения $F(x)$:

$$f(x) = dF(x)/dx = 1/4$$

Математическое ожидание.

$$M[x] = \int_a^b xf(x)dx$$

$$M[x] = \int_0^4 x \frac{1}{4} dx = \frac{1}{8} x^2 \Big|_0^4 = \frac{1}{8} \cdot 4^2 - \frac{1}{8} \cdot 0 = 2$$

Дисперсия.

$$D[x] = \int_a^b x^2 f(x)dx - M[x]^2$$

$$D[x] = \int_0^4 x^2 \frac{1}{4} dx - 2^2 = \frac{1}{12} x^3 \Big|_0^4 - 2^2 = \frac{1}{12} \cdot 4^3 - \frac{1}{12} \cdot 0^3 - 2^2 = \frac{4}{3}$$

Среднеквадратическое отклонение.

$$s = \sqrt{D[x]} = \frac{2}{\sqrt{3}}$$

Теоретические сведения к заданию №3:

Линейная корреляция. Если обе линии регрессии Y на X и X на Y – прямые, то корреляцию называют *линейной*.

Выборочное уравнение прямой линии регрессии Y на X имеет вид:

$$\bar{y}_x - \bar{y} = r_b \frac{\sigma_y}{\sigma_x} (x - \bar{x}), \quad (1)$$

где $\bar{y}_x$ – условная средняя; $\bar{y}$ и $\bar{x}$ – выборочные средние признаков X и Y ; σ_x и σ_y – выборочные средние квадратические отклонения; r_b – выборочный коэффициент корреляции, причем:

$$r_b = (\sum n_{xy} xy - n \bar{x} \bar{y}) / (n \sigma_x \sigma_y).$$

Выборочное уравнение прямой линии регрессии X на Y имеет вид:

$$x_y - \bar{x} = r_b \frac{\sigma_x}{\sigma_y} (y - \bar{y}). \quad (2)$$

Если данные наблюдений над признаками X и Y заданы в виде корреляционной таблицы с равноотстоящими вариантами, то целесообразно перейти к условным вариантам:

$$u_i = (x_i - C_1)/h_1; v_j = (y_j - C_2)/h_2.$$

где C_1 – «ложный нуль» вариант X (новое начало отсчета); в качестве ложного нуля выгодно принять варианту, которая расположена примерно в середине вариационного ряда (условимся принимать в качестве ложного нуля варианту,

имеющую наибольшую частоту); h_1 – шаг, т.е. разность между двумя соседними вариантами X; C_2 – ложный нуль вариант Y; h_2 – шаг вариант Y.

В этом случае выборочный коэффициент корреляции:

$$r_B = (\sum n_{uv}uv - n\bar{u}\bar{v}) / (n\sigma_u\sigma_v),$$

Величины $\bar{u}$, $\bar{v}$, σ_u , σ_v , могут быть найдены либо методом произведений (при большом числе данных), либо непосредственно по формулам:

$$\bar{u} = (\sum n_u u) / n, \quad \bar{v} = (\sum n_v v) / n, \quad \sigma_u = \sqrt{\bar{u}^2 - (\bar{u})^2}, \quad \sigma_v = \sqrt{\bar{v}^2 - (\bar{v})^2}.$$

Зная эти величины, можно определить входящие в уравнения регрессии (1) и (2) величины по формулам:

$$\bar{x} = \bar{u}h_1 + C_1, \quad \bar{y} = \bar{v}h_2 + C_2, \quad \sigma_x = \sigma_u h_1, \quad \sigma_y = \sigma_v h_2.$$

Для оценки силы линейной корреляционной связи служит выборочный коэффициент корреляции r_B .

Для обоснованного суждения о наличии связи между количественными признаками следует проверить, значим ли выборочный коэффициент корреляции.

Пример. Найти выборочное уравнение прямой линии регрессии Y на X по данным, приведенным в корреляционной табл.

Y	X					n_y
	20	25	30	35	40	
16	4	6	-	-	-	10
26	-	8	10	-	-	18
36	-	-	32	3	9	44
46	-	-	4	12	6	22
56	-	-	-	1	5	6
n_x	4	14	46	16	20	$n=100$

Решение. Составим корреляционную табл. в условных вариантах, выбрав в качестве ложных нулей $C_1=30$ и $C_2=36$ (каждая из этих вариант расположена в середине соответствующего вариационного ряда).

v	u					n_v
	-2	-1	0	1	2	
-2	4	6	-	-	-	10
-1	-	8	10	-	-	18
0	-	-	32	3	9	44
1	-	-	4	12	6	22
2	-	-	-	1	5	6
n_u	4	14	46	16	20	100

Найдем $\bar{u}$ и $\bar{v}$:

$$\bar{u} = (\sum n_u u) / n = (4 \cdot (-2) + 14 \cdot (-1) + 46 \cdot 0 + 16 \cdot 1 + 20 \cdot 2) / 100 = 0,34;$$

$$\bar{v} = (\sum n_v v) / n = (10 \cdot (-2) + 18 \cdot (-1) + 44 \cdot 0 + 22 \cdot 1 + 6 \cdot 2) / 100 = -0,04.$$

Найдем вспомогательные величины $\bar{u}^2$ и $\bar{v}^2$:

$$\bar{u}^2 = (\sum n_u u^2) / n = (4 \cdot 4 + 14 \cdot 1 + 16 \cdot 1 + 20 \cdot 4) / 100 = 1,26;$$

$$\bar{v}^2 = (\sum n_v v^2) / n = (10 \cdot 4 + 18 \cdot 1 + 22 \cdot 1 + 6 \cdot 4) / 100 = 1,04.$$

Найдем σ_u и σ_v :

$$\sigma_u = \sqrt{\overline{u^2} - (\bar{u})^2} = \sqrt{1,26 - 0,34^2} = 1,07;$$

$$\sigma_v = \sqrt{\overline{v^2} - (\bar{v})^2} = \sqrt{1,04 - 0,04^2} = 1,02.$$

Найдем $\sum n_{uv}uv$, для чего составим расчетную таблицу.

Суммируя числа последнего столбца таблицы, находим

$$\sum_v v \cdot U = \sum n_{uv}uv = 82.$$

Для контроля вычислений находим сумму чисел последней строки:

$$\sum_u u \cdot V = \sum n_{uv}uv = 82.$$

Таблица

$\begin{matrix} u \\ v \end{matrix}$	-2	-1	0	1	2	$U = \sum n_{uv} \cdot u$	$v \cdot U$
-2	-8 4 -8	-6 6 -12				-14	28
-1		-8 8 -8	0 10 -10			-8	8
0			0 32 0	3 12 0	9 18 0	21	0
1			4 4 12	12 12 6	6 12 10	24	24
2				2 1 10	1 5 10	11	22
$V = \sum n_{uv} \cdot v$	-8	-20	-6	14	16		$\sum_v v \cdot U = 82$
$u \cdot V$	16	20	0	14	32	$\sum_u u \cdot V = 82$	← Контроль

Криволинейная корреляция. Если график регрессии – кривая линия, то корреляцию называют *криволинейной*. В частности, в случае параболической корреляции второго порядка выборочное уравнение регрессии X на Y имеет вид:

$$\bar{y}_x = Ax^2 + Bx + C$$

Неизвестные параметры A , B и C находят (например, методом Гаусса) из системы уравнений:

$$\begin{aligned}(\sum n_x x^4) A + (\sum n_x x^3) B + (\sum n_x x^2) C &= \sum n_x \bar{y}_x x^3, \\(\sum n_x x^3) A + (\sum n_x x^2) B + (\sum n_x x) C &= \sum n_x \bar{y}_x x, \\(\sum n_x x^2) A + (\sum n_x x) B + nC &= \sum n_x \bar{y}_x.\end{aligned}$$

Аналогично находится выборочное уравнение регрессии X на Y :

$$\bar{x}_y = A_1 y^2 + B_1 y + C_1.$$

Идя оценки силы корреляции Y на X служит выборочное корреляционное отношение (отношение межгруппового среднего квадратического отклонения к общему среднему квадратическому отклонению признака Y)

$$\eta_{yx} = \sigma_{\text{межгр}} / \sigma_{\text{общ}},$$

или в других обозначениях

$$\eta_{yx} = \sigma_{\bar{y}_x} / \sigma_y.$$

Здесь

$$\begin{aligned}\bar{y}_x &= \sqrt{D_{\text{межгр}}} = \sqrt{(\sum n_x (\bar{y}_x - \bar{y})^2) / n}, \quad \sigma_y = \sqrt{D_{\text{общ}}} = \\&= \sqrt{(\sum n_y (y - \bar{y})^2) / n},\end{aligned}$$

где n – объем выборки (сумма всех частот); n_x – частота значения x признака X ; n_y – частота значения y признака Y ; $\bar{y}_x$ – условная средняя признака Y ; $\bar{y}$ – общая средняя признака Y .

Аналогично определяется выборочное корреляционное отношение X к Y :

$$\eta_{xy} = \sigma_{\bar{x}_y} / \sigma_x.$$

Ранговая корреляция.

Выборочный коэффициент ранговой корреляции Спирмена.

Пусть выборка объема n содержит независимые объекты, которые обладают двумя качественными признаками: A и B . Под качественным подразумевают признак, который невозможно измерить точно, но он позволяет сравнивать объекты между собой и, следовательно, расположить их в порядке убывания или возрастания качества. Для определенности условимся располагать объекты в порядке ухудшения качества.

Расположим сначала объекты в порядке ухудшения качества по признаку A . Припишем объекту, стоящему на i -м месте, число – ранг x_i , равный порядковому номеру объекта: $x_i = i$. Затем расположим объекты в порядке убывания качества по признаку B и припишем каждому из них ранг (порядковый номер) y_i , причем (для удобства сравнения рангов) индекс i при y по-прежнему равен порядковому номеру объекта по признаку A .

В итоге получим две последовательности рангов:

по признаку A $x_1 \ x_2 \ \dots \ x_n$

по признаку B $y_1 \ y_2 \ \dots \ y_n$

Для оценки степени связи признаков А и В служат, в частности, коэффициенты ранговой корреляции Спирмена и Кендалла.

Выборочный коэффициент ранговой корреляции Спирмена находят по формуле:

$$\rho_B = 1 - \frac{6 \sum d_i^2}{n^3 - n},$$

где $d_i = x_i - y_i$, n – объем выборки.

Абсолютная величина коэффициента ранговой корреляции Спирмена не превышает единицы: $|\rho_B| \leq 1$.

Для обоснованного суждения о наличии связи между качественными признаками следует проверить, значим ли выборочный коэффициент ранговой корреляции Спирмена.

Пример. Знания десяти студентов проверены по двум тестам: А и В. Оценки по стобалльной системе оказались следующими (в первой строке указано количество баллов по тесту А, а во второй – по тесту В):

95	90	86	84	75	70	62	60	57	50
92	93	83	80	55	60	45	72	62	70

(*)

Найти выборочный коэффициент ранговой корреляции Спирмена между оценками по двум тестам.

Решение. Присвоим ранги x_i оценкам по тесту А. Эти оценки расположены в убывающем порядке, поэтому их ранги x_i равны порядковым номерам:

ранги x_i	1	2	3	4	5	6	7	8	9	10
оценки по тесту А	95	90	86	84	75	70	62	60	57	50

Присвоим ранги y_i оценкам по тесту В, для чего сначала расположим эти оценки в убывающем порядке и пронумеруем их:

1	2	3	4	5	6	7	8	9	10
93	92	83	80	72	70	62	60	55	45

(**)

Напомним, что индекс i при y должен быть равен порядковому номеру оценки студента по тесту А.

Найдем ранг y_1 . Индекс $i=1$ указывает, что рассматривается оценка студента, который занимает по тесту А в ряду (*) первое место (эта оценка равна 95); из условия видно, что по тесту В студент получил оценку 92, которая в (**) расположена на втором месте. Таким образом, ранг $y_1=2$.

Найдем ранг y_2 . Индекс $i=2$ указывает, что рассматривается оценка студента, который занимает по тесту А в ряду (*) второе место; из условия видно, что студент получил по тесту В оценку 93, которая в (**) расположена на первом месте. Таким образом, ранг $y_2=1$.

Аналогично найдем остальные ранги: $y_3=3$, $y_4=4$, $y_5=9$, $y_6=8$, $y_7=10$, $y_8=5$, $y_9=7$, $y_{10}=6$.

Выпишем последовательности рангов x_i и y_i :

x_i	1	2	3	4	5	6	7	8	9	10
y_i	2	1	3	4	9	8	10	5	7	6

Найдем разности рангов: $d_1=x_1-y_1=1-2=-1$; $d_2=x_2-y_2=2-1=1$. Аналогично получим: $d_3=0$, $d_4=0$, $d_5=-4$, $d_6=-2$, $d_7=-3$, $d_8=3$, $d_9=2$, $d_{10}=4$.

Вычислим сумму квадратов разностей рангов:

$$24 = 1 + 1 + 16 + 4 + 9 + 9 + 4 + 16 = 60.$$

Найдем искомый коэффициент ранговой корреляции Спирмена, учитывая, что $n=10$:

$$\rho_s = 1 - \frac{6 \sum d_i^2}{n^3 - n} = 1 - \frac{6 \cdot 60}{10^3 - 10} = 0,64.$$

Итак, $\rho_s = 0,64$.

Выборочный коэффициент ранговой корреляции Кендалла.

