ГОСУДАРСТВЕННОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ПРОФЕССИОНАЛЬНОГО ОБРАЗОВАНИЯ
«БЕЛОРУССКО-РОССИЙСКИЙ УНИВЕРСИТЕТ»

Кафедра «Высшая математика»

ВЫСШАЯ МАТЕМАТИКА. МАТЕМАТИКА. МАТЕМАТИКА (СПЕЦГЛАВЫ). СПЕЦГЛАВЫ МАТЕМАТИКИ. ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. ТЕОРИЯ ВЕРОЯТНОСТЕЙ, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА И СЛУЧАЙНЫЕ ПРОЦЕССЫ

Методические рекомендации к практическим занятиям для студентов всех специальностей и направлений подготовки дневной и заочной форм обучения

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА: СТАТИСТИЧЕСКАЯ ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ, ЭЛЕМЕНТЫ ТЕОРИИ КОРРЕЛЯЦИИ

Могилёв 2018

УДК 519.22
ББК 22.172
В 93

Рекомендовано к изданию
учебно-методическим отделом
Белорусско-Российского университета

Одобрено кафедрой «Высшая математика» «29» июня 2018 г., протокол № 10

Составители: ст. преподаватель Т. Ю. Орлова;
ст. преподаватель С. Ф. Плешкунова;
доц. Д. В. Роголев

Рецензент И. Д. Камчицкая

Изложены теоретические вопросы, разобраны образцы решения примеров, приведены задачи для самостоятельного решения и задания для домашней работы.

Учебно-методическое издание

ВЫСШАЯ МАТЕМАТИКА. МАТЕМАТИКА. МАТЕМАТИКА (СПЕЦГЛАВЫ). СПЕЦГЛАВЫ МАТЕМАТИКИ. ТЕОРИЯ ВЕРОЯТНОСТЕЙ И МАТЕМАТИЧЕСКАЯ СТАТИСТИКА. ТЕОРИЯ ВЕРОЯТНОСТЕЙ, МАТЕМАТИЧЕСКАЯ СТАТИСТИКА И СЛУЧАЙНЫЕ ПРОЦЕССЫ

Ответственный за выпуск	В. Г. Замураев
Технический редактор	А. А. Подошевко
Компьютерная вёрстка	Н. П. Полевничая

Подписано в печать . Формат 60×84/16. Бумага офсетная. Гарнитура Таймс.
Печать трафаретная. Усл. печ. л. . Уч.-изд. л. . Тираж 105 экз. Заказ № .

Издатель и полиграфическое исполнение:
Государственное учреждение высшего профессионального образования
«Белорусско-Российский университет».
Свидетельство о государственной регистрации издателя,
изготовителя, распространителя печатных изданий
№ 1/156 от 24.01.2014.
Пр. Мира, 43, 212000, Могилёв.

Содержание

1 Статистические гипотезы. Проверка параметрических гипотез

1.1 Теоретическая часть

Основные понятия.

Статистической гипотезой называется любое предположение о генеральной совокупности, проверяемое по выборке.

Статистические гипотезы делятся на:

– *параметрические* – гипотезы о параметрах распределения известного вида;
– *непараметрические* – гипотезы о виде неизвестного распределения.

Нулевой (основной) называют выдвинутую гипотезу H_0.

Конкурирующей (альтернативной) называют гипотезу H_1, которая противоречит нулевой.

Простой называют гипотезу, содержащую только одно предположение.

Сложной называют гипотезу, которая состоит из конечного или бесконечного числа простых гипотез.

При проверке гипотезы могут быть допущены ошибки первого и второго рода.

Ошибка первого рода состоит в том, что будет отвергнута правильная нулевая гипотеза. Вероятность допустить ошибку первого рода называется ***уровнем значимости*** и обозначается α, т. е. $\alpha = P\left(H_1 \middle| H_0\right)$. Часто уровень значимости принимается равным 0,05, 0,01, 0,001.

Ошибка второго рода состоит в том, что будет принята неправильная нулевая гипотеза. Вероятность допустить ошибку второго рода равна β, т. е. $\beta = P\left(H_0 \middle| H_1\right)$, тогда $1-\beta = P\left(H_1 \middle| H_1\right)$. Величина $1-\beta$ называется ***мощностью критерия.***

Проверка статистических гипотез осуществляется на основе данных выборки. Для этого используют специальным образом подобранную случайную величину (выборочную статистику). Её обозначают U или Z – если случайная величина распределена по нормальному закону; F – если она имеет распределение Фишера – Снедекора; T – если она имеет распределение Стьюдента; χ^2 – если она имеет распределение «хи-квадрат» Пирсона. В общем случае будем обозначать эту случайную величину K.

Статистическим критерием называют случайную величину K, которая служит для проверки нулевой гипотезы.

Наблюдаемым значением $K_{набл}$ называют значение критерия, вычисленное по выборкам.

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

Критической областью называют совокупность значений критерия, при которых нулевую гипотезу отвергают.

Областью принятия гипотезы называют совокупность значений критерия, при которых гипотезу принимают.

Основной принцип проверки статистических гипотез: если наблюдаемое значение критерия принадлежит критической области, то гипотезу отвергают; если наблюдаемое значение критерия принадлежит области принятия гипотезы – гипотезу принимают.

Критическими точками $k_{кр}$ называют точки, отделяющие критическую область от области принятия гипотезы.

Правосторонней называют критическую область, определяемую неравенством $K > k_{кр}$, где $k_{кр} > 0$.

Левосторонней называют критическую область, определяемую неравенством $K < k_{кр}$, где $k_{кр} < 0$.

Это ***односторонние*** области.

Двусторонней называют критическую область, определяемую неравенствами $K < k_1, K > k_2$, где $k_2 > k_1$. В частности, если критические точки симметричны относительно нуля, то двусторонняя критическая область определяется неравенством $|K| > k_{кр}$, где $k_{кр} > 0$.

Мощность критерия – вероятность попадания критерия в критическую область при условии, что справедлива конкурирующая гипотеза. Таким образом, если уровень значимости выбран, то критическую область следует строить так, чтобы мощность критерия была максимальной. Выполнение этого требования должно обеспечить минимальную ошибку второго рода.

Этапы проверки гипотезы.

1 Формулируют нулевую H_0 и альтернативную H_1 гипотезы.

2 Выбирают критерий $K = K\left(X_1, X_2, ..., X_n\right)$.

3 По выборке $x_1, x_2, ..., x_n$ вычисляют наблюдаемое значение критерия $K_{набл} = K\left(x_1, x_2, ..., x_n\right)$.

4 Выделяют критическую область и область принятия гипотезы.

Для отыскания критической области задают уровень значимости α и ищут критические точки исходя из следующих соотношений:

– для правосторонней критической области

$$P\left(K > k_{кр}\right) = \alpha\text{, где } k_{кр} > 0;$$

– для левосторонней критической области

$$P\left(K < k_{кр}\right) = \alpha\text{, где } k_{кр} < 0;$$

– для двусторонней области

$$P(K<k_1)+P(K>k_2)=\alpha,\ \text{где}\ k_1<k_2;$$

– для двусторонней симметричной области

$$P(K>k_{кр})=P(K<-k_{кр})=\frac{\alpha}{2},\ \text{где}\ k_{кр}>0.$$

5 Если наблюдаемое значение критерия принадлежит области принятия гипотезы, то нет оснований отвергнуть нулевую гипотезу; если же оно принадлежит критической области, то нулевую гипотезу отвергают в пользу альтернативной.

Проверка параметрических гипотез

1 Проверка гипотезы о математическом ожидании нормально распределённой случайной величины.

Рассмотрим генеральную совокупность X нормального распределения с параметрами m и σ, т. е. $X\in N(m;\sigma)$. Нужно проверить нулевую гипотезу о том, что $m=m_0$ ($H_0: m=m_0$). Можно сформулировать две основные модели и построить для них соответствующие критерии значимости.

Модель 1. Среднее квадратическое отклонение σ известно.

Пусть генеральная совокупность имеет нормальное распределение: $X\in N(m;\sigma)$, причём σ известно. На основании случайной выборки $(x_1,\ldots,x_n)$ из этой генеральной совокупности требуется проверить нулевую гипотезу $H_0: m=m_0$ при заданном уровне значимости α.

Для проверки значимости отличия выборочной средней $\bar{x}_B$ в случайной выборке объёмом n от математического ожидания m_0 генеральной совокупности выбирается критерий

$$U=\frac{(\bar{X}-m_0)\sqrt{n}}{\sigma}.$$

$U\in N(0;1)$ – нормированная нормально распределённая случайная величина.

Критическую точку $u_{кр}$ находят по таблице значений функции Лапласа.

По выборке вычисляют наблюдаемое значение критерия

$$U_{набл}=\frac{(\bar{x}_B-m_0)\sqrt{n}}{\sigma}.$$

Критическая область строится в зависимости от вида альтернативной гипотезы.

Правило 1. Альтернативная гипотеза $H_1: m > m_0$.

Критическая область правосторонняя $\left(u_{кр};+\infty\right)$.

Определяем $u_{кр}$ из условия $\Phi\left(u_{кр}\right)=\dfrac{1-2\alpha}{2}$.

Если $U_{набл} < u_{кр}$, то нет оснований отвергнуть нулевую гипотезу.

Если $U_{набл} > u_{кр}$, то нулевую гипотезу отвергают.

Правило 2. Альтернативная гипотеза $H_1: m < m_0$.

Критическая область левосторонняя $\left(-\infty;-u_{кр}\right)$.

Определяем $u_{кр}$ из условия $\Phi\left(u_{кр}\right)=\dfrac{1-2\alpha}{2}$.

Если $U_{набл} > -u_{кр}$, то нет оснований отвергнуть нулевую гипотезу.

Если $U_{набл} < -u_{кр}$, то нулевую гипотезу отвергают.

Правило 3. Альтернативная гипотеза $H_1: m \neq m_0$.

Критическая область двусторонняя $\left(-\infty;-u_{кр}\right)\cup\left(u_{кр};+\infty\right)$.

Определяем $u_{кр}$ из условия $\Phi\left(u_{кр}\right)=\dfrac{1-\alpha}{2}$.

Если $\left|U_{набл}\right| < u_{кр}$, то нет оснований отвергнуть нулевую гипотезу.

Если $\left|U_{набл}\right| > u_{кр}$, то нулевую гипотезу отвергают.

Модель 2. Среднее квадратическое отклонение σ неизвестно.

Пусть генеральная совокупность имеет нормальное распределение: $X \in N\left(m;\sigma\right)$, причём σ неизвестно. На основании случайной выборки $\left(x_1,\ldots,x_n\right)$ из этой генеральной совокупности требуется проверить нулевую гипотезу $H_0: m = m_0$ при заданном уровне значимости α.

Для проверки нулевой гипотезы применяется случайная величина

$$T=\frac{\left(\bar{X}-m_0\right)\sqrt{n}}{S}.$$

Величина T имеет ***распределение Стьюдента*** с $k=n-1$ степенями свободы.

Вычисляется наблюдаемое значение критерия $T_{набл}=\dfrac{\left(\bar{x}_B-m_0\right)\sqrt{n}}{s}$.

Критическая область строится в зависимости от конкурирующей гипотезы.

Правило 1. Альтернативная гипотеза $H_1: m > m_0$.

Критическая область правосторонняя $\left(t_{кр};+\infty\right)$.

По таблице критических точек распределения Стьюдента находят критическую точку $t_{кр}=t_{\alpha,k}$, где $k=n-1$, α – заданный уровень значимости.

Если $T_{набл} < t_{кр}$, то нет оснований отвергнуть нулевую гипотезу.

Если $T_{набл} > t_{кр}$, то нулевую гипотезу отвергают.

Правило 2. Альтернативная гипотеза $H_1 : m < m_0$.

Критическая область левосторонняя $\left(-\infty; -t_{кр}\right)$.

$t_{кр}$ определяется, как и в случае правосторонней критической области.

Если $T_{набл} > -t_{кр}$, то нет оснований отвергнуть нулевую гипотезу.

Если $T_{набл} < -t_{кр}$, то нулевую гипотезу отвергают.

Правило 3. Альтернативная гипотеза $H_1 : m \neq m_0$.

Критическая область двусторонняя $\left(-\infty; -t_{кр}\right) \cup \left(t_{кр}; +\infty\right)$.

По таблице критических точек распределения Стьюдента находят критическую точку $t_{кр} = t_{\frac{\alpha}{2},k}$.

Если $\left|T_{набл}\right| < t_{кр}$, то нет оснований отвергнуть нулевую гипотезу.

Если $\left|T_{набл}\right| > t_{кр}$, то нулевую гипотезу отвергают.

2 Проверка гипотезы о дисперсии случайной величины, распределённой по нормальному закону.

Дисперсия характеризует такие важные технологические и конструкторские показатели, как точность машин, погрешность показаний контрольно-измерительных приборов, ритмичность производства, устойчивость работы автоматических линий и др.