Можно оценивать связь между двумя качественными признаками, используя коэффициент ранговой корреляции Кендалла. Пусть ранги объектов выборки объема n :

по признаку А $x_1 x_2 \dots x_n$

по признаку В $y_1 y_2 \dots y_n$

Допустим, что справа от y_1 имеется R_1 рангов, больших y_1 ; справа от y_2 имеется R_2 рангов, больших y_2 ; справа от y_{n-1} имеется R_{n-1} рангов, больших y_{n-1} . Введем обозначение суммы рангов:

$$R = R_1 + R_2 + \dots + R_{n-1}$$

Выборочный коэффициент ранговой корреляции Кендалла находят по формуле:

$$\tau_b = \frac{4R}{n(n-1)} - 1,$$

где n – объем выборки, R – сумма рангов R_i ($i = 1, 2, \dots, n-1$).

Абсолютная величина коэффициента ранговой корреляции Кендалла не превышает единицы: $|\tau_b| \leq 1$.

Для обоснованного суждения о наличии связи между качественными признаками следует проверить, значим ли выборочный коэффициент ранговой корреляции Кендалла.

Пример. Знания 10 студентов проверены по двум тестам: А и В. Оценки по стобалльной системе оказались следующими (в первой строке указано количество баллов по тесту А, а во второй – по тесту В):

95 90 86 84 75 70 62 60 57 50

92 93 83 80 55 60 45 72 62 70

Найти выборочный коэффициент ранговой корреляции Кендалла между оценками по двум тестам.

Решение. При решении задачи 540, условие которой совпадает с условием настоящей задачи, были получены две последовательности рангов (в первой строке приведены ранги по тесту А, во второй – по тесту В):

x_i 1 2 3 4 5 6 7 8 9 10

y_i 2 1 3 4 9 8 10 5 7 6

Справа от $y_1=2$ имеется $R_1=8$ рангов (3, 4, 9, 8, 10, 5, 7, 6), больших $y_1=2$; справа от $y_2=1$ имеется $R_2=8$ рангов, больших $y_2=1$. Аналогично получим: $R_3=7$, $R_4=6$, $R_5=1$, $R_6=1$, $R_7=0$, $R_8=2$, $R_9=0$. Следовательно, сумма рангов

$$R = R_1 + R_2 + \dots + R_9 = 8 + 8 + 7 + 6 + 1 + 1 + 2 = 33.$$

Найдем коэффициент ранговой корреляции Кендалла, учитывая, что $R=33$, $n=10$:

$$\tau_B = \frac{4R}{n(n-1)} - 1 = \frac{4 \cdot 33}{10 \cdot 9} - 1 = 0,47.$$

Итак, $\tau_B = 0,47$.

Теоретические сведения к заданию №4:

В ряде задач требуется не только найти для параметра θ подходящую оценку, но и указать, к каким ошибкам может привести замена параметра его оценкой. Другими словами, требуется оценить точность и надежность оценки. Такого рода задачи особенно актуальны при малом числе наблюдений, когда точечная оценка в значительной мере случайна и приближенная замена этой оценкой истинного значения параметра θ может привести к серьезным ошибкам.

Для определения точности оценки θ в математической статистике пользуются доверительными интервалами, а для определения надежности – доверительными вероятностями.

Интервал (θ_1, θ_2) называется α – доверительным интервалом или доверительным интервалом с доверительной вероятностью $1-\alpha$, если

$$P(\theta_1 \leq x \leq \theta_2) = 1 - \alpha.$$

Рассмотрим задачу построения доверительных интервалов для неизвестных математического ожидания и дисперсии. Как известно, α – доверительный интервал для математического ожидания а выглядит следующим образом:

$$\bar{x} - \frac{\Delta_{n-1, \alpha} s_1}{\sqrt{n}} \leq a \leq \bar{x} + \frac{\Delta_{n-1, \alpha} s_1}{\sqrt{n}},$$

где s_1 – среднееквадратическое отклонение (корень квадратный из исправленной выборочной дисперсии), $\Delta_{n-1, \alpha}$ находится из таблицы для вероятностей $P(t_{n-1} > \Delta_{n-1, \alpha}) = \alpha$ распределения t_{n-1} (распределение Стьюдента с $n-1$ степенью свободы).

Вводим в первый столбец, например, ячейки A1...A25 исходные данные. Задаем уровень значимости $\alpha=0.05$. Далее для получения результатов подписываем ячейки, как на рис. 1.

	F	G	H	I	
1		Уровень значимости		0,05	
2		Интервал	Левая граница	Правая граница	
3		Матожидание			
4		Дисперсия			

Рисунок 1 – Пример подписи ячеек для получения результатов

$$\frac{\Delta_{n-1, \alpha} s_1}{\sqrt{n}}$$

Для вычисления величины $\frac{\Delta_{n-1, \alpha} s_1}{\sqrt{n}}$ служит функция ДОВЕРИТ категории СТАТИСТИЧЕСКИЕ с тремя параметрами «Альфа» – уровень значимости α , «Станд. откл» – s_1 – среднееквадратическое отклонение (корень квадратный из исправленной выборочной дисперсии), «Размер» – объем выборки n .

Таким образом, вводим в ячейку H3 функцию:

=СРЗНАЧ(A1:A25)-ДОВЕРИТ(П1;СТАНДОТКЛОН(A1:A25);25)

а в ячейку I3 функцию:

=СРЗНАЧ(A1:A25)+ДОВЕРИТ(П1;СТАНДОТКЛОН(A1:A25);25).

Как известно, α – доверительный интервал для дисперсии σ^2 выглядит следующим образом:

$$\partial_{n-1, \frac{\alpha}{2}}^{-1} \sum_{i=1}^n (x_i - \bar{x})^2 \leq \sigma^2 \leq \partial_{n-1, 1-\frac{\alpha}{2}}^{-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

Нам потребуется функция ХИ2ОБР (категория СТАТИСТИЧЕСКИЕ), которая вычисляет обратное значение $\chi_{n-1,p}^2$ односторонней вероятности распределения хи-квадрат $P(\chi_{n-1}^2 > \chi_{n-1,p}^2) = p$.

$$P\left(\chi_{n-1}^2 > \partial_{n-1, \frac{\alpha}{2}}\right) = \frac{\alpha}{2}, P\left(\chi_{n-1}^2 > \partial_{n-1, 1-\frac{\alpha}{2}}\right) = 1 - \frac{\alpha}{2}.$$

В данном конкретном случае

ХИ2ОБР имеет два параметра: первый «Вероятность» содержит доверительную вероятность соответственно $\alpha/2$ и $1-\alpha/2$, второй – степень свободы $n-1$.

Вводим в соответствии с формулой для доверительного интервала σ^2 в ячейку H4 запись:

=ДИСП(A1:A25)*24/ХИ2ОБР(0,025;24),

а в ячейку I4 запись:

=ДИСП(A1:A25)*24/ХИ2ОБР(0,975;24).

Получаем значения границ доверительных интервалов для σ^2 .

Теоретические сведения к заданию №5:

Цель регрессионного анализа состоит в определении общего вида уравнения регрессии, построении оценок неизвестных параметров, входящих в уравнение регрессии, и проверке статистических гипотез о регрессии. В зависимости от формы связи между переменными различают линейную и нелинейную регрессию. Наиболее простым является случай, когда регрессия линейна.

Рассмотрим задачу наилучшей аппроксимации набора наблюдений $x_i, y_i, i = \overline{1, n}$ линейной функцией $f(X) = a + bX$ в смысле минимизации функционала

$$F = \sum_{i=1}^n (y_i - (a + bx_i))^2.$$

Запишем необходимые условия экстремума:

$$\frac{\partial F}{\partial a} = -2 \sum_{i=1}^n (y_i - a - bx_i) = 0$$

$$\frac{\partial F}{\partial b} = -2 \sum_{i=1}^n x_i (y_i - a - bx_i) = 0$$

или

$$\sum_{i=1}^n (y_i - a - bx_i) = 0$$

$$\sum_{i=1}^n x_i (y_i - a - bx_i) = 0.$$

Раскрыв скобки, получим:

$$an + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i$$

$$a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i.$$

Решая систему уравнений, находим неизвестные a и b :

$$b = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2},$$

$$a = \frac{1}{n} \sum_{i=1}^n y_i - \frac{b}{n} \sum_{i=1}^n x_i.$$

Добавим к постановке задачи некоторые статистические данные и запишем линейное регрессионное уравнение в виде:

$$Y_i = a + bX_i + \varepsilon_i, i = \overline{1, n},$$

где X_i – неслучайная (детерминированная) величина, Y_i , ε_i – случайные величины, ε_i – ошибки регрессии.

Основные гипотезы:

1. $Y_i = a + bX_i + \varepsilon_i, i = \overline{1, n}$ – спецификация модели.
2. X_i – детерминированная величина; вектор $(X_1, X_2, \dots, X_n)$ не коллинеарен вектору $(1, 1, \dots, 1)$.

3. $M(\varepsilon_i) = 0, D(\varepsilon_i) = \sigma^2, i = \overline{1, n}, M(\varepsilon_i, \varepsilon_j) = 0, i \neq j$.

Часто добавляется условие:

4. ε_i – нормально распределенная случайная величина, $M(\varepsilon_i) = 0, D(\varepsilon_i) = \sigma^2$.

Как утверждает теорема Гаусса-Маркова, в этих предположениях оценки неизвестных параметров модели

$$b'' = \frac{n \sum_{i=1}^n X_i Y_i - \left(\sum_{i=1}^n X_i \right) \left(\sum_{i=1}^n Y_i \right)}{n \sum_{i=1}^n X_i^2 - \left(\sum_{i=1}^n X_i \right)^2}$$

$$a'' = \frac{1}{n} \sum_{i=1}^n Y_i - \frac{\hat{b}}{n} \sum_{i=1}^n X_i,$$

полученные по МНК, имеют наименьшую дисперсию в классе всех линейных несмещенных оценок.

Непосредственно из 1) - 4) следует, что Y_i – нормально распределенная случайная величина, $M(Y_i) = a + bX_i$, $D(Y_i) = \sigma^2$.

Нетрудно

проверить,

что

$$M(b'') = b, M(a'') = a, D(b'') = S_b^2 = \frac{\sigma^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2}, D(a'') = S_a^2 = \sigma^2 \frac{\sum_{i=1}^n x_i^2}{n \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)} \quad \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i,$$

поэтому

$$a'' \in N \left(a, \sigma^2 \frac{\sum_{i=1}^n x_i^2}{n \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)} \right), b'' \in N \left(b, \frac{\sigma^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \right).$$

Обозначим через $e_i, i = \overline{1, n}$ разницу между действительным значением переменной Y и модельным значением этой переменной, то есть

$$e_i = Y_i - a - bX_i, i = \overline{1, n}.$$

Несмещенной оценкой дисперсии ошибок σ^2 является:

$$S^2 = \frac{1}{n-2} \sum_{i=1}^n e_i^2.$$

Нетрудно показать, что S^2 независима с a'' и b'' , а $\frac{(n-2)S^2}{\sigma^2} \in \chi_{n-2}^2$.

Построим статистику для проверки гипотезы $H_0: b = b_0$ против альтернативной гипотезы $H_1: b \neq b_0$.

$$b'' - b \in N \left(0, \frac{\sigma^2}{\sum_{i=1}^n x_i^2 - n\bar{x}^2} \right) \in N(0, \sigma_b^2), \text{ то } \frac{b'' - b}{\sigma_b} \in N(0, 1).$$

Поскольку

из условия

$$\frac{(n-2)S^2}{\sigma^2} \in \chi_{n-2}^2, \text{ следует что } \frac{\frac{b'' - b}{\sigma_b}}{\sqrt{\frac{(n-2)S^2}{\sigma^2(n-2)}}} = \frac{b'' - b}{\frac{\sigma_b S}{\sigma}} = \frac{b'' - b}{S_b} \in t_{n-2} \quad (\text{распределению})$$

Стьюдента с $n-2$ степенями свободы).

Таким образом, для проверки гипотезы $H_0: b = b_0$ против альтернативной

гипотезы $H_1: b \neq b_0$ будет использоваться статистика $\frac{b'' - b}{S_b}$.

Построим доверительный интервал для b , используя распределение t_{n-2} и его двусторонние квантили $t_{n-2,\alpha}$, которые находятся из таблицы для вероятностей $P(|t_{n-2}| \leq t_{n-2,\alpha}) = 1 - \alpha$, или $P(|t_{n-2}| > t_{n-2,\alpha}) = \alpha$:

$$P\left(-t_{n-2,\alpha} \leq \frac{b'' - b}{S_b} \leq t_{n-2,\alpha}\right) = 1 - \alpha,$$

откуда следует $b'' - t_{n-2,\alpha} S_b \leq b < b'' + t_{n-2,\alpha} S_b$.

Если b_0 принадлежит отрезку $[b'' - t_{n-2,\alpha} S_b, b'' + t_{n-2,\alpha} S_b]$, то принимается гипотеза H_0 , в противном случае принимается гипотеза H_1 .

Если требуется проверить наличие связи между переменными X и Y , то используется статистика b''/S_b , тем самым проверяется равенство нулю коэффициента b . Если в границы построенного при этом доверительного интервала попадает ноль, (то есть нижняя граница доверительного интервала отрицательна, а верхняя положительна), то коэффициент b принимается равным нулю и делается вывод об отсутствии связи между переменными X и Y . Другими

словами, при $\left|\frac{b''}{S_b}\right| > \hat{t}_{n-2,\alpha}$ делается вывод о достоверной связи между переменными X

и Y , при $-t_{n-2,\alpha} \leq \frac{b''}{S_b} \leq t_{n-2,\alpha}$ делается вывод об ее отсутствии.

Можно показать, что

$$\frac{a'' - a}{S_a} \in t_{n-2}$$

и использовать эту статистику для проверки аналогичных гипотез относительно коэффициента a .

$$\frac{Y - (a + bX)}{\sigma}$$

Рассмотрим статистику $\frac{Y - (a + bX)}{\sigma}$, которая принадлежит стандартному нормальному распределению $-N(0,1)$. При известной σ^2 (дисперсии ошибок) можно было бы использовать $N(0,1)$ для прогнозирования значений Y в виде доверительных интервалов.

Поскольку σ^2 неизвестно, то будем использовать ее оценку S^2 , для которой

известно, что $\frac{(n-2)S^2}{\sigma^2} \in \chi_{n-2}^2$.

Таким образом,

$$\frac{\frac{Y - (a + bX)}{\sigma}}{\frac{S}{\sigma}} = \frac{Y - (a + bX)}{S} \in t_{n-2}$$

и используется для построения доверительных интервалов с целью прогнозирования значений Y :

$$a + bX - t_{n-2,\alpha} S \leq Y \leq a + bX + t_{n-2,\alpha} S.$$

Количественным показателем качества построенной линейной модели является коэффициент детерминации

$$R^2 = 1 - \frac{\sum_{i=1}^n (\tilde{y}_i - y_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}.$$

Коэффициент детерминации показывает, какая доля общей дисперсии Y объясняется уравнением регрессии:

$$0 \leq R_2 \leq 1.$$

Чем ближе R_2 к 1, тем лучше построенная регрессионная модель согласуется с исходными данными.

Для построения регрессии в Excel, создаем файл исходных входных и выходных данных и начинаем с построения корреляционного поля, позволяющего визуализировать наличие связи между этими данными.

Выбираем меню ВСТАВКА/ДИАГРАММА, тип диаграммы: ТОЧЕЧНАЯ вид: ТОЧЕЧНАЯ ДИАГРАММА. Нажимаем кнопку ДАЛЕЕ. В появившемся диалоговом окне указываем диапазон значений и расположение данных: В СТОЛБЦАХ. Нажимаем кнопку ДАЛЕЕ. В следующем диалоговом окне указываем название диаграммы, наименование осей. Нажимаем ДАЛЕЕ и ГОТОВО. Построенная таким образом диаграмма рассеяния представляет собой совокупность пар точек, абсциссами которых являются значения переменной X , а ординатами значения переменной Y .

В меню СЕРВИС выбираем АНАЛИЗ ДАННЫХ и РЕГРЕССИЯ. Указываем входной интервал Y (для примера A2: A26) и входной интервал X (для примера B2: B26), а также параметры вывода, остатки, нормальную вероятность как показано на рис. 1.

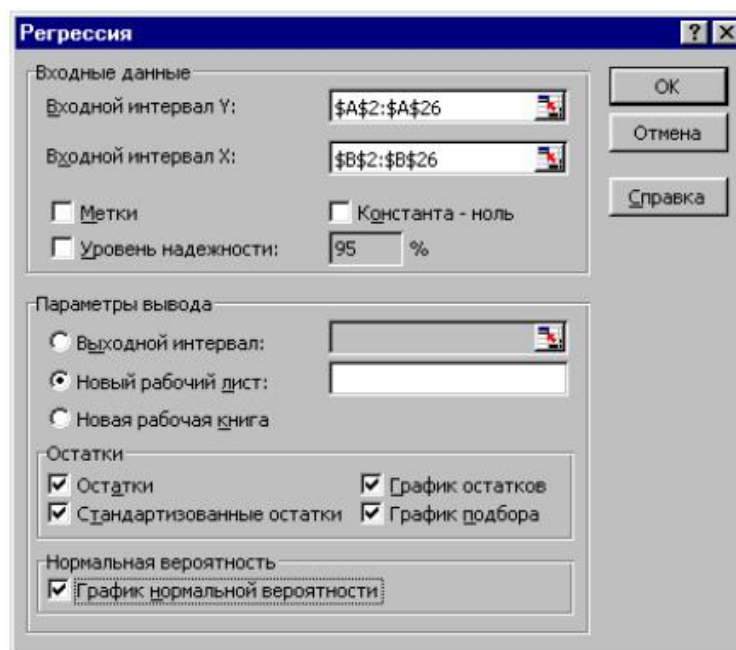


Рисунок 1 – Диалоговое окно Регрессия

В диалоговом окне задаются следующие параметры:

Входной интервал Y – диапазон ячеек, содержащий данные результативного признака;

Входной интервал X – диапазон ячеек, содержащий данные факторного признака;

Метки – флажок, который указывает, содержит ли первая строка названия столбцов или нет;

Константа-ноль – данный флажок необходимо установить, чтобы линия регрессии прошла через начало координат;

Уровень надежности – этот флажок необходимо использовать, если требуется уровень надежности отличный от 95%, принятый по умолчанию;

Выходной интервал – верхняя левая ячейка интервала, в который будут помещаться результаты вычислений.

Excel автоматически сгенерирует результаты по регрессионной статистике. Ниже в качестве примеров приведены возможные результаты и их расшифровки.

Регрессионная статистика

Множественный R 0,969525973

R² квадрат 0,939980612

Нормированный R² квадрат 0,935363736

Стандартная ошибка 14,22893673

Наблюдения 15.

Полученное значение коэффициента детерминации говорит об очень хорошей согласованности построенной регрессионной модели и исходных данных (соответственно об очень хорошей связи исследуемых факторов X и Y).

Результаты дисперсионного анализа будут представлены в виде:

Дисперсионный анализ					
	df	SS	MS	F	Значимость
Регрессия	1	41220,72106	41220,72106	203,5966782	2,55346E - 09
Остаток	13	2632,014326	202,4626405		
Итого	14	43852,73538			

	Коэффициенты	Стандартная ошибка	t-статистика	P-Значение	Нижние 95%	Верхние 95%
Y – пересечение	4,746	7,003	0,678	0,510	□10,384	19,876
Переменная X 1	9,595	0,672	14,269	0,000	8,142	11,048

df + ! степени свободы (degree of freedom);

SS + ! сумма квадратов отклонений (Sum of squares);

MS + ! средний квадрат отклонения (Mean square);

F + ! отношение дисперсий (факторной к остаточной).

$$F = \frac{S^2_{\text{факт}}}{S^2_{\text{ост}}} = \frac{\sum_{i=1}^n (\tilde{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \tilde{y}_i)^2}.$$

Значимость F – критическое значение квантиля распределения Фишера, которое используется для проверки нулевой гипотезы, состоящей в том, что факторная и остаточная дисперсии равны. По сути дела, нулевая гипотеза означает, что на результативный признак Y в равной степени влияют и независимая (факторная) переменная X и необъясненные факторы. В таком случае уравнение регрессии не значимо. Чтобы уравнение регрессии было значимым необходимо, чтобы факторная дисперсия превышала остаточную дисперсию в несколько раз.

В примере, приведенном выше, F больше, чем Значимость F (критическое значение), значит регрессионная модель адекватна.

Регрессионная сумма $SS=41220,72106$ (объясненная регрессией) намного больше остаточной $SS=2632,014326$ (не объясненной регрессией, вызванной случайными факторами), что тоже говорит о хорошей регрессии.

Коэффициенты – значения коэффициентов;

Стандартная ошибка – стандартная ошибка коэффициентов;

t-статистика – значение статистики критерия;

P-значение – уровень значимости отклонения гипотезы равенства коэффициента нулю (вероятность принять равенство коэффициента нулю);

Нижние 95% – нижняя граница доверительного интервала, в котором находится значение коэффициента;

Верхние 95% – верхняя граница доверительного интервала, в котором находится значение коэффициента.

Приведенные в качестве примера результаты позволяют проверить значимость коэффициентов регрессии: свободного члена и коэффициента при переменной X . Значение коэффициента при X 9,595 больше, чем его стандартная ошибка. К тому же этот коэффициент является значимым, о чем можно судить по значениям показателя P-значение в таблице, которые меньше заданного уровня значимости $\alpha=0,05$. Для свободного члена ситуация диаметрально противоположная. В построенный для него доверительный интервал попадает ноль, что говорит о том, что он незначим и может быть принят равным нулю.

Есть возможность вывести таблицу стандартных и простых остатков, где для каждого значения ряда выводится предсказанное значение, с которым сопоставляется остаток, представляющий разность между прогнозным и реальным значением.

Простым и наглядным способом проверки удовлетворительности регрессионной модели является графическое представление отклонений, которое Excel представляет в виде графика остатков. Если регрессионная модель близка к реальной зависимости, то отклонения будут носить случайный характер и их сумма будет близка к нулю. Если необходимо получить дополнительную

информацию и графики остатков, установите соответствующие флажки в диалоговом окне.

Теоретические сведения к заданию №6:

Статистической гипотезой называют гипотезу о виде неизвестного распределения генеральной совокупности или о параметрах известных распределений.

Нулевой (основной) называют выдвинутую гипотезу H_0 . Конкурирующей (альтернативной) называют гипотезу H_1 , которая противоречит нулевой.

Пример. Пусть H_0 заключается в том, что математическое ожидание генеральной совокупности $a=3$. Тогда возможные варианты H_1 : а) $a \neq 3$; б) $a > 3$; в) $a < 3$.

Проверка гипотез

Для проверки нулевой гипотезы используют специально подобранную случайную величину, точное или приближенное распределение которой известно. Эту величину называют статистическим критерием K . Для проверки гипотезы по данным выборок вычисляют точные значения входящих в критерий величин и таким образом получают наблюдаемое значение критерия $K_{\text{набл}}$.

После выбора определенного критерия множество всех его возможных значений разбивают два непересекающихся подмножества: W , содержащее значения критерия, при которых H_0 принимается – область принятия гипотезы, и $\bar{W}$ – при которых нулевая гипотеза H_0 отвергается, – критическая область. При использовании любого критерия возможны ошибки следующих видов:

- 1) ошибка первого рода – принять гипотезу H_1 , когда верна H_0 ;
- 2) ошибка второго рода – принять гипотезу H_0 , когда верна H_1 .

Для каждой из них могут быть заданы соответствующие вероятности:

$$\alpha = P\{K \in \bar{W} | H_0\}, \alpha' = P\{K \in W | H_1\}.$$

		Верная гипотеза	
		H_0	H_1
Результат применения критерия	H_0	H_0 верно принята	H_0 неверно принята (Ошибка второго рода)
	H_1	H_0 неверно отвергнута (Ошибка первого рода)	H_0 верно отвергнута

Вероятность ошибки первого рода называется критерием (или уровнем) значимости. Нулевую гипотезу отвергают, если для нее значение p ниже уровня значимости, т. е., если $p < \alpha$. Обычно назначают условное значение 0,05, тогда шанс допустить ошибку 1-го рода никогда не превысит выбранного уровня значимости, скажем $\alpha=0,05$, так как нулевую гипотезу отвергают только тогда, когда $p < 0,05$. Если обнаружено, что $p > 0,05$, то нулевую гипотезу не отвергнут и,

следовательно, не допустят ошибки 1-го рода. Величина $(1-\beta)$ называется мощностью критерия.

Односторонние и двусторонние критические области.

Точки, отделяющие критическую область от области принятия гипотезы называют критическими точками $k_{кр}$. Различают одностороннюю (правостороннюю и левостороннюю) и двустороннюю критические области:

1. Правосторонняя критическая область определяется неравенством $K > k_{кр}$, где $k_{кр} > 0$. Например, когда проверяется гипотеза о том, что среднее значение равно 10.

2. Левосторонняя критическая область $K < k_{кр}$, где $k_{кр} < 0$.

3. Двусторонняя критическая область $K < k_1$, $K > k_2$, где $k_2 > k_1$. В частности, если критические точки симметричны относительно нуля, то двусторонняя критическая область определяется как $|K| > k_{кр}$.

Для нахождения критической точки $k_{кр}$ правосторонней критической области при условии, что справедлива гипотеза H_0 :

1) задаются достаточно малой вероятностью — уровнем значимости α (связанной с доверительной вероятностью β соотношением $\alpha = 1 - \beta$);

2) ищут $k_{кр}$, исходя из требования: $P(K > k_{кр} = \alpha)$;

3) по таблицам соответствующих критериев находят $k_{кр}$ и сравнивают с наблюдаемым значением критерия $K_{набл}$;

4) если $K_{набл} > k_{кр}$, то H_0 отвергают; если $K_{набл} < k_{кр}$, нет оснований отвергать H_0 .

Отыскание левосторонней и двусторонней критических областей сводится к нахождению соответствующих критических точек:

$P(K < k_{кр}) = \alpha$ — требование для левосторонней критической области;

$P(K < k_1) + P(K > k_2) = \alpha$ — требование для двусторонней критической области.

Если выбрать симметричные относительно нуля точки: $-k_{кр}$ и $k_{кр}$ ($k_{кр} > 0$), то получим следующее соотношение для отыскания критических точек двусторонней области:

$$P(K > k_{кр}) = \frac{\alpha}{2}.$$

Для пересчета одностороннего значения уровня значимости в двустороннее значение можно воспользоваться простой формулой: $\alpha_{двустор} = 2 \cdot \alpha_{одностор}$.

Существенным моментом является то, что статистические методы не позволяют подтвердить гипотезу. Корректное утверждение, которое можно сформулировать в результате анализа, звучит, например: «проведенный анализ не позволяет с заданным уровнем значимости отвергнуть гипотезу».

Проверить гипотезу о равенстве среднего значения заданному можно двумя способами.

Первый способ:

Пусть генеральная совокупность X имеет нормальное распределение, и требуется проверить предположение о том, что ее математическое ожидание равно некоторому числу m_0 . Рассмотрим две возможности.

1) Известна дисперсия σ^2 генеральной совокупности. Тогда по выборке объема n найдем выборочное среднее $\bar{x}$ и проверим нулевую гипотезу $H_0: M(X)=m_0$. Критерием может служить величина

$$t = \frac{\bar{x} - m_0}{\sigma / \sqrt{n}}$$

где t – квантиль нормального распределения. Если же дисперсия неизвестна, то используется статистика

$$t = \frac{\bar{x} - m_0}{s / \sqrt{n}}$$

имеющая распределение Стьюдента с $(n-1)$ степенью свободы, t – квантиль распределения Стьюдента, s – среднее квадратическое отклонение выборки.