Пусть случайная величина X распределена по нормальному закону. Генеральная дисперсия неизвестна, но есть основания по теоретическим предположениям или по предыдущим опытам считать её равной σ_0^2. Из генеральной совокупности производится выборка объёмом n и вычисляется исправленная выборочная дисперсия s^2. Требуется по заданному уровню значимости α проверить нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2$.

В качестве критерия оценки нулевой гипотезы используют случайную величину

$$\chi^2 = \frac{(n-1)\,s^2}{\sigma_0^2},$$

которая имеет ***распределение «хи-квадрат» Пирсона*** с $k = n-1$ степенями свободы.

По выборке вычисляют наблюдаемое значение критерия

$$\chi^2_{набл} = \frac{(n-1)\,s^2}{\sigma_0^2}.$$

Критическая область строится в зависимости от альтернативной гипотезы.

Правило 1. Альтернативная гипотеза $H_1: \sigma^2 > \sigma_0^2$.

В этом случае правостороннюю критическую область строят из условия $P\left(\chi^2 > \chi^2_{кр}\right) = \alpha$. По таблице критических точек распределения χ^2, по заданному уровню значимости α и числу степеней свободы $k = n-1$ находят критическую точку $\chi^2_{кр} = \chi^2(\alpha, k)$.

Если $\chi^2_{набл} < \chi^2_{кр}$, то нет оснований отвергнуть нулевую гипотезу.

Если $\chi^2_{набл} > \chi^2_{кр}$, то нулевую гипотезу отвергают.

Правило 2. Альтернативная гипотеза $H_1: \sigma^2 < \sigma_0^2$.

В этом случае правостороннюю критическую область строят из условия $P\left(\chi^2 > \chi^2_{кр}\right) = 1-\alpha$. По таблице критических точек распределения χ^2, по заданному уровню значимости α и числу степеней свободы $k = n-1$ находят критическую точку $\chi^2_{\text{кр}} = \chi^2(1-\alpha, k)$.

Если $\chi^2_{набл} > \chi^2_{кр}$, то нет оснований отвергнуть нулевую гипотезу.

Если $\chi^2_{набл} < \chi^2_{кр}$, то нулевую гипотезу отвергают.

Правило 3. Альтернативная гипотеза $H_1: \sigma^2 \neq \sigma_0^2$.

В этом случае правостороннюю критическую область строят из условия $P\left(\chi^2 > \chi^2_{пр.\,кр}\right) = \dfrac{\alpha}{2}$. Левостороннюю критическую область строят из требования $P\left(\chi^2 > \chi^2_{лев.\,кр}\right) = 1-\dfrac{\alpha}{2}$. По таблице критических точек распределения χ^2 находят правостороннюю и левостороннюю критические точки $\chi^2_{пр.\,кр} = \chi^2\left(\dfrac{\alpha}{2}, k\right)$, $\chi^2_{лев.\,кр} = \chi^2\left(1-\dfrac{\alpha}{2}, k\right)$.

Если $\chi^2_{лев.\,кр} < \chi^2_{набл} < \chi^2_{пр.\,кр}$, то нет оснований отвергнуть нулевую гипотезу. В противном случае нулевую гипотезу отвергают.

Замечание **–** Если число степеней свободы $k > 30$, то критическую точку $\chi^2(\alpha, k)$ можно найти из ***равенства Уилсона-Гильферти***

$$\chi^2(\alpha, k) = k\left(1 - \frac{2}{9k} + z_\alpha \cdot \sqrt{\frac{2}{9k}}\right)^3,$$

где z_α находят, используя функцию Лапласа, из условия $\Phi(z_\alpha) = \dfrac{1-2\alpha}{2}$.

3 Проверка гипотез о равенстве математических ожиданий двух случайных величин, распределённых по нормальному закону.

Пусть исследуются две случайные величины X и Y, которые имеют нормальное распределение, т. е. $X \in N\left(m_X, \sigma_X\right)$, $Y \in N\left(m_Y, \sigma_Y\right)$, причём σ_X и σ_Y известны, а m_X и m_Y неизвестны. Производятся две независимые выборки объёмами n_1 и n_2 $\left(n_1 > 30,\ n_2 > 30\right)$, взятые из двух независимых генеральных совокупностей, и по ним находятся выборочные средние $\overline{x}_B$ и $\overline{y}_B$. Выдвигается нулевая гипотеза $m_X = m_Y$.

Для проверки нулевой гипотезы используется нормированная нормальная случайная величина

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\dfrac{D(X)}{n_1} + \dfrac{D(Y)}{n_2}}}.$$

$Z \in N\left(0;1\right)$ – нормированная нормально распределённая случайная величина.

Критическую точку $z_{кр}$ находят по таблице функции Лапласа.

По выборке вычисляют наблюдаемое значение критерия

$$Z_{набл} = \frac{\overline{x}_B - \overline{y}_B}{\sqrt{\dfrac{D(X)}{n_1} + \dfrac{D(Y)}{n_2}}}.$$

Критическая область строится в зависимости от вида альтернативной гипотезы.

Правило 1. Альтернативная гипотеза $H_1 : m_X > m_Y$.

Критическая область правосторонняя.

Определяем $z_{пр.\,кр}$ из условия $\Phi\left(z_{кр}\right) = \dfrac{1 - 2\alpha}{2}$.

Правило 2. Альтернативная гипотеза $H_1 : m_X < m_Y$.

Критическая область левосторонняя, $z_{лев.\,кр} = -z_{пр.\,кр}$.

Правило 3. Альтернативная гипотеза $H_1 : m_X \neq m_Y$.

Критическая область двусторонняя.

Определяем $z_{пр.кр}$ из условия $\Phi\left(z_{кр}\right) = \dfrac{1 - \alpha}{2}$, $z_{лев.\,кр} = -z_{пр.\,кр}$.

Если средние квадратические отклонения неизвестны, но можно предположить, что они равны, т. е. $\sigma_1 = \sigma_2$, то наблюдаемое значение критерия находят по формуле

$$T_{набл} = \frac{\overline{x}_B - \overline{y}_B}{\sqrt{\left(n_1 - 1\right) s_X^2 + \left(n_2 - 1\right) s_Y^2}} \sqrt{\frac{n_1 n_2 \left(n_1 + n_2 - 2\right)}{n_1 + n_2}},$$

где s_X^2 и s_Y^2 – исправленные выборочные дисперсии. По таблице критических точек распределения Стьюдента, по заданному уровню α находят критическую точку в зависимости от вида альтернативной гипотезы: $t_{пр.кр} = t(\alpha, k)$ – для правосторонней критической области, $t_{лев.кр} = -t(\alpha, k)$ – для левосторонней критической области, $t_{пр.кр} = t\left(\dfrac{\alpha}{2}, k\right)$ – для двусторонней критической области, где число степеней свободы $k = n_1 + n_2 - 2$.

Замечание – Если равенство $\sigma_1 = \sigma_2$ не может быть обосновано из общих соображений, то оно подлежит предварительной проверке с помощью критерия Фишера–Снедекора (см. далее).

4 Проверка гипотез о равенстве дисперсий двух случайных величин, распределённых по нормальному закону.

На практике задача сравнения дисперсий возникает, если требуется сравнить точность приборов, инструментов, самих методов измерений и т. д. Очевидно, предпочтительнее тот прибор, инструмент и метод, который обеспечивает наименьшее рассеяние результатов измерений, т. е. наименьшую дисперсию.

Пусть генеральные совокупности X и Y распределены нормально. По независимым выборкам с объёмами, соответственно равными n_1 и n_2, извлечённым из этих совокупностей, найдены исправленные выборочные дисперсии s_X^2 и s_Y^2. Требуется по исправленным дисперсиям при заданном уровне значимости α проверить нулевую гипотезу $H_0 : D(X) = D(Y)$.

В качестве критерия проверки нулевой гипотезы принимают случайную величину $F = \dfrac{S_б^2}{S_м^2}$.

Величина F имеет ***распределение Фишера–Снедекора*** со степенями свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$, где n_1 – объём выборки, по которой вычислена большая дисперсия $S_б^2$; n_2 – объём выборки, по которой вычислена меньшая дисперсия $S_м^2$.

Наблюдаемое значение критерия находят по формуле

$$F_{набл} = \frac{s_б^2}{s_м^2}.$$

Критическая область строится в зависимости от вида конкурирующей гипотезы.

Правило 1. Альтернативная гипотеза $H_1 : D(X) > D(Y)$.

Критическая область правосторонняя, строится исходя из требования $P\left(F > F_{кр}\left(\alpha, k_1, k_2\right)\right) = \alpha$.

Критическую точку $F_{кр} = F\left(\alpha, k_1, k_2\right)$ находят по таблице критических точек распределения Фишера – Снедекора.

Если $F_{набл} < F_{кр}$, нет оснований отвергать нулевую гипотезу.

Правило 2. Альтернативная гипотеза $H_1 : \mathrm{D}(X) \neq D(Y)$.

Критическая область двусторонняя. Правостороннюю критическую точку $F_{кр} = F\left(\dfrac{\alpha}{2}, k_1, k_2\right)$ находят по таблице критических точек распределения Фишера–Снедекора.

Если $F_{набл} < F_{кр}$, нет оснований отвергать нулевую гипотезу.

1.2 Образцы решения примеров

Пример 1 – Для проверки партии электроламп, средний срок горения которых должен быть не меньше, чем 1250 ч, отобрано 50 шт. Оказалось, что средний срок горения для них $\overline{x}_B = 1244$, а среднее квадратическое отклонение продолжения горения 85 ч. Есть ли основания забраковать эту партию электроламп? Уровень значимости принять равным 0,05.

Решение

Выдвигается нулевая гипотеза $H_0 : a = 1250$ против альтернативной $H_1 : a < 1250$. По выборке получили $\overline{x}_в = 1244$.

Вычисляем $U_{набл} = \dfrac{\left(\overline{x}_B - m_0\right)\sqrt{n}}{\sigma} = \dfrac{(1244 - 1250)\sqrt{50}}{85} = -0,5$.

Критическую точку $u_{кр}$ находим из условия $\Phi\left(u_{кр}\right) = \dfrac{1 - 2\alpha}{2}$.

$\Phi\left(u_{кр}\right) = 0,45$. По таблице функции Лапласа найдём $u_{кр} = 1,65$. Для левосторонней критической области $u_{кр} = -1,65$. Так как $U_{набл} > -u_{кр}$, то нет оснований отклонить нулевую гипотезу.

Таким образом, срок горения электроламп не меньше, чем 1250 ч.

Пример 2 – При разработке норм выработки на предприятии проведено 25 независимых измерений производительности труда рабочих, выполняющих определённую операцию. Средняя производительность труда составила $\overline{x}_B = 5,2$ единицы товара за 1 чел./ч, а исправленное среднее квадратическое отклонение $s = 0,4$ единицы товара за 1 чел./ч. Требуется проверить гипотезу, что при массовом выпуске этой продукции средняя производительность труда m составляет 5,5 единицы товара за 1 чел./ч при альтернативной гипотезе $m < 5,5$ и уровне значимости $\alpha = 0,01$.

Решение

$H_0 : m = 5,5$, $H_1 : m < 5,5$.

Так как среднее квадратическое отклонение генеральной совокупности неизвестно, наблюдаемое значение критерия будем искать по формуле

$$T_{набл} = \frac{(\overline{x}_B - m_0)\sqrt{n}}{s} = \frac{(5,2-5,5)\cdot\sqrt{25}}{0,4} = -3,75.$$

По виду альтернативной гипотезы строим левостороннюю критическую область.

По таблице критических точек распределения Стьюдента находим критическую точку $t_{кр} = t(0,01,24) = 2,492$. Так как $T_{набл} < -t_{кр}$, то нулевую гипотезу отвергаем в пользу альтернативной, т. е. средняя производительность труда меньше 5,5 единицы товара за 1 чел./ч.

Пример 3 – Погрешность показаний некоторого прибора не должна превышать 0,2. Для проверки этого прибора проведено 16 замеров, найдено $s^2 = 0,05$. При уровне значимости $\alpha = 0,02$ проверить гипотезу о точной работе прибора.

Решение

$H_0 : \sigma^2 = \sigma_0^2 = 0,04$, $H_1 : \sigma^2 \neq 0,04$.

Найдём наблюдаемое значение критерия

$$\chi^2_{набл} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{15\cdot 0,05}{0,04} = 18,75.$$

Критическая область двусторонняя. По таблице критических точек распределения χ^2 находим критические точки:

$$\chi^2_{лев.\,кр} = \chi^2\left(1-\frac{\alpha}{2}, k\right) = \chi^2(0,99;15) = 5,23;\ \chi^2_{пр.\,кр} = \chi^2\left(\frac{\alpha}{2}, k\right) = \chi^2(0,01;15) = 30,6.$$

Так как $\chi^2_{\text{лев. кр}} < \chi^2_{\text{набл}} < \chi^2_{\text{пр. кр}}$ $(5,23 < 18,75 < 30,6)$, то нет оснований отклонять нулевую гипотезу.