Найденное значение t сравнивается с критическим $t_{кр}$, которое определяется по таблице квантилей в зависимости от уровня значимости или вычисляется встроенными функциями Excel. При попадании выборочного значения в критическую область гипотеза отвергается.

Второй способ:

Для проверки гипотезы о равенстве среднего заданному значению может быть использована функция Z.ТЕСТ, которая вычисляет двухстороннюю вероятность значений z-теста при стандартном распределении.

Синтаксис:

Z.ТЕСТ (Данные; Число; Сигма)

Данные: массив данных.

Число: значение для теста.

Сигма: необязательное стандартное отклонение генеральной совокупности.

Если аргумент не указан, используется стандартное отклонение выборки.

Пример:

=Z.ТЕСТ (A1:A50; 12) вычисляет вероятность того, что значение 12 относится к стандартному распределению генеральной совокупности данных в A1:A50.

При использовании этой функции вычисляется вероятность того, что для генеральной совокупности справедлива гипотеза $H_0: m=m_0$. Используется двухсторонний критерий, то есть альтернативная гипотеза $H_1: m \neq m_0$. Если эта вероятность меньше заданного уровня значимости, гипотеза отклоняется.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1					Проверяется гипотеза H_0 о том, что среднее выборки равно 10										
2	Результаты наблюдений				Дисперсия D	0,1									
3	1	10,16			$\sigma=\sqrt{D}$	0,31623									
4	2	9,99			среднее выборки	10,1295									
5	3	9,8			среднее ген.совокупности	10									
6	4	10,27													
7	5	10,13			Стандартный метод										
8	6	9,97													
9	7	10,04			Уровень значимости α (вероятность принять гипотезу H_1 , когда верна H_0)	0,05									
10	8	10,16			Вероятность принятия гипотезы H_0 1- α	0,95									
11	9	10,19			Квантиль нормального распределения для 0,95 ткр	1,64485									
12	10	10,26			Вычисляем действительную ошибку $\Delta = X_{\text{ср. ген}} - X_{\text{ср. выб}} $	0,1295									
13	11	9,96			С другой стороны $\Delta = t \cdot \mu$, где $\mu = \sigma/\sqrt{n}$, n - объем выборки, t - квантиль распределения										
14	12	9,89			Выразим значение t для данной ошибки $t_{\text{наб}} = \Delta/\mu/\sigma$	1,83141									
15	13	10,11			$t_{\text{наб}} > t_{\text{кр}}$										
16	14	10,3													
17	15	10,15													
18	16	10,04													
19	17	10,39													
20	18	10,43			Решение задачи с помощью функции Z.TEST										
21	19	10,03													
22	20	10,32			Z-тест	=Z.TEST									
23					Вероятность принятия гипотезы H_0 (Шарики, изготовленные станком-автоматом, должны иметь диаметр 10 мм) n=22 шарика в пс. Гипотеза отвергается										

Аргументы функции

Z-ТЕСТ

Массив: B3:B22 = {10,16;9,99;9,8;10,27;10,13;9,97;10,04;10,16;10,19;10,26;9,96;9,89;10,11;10,3;10,15;10,04;10,39;10,43;10,03;10,32}

X: 10 = 10

Сигма: F3 = 0,316227766

Возвращает одностороннее P-значение z-теста.

X - проверяемое значение.

Значение: 0,033519941

[Справка по этой функции](#)

OK Отмена

Аргументы функции

НОРМ.СТ.ОБР

Вероятность: 0,95 = 0,95

= 1,644853627

Возвращает обратное значение стандартного нормального распределения.

Вероятность - вероятность, соответствующая нормальному распределению.

Значение: 1,644853627

[Справка по этой функции](#)

OK Отмена

Аргументы функции

СТЮДЕНТ.ОБР

Вероятность: 0,95 = 0,95

Степени_свободы: 19 = 19

= 1,729132812

Возвращает левостороннее обратное распределение Стюдента.

Степени_свободы - положительное целое число степеней свободы, характеризующее распределение.

Значение: 1,729132812

[Справка по этой функции](#)

OK Отмена

	A	B	C	D	E	F	G	H	I	J	K
1					Проверяется гипотеза H0 о том, что среднее выборки равно 10						
2	Результаты наблюдений				среднее выборки	10,1295					
3	1	10,16			среднее ген.совокупности m0	10					
4	2	9,99			дисперсия выборки Dв	0,02801					
5	3	9,8			s=√Dв	0,16735					
6	4	10,27									
7	5	10,13			Стандартный метод						
8	6	9,97			вычислим квантиль распределения Стьюдента по формуле						
9	7	10,04			$u = \frac{\bar{x} - m_0}{s / \sqrt{n}}$						
10	8	10,16									
11	9	10,19									
12	10	10,26			tнаб	3,46072					
					Найдем tкр квантиль распределения Стьюдента для вероятности 0,95 и n-1=19	1,72913					
13	11	9,96									
14	12	9,89									
15	13	10,11			tнаб>tкр	гипотеза H0 отвергается					
16	14	10,3									
17	15	10,15									
18	16	10,04									
19	17	10,39									
20	18	10,43			Решение задачи с помощью функции Z.ТЕСТ						
21	19	10,03									
22	20	10,32			Z тест	0,00027					
23											
24					Вероятность принятия гипотезы H0 (Шарики, изготовленные станком-автоматом, должны иметь диаметр 10 мм) меньше 0,05. Гипотеза отклоняется						

Проверка гипотез о равенстве дисперсий

При проверке гипотеза о равенстве дисперсий двух нормально распределенных совокупностей $H_0: \sigma_1^2 = \sigma_2^2$ при неизвестных математических ожиданиях m_1 и m_2 используется статистика

$$F = \frac{s_1^2}{s_2^2} \quad (1)$$

которая имеет F -распределение Фишера с числом степеней свободы $(n_1 - 1)$ и $(n_2 - 1)$; здесь n_1 и n_2 – объемы выборок, s_1^2 и s_2^2 – соответствующие несмещенные дисперсии; при этом предполагается, что $s_1^2 > s_2^2$.

Для проверки этой гипотезы в Excel есть функция F.ТЕСТ, которая возвращает результат F -теста. Синтаксис функции:

F.ТЕСТ (Данные1; Данные2)

Данные1: первый массив записей.

Данные2: второй массив записей.

Пример:

=F.ТЕСТ (A1:A30; B1:B12) вычисляет различие дисперсий для двух множеств данных и возвращает вероятность того, что оба множества представляют собой выборку из общей совокупности.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	Проверяется гипотеза об одинаковой точности работы станков по результатам измерений(точность характеризуется дисперсией соответствующего размера) на уровне												
2	значимости 0,05												
3													
4	№	Станок1	Станок2										
5	1	12,05	12,36			Дисперсия 1 выборки		0,044761					
6	2	12,08	12,45			Дисперсия 2 выборки		0,017126					
7	3	12,33	12,48			F=		2,613645					
8	4	12,34	12,56			zкр=		5;22;24)					
9	5	12,75	12,63										
10	6	12,32	12,25			$z > z_{кр}$ гипотеза отклоняется							
11	7	12,12	12,54										
12	8	12,05	12,35			Использование F теста							
13	9	12,08	12,54										
14	10	12,33	12,33			F=		0,027323					
15	11	12,08	12,85										
16	12	12,75	12,42			Вероятность принятия гипотезы о равенстве дисперсий меньше 0,05							
17	13	12,05	12,47			Гипотеза отвергается							
18	14	12,08	12,41										
19	15	12,33	12,34										
20	16	12,05	12,51										
21	17	12,08	12,45										
22	18	12,31	12,24										
23	19	12,34	12,55										
24	20	12,42	12,32										
25	21	12,42	12,44										
26	22	12,12	12,41										
27	23		12,38										
28	24		12,51										
29													
30													
31													

Аргументы функции

F.ОБР

Вероятность 0,95 - 0,95

Степени_свободы1 22 - 22

Степени_свободы2 24 - 24

- 2,003481506

Возвращает обратное значение для (левостороннего) F-распределения вероятностей: если p = F.РАСП(x,...), то F.ОБР(p,...) = x.

Степени_свободы1 числитель степеней свободы - число от 1 до 10¹⁰, исключая 10¹⁰.

Значение: 2,003481506

Проверка гипотез о виде распределения

Другой группой статистических гипотез являются гипотезы о проверке вида распределения: неизвестен вид распределения генеральной совокупности, и в частности, неизвестна функция распределения $F(x)$.

Пусть $x_1, x_2, \dots, x_n$ – выборка наблюдений случайной величины X . Проверяется гипотеза H_0 о том, что случайная величина X имеет функцию распределения $F(x)$. Разобьем область возможных значений X на r интервалов $\Delta_1, \Delta_2, \dots, \Delta_r$. Пусть n_i – число элементов выборки, принадлежащих интервалу Δ_i ($i=1, \dots, r$); при малых значениях n_i интервалы объединяют таким образом, чтобы в каждом из них было $n_i \geq 5$. Используя предполагаемый закон распределения – с функцией $F(x)$, с учетом оценок параметров этого закона, найденных по выборке, находят вероятности того, что значения X принадлежат интервалу Δ_i , то есть

$$p_i = P\{X \in \Delta_i\}, i = \overline{1, r}.$$

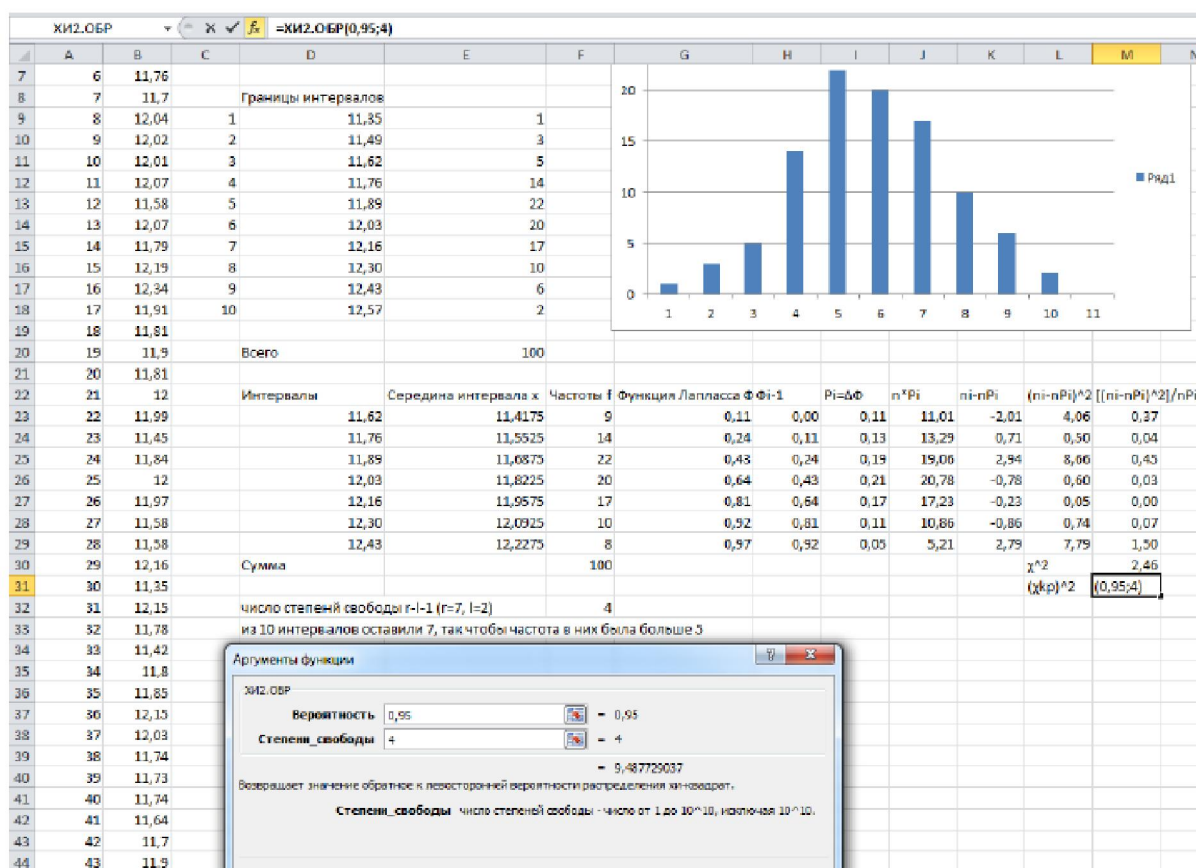
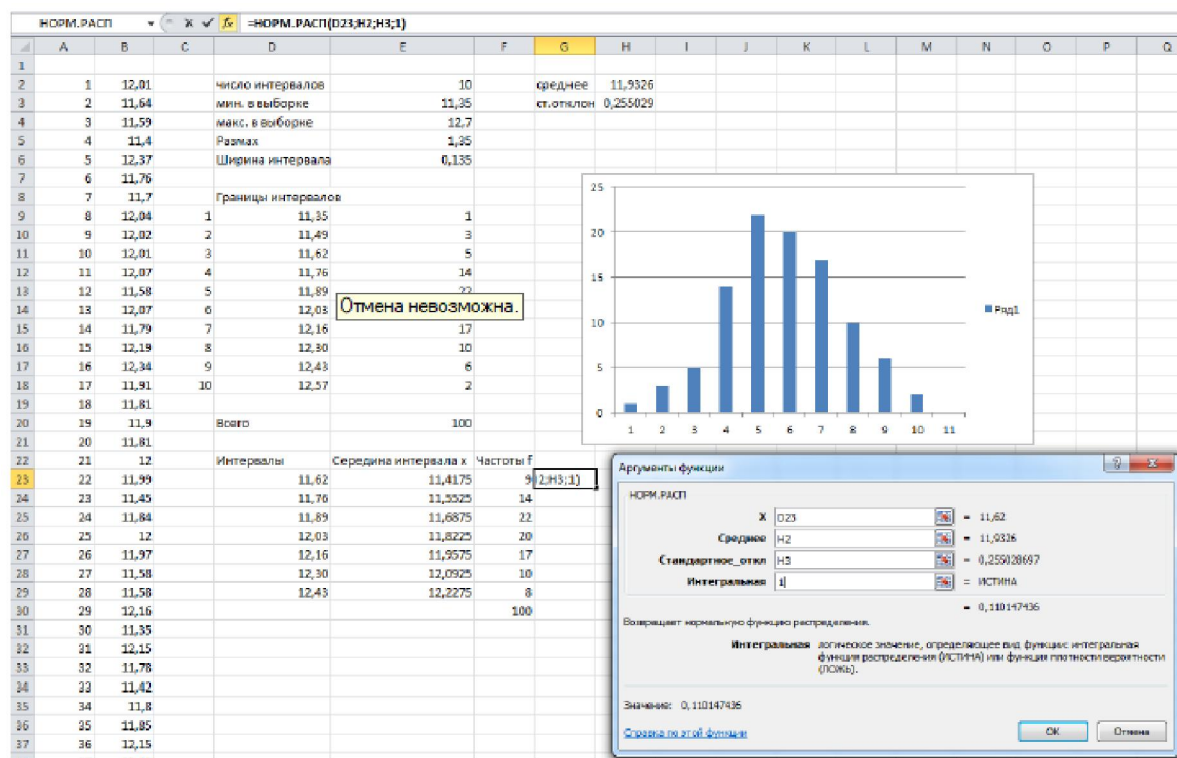
Статистика