Значит, исправленная выборочная дисперсия незначительно отличается от $\sigma_0^2 = 0,04$, т. е. прибор обеспечивает заданную точность.

Пример 4 – Выполнены две выборки по 50 электроламп заводов *A* и *B*. Средняя продолжительность работы электролампы соответственно $\overline{x}_B = 1282$ ч, $\overline{y}_B = 1208$ ч со средними квадратическими отклонениями $\sigma_X = 80$ ч и $\sigma_Y = 94$ ч.

Проверить гипотезу о том, что заводы выпускают лампы одинакового качества при уровне значимости $\alpha = 0,01$.

Решение

Нулевая гипотеза $H_0 : m_X = m_Y$. В качестве альтернативной выберем гипотезу $H_1 : m_X \neq m_Y$.

Вычисляем наблюдаемое значение критерия

$$Z_{набл} = \frac{\overline{x}_B - \overline{y}_B}{\sqrt{\dfrac{\sigma_X^2}{n_1} + \dfrac{\sigma_Y^2}{n_2}}} = \frac{1282 - 1208}{\sqrt{\dfrac{80^2}{50} + \dfrac{94^2}{50}}} = 13,5.$$

По таблице функции Лапласа найдём критическую точку из условия

$$\Phi\left(z_{кр}\right) = \frac{1-\alpha}{2} = 0,495,\ z_{кр} = 2,58.$$

Так как $\left|Z_{набл}\right| > z_{кр}$, то нулевую гипотезу отвергаем. Таким образом, выборочные средние имеют существенные отклонения и качество выпускаемых заводами ламп разное.

Пример 5 – Пусть на предприятии разработаны два метода изготовления изделия. Чтобы проверить, одинаково ли материалоёмки эти методы, в таблице 1 собраны статистические данные о расходе сырья на единицу готовой продукции в процессе работы обоими методами.

Таблица 1

X (I метод)	2,0	2,7	2,5	2,9	2,3	2,6
Y (II метод)	2,5	3,2	3,5	3,8	3,5	–

Предполагая, что средние квадратические отклонения обоих методов равны, проверить нулевую гипотезу $H_0 : m_X = m_Y$ при альтернативной гипотезе $H_1 : m_X < m_Y$. Принять уровень значимости равным 0,05.

Решение

Используя экспериментальные данные, получаем

$$n_1 = 6;\ n_2 = 5;\ k = n_1 + n_2 - 2 = 9;\ \overline{x}_B = \frac{1}{n_1}\sum_i x_i = 2,5;\ \overline{y}_B = \frac{1}{n_2}\sum_i y_i = 3,3;$$

$$s_X^2 = \frac{1}{n_1 - 1}\sum_i \left(x_i - \overline{x}_B\right)^2 = 0{,}1;\ s_Y^2 = \frac{1}{n_2 - 1}\sum_i \left(y_i - \overline{y}_B\right)^2 = 0{,}245;$$

$$T_{набл} = \frac{2{,}5 - 3{,}3}{\sqrt{5 \cdot 0{,}1 + 4 \cdot 0{,}245}}\sqrt{\frac{6 \cdot 5 \cdot (6 + 5 - 2)}{6 + 5}} = -3{,}258.$$

Критическая область левосторонняя. По таблице критических точек распределения Стьюдента найдём $t_{лев.кр} = t\left(0{,}05;9\right) = 1{,}83$.

Так как $T_{набл} < t_{лев.кр}$, то гипотезу H_0 отвергаем, т. е. можно сделать вывод, что материалоёмкость первого метода меньше.

Пример 6 – До наладки станка была проверена точность изготовления 10 втулок и найдено значение оценки дисперсии диаметра $s_X^2 = 9{,}6$ мкм^2. После наладки подверглись контролю ещё 15 втулок и получено новое значение оценки дисперсии $s_Y^2 = 5{,}7$ мкм^2. Можно ли считать, что в результате наладки станка точность изготовления деталей увеличилась? Принять $\alpha = 0{,}05$.

Решение

Выдвигаем нулевую гипотезу $H_0 : D(X) = D(Y)$ и альтернативную $H_1 : D(X) > (Y)$.

Вычисляем наблюдаемое значение критерия $F_{набл} = \dfrac{s_б^2}{s_м^2} = \dfrac{9{,}6}{5{,}7} = 1{,}68$.

Строим правостороннюю критическую область.

По таблице критических точек распределения Фишера–Снедекора находим $F_{кр} = F\left(0{,}05;9;14\right) = 2{,}65$.

Так как $F_{набл} < F_{кр}$, то нет оснований отклонять нулевую гипотезу. Следовательно, в результате наладки станка точность изготовления детали не увеличилась.

1.3 Задачи для самостоятельной работы

1 По результатам девяти замеров установлено, что выборочное среднее время (в секундах) изготовления детали $\overline{x}_B = 48$. Предполагая, что время изготовления – нормально распределённая случайная величина с дисперсией $\sigma^2 = 9$, рассмотреть на уровне значимости 0,05 гипотезу $H_0 : m = 49$ против альтернативной гипотезы $H_1 : m \neq 49$. (Ответ: $U_{набл} = -1$, $u_{кр} = 1{,}96$. Нулевая гипотеза принимается.)

2 По техническим условиям средняя прочность троса составляет 2000 кг. В результате испытаний 20 кусков троса было установлено, что средняя прочность на разрыв равна 1955 кг при средней ошибке 25 кг. Удовлетворяет ли об-

разец троса техническим условиям? (Ответ: $U_{набл} = -8{,}05,\ u_{кр} = 1{,}96$. Образец троса не удовлетворяет техническим условиям.)

3 По утверждению руководства фирмы, средний размер дебиторского счета равен 187,5 тыс. р. Ревизор составляет случайную выборку из 10 счетов и обнаруживает, что средняя арифметическая выборки равна 175 тыс. р. при среднем квадратическом отклонении 35 тыс. р. Может ли оказаться в действительности правильным объявленный размер дебиторского счета? Принять уровень значимости равным 0,05. (Ответ: $T_{набл} = -1{,}129$, $t_{кр}\left(0{,}05;\ 9\right) = 2{,}26$. Нет оснований считать неправильным объявленный размер дебиторского счета.)

4 Хронометраж затрат времени на сборку узла машины 25 слесарями показал, что среднее время сборки узла $\overline{x}_B = 16$ мин, а исправленное среднее квадратическое отклонение $s = 4{,}5$. В предположении о нормальном распределении решить вопрос о том, можно ли на уровне значимости $\alpha = 0{,}05$ считать 15 мин нормативом. (Ответ: $T_{набл} = 1{,}11,\ t_{кр} = t\left(0{,}25;\ 24\right) = 2{,}064$. Нет оснований не считать 15 мин нормативом.)

5 Из нормальной генеральной совокупности извлечена выборка объёма 21 и по ней найдена исправленная выборочная дисперсия 16,2. Требуется при уровне значимости 0,01 проверить нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2 = 15$, приняв в качестве конкурирующей гипотезы $H_1 : \sigma^2 > 15$. (Ответ: $\chi^2_{набл} = 21{,}6$, $\chi^2_{кр} = \chi^2\left(0{,}05;\ 20\right) = 31{,}671$. Нет оснований отвергать нулевую гипотезу.)

6 Новый метод измерения длины деталей был опробован на эталоне, причём дисперсия результатов измерений, определённая по 10 замерам, составила 100 мкм^2. Согласуется ли этот результат с утверждением: «дисперсия результатов измерений по предложенному методу не превосходит 50 мкм^2»? Принять $\alpha = 0{,}05$. (Ответ: $\chi^2_{набл} = 20$, $\chi^2_{кр} = \chi^2\left(0{,}05;\ 9\right) = 16{,}9$. Дисперсия результатов измерений по предложенному методу превосходит 50 мкм^2.)

7 Точность наладки станка-автомата, производящего некоторые детали, характеризуется дисперсией длины деталей. Если эта величина будет больше 400 мкм^2, станок останавливается для наладки. Выборочная дисперсия длины 15 случайно отобранных деталей из продукции станка оказалась равной 680 мкм^2. Нужно ли производить наладку станка, если уровень значимости $\alpha = 0{,}01$? (Ответ: $\chi^2_{набл} = 25{,}5$, $\chi^2_{пр.кр} = \chi^2\left(0{,}01;\ 14\right) = 29{,}1$. Наладку станка производить не нужно.)

8 В результате длительного хронометража времени сборки узла различными сборщиками установлено, что дисперсия этого времени $\sigma_0^2 = 2$ мин^2. Результаты 20 наблюдений за работой новичка собраны в таблице 2.

Таблица 2

Время сборки одного узла x_i, мин	56	58	60	62	64
Частота n_i	1	4	10	3	2

Можно ли при уровне значимости 0,05 считать, что новичок работает ритмично (в том смысле, что дисперсия затрачиваемого им времени существенно не отличается от дисперсии времени остальных сборщиков)? (Ответ: $s^2 = 3,99$, $\chi^2_{набл} = 37,91$, $\chi^2_{лев.кр} = \chi^2(0,975; 19) = 8,91$, $\chi^2_{пр.кр} = \chi^2(0,025; 19) = 32,9$. Новичок работает неритмично.)

9 По выборке объёма $n_1 = 50$ найден средний размер $\overline{x}_B = 20{,}1$ мм диаметра валиков, изготовленных станком-автоматом № 1; по выборке объёма $n_2 = 50$ найден средний размер $\overline{y}_B = 19{,}8$ мм диаметра валиков, изготовленных станком-автоматом № 2. Генеральные дисперсии известны: $D(X) = 1{,}750$ мм^2 и $D(Y) = 1{,}375$ мм^2. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0: m_X = m_Y$ при конкурирующей гипотезе $H_1: m_X \neq m_Y$. Предполагается, что СВ X и Y распределены нормально и выборки независимы. (Ответ: $Z_{набл} = 1{,}2$, $z_{кр} = 1{,}96$. Нулевая гипотеза принимается.)

10 Для проверки новой технологии были выбраны две группы рабочих численностью $n_1 = 40$ человек и $n_2 = 50$ человек. В первой группе при применении старой технологии средняя выработка составила $\overline{x}_1 = 85$ изделий, во второй, где применялась новая технология, $\overline{x}_2 = 95$. Дисперсии по группам $\sigma_1^2 = 100$, $\sigma_1^2 = 75$ были известны заранее. Выяснить на уровне значимости $\alpha = 0{,}05$ влияние новой технологии на производительность. (Ответ: $Z_{набл} = -5$, $z_{кр} = 1{,}96$. Новая технология повышает производительность.)

11 Средний годовой оборот пяти компаний в регионе А составил 4900 усл. ед., средний оборот 10 компаний в регионе В составил 5000 усл. ед. Выборочная дисперсия оборота компаний в регионе А оказалась равной 1000, а в регионе В – 4000. Считая, что дисперсии среднегодовых оборотов одинаковы, проверить на уровне значимости 0,05 гипотезу о равенстве средних значений в регионах А и В. (Ответ: $T_{набл} = -3{,}103$, $t_{кр}(0{,}05; 23) = 1{,}7$. Средний годовой оборот компаний в регионе В больше, чем в регионе А.)

12 В результате проверки 10 продавцов одной из торговых точек города были обнаружены недовесы со средним значением 150 г и выборочной дисперсией 2500. В другой точке недовесы характеризовались средним значением 125 г и выборочной дисперсией 1600 среди выборки из 15 продавцов. Выяснить на уровне доверия 0,95, в какой точке предпочтительнее покупать продукцию.

Указание – Предварительно проверить равенство дисперсий, используя критерий Фишера–Снедекора. (Ответ: $T_{набл} = 1{,}326$, $t_{кр}(0{,}05; 23) = 1{,}7$. На данном уровне доверия недовесы одинаковы.)

13 Точность работы двух станков оценивается отклонением от номинала производимой продукции. Из 10 единиц продукции первого станка отклонение от номинала (выборочная дисперсия) составило 3,5; из 15 единиц продукции второго станка – 4,5. Можно ли считать на уровне значимости 0,1, что станки имеют одинаковые точности? (Ответ: $F_{набл} = 1{,}06$, $f_{кр}(0{,}05; 9; 9) = 3{,}18$. Нет оснований считать, что станки имеют разную точность.)

14 Для сравнения точности двух станков-автоматов взяты две пробы (выборки), объёмы которых $n_1 = 10$ и $n_2 = 8$. В результате измерения контролируемого размера отобранных изделий составлена таблица 3, x_i и y_i – размеры изделий с первого и второго станков соответственно.

Таблица 3

x_i	1,08	1,10	1,12	1,14	1,15	1,25	1,36	1,38	1,40	1,42
y_i	1,11	1,12	1,18	1,22	1,33	1,35	1,36	1,38	–	–

Можно ли считать, что станки обладают одинаковой точностью $\left(H_0 : D(X) = D(Y)\right)$, если принять уровень значимости $\alpha = 0{,}1$ и в качестве конкурирующей гипотезы $H_1 : D(X) \neq D(Y)$? (Ответ: $s_X^2 = 188{,}67$, $s_Y^2 = 124{,}84$, $F_{набл} = 1{,}51$, $f_{кр}\left(0{,}05; 9; 7\right) = 3{,}68$. Нулевая гипотеза принимается. Нет оснований считать точность станков различной.)