$$\chi^2 = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i}$$

имеет распределение χ^2 с числом степеней свободы $(r-l-1)$, где r – число интервалов, l – число неизвестных параметров распределения. Например, для нормального распределения $l=2$ (неизвестные параметры m и σ). Считается, что гипотеза H_0 согласуется с опытом, если

$$\chi^2 < \chi_{1-\alpha}^2(r-l-1),$$

где χ^2 – выборочное значение статистики, $\chi^2_{1-\alpha}(r-l-1)$ – квантиль порядка $(1-\alpha)$ распределения χ^2 с числом степеней свободы $(r-l-1)$. Рассмотренный метод проверки гипотезы вида распределения называется критерием хи-квадрат или критерием согласия Пирсона.



Вариант 1

1. Производитель утверждает, что доля бракованных изделий не превосходит 7 %. В случайной выборке объема $n = 150$ изделий оказалось 16 бракованных изделий. Не противоречит ли это утверждению производителя? Доверительная вероятность $\gamma = 99 \%$. Ответ: выборка не противоречит утверждению производителя.

2. Инвестиция 1 рассчитана на $n_1=12$ лет, дисперсия ежегодных прибылей $\text{var}1= 20 \%$. Инвестиция 2 рассчитана на $n_2=10$ лет, дисперсия ежегодных прибылей $\text{var}2= 30 \%$. Предполагается, что распределение ежегодных прибылей на инвестиции подчиняется нормальному закону распределения. Равны ли риски 1 и 2? Доверительная вероятность $\gamma = 95 \%$.

3. Проводились испытания нового лекарства. В эксперименте участвовали $n_1 = 2000$ мужчин и $n_2 = 2500$ женщин. У 40 мужчин и 70 женщин наблюдались побочные эффекты. Можно ли утверждать, что побочные эффекты от нового лекарства у женщин возникают чаще, чем у мужчин? Доверительная вероятность $\gamma = 99 \%$.

Вариант 2

1. Автомат, работающий со стандартным отклонением $\sigma = 1,5$ г, фасует чай в пачки со средним весом $a = 80$ г. В случайной выборке объема $n = 16$ пачек средний вес $X = 78,5$ г. Надо ли отрегулировать автомат? Доверительная вероятность $\gamma = 99 \%$. Ответ: автомат нужно отрегулировать.

2. Станок, работающий со стандартным отклонением $\sigma = 0,5$ мм, производит детали средней длины $a = 20$ мм. В случайной выборке объема $n = 16$ деталей средняя длина $X = 19,8$ мм. Правильно ли настроен станок? Доверительная вероятность $\gamma = 99 \%$

3. Производитель утверждает, что средний вес пачки чая не меньше $a = 100$ г. Инспектор отобрал 10 пачек чая и взвесил. Их вес оказался 97, 102, 103, 98, 96, 105, 98, 100, 101, 99 г. соответственно. Не противоречит ли это утверждению производителя? Предполагается, что вес пачек чая распределен нормально. Доверительная вероятность $\gamma = 99 \%$.

Вариант 3

1. На двух токарных станках обрабатываются втулки. Отобраны две пробы: из втулок, сделанных на первом станке, $n_1 = 15$ шт., на втором станке - $n_2 = 18$ шт. По данным этих выборок рассчитаны выборочные смещенные дисперсии $S_1^2 = \text{var}1= 8,5$ (для первого станка) и $S_2^2 = \text{var} 2= 6,3$ (для второго станка). Полагая, что размеры втулок подчиняются нормальному закону распределения, выяснить, можно ли считать, что станки обладают различной точностью. Доверительная вероятность $\gamma = 95\%$.

2. Автомат 1 и автомат 2 фасуют чай в пачки. Стандартные отклонения $\sigma_1 = 1$ г и $\sigma_2 = 2$ г соответственно. В случайной выборке объема $n_1 = 20$ пачек для автомата 1 средний вес $\bar{X}_1 = 101$ г. В случайной выборке объема $n_2 = 15$ пачек для

автомата 2 средний вес $\bar{X}_2 = 98$ г. Верно ли, что оба автомата фасуют чай в пакеты одинакового среднего веса? Доверительная вероятность $\gamma = 95\%$. Верно ли, что оба автомата фасуют чай в пакеты одинакового среднего веса? Доверительная вероятность $\gamma = 95\%$.

3. Рассмотрим две выборки x и y из двух нормальных распределений X и Y с дисперсиями $D(X)=2$, $D(Y)=3$ соответственно:

$x = 5,393; 4,431; 6,841; 3,051; 5,538; 5,619; 7,49; 4,085; 5,779; 7,187; 2,424; 6,85; 3,517; 5,649; 3,512; 3,243; 5,6; 2,29; 4,825; 5,584;$

$y = 8,539; 3,871; 5,334; 6,825; 6,322; 6,34; 6,902; 7,425; 7,613; 7,554; 6,451; 4,799; 6,806; 5,739; 4,505.$

Проверим нулевую гипотезу $H_0: M(X) = M(Y)$ против альтернативной гипотезы $H_1: M(X) \neq M(Y)$. Доверительная вероятность $\gamma = 99\%$.

Вариант 4

1. Для производства каждой из $n_1 = 10$ деталей по первой технологии было затрачено в среднем $X_1=30$ с (выборочная дисперсия $var1=1$ с²). Для производства каждой из $n_2 = 16$ деталей по второй технологии было затрачено в среднем $X_2 = 28$ с (выборочная дисперсия $var2=2$ с²). Можно ли сделать вывод, что первой технологии требуется в среднем больше времени для производства одной детали? Доверительная вероятность $\gamma = 95\%$.

2. Рассмотрим две выборки из двух нормальных распределений с одинаковыми дисперсиями:

$x = 4,379; 4,039; 4,331; 3,654; 2,616; 5,062; 4,829; 5,787; 8,1; 6,144; 6,393; 6,219; 6,295; 5,952; 3,523; 5,098; 3,931; 5,985; 4,743; 4,089;$

$y = 5,14; 6,09; 5,433; 5,016; 3,919; 4,421; 4,047; 4,813; 6,44; 5,255; 5,726; 6,046; 4,977; 5,394; 5,229.$

Проверим гипотезу H_0 о равенстве математических ожиданий исследуемых распределений против альтернативной гипотезы о том, что они не равны. Таким образом, $H_0: M(X) = M(Y)$, $H_1: M(X) \neq M(Y)$. Доверительная вероятность $\gamma = 99\%$.

3. Для производства каждой из $n_1=51$ детали по первой технологии было затрачено в среднем $X_1=30$ с (выборочная дисперсия $var1 = 6$ с²). Для производства каждой из $n_2 = 41$ детали по второй технологии было затрачено в среднем $X_2 = 25$ с (выборочная дисперсия $var2 = 3$ с²). Можно ли сделать вывод, что первой технологии требуется в среднем больше времени для производства одной детали? Доверительная вероятность $\gamma = 95\%$.

Вариант 5

1. Для производства каждой из $n_1 = 51$ детали по первой технологии было затрачено в среднем $X_1 = 32$ с (выборочная дисперсия $var1 = 9$ с²). Для производства каждой из $n_2 = 41$ детали по второй технологии было затрачено в среднем $X_2 = 28$ с (выборочная дисперсия $var2 = 4$ с²). Можно ли сделать вывод, что первой технологии требуется в среднем больше времени для производства одной детали? Доверительная вероятность $\gamma = 90\%$.

2. Проводились испытания нового лекарства. В эксперименте участвовали $n_1 = 3000$ мужчин и $n_2 = 3500$ женщин. У 50 мужчин и 110 женщин наблюдались побочные эффекты. Можно ли утверждать, что побочные эффекты от нового лекарства у женщин возникают чаще, чем у мужчин? Доверительная вероятность $\gamma = 95\%$.

3. Станок, работающий со стандартным отклонением $\sigma = 0,4$ мм, производит детали средней длины $a = 30$ мм. В случайной выборке объема $n = 25$ деталей средняя длина $X = 30,1$ мм. Правильно ли настроен станок? Доверительная вероятность $\gamma = 95\%$. Ответ: станок настроен правильно.

Вариант 6

1. Утверждается, что шарики для подшипников, изготовленные автоматическим станком, имеют средний диаметр 10 мм. Используя односторонний критерий с $\alpha = 0,05$, проверить эту гипотезу, если в выборке из n шариков средний диаметр оказался равным 10,3 мм, а дисперсия известна и равна 1 мм.

2. Какой минимальный объем выборки следует взять в условиях задачи 2 (вариант 9), чтобы при проверке гипотезы $H_0: a=10$ против альтернативной $H_1: a=9$ вероятность ошибки первого рода равнялась 0,01, а вероятность ошибки второго рода не превышала 0,1? Указать выражение для соответствующей критической области.

3. При измерении производительности двух агрегатов получены следующие результаты (в килограммах вещества за час работы):

№ измерения	1	2	3	4	5
Агрегат 1	14,1	10,1	14,7	13,7	14,0
Агрегат 2	14,0	14,5	13,7	12,7	14,1

Можно ли считать, что производительности агрегатов одинаковы, в предположении, что обе выборки получены из нормальных распределений с одинаковой дисперсией. Принять уровень значимости равным 0,05.

Вариант 7

1. Из 200 задач первого раздела курса математики, предложенных для решения, абитуриенты решили 130, а из 300 задач второго раздела абитуриенты решили 120. Можно ли при $\alpha = 0,01$ утверждать, что первый раздел школьного курса абитуриенты усвоили лучше, чем второй.

2. Проверка правильности функционирования устройства осуществляется специальным тестом. Если устройство функционирует правильно, то вероятность прохождения теста равна 0,99; в противном случае вероятность прохождения теста равна 0,4. Устройство допускается к работе, если тест проходит пять раз при пяти испытаниях. В предположении, что число прохождений теста в пяти испытаниях подчиняется биномиальному распределению, указать критическую область для проверки гипотезы о том, что устройство функционирует правильно, и вычислить вероятности ошибок первого и второго рода.

3. Два штурмана определяли пеленг маяка по нескольким замерам, используя различные пеленгаторы. Первый штурман произвел 4 замера, при этом

выборочное среднее полученных результатов оказалось равным $70,2^\circ$. Второй штурман произвел 9 замеров, выборочное среднее оказалось равным $70,5^\circ$. В предположении, что результаты измерений имеют нормальное распределение со стандартным отклонением $0,5^\circ$, проверить гипотезу о том, что различие результатов вызвано только случайными ошибками. Принять уровень значимости равным 0,05.

Вариант 8

1. Большая партия изделий содержит некоторую долю дефектных. Поставщик утверждает, что эта доля составляет 5%, а покупатель предполагает, что 10%. Условия поставки: из партии случайным образом отбирается и проверяется 10 изделий; партия принимается на условиях поставщика, если при проверке обнаружено не более одного дефектного изделия; в противном случае партия принимается на условиях покупателя. Найти вероятности ошибок первого и второго рода при использовании указанных условий поставки.

2. По соответствующим нормативам средний вес таблетки лекарства сильного действия должен быть равен 0,5 мг. Выборочная проверка 121 таблетки полученной партии лекарств показала, что средний вес таблетки в выборке равен 0,53 мг. Путем многократного взвешивания таблеток было установлено, что их вес имеет нормальное распределение со стандартным отклонением $\sigma=0,11$ мг. На уровне значимости $\alpha=0,01$ проверить гипотезу о том, что средний вес таблетки во всей партии равен 0,5 мг против альтернативной гипотезы, состоящей в том, что он равен $a_1(a_1>0,5)$.

3. Сырье, поступающее со складов a и b , номинально расфасовано по 1 кг в пакеты, маркированные буквами А и В соответственно. Из опыта известно, что веса наудачу взятых пакетов распределены нормально со стандартным отклонением 0,07кг и 0,03кг соответственно для складов a и b . Случайным образом отобрали 100 пакетов со склада a , их средний вес оказался равным 0,99кг. Средний вес четырехсот наудачу выбранных пакетов со склада b оказался равным 1,01кг. При уровне значимости 0,05 проверить гипотезу о том, что пакеты А и В имеют одинаковый средний вес.

Вариант 9

1. Из продукции автомата, обрабатывающего болты с номинальным значением контролируемого размера $m=40$ мм, была взята выборка объема 36. Результаты предыдущих измерений дают основания полагать, что действительные контролируемые размеры болтов подчиняются нормальному распределению с дисперсией, равной 1 (мм)^2 . Партия болтов бракуется, если выборочное среднее контролируемого размера больше 40,1 мм. Найти вероятность ошибки первого рода при использовании такого правила.

2. По повторной выборке объема 16 из нормального распределения с параметрами θ , $\sigma=4$ при уровне значимости 0,05 проверяется гипотеза $H_0: \theta = 2$ при конкурирующей гипотезе $H_1: \theta > 2$. Построить наиболее мощный критерий для проверки основной гипотезы и определить мощность критерия при альтернативном значении $\theta=3$. Найти объем выборки, при котором мощность критерия равна 0,6, если альтернативная гипотеза имеет вид $H_1: \theta=3$.

3. Из долгого опыта по производству безалкогольного напитка известно, что количество напитка, производимого за единицу времени, имеет нормальный закон распределения с математическим ожиданием 500 единиц и стандартным отклонением 96 единиц. Планируется осуществить модернизацию производства, в результате которой может измениться количество производимого напитка (стандартное отклонение останется прежним). По 50 наблюдениям построить критическую область для проверки гипотезы о том, что количество производимого напитка после модернизации не изменилось. Принять уровень значимости равным 0,05.

Вариант 10

1. По паспортным данным автомобильного двигателя расход топлива на 100 км пробега составляет 10 л. В результате изменения конструкции двигателя ожидается, что расход топлива на 100 км пробега уменьшится до 9 л. Для проверки этого факта были проведены испытания 25 случайно отобранных автомобилей с модернизированным двигателем. Выборочное среднее расходов топлива на 100 км пробега по результатам испытаний составило 9,3 л. В предположении, что расход топлива на 100 км пробега имеет нормальное распределение с дисперсией 4 (л)² проверить гипотезу, утверждающую, что изменение конструкции двигателя не повлияло на расход топлива.

2. Случайная величина ξ имеет показательное распределение с неизвестным параметром θ ($\theta > 0$). По одному наблюдению над случайной величиной ξ построить наиболее мощный критерий для проверки гипотезы $H_0: \theta = \theta_0$ ($\theta_0 = 1$) против альтернативной гипотезы $H_1: \theta = \theta_1$ ($\theta_1 > \theta_0$) при уровне значимости, равном 0,1. Найти мощность критерия, если $\theta_1 = 2$.

3. Из большой партии резисторов одного типа и номинала случайным образом отобрано 36 штук. Выборочное среднее величины сопротивления при этом оказалось равным 9,3 кОм. При уровне значимости 0,05 проверить гипотезу о том, что выборка взята из партии с номиналом 10 кОм, если дисперсия величины сопротивления известна и равна 4 (кОм)². Предполагается, что величина сопротивления наудачу взятого резистора имеет нормальный закон распределения.

Вариант 1

1. Непрерывная случайная величина X распределена равномерно в интервале (3; 6). Найти: а) дифференциальную и интегральную функции распределения, построить их графики; б) характеристики случайной величины; в) вероятность попадания случайной величины в интервал (4; 5), показать эту вероятность на графике.

2. Непрерывная случайная величина X имеет заданное показательное распределение с параметром $\lambda=2$. Найти: а) дифференциальную и интегральную функции распределения, построить их графики; б) характеристики случайной величины; в) вероятность попадания случайной величины в интервал (2; 8), показать эту вероятность на графике.

3. Какое из написанных ниже законов распределения является нормальным? Определите для него $M(X)$ и $\sigma(X)$ и постройте график этого распределения.

$$\text{а) } f(x) = \sqrt{\frac{2}{\pi}} \cdot e^{-2(x-5)^2}; \quad \text{б) } f(x) = \frac{1}{\sqrt{6\pi}} \cdot e^{\frac{-(x-2)^2}{6}}$$

4. Станок-автомат изготавливает валики, причем контролируется их диаметр X . Считая, что X – нормально распределенная случайная величина с математическим ожиданием, равным 10 мм и дисперсией, равной 0,25 мм², найти интервал, в котором с вероятностью 0,9973 будут заключены диаметры изготовленных валиков.

Вариант 2

1. Дана непрерывная случайная величина X имеет заданное показательное распределение с параметром $\lambda=5$. Найти: а) дифференциальную и интегральную функции распределения, построить их графики; б) характеристики случайной величины; в) вероятность попадания случайной величины в интервал (0,4;1), показать эту вероятность на графике.

2. Какое из написанных ниже законов распределения является нормальным? Определите для него $M(X)$ и $\sigma(X)$ и постройте график этого распределения.

$$\text{а) } f(x) = \frac{1}{4\sqrt{2\pi}} \cdot e^{\frac{-(x-3)^2}{8}}; \quad \text{б) } f(x) = \frac{1}{\sqrt{\pi}} \cdot e^{x^2-2x+1}$$

3. Случайная величина X распределена по нормальному закону с параметрами $\mu=M(X)=5$, $\sigma=\sigma(X)=2$. Найти вероятность того, что случайная величина X окажется в интервале (1; 10).

4. На автомате изготавливаются заклепки. Диаметр их головок представляет собой случайную величину, распределенную по нормальному закону и имеет среднее значение, равное 2 мм и дисперсию, равную 0,01 мм². Найти а) вероятность того, что диаметр головки заклепки будет от 2,1 до 2,3 мм; б) какие размеры диаметра головок заклепки можно гарантировать с вероятностью 0,95.

Вариант 3

1. Вычислить математическое ожидание, дисперсию и среднее квадратическое отклонение случайной величины X , равномерно распределенной в интервале (9;15). Найти вероятность того, что случайная величина попадет в интервал (11;14), ее на графике. Записать интегральную функцию и построить ее график.

2. Какое из написанных ниже законов распределения является нормальным? Определите для него $M(X)$ и $\sigma(X)$ и постройте график этого распределения.

$$\text{а) } f(x) = \frac{10}{\sqrt{\pi}} \cdot e^{-400(x-20)^2}; \quad \text{б) } f(x) = \frac{1}{\sqrt{10\pi}} \cdot e^{-\frac{(x-10)^2}{10}}$$

3. Случайная величина X распределена нормально. Математическое ожидание и среднее квадратическое отклонение этой величины, соответственно, равны 6 и 2. Найти вероятность того, что в результате испытания случайная величина X окажется в интервале (4; 8).

4. Диаметр подшипников, выпускаемых заводом, представляет случайную величину, распределенную по нормальному закону с $M(X)=15$ мм, $\sigma(X)=0,4$ мм. Какую точность диаметра подшипника можно гарантировать с вероятностью 0,9?

Вариант 4

1. Случайная величина X распределена равномерно в интервале (5; 9). Составить дифференциальную и интегральную функции распределения данной случайной величины, построить их графики. Найти характеристики и вероятность попадания случайной величины в интервал (3;8). Показать ее на графике.

2. Найти дисперсию и среднее квадратическое отклонение показательного распределения, заданного интегральной функцией:

$$F(x) = \begin{cases} 0, & \text{если } x < 0, \\ 1 - e^{-0,4x}, & \text{если } x \geq 0. \end{cases}$$

3. Математическое ожидание и среднее квадратическое отклонение нормально распределенной случайной величины X , соответственно, равны 20 и 5. Найти вероятность того, что в результате испытания X примет значение в интервале (15; 25).

4. Детали, выпускаемые цехом, по размеру диаметра распределяются согласно нормальному закону. Средний диаметр детали равен 5см, дисперсия – $0,64\text{см}^2$. Найти вероятность того, что диаметр деталей, выпускаемых цехом, будет от 4,2 до 6,6 см.

Вариант 5

1. Непрерывная случайная величина X распределена равномерно в интервале (16;20). Найти: а) дифференциальную и интегральную функции, построить их графики; б) характеристики случайной величины; в) вероятность попадания случайной величины в интервал (17;22).

2. Непрерывная случайная величина X распределена по показательному закону, заданному интегральной функцией:

$$F(x) = \begin{cases} 0, & \text{если } x < 0, \\ 1 - e^{-0,6x}, & \text{если } x \geq 0. \end{cases}$$

Найти вероятность того, что в результате испытания случайная величина попадет в интервал (2; 5).

3. Какое из написанных ниже законов распределения является нормальным? Определите для него $M(X)$ и $\sigma(X)$ и постройте график этого распределения.

$$\text{а) } f(x) = \frac{1}{\sqrt{2\pi}} \cdot e^{\frac{-(x-1)^2}{2}}; \quad \text{б) } f(x) = \sqrt{\frac{2}{\pi}} \cdot e^{-2(x-5)^2}$$

4. На станке изготавливается деталь. Оказывается, что ее длина представляет случайную величину, распределенную по нормальному закону, и имеет математическое ожидание 20см, а дисперсию 0,25см². Найти вероятность того, что отклонение длины детали от ее математического ожидания не превзойдет 0,3.

Вариант 6

1. Случайная величина X имеет равномерное распределение в интервале (2; 8). Написать дифференциальную и интегральную функции, построить их графики. Вычислить характеристики и вероятность попадания случайной величины в интервал (1; 6), показать эту вероятность на графике.

2. Время между двумя сбоями электронной вычислительной машины распределено по показательному закону. Интенсивность равна 4. Найти: а) дифференциальную и интегральную функции распределения, построить их графики; б) характеристики случайной величины; в) вероятность того, что в результате испытания случайная величина попадет в интервале (0,4; 1), показать эту вероятность на графике.

3. Какое из написанных ниже законов распределения является нормальным? Определите для него $M(X)$ и $\sigma(X)$ и постройте график этого распределения.

$$\text{а) } f(x) = \frac{10}{\sqrt{\pi}} \cdot e^{-400(x-20)^2}; \quad \text{б) } f(x) = \frac{1}{4\sqrt{2\pi}} \cdot e^{\frac{-(x-3)^2}{8}}$$

4. Диаметр изготавливаемой детали является случайной величиной, распределенной по нормальному закону. Средний диаметр равен 4,5см, дисперсия – 0,0025см². Найти интервал, в который с вероятностью 0,9545 попадает диаметр взятой наудачу детали.

Вариант 7

1. Случайная величина X распределена нормально с математическим ожиданием $M(X) = 25$. Вероятность попадания X в интервал (10;15) равна 0,2. Найти вероятность попадания X в интервал (35;40).

2. Случайная величина X распределена по нормальному закону с параметрами $a=6$, $\sigma =2$. Найти вероятность того, что случайная величина X окажется в интервале (3;7).

3. Случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при } x < 0 \\ x & \text{при } 0 < x \leq 1 \\ 1 & \text{при } x > 1 \end{cases}$$

Найти вероятность того, что в результате четырех испытаний случайная величина X ровно три раза примет значения из интервала $(0,3; 0,7)$.

4. Автомат штампует детали. Контролируется длина детали X , которая является случайной величиной, распределённой по нормальному закону с математическим ожиданием (проектная длина) равным 50 мм. Фактически длина изготовленных деталей не менее 32 мм и не более 62 мм. Найти вероятность того, что длина наудачу взятой детали: а) более 55 мм; б) менее 40 мм.

Вариант 8

1. Диаметр деталей, изготовленных цехом, является случайной величиной, распределенной по нормальному закону. Дисперсия ее равна $0,0001 \text{ мм}^2$, а математическое ожидание 2,5 мм. В каких границах с вероятностью 0,9973 можно гарантировать диаметр детали?

2. Случайная величина X задана рядом распределения

X_n	3	5	7	11
P_n	0,14	0,20	0,49	0,17

Найти функцию распределения $F(x)$ случайной величины X и построить ее график. Вычислить для X ее математическое ожидание и дисперсию.

3. Случайная величина X задана функцией распределения

$$F(x) = \begin{cases} 0 & \text{при } x < 1 \\ (x - 1)/2 & \text{при } 1 \leq x \leq 3 \\ 1 & \text{при } x > 3 \end{cases}$$

Вычислить вероятности попадания случайной величины X в интервалах $(1,5; 2,5)$ и $(2,5; 3,5)$.

4. Составьте закон распределения вероятностей дискретной случайной величины X – числа k выпадений хотя бы одной «шестерки» в $n=8$ бросаниях пары игральных кубиков. Постройте многоугольник распределения. Найдите числовые характеристики распределения (моду распределения, математическое ожидание $M(X)$, дисперсию $D(X)$, среднее квадратическое отклонение $s(X)$).

Вариант 9

1. Какое из написанных ниже законов распределения является нормальным? Определите для него $M(X)$ и $\sigma(X)$ и постройте график этого распределения.

$$\text{а) } f(x) = \frac{1}{5\sqrt{2\pi}} \cdot e^{-\frac{(x-7)^2}{50}}; \quad \text{б) } f(x) = \frac{1}{\sqrt{\pi}} \cdot e^{x^2-2x+1}$$

2. Непрерывная случайная величина X распределена равномерно в интервале $(-5; 15)$. Найти: а) дифференциальную и интегральную функции распределения, построить их графики; б) характеристики случайной величины; в) вероятность попадания случайной величины в интервал $(0; 5)$, показать эту вероятность на графике.

3. Производятся независимые испытания с одинаковой вероятностью появления события A в каждом испытании. Найти вероятность появления события A , если дисперсия числа появлений события в трех независимых испытаниях равна 0,63.