1.4 Домашнее задание

1 На станке-автомате изготавливается деталь с контролируемым размером $m = 12$ мм. Известно, что распределение контролируемого размера является нормальным с параметрами m и $\sigma = 0{,}5$. Отдел технического контроля в течение смены произвёл 36 случайно отобранных деталей и подсчитал средний размер контролируемого параметра $\overline{x}_B = 11{,}7$ мм. Можно ли утверждать, что станок-автомат изготавливает детали уменьшенного размера и потому требуется произвести наладку станка? Принять уровень значимости 0,05. (Ответ: $U_{набл} = -3{,}6$, $u_{кр} = -1{,}65$. Станок-автомат изготавливает детали уменьшенного размера, требуется произвести наладку станка.)

2 Установлено, что средний вес таблетки лекарства сильного действия должен быть равен 0,05 мг. Выборочная проверка 121 таблетки полученной партии лекарства показала, что средний вес таблетки этой партии 0,53 мг. При уровне значимости 0,01 проверить нулевую гипотезу $H_0 : m = 0{,}50$ при альтернативной гипотезе $H_1 : m > 0{,}50$. Многократными предварительными опытами по взвешиванию таблеток, поставляемых фармацевтическим заводом, было установлено, что вес таблеток распределён нормально со средним квадратическим отклонением 0,11 мг. (Ответ: $U_{набл} = 3$, $u_{кр} = 2{,}33$. Нулевая гипотеза отвергается.)

3 Точность работы станка-автомата проверяется по дисперсии контролируемого размера изделий, которая не должна превышать $\sigma_0^2 = 0{,}18$. Взята проба из 25 случайно отобранных изделий, результаты измерений записаны в таблице 4.

При уровне значимости $\alpha = 0{,}05$ проверить, обеспечивает ли станок требуемую точность. (Ответ: $s^2 = 19{,}75$, $\chi^2_{набл} = 48$, $\chi^2_{кр} = \chi^2\left(0{,}05; 24\right) = 36{,}4$. Станок не обеспечивает требуемую точность.)

Таблица 4

x_i	3,0	3,5	3,8	4,4	4,5
n_i	2	6	8	7	1

4 Партия изделий принимается, если дисперсия контролируемого размера значимо не превышает 0,2. Исправленная выборочная дисперсия, найденная по выборке объёма $n=121$, оказалась $s^2=0,3$. Можно ли принять партию при уровне значимости $\alpha=0,01$? (Ответ: $\chi^2_{набл}=180$, $\chi^2_{кр}=\chi^2\left(0,01;\,120\right)=158,9$. Партию изделий принять нельзя.)

5 По выборке объёма $n=30$ найден средний вес $\overline{x}_B=130$ г изделий, изготовленных на первом станке, по выборке объёма $m=40$ найден средний вес $\overline{y}_B=125$ г изделий, изготовленных на втором станке. Генеральные дисперсии известны: $D(X)=60$ г^2, $D(Y)=80$ г^2. Требуется при уровне значимости 0,05 проверить нулевую гипотезу $H_0: m_X=m_Y$ при конкурирующей гипотезе $H_1: m_X\neq m_Y$. Предполагается, что случайные величины X и Y распределены нормально и выборки независимы. (Ответ: $Z_{набл}=2,5$, $z_{кр}=1,96$. Нулевая гипотеза отвергается.)

6 По двум независимым выборкам, объёмы которых $n_1=9$ и $n_2=6$, извлечённым из нормальных генеральных совокупностей X и Y, найдены выборочные дисперсии $D_B(X)=14,4$ и $D_B(Y)=20,5$. При уровне значимости 0,1 проверить нулевую гипотезу $H_0: D(X)=D(Y)$ о равенстве генеральных дисперсий при конкурирующей гипотезе $H_1: D(X)\neq D(Y)$. (Ответ: $F_{набл}=1,52$, $f_{кр}\left(0,05;\,5;\,8\right)=3,69$. H_0 принимаем.)

2 Проверка непараметрических гипотез

2.1 Теоретическая часть

Если закон распределения неизвестен, но есть основания предположить, что он имеет определённый вид (назовём его A), то проверяют нулевую гипотезу: генеральная совокупность распределена по закону A.

Проверка гипотезы о предполагаемом законе неизвестного распределения производится так же, как и проверка параметрических гипотез, т. е. при помощи специально подобранной случайной величины – критерия согласия.

Критерием согласия называется критерий проверки гипотезы о предполагаемом законе неизвестного распределения.

Проверка гипотезы о нормальном распределении генеральной совокупности с помощью критерия Пирсона.

Гипотеза H_0: случайная величина X имеет функцию распределения $F(x)$. В качестве критерия проверки нулевой гипотезы принимают случайную величину

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(n_i - n_i'\right)^2}{n_i'},$$

где n_i – эмпирические (наблюдаемые) частоты;

n_i' – теоретические частоты, $n_i' = np_i$;

p_i – вероятность попадания случайной величины в i-й интервал при гипотезе H_0, $p_i = P\left(x_i < X < x_{i+1}\right) = F\left(x_{i+1}\right) - F\left(x_i\right)$.

Доказано, что при $n \to \infty$ закон распределения этой случайной величины независимо от того, какому закону подчинена генеральная совокупность, стремится к закону распределения χ^2 с k степенями свободы. Число степеней свободы находят по равенству $k = s - r - 1$, где s – число групп (частичных интервалов) выборки; r – число параметров предполагаемого распределения, которые оценены по данным выборки.

В частности, если предполагаемое распределение нормальное, то оценивают два параметра (математическое ожидание и среднее квадратическое отклонение), поэтому $r = 2$ и число степеней свободы $k = s - 3$.

Поскольку односторонний критерий более «жёстко» отвергает нулевую гипотезу, чем двусторонний, строят правостороннюю критическую область исходя из требования $P\left(\chi^2 > \chi^2_{\text{кр}}(\alpha;k)\right) = \alpha$.

Таким образом, чтобы при заданном уровне значимости α проверить нулевую гипотезу H_0: генеральная совокупность распределена нормально, надо вычислить теоретические частоты, затем наблюдаемое значение критерия $\chi^2_{набл} = \sum_{i=1}^{k} \frac{\left(n_i - n_i'\right)^2}{n_i'}$ и по таблице критических точек распределения χ^2, по заданному уровню значимости α и числу степеней свободы $k = s - 3$ найти критическую точку $\chi^2_{кр}(\alpha;k)$.

Если $\chi^2_{набл} < \chi^2_{кр}$ – нет оснований отвергать нулевую гипотезу.

Если $\chi^2_{набл} > \chi^2_{кр}$ –нулевую гипотезу отвергают.

Замечание 1 **–** Объем выборки должен быть достаточно велик (не менее 50). Каждая группа должна содержать не менее 5–8 вариант. Малочисленные группы следует объединять в одну, суммируя частоты.

Замечание 2 **–** ля контроля вычислений можно использовать формулу

$$\chi^2_{набл} = \sum \frac{n_i^2}{n_i'} - n.$$

Методика вычисления теоретических частот нормального распределения.

Правило 1. Эмпирическое распределение задано в виде последовательности равноотстоящих вариант и соответствующих им частот.

Рассмотрим выборку с равноотстоящими значениями вариант (таблица 5).

Таблица 5

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

Теоретические частоты находят по формуле

$$n_i' = \frac{nh}{\sigma_B}\varphi(u_i),$$

где $n-$ объём выборки;

$h-$ шаг (разность между соседними вариантами);

$$u_i = \frac{x_i - \overline{x}_B}{\sigma_B},\ \varphi(u) = \frac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}.$$

Значения $\varphi(u)$ находят по таблице функции Гаусса.

Правило 2. Эмпирическое распределение задано в виде последовательности интервалов одинаковой длины и соответствующих им частот.

Пусть эмпирическое распределение задано в виде последовательности интервалов равной длины (таблица 6).

Таблица 6

$[x_i;x_{i+1})$	$[x_1;x_2)$	$[x_2;x_3)$	…	$[x_k;x_{k+1})$
n_i	n_1	n_2	…	n_k

Теоретические частоты вычисляем по формуле

$$n_i' = np_i.$$

Для нахождения теоретических вероятностей p_i нужно перейти к нормированным интервалам $[u_i;u_{i+1})$. Концы интервалов находят по формуле

$$u_i = \frac{x_i - \overline{x}_B}{\sigma_B},$$

причём левый конец первого интервала устремить к $-\infty$, а правый конец последнего – к $+\infty$.

Тогда $p_i = \Phi\left(u_{i+1}\right) - \Phi\left(u_i\right)$, где $\Phi\left(u_i\right)$ – функция Лапласа.

Проверка гипотезы о нормальном распределении генеральной совокупности с помощью критерия Колмогорова.

Критерий Колмогорова предназначен для проверки гипотезы H_0 о законе распределения только непрерывных случайных величин. Этот критерий можно применять только в том случае, когда гипотетическое распределение $F(x)$ полностью известно заранее из каких-либо теоретических соображений, т. е. когда известен не только вид функции $F(x)$, но и все входящие в неё параметры.

Схема применения критерия Колмогорова:

$H_0 : \tilde{F}(x) = F(x), \quad H_1 : \tilde{F}(x) \neq F(x)$.

1 По результатам n независимых опытов определяют эмпирическую функцию распределения $\tilde{F}(x)$.

2 Вычисляют, пользуясь гипотетической функцией распределения, значения функции распределения $F(x)$ случайной величины X.

3 Определяют величину D критерия Колмогорова:

$$D = \max\left|\tilde{F}(x) - F(x)\right|.$$

4 Вычисляют $\lambda_{\text{набл}} = D\sqrt{n}$.

5 По таблице критических точек распределения Колмогорова $\left(P\left(\lambda > \lambda_\alpha\right) = = 1 - K(\lambda) = 1 - \sum\limits_{k=-\infty}^{+\infty} (-1)^k e^{-k^2\lambda^2} = \alpha\right)$ находят $\lambda_{кр} = \lambda(\alpha)$. Если $\lambda_{набл} \geq \lambda_{кр}$, то H_0 отвергается. Если $\lambda_{набл} < \lambda_{кр}$, то H_0 принимается.

Замечание **–** Величину D можно определить, либо построив предварительно в одной и той же системе координат графики функций $\tilde{F}(x)$ и $F(x)$, либо путём сравнения величин $\left|\tilde{F}(x) - F(x)\right|$ в заданных точках.

2.2 Образцы решения примеров

Пример 1 – При уровне значимости $\alpha = 0,05$ проверить гипотезу о нормальном распределении генеральной совокупности с помощью критерия Пирсона, если известны эмпирические и теоретические частоты (таблица 7).

Таблица 7

Эмпирическая частота n_i	6	12	16	40	13	8	5
Теоретическая частота n_i'	4	11	15	43	15	6	6

Решение

Найдём наблюдаемое значение критерия

$$\chi^2_{набл} = \sum_{i=1}^{k} \frac{\left(n_i - n'_i\right)^2}{n'_i} = 2,47 .$$

По таблице критических точек распределения χ^2, по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = 7 - 3 = 4$ находим $\chi^2_{кр} = \chi^2\left(0,05;4\right) = 9,5$.

Так как $\chi^2_{набл} < \chi^2_{кр}$, то нет оснований отвергать нулевую гипотезу. Следовательно, расхождение эмпирических и теоретических частот незначимое и, значит, данные наблюдений согласуются с гипотезой о нормальном распределении.

Пример 2 – Используя критерий Пирсона, при уровне значимости $\alpha = 0,05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением данной выборки объёмом $n = 200$ (таблица 8).

Таблица 8

x_i	5	7	9	11	13	15	17	19	21
n_i	15	26	25	30	26	21	24	20	13

Решение

Найдём числовые характеристики выборки.

$$\overline{x}_B = \frac{1}{n}\sum_{i=1}^{s} x_i n_i = 12,63 ; \qquad D_B = \frac{1}{n}\sum_{i=1}^{s} x_i^2 n_i - \overline{x}_B^2 = 22,044 ;$$

$$\sigma_B = \sqrt{D_B} = \sqrt{22,044} = 4,695 .$$

Вычислим теоретические частоты, учитывая, что $n = 200$, $h = 2$.

$$n'_i = \frac{nh}{\sigma_в}\varphi\left(u_i\right) = \frac{200 \cdot 2}{4,695}\varphi\left(u_i\right) = 85,2\varphi\left(u_i\right),$$

где $u_i = \dfrac{x_i - \overline{x}_B}{\sigma_B}$. Значения $\varphi\left(u_i\right)$ находим по таблице значений функции Гаусса.

Последовательность вычислений приведём в таблице 9.