4. Заданы две случайных величины с законами распределения:

X 1 2 3

p 0,2 0,5 0,3

Y -1 0 0,5 2

p 0,1 0,2 0,4 0,3

Найти математическое ожидание и дисперсию каждой из них.

Вариант 10

1. Непрерывная случайная величина X имеет заданная показательное распределение с параметром $\lambda=0,6$. Найти дифференциальную и интегральную функции распределения, характеристики этой случайной величины и вероятность того, что случайная величина принимает значения в интервале (5; 10).

2. Определить математическое ожидание числа приборов, отказавших в работе за время испытаний, если: а) вероятность отказа для всех приборов одна и та же равна p , а число испытываемых приборов равно n ; б) вероятность отказа для i -го прибора равна p_i , $i=1, 2, \dots, n$.

3. Функция распределения случайного времени X безотказной работы прибора имеет вид

$$F(x) = \begin{cases} 0 & \text{при } x < 0 \\ \frac{1}{T} e^{-\frac{x}{T}} & \text{при } x \geq 0 \end{cases}$$

где $T>0$. Найти: а) вероятность безотказной работы прибора в течение времени $t=T$; б) плотность распределения случайной величины X ; в) математическое ожидание $M(X)$ и дисперсию $D(X)$ случайной величины X .

4. Известно, что некоторая случайная величина может принимать значения 0, 2 и

4. Известно, что математическое ожидание равно 2, а дисперсия – 0,8. Найти закон распределения случайной величины.

Приложение 3

Вариант 1

1. Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведенным в корреляционной таблице:

Y	X								n _y
	5	10	15	20	25	30	35	40	
100	2	1	-	-	-	-	-	-	3
120	3	4	3	-	-	-	-	-	10
140	-	-	5	10	8	-	-	-	23
160	-	-	-	1	-	6	1	1	9
180	-	-	-	-	-	-	4	1	5
n _x	5	5	8	11	8	6	5	2	n=50

2. Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ и выборочное корреляционное отношение η_{yx} по данным, приведенным в корреляционной таблице:

Y	X					n _y
	0	1	2	3	4	
0	18	1	1	-	-	20
3	1	20	-	-	-	21
5	3	5	10	2	-	20
10	-	-	7	12	-	19
17	-	-	-	-	20	20
n _x	22	26	18	14	20	n=100

3. Тринадцать цветных полос расположены в порядке убывания окраски от темной к светлой и каждой полосе присвоен ранг – порядковый номер. В итоге получена последовательность рангов

x_i 1 2 3 4 5 6 7 8 9 10 11 12 13

При проверке способности различать оттенки цветов, испытуемый расположил полосы в следующем порядке:

y_i 6 3 4 2 1 10 7 8 9 5 11 13 12

Найти коэффициент ранговой корреляции Спирмена между «правильными» рангами x_i и рангами y_i , которые присвоены полосам испытуемым.

Вариант 2

1. Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведенным в корреляционной таблице:

Y	X							n _y
	18	23	28	33	38	43	48	
125	-	1	-	-	-	-	-	1
150	1	2	5	-	-	-	-	8
175	-	3	2	12	-	-	-	17
200	-	-	1	8	7	-	-	16

225	-	-	-	-	3	3	-	6
250	-	-	-	-	-	1	1	2
n_x	1	6	8	20	10	4	1	$n=50$

2. Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ и выборочное корреляционное отношение η_{yx} по данным, приведенным в корреляционной таблице:

Y	X					n_y
	0	4	6	7	10	
7	19	1	1	-	-	21
13	2	14	-	-	-	16
40	-	3	22	2	-	27
80	-	-	-	15	-	15
200	-	-	-	-	21	21
n_x	21	18	23	17	21	$n=100$

3. Два товароведов расположили девять мотков пряжи в порядке убывания толщины нити. В итоге были получены две последовательности рангов:

x_i 1 2 3 4 5 6 7 8 9

y_i 4 1 5 3 2 6 9 8 7

Найти коэффициент ранговой корреляции Спирмена между рангами x_i и y_i .

Вариант 3

1. Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведенным в корреляционной таблице:

Y	X							n_y
	5	10	15	20	25	30	35	
100	-	-	-	-	-	6	1	7
120	-	-	-	-	-	4	2	6
140	-	-	3	10	5	-	-	23
160	3	4	8	-	-	-	-	10
180	2	1	-	1	-	-	-	4
n_x	5	5	11	11	5	10	3	$n=50$

2. Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ и выборочное корреляционное отношение η_{yx} по данным, приведенным в корреляционной таблице:

Y	X					n_y
	0	1	2	3	4	
10	20	5	-	-	-	21
11	7	15	3	1	-	16
20	-	3	17	4	-	27
35	-	-	8	13	7	15
50	-	-	-	5	42	21
n_x	27	23	28	23	49	$n=150$

3. Специалисты двух заводов проранжировали 11 факторов, влияющих на ход технологического процесса.

В итоге были получены две последовательности рангов:

x_i 1 2 3 4 5 6 7 8 9 10 11

y_i 1 2 3 5 4 9 8 11 6 7 10

Определить, согласуются ли мнения специалистов различных заводов, используя коэффициент ранговой корреляции Спирмена.

Вариант 4

1. Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведенным в корреляционной таблице:

Y	X						n_y
	18	23	28	33	38	43	
125	-	-	-	-	-	3	3
150	-	-	5	-	-	3	8
175	-	-	2	10	-	-	12
200	-	1	5	8	7	-	21
225	-	2	-	-	-	-	2
250	1	3	-	-	-	-	4
n_x	1	6	12	18	7	6	$n=50$

2. Найти выборочное уравнение регрессии $\bar{y}_x = Ax^2 + Bx + C$ и выборочное корреляционное отношение η_{yx} по данным, приведенным в корреляционной таблице:

Y	X			n_y
	0	4	5	
1	50	5	1	56
35	-	44	-	44
50	-	5	45	50
n_x	21	18	23	$n=150$

3. Три арбитра оценили мастерство 10 спортсменов, в итоге были получены три последовательности рангов (в первой строке приведены ранги арбитра А, во второй – ранги арбитра В, в третьей – ранги арбитра С):

x_i 1 2 3 4 5 6 7 8 9 10

y_i 3 10 7 2 8 5 6 9 1 4

z_i 6 2 1 3 9 4 5 7 10 8

Определить пару арбитров, оценки которых наиболее согласуются, используя коэффициент ранговой корреляции Спирмена.

Вариант 5

1. Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведенным в корреляционной таблице:

Y	X						n_y
	18	22	26	30	34	38	
70	5	-	-	-	-	-	5
75	7	19	1	-	-	-	27
80	-	14	5	-	-	3	22

85	-	-	10	8	-	-	18
90	-	-	-	3	7	-	10
95	-	-	-	-	9	9	18
n_x	12	33	16	11	16	12	$n=100$

2. Найти выборочное уравнение регрессии $\bar{x}_y=Ay^2+By+C$ и выборочное корреляционное отношение η_{xy} по данным, приведенным в корреляционной таблице:

Y	X			n_y
	6	30	50	
1	15	-	-	15
3	1	14	-	15
4	-	2	18	20
n_x	16	16	18	$n=50$

3. Два инспектора А и В проверили 12 водителей на быстроту реакции и расположили их в порядке ухудшения реакции (в скобках помещены порядковые номера водителей с одинаковой реакцией):

(А) 1 (2, 3, 4) 5 (6, 7, 8) 9 10 11 12

(В) 3 1 2 6 4 5 7 8 11 10 9 12

Найти выборочный коэффициент ранговой корреляции Спирмена между рангами водителей, присвоенными им двумя инспекторами.

Вариант 6

1. Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведенным в корреляционной таблице:

Y	X						n_y
	10	15	20	25	30	35	
30	2	6	-	-	-	-	8
40	-	4	4	-	-	-	8
50	-	-	7	35	8	-	50
60	-	-	2	10	8	-	20
70	-	-	-	5	6	3	14
n_x	2	10	13	50	22	3	$n=100$

2. Найти выборочное уравнение регрессии $\bar{x}_y=Ay^2+By+C$ и выборочное корреляционное отношение η_{xy} по данным, приведенным в корреляционной таблице:

Y	X			n_y
	1	9	19	
0	13	-	-	13
2	2	10	-	12
3	1	1	23	25
n_x	16	11	23	$n=50$

3. Два контролера расположили 10 деталей в порядке ухудшения их качества. В итоге были получены две последовательности рангов:

x_i 1 2 3 4 5 6 7 8 9 10

y_i 1 2 4 3 6 5 7 10 9 8

Используя коэффициент ранговой корреляции Кендалла, определить, согласуются ли оценки контролеров.

Вариант 7

1. Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведенным в корреляционной таблице:

Y	X						n_y
	10	15	20	25	30	35	
20	4	2	-	-	-	-	6
30	-	5	3	-	-	-	8
40	-	-	5	45	5	-	55
50	-	-	2	8	7	-	17
60	-	-	-	4	7	3	14
n_x	2	7	10	57	19	3	$n=100$

2. Найти выборочное уравнение регрессии $\bar{x}_y = Ay^2 + By + C$ и выборочное корреляционное отношение η_{xy} по данным, приведенным в корреляционной таблице:

Y	X			n_y
	2	3	5	
25	20	-	-	20
45	-	30	1	31
110	-	1	48	49
n_x	20	31	49	$n=100$

3. Тринадцать цветных полос расположены в порядке убывания окраски от темной к светлой и каждой полосе присвоен ранг – порядковый номер. В итоге получена последовательность рангов

x_i 1 2 3 4 5 6 7 8 9 10 11 12 13

При проверке способности различать оттенки цветов, испытуемый расположил полосы в следующем порядке:

y_i 6 3 4 2 1 10 7 8 9 5 11 13 12

Найти выборочный коэффициент ранговой корреляции Кендалла.

Вариант 8

1. Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведенным в корреляционной таблице:

Y	X						n_y
	65	95	125	155	185	215	
30	5	-	-	-	-	-	5
40	4	12	-	-	-	-	16
50	-	8	5	4	-	-	17
60	-	1	5	7	2	-	15
70	-	-	-	-	1	1	2
n_x	9	21	10	11	3	1	$n=55$

2. Найти выборочное уравнение регрессии $\bar{x}_y = Ay^2 + By + C$ и выборочное корреляционное отношение η_{xy} по данным, приведенным в корреляционной таблице:

Y	X						n_y
	5	10	15	20	25	30	
10	4	6	-	-	-	-	10
20	-	2	8	-	-	-	10
35	-	-	-	10	2	8	20
40	-	-	-	5	8	2	15
50	-	-	-	-	10	5	15
n_x	4	8	8	15	20	15	$n=70$

3. Специалисты двух заводов проранжировали 11 факторов, влияющих на ход технологического процесса.

В итоге были получены две последовательности рангов:

x_i 1 2 3 4 5 6 7 8 9 10 11

y_i 1 2 3 5 4 9 8 11 6 7 10

Используя коэффициент ранговой корреляции Кендалла, определить, согласуются ли мнения специалистов различных заводов.

Вариант 9

1. Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведенным в корреляционной таблице:

Y	X							n_y
	-3	-2	-1	0	1	2	3	
-2	2	-	2	-	-	-	-	4
-1	-	1	4	-	-	-	-	5
0	-	4	3	10	-	-	-	17
1	-	2	-	2	3	6	-	13
2	-	-	-	-	5	4	-	9
3	-	-	-	-	-	1	1	2
n_x	2	7	10	12	8	11	1	$n=50$

2. Найти выборочное уравнение регрессии $\bar{x}_y = Ay^2 + By + C$ и выборочное корреляционное отношение η_{xy} по данным, приведенным в корреляционной таблице:

Y	X			n_y
	1	2	3	
35	10	-	-	10
75	-	22	1	23
95	-	-	16	16
n_x	10	22	17	$n=50$

3. Три арбитра оценили мастерство 10 спортсменов, в итоге были получены три последовательности рангов (в первой строке приведены ранги арбитра А, во второй – ранги арбитра В, в третьей – ранги арбитра С):

x_i 1 2 3 4 5 6 7 8 9 10
 y_i 3 10 7 2 8 5 6 9 1 4
 z_i 6 2 1 3 9 4 5 7 10 8

Используя коэффициент ранговой корреляции Кендалла, определить пару арбитров, оценки которых наиболее совпадают.

Вариант 10

1. Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведенным в корреляционной таблице:

Y	X						n_y
	10	20	30	40	50	60	
15	5	7	-	-	-	-	12
25	-	20	23	-	-	-	43
35	-	-	30	47	2	-	79
45	-	-	10	11	20	6	47
55	-	-	-	9	7	3	19
n_x	5	27	63	67	29	9	$n=200$

2. Найти выборочное уравнение регрессии $\bar{x}_y = Ay^2 + By + C$ и выборочное корреляционное отношение η_{xy} по данным, приведенным в корреляционной таблице:

Y	X			n_y
	0	4	5	
1	30	5	1	36
20	-	14	-	14
50	-	5	45	50
n_x	30	24	46	$n=100$

3. Два преподавателя оценили знания 12 учащихся по стобалльной системе и выставили им следующие оценки (в первой строке указано количество баллов, выставленных первым преподавателем, а во второй – вторым):

98 94 88 80 76 70 63 61 60 58 56 51
 99 91 93 74 78 65 64 66 52 53 48 62

Найти выборочный коэффициент ранговой корреляции Спирмена между оценками двух преподавателей.

Приложение 4

1. Используя MSExcel (Ooo Calc) выполнить задание: станок производит детали, измерения которых приведено ниже. С доверительной вероятностью 0.95 построить доверительные интервалы для математического ожидания и дисперсии размера деталей.