Таблица 9

i	x_i	n_i	$u_i = \dfrac{x_i - \overline{x}_B}{\sigma_B}$	$\varphi(u_i)$	$n_i' = 85{,}2\varphi(u_i)$	$(n_i - n_i')^2 / n_i'$
1	5	15	–1,62	0,1074	9,1	3,83
2	7	26	–1,20	0,1942	16,5	5,47
3	9	25	–0,77	0,2966	25,3	0,00
4	11	30	–0,35	0,3752	32,0	0,13
5	13	26	0,08	0,3977	33,9	1,84
6	15	21	0,51	0,3503	29,8	2,60
7	17	24	0,93	0,2589	22,0	0,18
8	19	20	1,36	0,1582	13,5	3,13
9	21	13	1,78	0,0818	7,0	5,14
						$\chi^2_{набл} = 22{,}32$

$$\chi^2_{набл} = \sum_{i=1}^{s} \frac{(n_i - n_i')^2}{n_i'} = 22{,}32.$$

Из таблицы критических точек распределения χ^2, по уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $k = s - 3 = 9 - 3 = 6$ находим $\chi^2_{кр}(0{,}05;\ 6) = 12{,}6$.

Так как $\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$, то гипотезу о нормальном распределении генеральной совокупности отвергаем.

Пример 3 – Используя критерий Пирсона, при уровне значимости $\alpha = 0{,}05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности *X* с эмпирическим распределением выборки (таблица 10).

Таблица 10

i	1	2	3	4	5	6	7	8	9	10	11	
$[x_i; x_{i+1})$	$[1;3)$	$[3;5)$	$[5;7)$	$[7;9)$	$[9;11)$	$[11;13)$	$[13;15)$	$[15;17)$	$[17;19)$	$[19;21)$	$[21;23]$	
n_i	2	4	6	10	18	20	16	11	7	5	1	$n = 100$

Решение

Найдём выборочное среднее и выборочное среднее квадратическое отклонение.

$$\overline{x}_B = \frac{1}{n}\sum x_i^* n_i = 12{,}04, \qquad D_B = \frac{1}{n}\sum \left(x_i^*\right)^2 n_i - \overline{x}_B^2 = 18{,}16, \qquad \sigma_B = 4{,}26.$$

Здесь x_i^* – середина i-го частичного интервала.

Для проверки нулевой гипотезы H_0 о нормальном распределении признака составим расчётную таблицу 11, объединив два первых и два последних интервала, так как частоты этих интервалов меньше 5.

Теоретические вероятности p_i находим по формуле

$$p_i = \Phi(u_{i+1}) - \Phi(u_i),$$

где $\Phi(u_i)$ – функция Лапласа.

Таблица 11

i	$[x_i;x_{i+1})$	$[u_i;u_{i+1})$	n_i	p_i	n_i'	$(n_i - n_i')^2 / n_i'$
1	$[1;5)$	$(-\infty;-1,65)$	6	0,0495	4,95	0,2227
2	$[5;7)$	$[-1,65;-1,18)$	6	0,0695	6,95	0,1299
3	$[7;9)$	$[-1,18;-0,71)$	10	0,1199	11,99	0,3303
4	$[9;11)$	$[-0,71;-0,24)$	18	0,1663	16,63	0,1129
5	$[11;13)$	$[-0,24;0,23)$	20	0,1858	18,58	0,1085
6	$[13;15)$	$[0,23;0,69)$	16	0,1639	16,39	0,0093
7	$[15;17)$	$[0,69;1,16)$	11	0,1221	12,21	0,1199
8	$[17;19)$	$[1,16;1,63)$	7	0,0714	7,14	0,0027
9	$[19;23]$	$(1,63;+\infty)$	6	0,0516	5,16	0,1367
Σ			100	1	100	$\chi^2_{набл} = 1,1729$

Итак, $\chi^2_{набл} = 1,1729$. По таблице критических точек распределения χ^2, по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = s - 3 = 9 - 3 = 6$ находим $\chi^2_{кр}(0,05;6) = 12,6$.

Так как $\chi^2_{набл} < \chi^2_{кр}$, то нет оснований отвергать нулевую гипотезу о нормальном распределении генеральной совокупности.

Пример 4 – Дано распределение 1000 деревьев по диаметру ствола (таблица 12).

По критерию Колмогорова при уровне значимости 0,05 проверить, согласуются ли эмпирические данные с тем предположением, что диаметр ствола имеет нормальное распределение.

Решение

H_0 : диаметр ствола имеет нормальное распределение.

Найдём выборочное среднее и выборочное среднее квадратическое отклонение.

$$\bar{x}_B = \frac{1}{n}\sum x_i^* n_i = 33{,}356, \quad D_B = \frac{1}{n}\sum \left(x_i^*\right)^2 n_i - \bar{x}_B^2 = 57{,}61, \quad \sigma_B = 7{,}59.$$

Составим расчётную таблицу 13.

Таблица 12

Диаметр ствола $x_i - x_{i+1}$, см	Количество деревьев n_i	Диаметр ствола $x_i - x_{i+1}$, см	Количество деревьев n_i
14…18	16	38…42	115
18…22	35	42…46	71
22…26	109	46…50	36
26…30	183	50…54	19
30…34	214	54…58	5
34…38	197		

Таблица 13

$[x_i; x_{i+1})$	n_i	$F^*(x)$	u_{i+1}	$\Phi(u_{i+1})$	$F(x) = 0{,}5 + \Phi(u_{i+1})$	$\lvert F^*(x) - F(x)\rvert$
Менее 14	0	0	–4,3947	–0,4999	0	0
$[14;18)$	16	0,016	–2,0232	–0,4783	0,0217	0,0057
$[18;22)$	35	0,051	–1,4962	–0,4332	0,0668	0,0158
$[22;26)$	109	0,160	–0,9692	–0,3340	0,166	0,006
$[26;30)$	183	0,343	–0,4422	–0,1700	0,3300	0,013
$[30;34)$	214	0,557	–0,0848	0,0319	0,5319	0,0251
$[34;38)$	197	0,754	0,612	0,2291	0,7291	0,0249
$[38;42)$	115	0,869	1,139	0,3729	0,8729	0,0039
$[42;48)$	71	0,940	1,666	0,4525	0,9525	0,0125
$[48;52)$	36	0,976	2,193	0,4858	0,9858	0,0098
$[52;54)$	19	0,995	2,720	0,4967	0,9967	0,0017
$[54;58]$	5	1	3,247	0,4993	0,9993	0,0007

Из таблицы 13 следует, что

$$D = \max\left|\tilde{F}(x) - F(x)\right| = 0{,}0251, \quad \lambda_{набл} = D\sqrt{n} = 0{,}0251 \cdot \sqrt{1000} = 0{,}7937.$$

По таблице критических точек распределения Колмогорова, по уровню значимости $\alpha = 0{,}05$ находим $\lambda_{кр} = 1{,}358$.

Так как $\lambda_{набл} < \lambda_{кр}$, нет оснований отвергать нулевую гипотезу, т. е. расхождения между эмпирической и теоретической функциями распределения случайные.

Пример 5 – В результате испытания 200 элементов на длительность работы получено эмпирическое распределение (таблица 14), в первой строке указаны интервалы времени в часах, во второй – частоты.

Таблица 14

$[x_i; x_{i+1})$	$[0;5)$	$[5;10)$	$[10;15)$	$[15;20)$	$[20;25)$	$[25;30]$
n_i	133	45	15	4	2	1

Требуется при уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что время работы элементов распределено по показательному закону с помощью:

1) критерия согласия Пирсона;

2) критерия согласия Колмогорова.

Решение

Найдём среднее время работы всех элементов:

$$\overline{x}_B = \frac{1}{n}\sum x_i^* n_i = 5.$$

Найдём оценку параметра предполагаемого показательного распределения:

$$\lambda = 1/\overline{x}_B = 1/5 = 0,2.$$

Таким образом, $F(x) = 1 - e^{-\lambda x} = 1 - e^{-0,2x}$, $(x > 0)$ – функция предполагаемого показательного распределения.

Найдём вероятности попадания X в каждый из интервалов по формуле

$$p_i = P(x_i < X < x_{i+1}) = F(x_{i+1}) - F(x_i) = e^{-\lambda x_i} - e^{-\lambda x_{i+1}}.$$

$$p_1 = P(0 < X < 5) = e^{-0,2\cdot 0} - e^{-0,2\cdot 5} = 1 - 0,3679 = 0,6321.$$

Аналогично вычислим вероятности попадания X в остальные интервалы.

Найдём теоретические частоты по формуле

$$n_i' = np_i = 200 p_i.$$

Результаты вычислений занесём в таблицу 15.

1 Проведём проверку гипотезы H_0 о том, что время работы элементов распределено по показательному закону с помощью критерия Пирсона. Для этого составим расчётную таблицу 16, причём объединим интервалы с малочисленными частотами.

Таблица 15

$[x_i; x_{i+1})$	n_i	p_i	n_i'
$[0;5)$	133	0,6321	126,42
$[5;10)$	45	0,2326	46,52
$[10;15)$	15	0,0855	17,10
$[15;20)$	4	0,0315	6,30
$[20;25)$	2	0,0116	2,32
$[25;30]$	1	0,0042	0,84

Таблица 16

i	n_i	n_i'	$(n_i - n_i')^2 / n_i'$
1	133	126,42	0,3425
2	45	46,52	0,0497
3	15	17,10	0,2579
4	7	9,46	0,6397
Σ	200	199,5	$\chi^2_{набл} = 1,29$

Итак, $\chi^2_{набл} = 1,29$. По таблице критических точек распределения χ^2, по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = 4 - 1 - 1 = 2$ находим $\chi^2_{кр}(0,05;\ 2) = 6$.

Так как $\chi^2_{набл} < \chi^2_{кр}$, то нет оснований отвергать нулевую гипотезу о распределении генеральной совокупности по показательному закону.

2 Проведём проверку гипотезы H_0 о том, что время работы элементов распределено по показательному закону с помощью критерия Колмогорова. Составим расчётную таблицу 17.

Таблица 17

$x_i - x_{i+1}$	n_i	$\tilde{F}(x) = \frac{n_x}{n}$	p_i	$F(x)$	$\lvert \tilde{F}(x) - F(x) \rvert$
Менее 0	0	0	0	0	0
$[0;5)$	133	0,665	0,6321	0,6321	0,0329
$[5;10)$	45	0,89	0,2326	0,8647	0,0253
$[10;15)$	15	0,965	0,0855	0,9502	0,0148
$[15;20)$	4	0,985	0,0315	0,9817	0,0033
$[20;25)$	2	0,995	0,0116	0,9933	0,0017
$[25;30]$	1	1	0,0042	0,9975	0,0025

$$D = \max \lvert \tilde{F}(x) - F(x) \rvert = 0,0329. \quad \lambda_{набл} = D\sqrt{n} = 0,0329 \cdot \sqrt{200} = 0,4653.$$

По таблице критических точек распределения Колмогорова, по уровню значимости $\alpha = 0{,}05$ находим $\lambda_{кр} = 1{,}358$.

Так как $\lambda_{набл} < \lambda_{кр}$, нет оснований отвергать нулевую гипотезу, т. е. расхождения между эмпирической и теоретической функциями распределения случайные.

2.3 Задачи для самостоятельной работы

1 Используя критерий Пирсона, при уровне значимости $\alpha = 0{,}05$ установить, случайно или значимо расхождение между эмпирическими n_i и теоретическими n_i' частотами (таблица 18), которые вычислены исходя из гипотезы о нормальном распределении генеральной совокупности *X*.

Таблица 18

n_i	5	10	20	8	7
n_i'	6	14	18	7	5

(Ответ: $\chi^2_{набл} = 2{,}47$, $\chi^2_{кр} = 6{,}0$. Эмпирические и теоретические частоты расходятся случайно.)

2 Пользуясь критерием Пирсона, при уровне значимости $\alpha = 0{,}05$ установить, согласуется ли гипотеза о нормальном распределении генеральной совокупности с данными выборки объёма $n = 200$ (таблица 19).

Таблица 19

x_i	3	5	7	9	11	13	15	17	19	21	23
n_i	6	9	26	25	30	26	21	24	20	8	5

(Ответ: $\overline{x}_B = 12{,}6$, $\sigma_B = 0{,}491$, $\chi^2_{набл} = 12{,}222$, $\chi^2_{кр}\left(0{,}05;8\right) = 15{,}5$. Эмпирические и теоретические частоты расходятся случайно.)

3 Дана выборка (таблица 20).

Таблица 20

$[x_i; x_{i+1})$	$[-20;-10)$	$[-10;0)$	$[0;10)$	$[10;20)$	$[20;30)$	$[30;40)$	$[40;50]$
n_i	20	47	80	89	40	16	8

При уровне значимости $\alpha = 0{,}05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности *X* с заданным эмпирическим распределением с помощью:

1) критерия Пирсона;

2) критерия Колмогорова.

(Ответ: $\overline{x}_B = 10,4$, $\sigma_B = 13,67$, 1) $\chi^2_{набл} = 1,52$, $\chi^2_{кр}\left(0,05;4\right) = 9,5$, 2) $\lambda_{набл} = 0,497$, $\lambda_{кр}\left(0,05\right) = 1,36$. Нет оснований отвергать гипотезу о нормальном распределении генеральной совокупности.)

4 Произведено $n = 100$ опытов. Любой опыт состоял из $l = 5$ испытаний, в каждом из которых вероятность p появления события A равна 0,3. В итоге получено эмпирическое распределение (таблица 21), x_i – число появлений события A в каждом опыте; n_i – частота числа появлений события A в i-м опыте.

Таблица 21

x_i	0	1	2	3	4	5
n_i	2	10	27	32	23	6

Требуется при уровне значимости $\alpha = 0,05$ проверить гипотезу о том, что дискретная случайная величина X распределена по биномиальному закону.

(Ответ: $\chi^2_{набл} = 4,44$, $\chi^2_{кр}\left(0,05;4\right) = 9,5$. Нет оснований отвергать нулевую гипотезу.)

5 В результате взвешивания 800 стальных шариков получено эмпирическое распределение, приведённое в таблице 22.

Таблица 22

$[x_i; x_{i+1})$	n_i	$[x_i; x_{i+1})$	n_i
$[20;20,5)$	91	$[22,5;23)$	83
$[20,5;21)$	76	$[23;23,5)$	79
$[21;21,5)$	75	$[23,5;24)$	73
$[21,5;22)$	74	$[24;24,5)$	80
$[22;22,5)$	92	$[24,5;25]$	77

При уровне значимости 0,01 проверить гипотезу о том, что вес шариков X распределён равномерно, используя:

1) критерий Пирсона;

2) критерий Колмогорова.

Указание **–** По методу моментов оценки параметров равномерного распределения находят из системы уравнений

$$\begin{cases} \left(\tilde{b} - \tilde{a}\right)/2 = \overline{x}_B; \\ \left(\tilde{b} - \tilde{a}\right)/2\sqrt{3} = \sigma_B. \end{cases}$$

(Ответ: $\overline{x}_B = 22,47$, $\sigma_B = 1,44$, $\tilde{a} = 19,98$, $\tilde{b} = 24,96$, 1) $\chi^2_{набл} = 4,39$,

$\chi^2_{кр}\left(0,01;7\right)=18,5$; 2) $\lambda_{набл}=0,28$, $\lambda_{кр}=1,63$. Нет оснований отвергать гипотезу о равномерном распределении веса шариков.)

6 В итоге испытаний 1000 элементов на время безотказной работы получено эмпирическое распределение (таблица 23), в первом столбце указаны интервалы времени в часах, во втором – количество отказавших элементов в интервале.

Таблица 23

$[x_i;x_{i+1})$	n_i	$[x_i;x_{i+1})$	n_i
$[0;10)$	365	$[40;50)$	70
$[10;20)$	245	$[50;60)$	45
$[20;30)$	150	$[60;70]$	25
$[30;40)$	100		

Требуется при уровне значимости 0,01 проверить гипотезу о том, что время безотказной работы элементов распределено по показательному закону, используя:

1) критерий Пирсона;

2) критерий Колмогорова.

(Ответ: $\bar{x}_B=20$, $\lambda=0,05$, 1) $\chi^2_{набл}=15,88$, $\chi^2_{кр}\left(0,01;5\right)=15,1$; 2) $\lambda_{набл}=9,78$, $\lambda_{кр}=1,63$. Гипотеза о показательном распределении времени безотказной работы элементов отвергается.)

2.4 Домашнее задание

1 Используя критерий Пирсона, при уровне значимости $\alpha=0,05$ установить, случайно или значимо расхождение между эмпирическими n_i и теоретическими n_i' частотами (таблица 24), которые вычислены исходя из гипотезы о нормальном распределении генеральной совокупности *X*.

Таблица 24

n_i	14	18	32	70	20	36	10
n_i'	10	24	34	80	18	22	12

(Ответ: $\chi^2_{набл}=13,93$, $\chi^2_{кр}\left(0,05;4\right)=9,5$. Эмпирические и теоретические частоты расходятся значимо.)

2 Используя критерий Пирсона, при уровне значимости $\alpha=0,05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности *X* с эмпирическим распределением выборки объёмом $n=115$ (таблица 25).

Таблица 25

x_i	3	5	7	9	11	13	15
n_i	10	15	20	30	25	10	5

(Ответ: $\chi^2_{набл} = 5{,}65$, $\chi^2_{кр}\left(0{,}05;4\right) = 9{,}5$. Нет оснований отвергать гипотезу о нормальном распределении генеральной совокупности.)

3 Результаты исследования прочности на сжатие (СВ X) 200 образцов бетона представлены в виде сгруппированного статистического ряда (таблица 26).

Таблица 26

$\left[x_i; x_{i+1}\right)$	$\left[19;20\right)$	$\left[20;21\right)$	$\left[21;22\right)$	$\left[22;23\right)$	$\left[23;24\right)$	$\left[24;25\right]$
n_i	10	26	56	64	30	14

Проверить нулевую гипотезу о том, согласуются ли результаты эксперимента по определению прочности на сжатие образца бетона с предположением о нормальном распределении СВ X (прочности на сжатие) с помощью:

1) критерия Пирсона (уровень значимости $\alpha = 0{,}05$);

2) критерия Колмогорова (уровень значимости $\alpha = 0{,}1$).

(Ответ: $\overline{x}_B = 22{,}1$, $\sigma_B = 1{,}233$, 1) $\chi^2_{набл} = 1{,}35$, $\chi^2_{кр}\left(0{,}05;3\right) = 7{,}815$, 2) $\lambda_{набл} = 0{,}183$, $\lambda_{кр}\left(0{,}1\right) = 1{,}224$. Нет оснований отвергать гипотезу о нормальном распределении генеральной совокупности.)

3 Линейная корреляция

3.1 Теоретическая часть

Во многих задачах требуется установить и оценить зависимость изучаемой случайной величины Y от одной или нескольких других случайных величин. Будем рассматривать зависимость Y от одной случайной величины X.

Две случайные величины могут быть связаны либо функциональной зависимостью, либо статистической.

Строгая функциональная зависимость реализуется редко, так как обе случайные величины подвержены ещё действию случайных факторов. В этом случае возникает статистическая зависимость.

Статистической называют зависимость, при которой изменение одной из величин влечёт изменение распределения другой. В частности, статистическая зависимость проявляется в том, что при изменении одной из величин изменяется среднее значение другой. В этом случае статистическая зависимость называется ***корреляционной***.

Условным средним $\overline{y}_x$ называют среднее арифметическое наблюдавшихся значений *Y*, соответствующих $X = x$.

Аналогично ***условное среднее*** $\overline{x}_y$ – это среднее арифметическое наблюдавшихся значений *X*, соответствующих $Y = y$.

Условное математическое ожидание $M(Y|x) = f(x)$ есть функция от *x*, которую называют ***функцией регрессии Y на X.*** Оценка условного математического ожидания, т. е. условное среднее $\overline{y}_x$, также функция от *x*. Обозначив эту функцию через $\tilde{f}(x)$, получим уравнение

$$\overline{y}_x = \tilde{f}(x).$$

Это уравнение называют ***выборочным уравнением регрессии Y*** на ***X***, функцию $\tilde{f}(x)$ – ***выборочной регрессией Y*** на ***X***, а её график – ***выборочной линией регрессии Y*** на ***X***.

Аналогично уравнение

$$\overline{x}_y = \tilde{\varphi}(y)$$

называют ***выборочным уравнением регрессии X*** на ***Y***, функцию $\tilde{\varphi}(y)$ – ***выборочной регрессией X*** на ***Y***, а её график – ***выборочной линией регрессии X*** на ***Y***.

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

Выборочное уравнение регрессии *Y* на *X* имеет следующий вид:

$$\overline{y}_x - \overline{y} = r_B \frac{\sigma_Y}{\sigma_X}\left(x - \overline{x}\right),$$

где r_B – выборочный коэффициент корреляции, $r_B = \dfrac{\left(\dfrac{1}{n}\sum\limits_i \sum\limits_j n_{ij} x_i y_j - \overline{x}\,\overline{y}\right)}{\left(\sigma_X \sigma_Y\right)}$.

Выборочное уравнение регрессии *X* на *Y*

$$\overline{x}_y - \overline{x} = r_B \frac{\sigma_X}{\sigma_Y}\left(y - \overline{y}\right).$$

3.2 Образцы решения примеров

Пример 1 – Найти выборочное уравнение прямой линии регрессии *Y* на *X* и *X* на *Y* по данным, приведённым в корреляционной таблице 27.

Таблица 27

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

Решение

Найдём выборочные средние, выборочные средние квадратические отклонения, выборочный коэффициент корреляции.

$$\overline{x}=\frac{1}{n}\sum x_i n_i=\frac{1}{100}\left(20\cdot 4+25\cdot 14+30\cdot 46+35\cdot 16+40\cdot 20\right)=31,7\,.$$

$$\overline{y}=\frac{1}{n}\sum y_j n_j=\frac{1}{100}\left(16\cdot 10+26\cdot 18+36\cdot 44+46\cdot 22+56\cdot 6\right)=35,6\,.$$

$$\sigma_x=\sqrt{\frac{1}{n}\sum x_i^2 n_i-\overline{x}^2}=$$

$$=\sqrt{\frac{1}{100}\left(20^2\cdot 4+25^2\cdot 14+30^2\cdot 46+35^2\cdot 16+40^2\cdot 20\right)-31,7^2}=5,35.$$

$$\sigma_y=\sqrt{\frac{1}{n}\sum y_j^2 n_j-\overline{y}^2}=$$

$$=\sqrt{\frac{1}{100}^2\left(16^2\cdot 10+26^2\cdot 18+36^2\cdot 44+46^2\cdot 22+56^2\cdot 6\right)-35,6^2}=10,2.$$

$$r_в=\frac{\left(\frac{1}{n}\sum_i\sum_j n_{ij}x_i y_j-\overline{x}\,\overline{y}\right)}{\left(\sigma_x\sigma_y\right)}=\frac{\sum_i\sum_j n_{ij}x_i y_j-n\overline{x}\,\overline{y}}{n\sigma_x\sigma_y}=$$

$$=\frac{16\cdot 20\cdot 4+16\cdot 25\cdot 6+26\cdot 25\cdot 8+26\cdot 30\cdot 10+\ldots+56\cdot 40\cdot 5-100\cdot 31,7\cdot 35,6}{100\cdot 5,35\cdot 10,2}=0,76.$$

Искомое уравнение прямой линии регрессии Y на X

$$\overline{y}_x-35,6=0,76\cdot\frac{10,2}{5,35}\left(x-31,7\right),$$

$$\overline{y}_x = 1,45x - 10,36.$$

Искомое уравнение прямой линии регрессии *X* на *Y*

$$\overline{x}_y - 31,7 = 0,76 \cdot \frac{5,35}{10,2}\left(y - 35,6\right),$$

$$\overline{x}_y = 0,4y + 17,5.$$

Построим случайные точки и графики линий регрессий (рисунок 1).

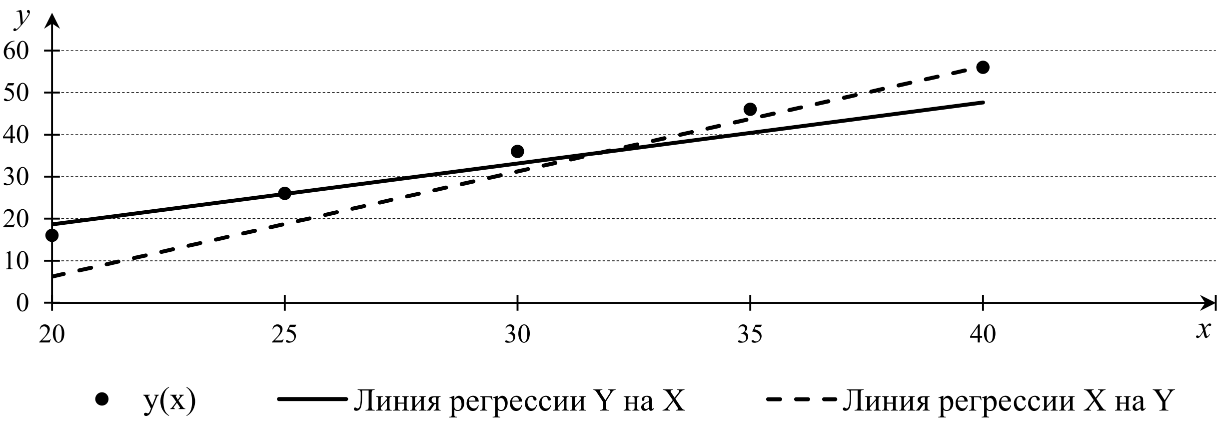

Рисунок 1

3.3 Задачи для самостоятельной работы

1 Дано распределение 100 фабрик по основным фондам *Y*, млн р., и по готовой продукции *X*, млн р. (таблица 28). Известно, что между *X* и *Y* существует линейная корреляционная зависимость. Найти уравнение прямой регрессии *Y* на *X*, построить линию регрессии и случайные точки. (Ответ: $\overline{y}_x = 0,32x + 12,36$.)