Вариант 1

43.8	43.9	46.3	44.6	47.5	42.0	44.5	45.0	46.8	45.3	41.8	42.3	47.9	45.5	44.4	43.1	42.8
41.9	42.8	46.0	45.3	41.8	42.3	47.9	45.5	46.3	44.6	47.5	42.0	44.5	43.8	43.9	46.3	44.6

Вариант 2

49.0	48.8	49.2	50.2	49.5	49.8	49.9	49.3	49.6	49.5	49.7	49.0	48.8	51.8	49.1	48.3	50.0
49.0	48.4	48.5	49.6	49.5	49.7	49.0	48.8	51.8	49.5	49.8	49.9	49.3	49.6	48.8	49.2	50.2

Вариант 3

42.1	41.9	42.3	43.1	42.5	42.7	42.9	42.3	42.6	42.5	42.7	42.0	41.9	44.5	42.2	41.5	42.9
42.1	41.5	41.6	41.9	42.3	43.1	42.5	42.7	42.9	42.3	42.6	42.5	42.7	42.0	42.3	43.1	42.5

Вариант 4

25.6	25.5	25.5	24.7	25.0	25.8	25.2	25.0	25.0	25.3	25.4	25.1	25.2	25.3	24.8	25.1	24.6
25.2	25.8	24.8	25.5	24.7	25.0	25.8	25.2	25.0	25.0	25.3	25.4	25.1	25.2	25.0	25.0	25.3

Вариант 5

51.2	49.9	50.7	52.3	51.3	51.2	52.4	51.6	51.5	51.0	51.8	50.9	50.7	52.0	50.2	51.1	51.0
51.2	52.0	49.9	50.7	52.3	51.3	51.2	52.4	51.6	51.5	51.0	51.8	50.9	50.7	52.4	51.6	51.5

Вариант 6

21.8	21.6	21.7	20.8	20.8	20.8	21.2	20.7	19.8	22.2	21.9	21.5	20.9	20.9	20.9	21.6	21.2
22.2	20.6	21.7	20.8	20.8	20.8	21.2	20.7	19.8	22.2	21.9	21.5	20.9	20.9	20.9	21.8	21.6

Вариант 7

22.7	25.7	25.5	25.2	22.7	23.6	26.4	24.2	23.7	23.5	24.7	24.8	24.0	24.3	24.7	23.1	24.0
22.1	24.4	26.5	25.7	25.5	25.2	22.7	23.6	26.4	24.2	23.7	23.5	22.7	25.7	25.5	25.2	22.7

Вариант 8

49.5	49.5	46.7	47.2	49.1	48.7	51.2	46.1	50.5	48.9	49.3	50.4	47.2	48.5	49.4	48.8	50.1
46.5	51.2	46.2	46.1	50.5	49.5	48.7	51.2	47.2	49.1	49.5	46.7	47.2	49.1	48.7	47.2	49.1

Вариант 9

31.6	31.4	31	31.4	33.1	33.0	32.9	31.4	31.9	33.5	32.3	32.0	31.9	32.6	32.7	32.2	32.3
32.6	31.6	32.1	32.9	31.4	31.9	33.5	32.3	32.0	31.6	31.4	31	31.4	33.1	33.0	32.9	31.4

Вариант 10

36.2	35.9	36.1	34.5	34.5	34.6	35.2	34.3	32.8	36.9	36.4	35.8	34.6	34.7	34.8	35.9	35.3
37	34.3	36.0	36.2	35.9	36.1	34.5	34.5	34.6	35.2	34.3	32.8	36.9	36.4	34.5	34.6	35.2

Приложение 5

Используя MSExcel (Ooo Calc) выполнить задание:

Построить уравнение регрессии $Y=a+bx$.

Вариант 1			Вариант 2			Вариант 3		
№	X	Y	№	X	Y	№	X	Y
1	-1.132	1.554	1	-0.132	1.791	1	-0.332	-1.65
2	-0.204	4.601	2	0.796	1.51	2	0.596	2.31
3	0.858	2.943	3	1.858	4.17	3	1.658	4.953
4	1.715	-1.157	4	2.715	3.007	4	2.515	8.62
5	2.494	-6.048	5	3.494	5.875	5	3.294	7.49
6	4.013	-1.194	6	5.013	7.187	6	4.813	6.09
7	4.964	11.465	7	5.964	9.005	7	5.764	11.958
8	6.167	-6.257	8	7.167	14.865	8	6.967	14.975
9	7.658	-11.07	9	8.658	12.008	9	8.458	18.518
10	8.243	10.243	10	9.243	11.718	10	9.043	21.794
11	9.296	10.995	11	10.296	16.744	11	10.096	17.245
12	10.259	-11.17	12	11.259	18.789	12	11.059	20.881
13	11.275	-10.84	13	12.275	12.863	13	12.075	18.787
14	12.202	-9.78	14	13.202	20.862	14	13.002	16.334
15	12.687	15.066	15	13.687	15.309	15	13.487	28.613

Вариант 4			Вариант 5			Вариант 6		
№	X	Y	№	X	Y	№	X	Y
1	-3.132	0.412	1	1.868	12.669	1	-4.132	-4.539
2	-2.204	2.204	2	2.796	-9.23	2	-3.204	-3.306
3	-1.142	-1.282	3	3.858	0.753	3	-2.142	-2.018
4	-0.285	-2.529	4	4.715	-2.106	4	-1.285	-1.223
5	0.494	-2.995	5	5.494	-4.044	5	-0.506	-0.739
6	2.013	-0.325	6	7.013	1.903	6	1.013	-0.823
7	2.964	-5.018	7	7.964	5.065	7	1.964	-1.873
8	4.167	-6.295	8	9.167	3.807	8	3.167	1.055
9	5.658	-3.773	9	10.658	3.374	9	4.658	0.989
10	6.243	-6.772	10	11.243	0.563	10	5.243	0.116
11	7.296	-5.079	11	12.296	2.511	11	6.296	6.213
12	8.259	-5.519	12	13.259	1.934	12	7.259	0.856
13	9.275	-6.505	13	14.275	4.609	13	8.275	0.743
14	10.202	-7.864	14	15.202	9.673	14	9.202	5.137
15	10.687	-8.578	15	15.687	6.565	15	9.687	2.209

Вариант 7			Вариант 8			Вариант 9		
№	X	Y	№	X	Y	№	X	Y
1	1.868	4.322	1	-1.118	7.252	1	1.132	2.573

2	2.796	11.126	2	-1.062	10.683	2	3.204	14.963
3	3.858	4.536	3	0.054	4.075	3	5.142	4.317
4	4.715	11.52	4	2.359	5.463	4	7.285	17.39
5	5.494	15.855	5	3.561	11.377	5	9.506	26.456
6	7.013	19.365	6	3.56	13.401	6	10.987	30.205
7	7.964	20.027	7	6.348	12.638	7	13.036	32.238
8	9.167	15.986	8	6.617	18.804	8	14.833	25.086
9	10.658	23.784	9	8.108	8.975	9	16.342	36.133
10	11.243	22.948	10	8.538	26.612	10	18.757	37.62
11	12.296	18.02	11	8.312	23.302	11	20.704	29.791
12	13.259	31.75	12	10.895	21.718	12	22.741	53.538
13	14.275	29.538	13	10.381	20.979	13	24.725	50.432
14	15.202	24.758	14	12.18	17.243	14	26.798	43.52
15	15.687	27.018	15	12.973	19.821	15	29.313	50.701

Вариант 10		
№	X	Y
1	-1.121	-4.575
2	0.391	3.839
3	0.587	1.864
4	2.114	2.832
5	3.131	17.286
6	4.528	7.376
7	4.806	6.239
8	6.165	17.959
9	7.464	13.944
10	7.454	17.99
11	9.392	27.978
12	9.685	22.938
13	11.138	25.206
14	11.684	26.74
15	12.627	36.957

Приложение 6

Используя MSExcel (Ooo Calc) выполнить задания:

1. Шарики, изготовленные станком-автоматом, должны иметь диаметр 10 мм; проверить эту гипотезу по заданной выборке на уровне значимости 0.05, если:

А) Дисперсия известна и равна 0,1 мм²

В) Дисперсия неизвестна.

Результаты наблюдений приведены в таблице.

10,16	9,99	9,8	10,27	10,13	9,97	10,04	10,16	10,19	10,26
9,96	9,89	10,11	10,3	10,15	10,04	10,39	10,43	10,03	10,32

Воспользоваться стандартным алгоритмом проверки гипотез и встроенной функцией Z.ТЕСТ.

2. Проверить гипотезу об одинаковой точности работы станков по результатам измерений (точность характеризуется дисперсией соответствующего размера) на уровне значимости 0,05 с использованием формулы (1) и функции F.ТЕСТ. Результаты измерений контролируемого параметра на двух станках приведены в таблице.

№	Станок1	Станок2	№	Станок1	Станок2
1	12,05	12,36	13	12,05	12,47
2	12,08	12,45	14	12,08	12,41
3	12,33	12,48	15	12,33	12,34
4	12,34	12,56	16	12,05	12,51
5	12,75	12,63	17	12,08	12,45
6	12,32	12,25	18	12,31	12,24
7	12,12	12,54	19	12,34	12,55
8	12,05	12,35	20	12,42	12,32
9	12,08	12,54	21	12,42	12,44
10	12,33	12,33	22	12,12	12,41
11	12,08	12,85	23		12,38
12	12,75	12,42	24		12,51

3. Дана выборка из 100 наблюдений; определить числовые характеристики, построить гистограмму частот, проверить нормальность распределения по критерию хи-квадрат.

12,01	12	11,64	12,09	11,79
11,64	11,99	11,7	11,79	12,16
11,59	11,45	11,9	11,86	12,25
11,4	11,84	12,15	11,79	11,92
12,37	12	11,93	11,98	11,72
11,76	11,97	12,25	11,76	11,9
11,7	11,58	12,1	12,39	11,74
12,04	11,58	11,81	12,13	12,09
12,02	12,16	11,94	12,2	11,66
12,01	11,35	12,2	11,84	11,84

12,07	12,15	12,1	11,52	11,84
11,58	11,78	11,79	11,78	11,83
12,07	11,42	12,08	12,03	12,03
11,79	11,8	12,7	11,65	11,96
12,19	11,85	12,42	11,72	12,4
12,34	12,15	11,65	12,27	11,81
11,91	12,03	12,16	12,11	11,92

Для выполнения задания необходимо рассчитать и внести в таблицу следующие данные:

n_i – частота попаданий элементов выборки в i -й интервал;

x_i – верхняя граница i -го интервала;

$F(x_i)$ – значение функции нормального распределения;

ΔF_i – теоретическое значение вероятности попадания случайной величины в i -й интервал

$$\Delta F_i = \frac{1}{\sigma \sqrt{2\pi}} \int_{x_{i-1}}^{x_i} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_1)^2\right) dx = F_{\Phi}(x_i) - F_{\Phi}(x_{i-1}) =$$

$$= \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{x_i} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_1)^2\right) dx - \frac{1}{\sigma \sqrt{2\pi}} \int_{-\infty}^{x_{i-1}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu_1)^2\right) dx;$$

$F_i = \Delta F_i * n$ – теоретическая частота попадания случайной величины в i -й интервал; $(n_i - F_i)^2 / F_i$ – взвешенный квадрат отклонения.

Используемая литература

Основная литература

1. Энатская Н. Ю. Теория вероятностей и математическая статистика для инженерно-технических направлений: Учебник и практикум / Энатская Н. Ю., Хакимуллин Е. Р. – Электрон. дан. – Москва: Издательство Юрайт, 2019. – 399.
2. Гмурман В. Е. Теория вероятностей и математическая статистика: Учебник / Гмурман В. Е. – 12-е изд. – Электрон. дан. – Москва: Издательство Юрайт, 2019. – 479. – (Высшее образование). – Режим доступа: <https://www.biblio-online.ru>.
3. Севастьянов Б. А. Курс теории вероятностей и математической статистики / Б. А. Севастьянов. – Москва, Ижевск: Институт компьютерных исследований, 2019. – 272 с.
4. Махова Н. Б. Теория вероятностей и основы математической статистики: курс лекций / Н. Б. Махова. – Москва: Московская государственная академия водного транспорта, 2019. – 87 с.
5. Шилова З. В. Теория вероятностей и математическая статистика: учебное пособие / З. В. Шилова, О. И. Шилов. – Саратов: Ай Пи Ар Букс, 2015. – 158 с.

Дополнительная литература

1. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике: учеб. пособие для вузов / В.Е. Гмурман. – 11-е изд., перераб. – М.: Высш. образование, 2008. – 404с.
2. Гмурман В.Е. Теория вероятностей и математическая статистика: учеб. пособие для вузов / В.Е. Гмурман. – 12-е изд., перераб. – М.: Высш. образование, 2008. – 479с.
3. Симогин А. А. Специальные разделы высшей математики. Теория вероятностей и математическая статистика для инженера-исследователя: учебное пособие / А. А. Симогин. – Макеевка: Донбасская национальная академия строительства и архитектуры, ЭБС АСВ, 2019. – 169 с.
4. Энатская Н. Ю. Теория вероятностей и математическая статистика для инженерно-технических направлений: Учебник и практикум / Энатская Н. Ю., Хакимуллин Е. Р. – Электрон. дан. – Москва: Издательство Юрайт, 2019. – 399.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля)

1. Электронный читальный зал “БИБЛИОТЕХ”: учебники авторов ТулГУ по всем дисциплинам.- Режим доступа: <https://tsutula.bibliotech.ru/>
2. Единое окно доступа к образовательным ресурсам: портал [Электронный ресурс]. - Режим доступа : <http://window.edu.ru>.
3. ЭБС IPRBooks универсальная базовая коллекция изданий.- Режим доступа: <http://www.iprbookshop.ru/>

4. Научная Электронная Библиотека eLibrary – библиотека электронной периодики, режим доступа: <http://elibrary.ru/>