Таблица 28

X	*Y*					
	12	18	24	30	36	n_i
20	4	6				10
30		10	8		1	19
40		2	13	7	2	24
50			1	9	3	13
60			1	3	12	16
70				4	4	8
80				2	8	10
n_j	4	18	23	25	30	$n = 100$

2 Найти выборочные уравнения прямых линий регрессии Y на X и X на Y по данным, приведённым в корреляционных таблицах 29 (вариант 1) и 30 (вариант 1).

Таблица 29

Y	X								
	5	10	15	20	25	30	35	40	n_j
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_i	5	5	8	11	8	6	5	2	$n=50$

Таблица 30

Y	X							
	18	23	28	33	38	43	48	n_j
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_i	1	6	8	20	10	4	1	$n=50$

(Ответ: 1) $\overline{y}_x = 1{,}92x + 100{,}9;$ $\overline{x}_y = 0{,}42y - 38{,}3;$ 2) $\overline{y}_x = 3{,}69x + 66;$ $\overline{x}_y = 0{,}19y - 3{,}1.$)

4 Обработка экспериментальных данных

4.1 Теоретическая часть

Для полного и глубокого изучения признака X в генеральной совокупности полезно использовать примерную схему для исследования.

1 Выполнить выборку объёмом n.

2 Составить ряд распределения.

3 Изучить ряд распределения графически.

4 Найти эмпирическую функцию распределения и построить её график.

5 Вычислить числовые характеристики выборки (выборочную среднюю, выборочную дисперсию, среднее квадратическое отклонение, выборочные коэффициенты асимметрии, эксцесса и вариации).

6 Выбрать закон распределения изучаемого признака в генеральной совокупности.

7 Найти точечные оценки параметров предполагаемого закона распределения. Если распределение непрерывное, то найти функцию распределения и плотность вероятности.

8 Найти теоретические частоты распределения, определить согласие теоретической и эмпирической функций распределения, применив соответствующий критерий согласия.

9 Найти доверительные интервалы параметров распределения.

4.2 Образец решения

В результате эксперимента получены данные:

27,55	29,589	25,921	28,246	27,725	26,612	28,062	28,433	28,968	29,626
28,218	28,592	29,285	26,263	29,111	27,328	28,281	27,558	26,432	27,061
29,166	28,42	27,792	28,389	28,082	28,147	27,218	28,753	29,2	27,4
28,965	27,001	28,377	29,134	28,565	27,28	29,367	28,57	28,702	29,243
28,514	28,012	30,196	26,691	27,913	28,087	26,684	28,265	29,161	28,27
28,42	27,853	28,766	27,249	29,047	27,23	27,842	30,081	27,939	29,061
27,394	28,864	28,22	28,482	28,032	27,223	27,666	27,389	27,387	27,204
28,014	28,562	28,353	28,61	27,871	26,912	29,362	25,562	27,235	27,61
28,9	27,374	26,904	29,358	29,622	27,295	28,909	27,473	27,529	28,945
26,931	28,559	27,39	26,168	26,058	28,16	28,796	27,111	28,417	28,811

Требуется:

1) определить размах варьирования;

2) составить статистический ряд распределения частот для случайной величины X;

3) составить интервальный статистический ряд относительных частот;

4) построить гистограмму и полигон относительных частот;

5) найти эмпирическую функцию распределения и построить её график;

6) вычислить числовые характеристики выборки: среднее арифметическое $\overline{x}_B$, выборочную дисперсию D_B, выборочное среднее квадратическое отклонение σ_B, выборочные моду и медиану, выборочные асимметрию A_B и эксцесс E_B;

7) по виду гистограммы и полигона относительных частот сделать предварительный вывод о законе распределения;

8) найти точечные сценки параметров распределения, записать плотность вероятности, функцию распределения СВ X;

9) найти теоретические частоты распределения, проверить согласие эмпирической функции распределения $\tilde{F}(x)$ с теоретической $F(x)$ при помощи критерия согласия χ^2 и λ – Колмогорова, уровень значимости принять 0,05.

В случае нормального распределения:

10) найти интервальные оценки параметров распределения, взяв надёжность $\gamma = 0{,}95$;

11) проверить нулевую гипотезу H_0: $m = m_0 = 28$ о математическом ожи-

дании при альтернативной гипотезе $H_a: m \neq 28$;

12) проверить нулевую гипотезу $H_0: D(X) = \sigma^2 = \sigma_0^2 = 1$ о дисперсии против альтернативной $H_a: \sigma^2 < 1$.

Решение

1 Найдём размах варьирования:

$$R = x_{\max} - x_{\min};\ x_{\min} = 25,562;\ x_{\max} = 30,196;\ R = 30,196 - 25,562 = 4,634.$$

2 Статистический ряд распределения составлять не имеет смысла, так как значения в выборке практически не повторяются. Перейдём к составлению интервального статистического ряда.

3 Число интервалов рассчитаем по формуле

$$k = 1 + 3,2 \cdot \lg n.$$

$$k = 1 + 3,2 \cdot \lg 100 = 1 + 3,2 \cdot 2 = 7,4.$$

Возьмём $k = 8$ (можно $k = 7$).

Длина частичного интервала вычисляется по формуле

$$h = \frac{R}{k} = \frac{4,634}{8} \approx 0,579.$$

Составим интервальный статистический ряд абсолютных и относительных частот. Полученные значения запишем в таблицу 31.

Таблица 31

$[x_i; x_{i+1})$	m_i	W_i
$[25,562; 26,141)$	3	0,03
$[26,141; 26,720)$	6	0,06
$[26,720; 27,299)$	14	0,14
$[27,299; 27,878)$	18	0,18
$[27,878; 28,457)$	23	0,23
$[28,457; 29,036)$	19	0,19
$[29,036; 29,615)$	13	0,13
$[29,615; 30,196]$	4	0,04

4 Для построения полигона и гистограммы составим вспомогательную таблицу 32.

Таблица 32

$[x_i; x_{i+1})$	m_i	W_i	x_i^*	W_i/h
$[25,562; 26,141)$	3	0,03	25,8515	0,052
$[26,141; 26,720)$	6	0,06	26,4305	0,104
$[26,720; 27,299)$	14	0,14	27,0095	0,242
$[27,299; 27,878)$	18	0,18	27,5885	0,311
$[27,878; 28,457)$	23	0,23	28,1675	0,397
$[28,457; 29,036)$	19	0,19	28,7465	0,328
$[29,036; 29,615)$	13	0,13	29,3255	0,225
$[29,615; 30,196]$	4	0,04	29,9055	0,069

Построим полигон относительных частот (рисунок 2) и гистограмму (рисунок 3).

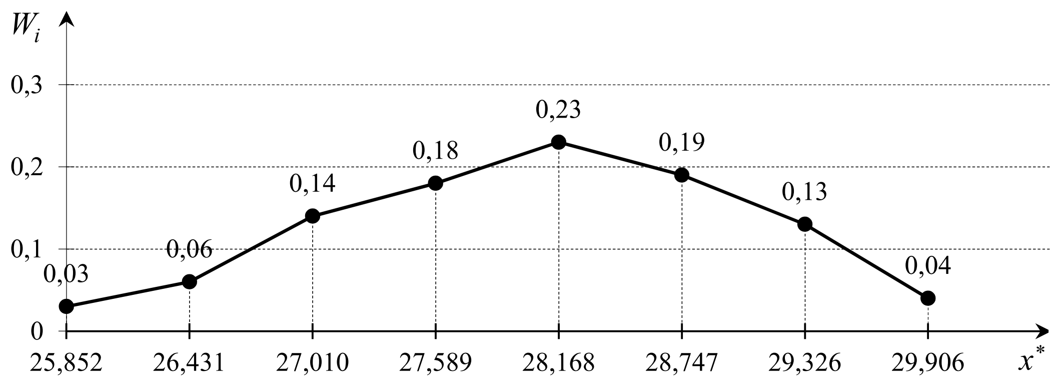

Рисунок 2

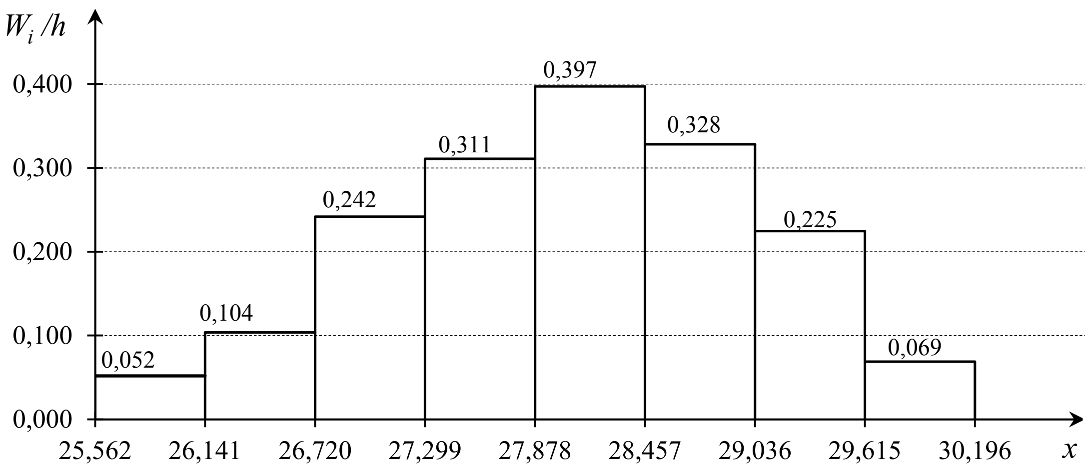

Рисунок 3

5 Составим эмпирическую функцию распределения.

$$\tilde{F}(x)=\begin{cases}0, \quad x\le 25,562;\\ 0,03, \quad x=26,141;\\ 0,03+0,06=0,09, \quad x=26,72;\\ 0,09+0,14=0,23, \quad x=27,299;\\ 0,23+0,18=0,41, \quad x=27,878;\\ 0,41+0,23=0,64, \quad x=28,457;\\ 0,64+0,19=0,83, \quad x=29,036;\\ 0,83+0,13=0,96, \quad x=29,615;\\ 0,96+0,04=1, \quad x\ge 30,196.\end{cases}$$

Строим график эмпирической функции распределения (рисунок 4).

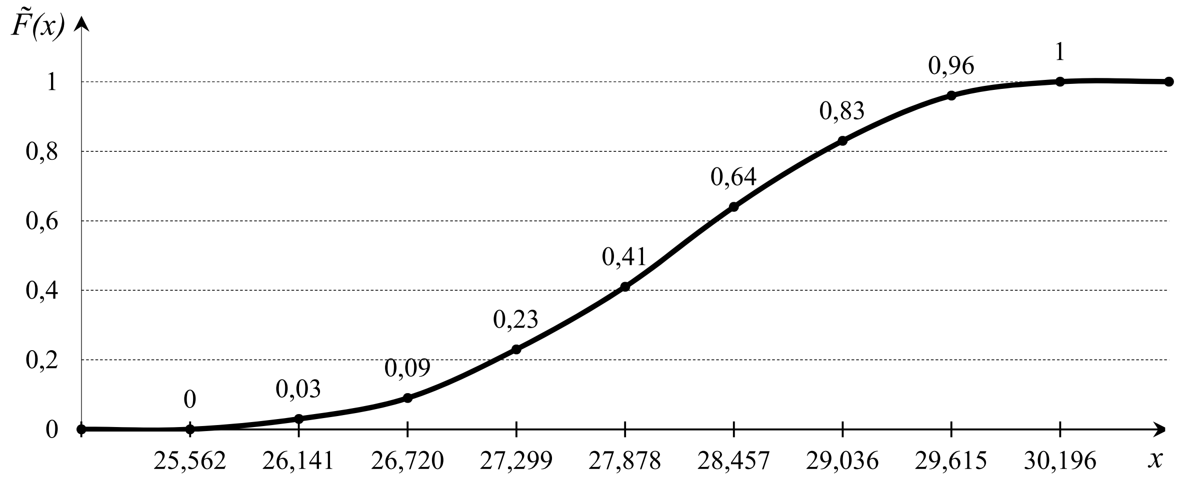

Рисунок 4

6 Вычислим числовые характеристики выборки.

$$\overline{x}_B=\frac{1}{n}\sum_{i=1}^{8}x_i^*\cdot m_i=\frac{1}{100}\left(25,8515\cdot 3+...+29,9055\cdot 4\right)\approx 28,0575;$$

$$D_B=\frac{1}{n}\sum_{i=1}^{8}\left(x_i^*\right)^2\cdot m_i-\overline{x}_B^2=\frac{1}{100}(25,8515^2\cdot 3+...+29,9055^2\cdot 4)-28,0575^2\approx 0,9368;$$

$$\sigma_B=\sqrt{D_B}=\sqrt{0,9368}\approx 0,9679;$$

$$Me_B=\frac{27,5885+28,1675}{2}=27,878;$$

$$Mo_B=28,1675 \quad (m=23);$$

$$A_B = \frac{\tilde{\mu}_3}{\sigma_B^3} = \frac{1}{n\sigma_B^3}\sum_{i=1}^{8}\left(x_i^* - \overline{x}_B\right)^3 \cdot m_i =$$

$$= \frac{1}{100 \cdot 0{,}9678^3}((25{,}8515 - 28{,}0575)^3 \cdot 3 + \ldots + (29{,}9055 - 28{,}0575)^3 \cdot 4) \approx -0{,}1989;$$

$$E_B = \frac{\tilde{\mu}_4}{\sigma_B^4} - 3 = \frac{1}{n\sigma_B^4}\sum_{i=1}^{8}\left(x_i^* - \overline{x}_B\right)^4 \cdot m_i - 3 =$$

$$= \frac{1}{100 \cdot 0{,}9678^4}((25{,}8515 - 28{,}0575)^4 \cdot 3 + \ldots + (29{,}9055 - 28{,}0575)^4 \cdot 4) - 3 \approx$$

$$\approx -0{,}5455.$$

7 Вид гистограммы и полигона относительных частот (рисунки 2,3) напоминает кривую Гаусса. Поэтому можно предположить, что распределение является нормальным.

8 Плотность вероятности и функция нормального распределения имеют следующий вид:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}};$$

$$F(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \frac{1}{2} + \Phi\left(\frac{x-m}{\sigma}\right).$$

Найдём точечные оценки параметров m и σ нормального распределения методом моментов.

$$\tilde{m} = \overline{x}_B = 28{,}0575;$$

$$\tilde{\sigma} = \sigma_B = 0{,}9679.$$

Несмещённая оценка среднего квадратического отклонения

$$S_X = \sqrt{\frac{n}{n-1}}\sigma_B = \sqrt{\frac{100}{100-1}} \cdot 0{,}9679 \approx 0{,}9727.$$

Следовательно, плотность вероятности предполагаемого нормального распределения имеет вид:

$$f(x) = \frac{1}{0{,}9679\sqrt{2\pi}} e^{-\frac{(x-28{,}0575)^2}{1{,}8736}}, \quad f(x) = 0{,}412 \cdot e^{-\frac{(x-28{,}0575)^2}{1{,}8736}}.$$

Функция распределения предполагаемого нормального распределения

$$F(x)=\frac{1}{0{,}9679\sqrt{2\pi}}\int_{-\infty}^{x} e^{-\frac{(x-28{,}0575)^2}{1{,}8736}}dx, \qquad F(x)=\frac{1}{2}+\Phi\left(\frac{x-28{,}0575}{0{,}9679}\right).$$

Значения плотности распределения занесём в таблицу 33.

Таблица 33

x_i^*	25,8515	26,4305	27,0095	27,5885	28,0575	28,1675	28,7465	29,3255	29,9055
$f\left(x_i^*\right)$	0,0307	0,1003	0,2293	0,3665	0,4122	0,4095	0,3199	0,1747	0,0666

Построим график плотности распределения теоретической СВ X на одной координатной плоскости с гистограммой (рисунок 5).

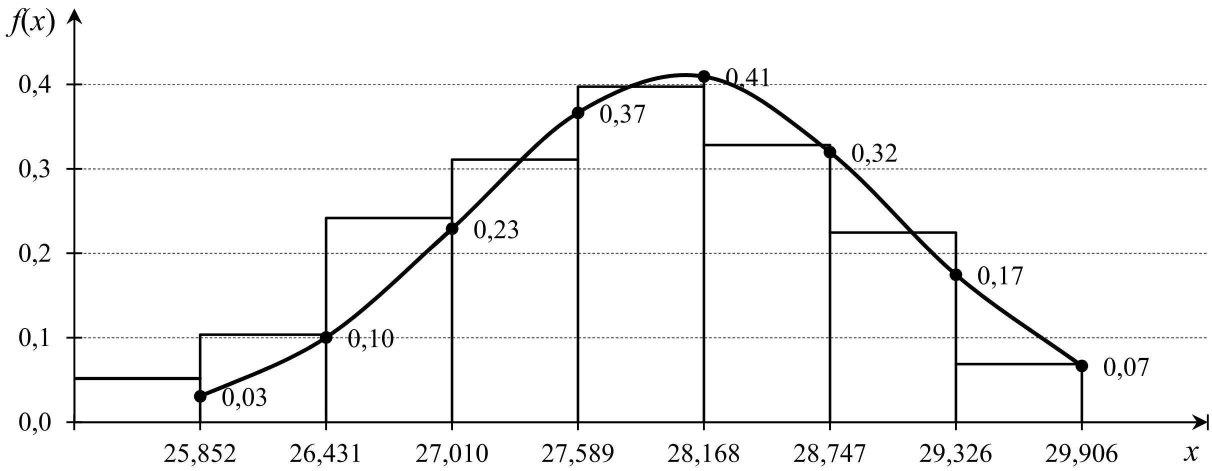

Рисунок 5

9 Проверим гипотезу о нормальном распределении СВ X при помощи критерия согласия χ^2: $H_o: F(x)=\tilde{F}(x),\ H_a: F(x)\neq\tilde{F}(x)$.

Согласно критерию Пирсона, необходимо сравнить эмпирические и теоретические частоты. Эмпирические частоты m_i даны. Находим теоретические частоты m'.

Объединяем интервалы, для которых $m_i<5$.

Нормируем СВ X, т. е. переходим к СВ $U=\dfrac{X-\overline{X}}{\sigma_B}$ (т. к. χ^2 – распределение для стандартизированной СВ $N(0;1)$).

Наименьший конец интервала устремляем к $-\infty$, наибольший – к $+\infty$.

Вероятности попадания в интервалы определяем по формуле

$$p_i=P\left(u_i<U<u_{i+1}\right)=\Phi\left(u_{i+1}\right)-\Phi\left(u_i\right), \text{ где } \Phi\left(u_i\right)=\frac{1}{\sqrt{2\pi}}\int_0^{u_i} e^{-\frac{t^2}{2}}dt.$$

Имеем

$$u_2 = \frac{x_2 - \overline{x}_B}{\sigma_B} = \frac{26,720 - 28,0575}{0,9679} \approx -1,38 \text{ и т. д.}$$

$$p_1 = P(-\infty < U < -1,38) = \Phi(-1,38) - \Phi(-\infty) = 0,5 - 0,4162 = 0,0838;$$

$$p_2 = P(-1,38 < U < -0,78) = \Phi(-0,78) - \Phi(-1,38) = 0,4162 - 0,2823 = 0,1339 \text{ и т. д.}$$

Полученные значения заносим в таблицу 34.

Таблица 34

$[x_i; x_{i+1})$	m_i	$[u_i; u_{i+1})$	p_i	$m' = np_i$	$\frac{(m_i - m')^2}{m'}$
$[25,562; 26,720)$	9	$(-\infty; -1,38)$	0,0838	8,38	0,0459
$[26,720; 27,299)$	14	$[-1,38; -0,78)$	0,1339	13,39	0,0278
$[27,299; 27,878)$	18	$[-0,78; -0,19)$	0,207	20,7	0,3522
$[27,878; 28,457)$	23	$[-0,19; 0,41)$	0,2344	23,44	0,0083
$[28,457; 29,036)$	19	$[0,41; 1,01)$	0,1847	18,47	0,0152
$[29,036; 30,196]$	17	$[1,01; +\infty)$	0,1562	15,62	0,1219
	$\sum_i m_i = 100$		$\sum_i p_i = 1$		$\chi^2 = 0,5712$

Находим наблюдаемое $\chi^2 = \sum_{i=1}^{k} \frac{(m_i - m')^2}{m'}$.

По таблице критических точек распределения χ^2 находим критическое значение $\chi^2(\alpha, k)$, где α – заданный уровень значимости; $k = s - r - 1 = 6 - 2 - 1 = 3$ – число степеней свободы ($s = 4$ – число интервалов; $r = 2$ – число параметров распределения): $\chi^2_{кр} = \chi^2(0,05; 3) = 7,8$.

Так как $\chi^2_{набл} = 0,5712 < \chi^2_{кр} = 7,8$, то гипотеза о нормальном распределении принимается.

Проверим гипотезу о нормальном распределении с помощью λ-критерия Колмогорова. Все вспомогательные расчёты сведём в таблицу 35.

$$F(x_1) = p_1 = 0,0239; \quad F(x_2) = p_1 + p_2 = 0,0239 + 0,0599 = 0,0838 \text{ и т. д.}$$

$$D = \max\left|\tilde{F}(x) - F(x)\right| = 0,0677; \; \lambda_{набл} = D \cdot \sqrt{n} = 0,0677 \cdot \sqrt{100} = 0,677.$$

По таблице критических точек распределения Колмогорова, по уровню значимости $\alpha = 0,05$ находим критическое значение $\lambda_{кр} = \lambda(0,05) = 1,358$.

Так как $\lambda_{набл} = 0,677 < \lambda_{кр} = 1,358$, то нет оснований отвергать гипотезу о нормальном распределении.

Таблица 35

$[x_i;x_{i+1})$	m_i	$[u_i;u_{i+1})$	p_i	$\tilde{F}(x)$	$F(x)$	$\lvert\tilde{F}(x)-F(x)\rvert$
$[25,562;26,141)$	3	$(-\infty;-1,98)$	0,0239	0,03	0,0239	0,0139
$[226,141;26,720)$	6	$[-1,98;-1,38)$	0,0599	0,09	0,0838	0,0538
$[26,720;27,299)$	14	$[-1,38;-0,78)$	0,1339	0,23	0,2177	0,0677
$[27,299;27,878)$	18	$[-0,78;-0,19)$	0,207	0,41	0,4247	0,0247
$[27,878;28,457)$	23	$[-0,19;0,41)$	0,2344	0,64	0,6591	0,0209
$[28,457;29,036)$	19	$[0,41;1,01)$	0,1847	0,83	0,8438	0,0362
$[29,036;29,615)$	13	$[1,01;1,61)$	0,1025	0,96	0,9463	0,0237
$[29,615;30,196]$	4	$[1,61;+\infty)$	0,0537	1	1	0
	$\sum_i m_i = 100$		$\sum_i p_i = 1$			

10 Доверительный интервал для математического ожидания (при неизвестном σ) определяем по формуле

$$\overline{x}_B - t_\gamma \frac{S}{\sqrt{n}} < m < \overline{x}_B + t_\gamma \frac{S}{\sqrt{n}}.$$

По таблице значений $t_\gamma = t(\gamma, n)$ находим $t_\gamma = t(0{,}95;100) = 1{,}984$. Имеем

$$28{,}0575 - 1{,}984 \cdot \frac{0{,}9727}{\sqrt{100}} < m < 28{,}0575 + 1{,}984 \cdot \frac{0{,}9727}{\sqrt{100}};\ 27{,}8645 < m < 28{,}2505.$$

Доверительный интервал для среднеквадратического отклонения находим по одной из формул

$$S(1-q) < \sigma < S(1+q), \quad q < 1;$$

$$0 < \sigma < S(1+q), \quad q > 1.$$

По таблице значений $q = q(\gamma, n)$ находим $q = q(0{,}95;100) = 0{,}143 < 1$.

$$0{,}9727 \cdot (1 - 0{,}143) < \sigma < 0{,}9727 \cdot (1 + 0{,}143);\ 0{,}8336 < \sigma < 1{,}1119.$$

11 Проверим нулевую гипотезу $\alpha =: m = m_0 = 28$ при альтернативной гипотезе $H_1 : m \neq 28$.

$$T_{набл.} = \frac{\overline{x}_B - m_0}{S}\sqrt{n} = \frac{28{,}0575 - 28}{0{,}9727}\sqrt{30} = 0{,}324.$$

Критическая область двусторонняя. По таблице критических точек распределения Стьюдента, по заданному уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = 30 - 1 = 29$ находим $t_{пр.кр} = t(0,025;29) = 2,045$. Так как $|T_{набл}| < t_{кр}$, то нет оснований отвергать нулевую гипотезу, т. е. $m = 28$.

12 Проверим нулевую гипотезу $H_0 : \sigma^2 = \sigma_0^2 = 1$ против альтернативной $H_1 : \sigma^2 < 1$.

$$\chi^2_{набл} = \frac{(n-1)s^2}{\sigma_0^2} = \frac{29 \cdot 0,9727^2}{1} = 27,44\,.$$

Критическая область левосторонняя. По таблице критических точек распределения χ^2, по уровню значимости $\alpha = 0,05$ и числу степеней свободы $k = 29$ находим $\chi^2_{кр} = \chi^2(1-\alpha, k) = \chi^2(0,95;29) = 17,7$. Так как $\chi^2_{набл} > \chi^2_{кр}$, нет оснований отвергать нулевую гипотезу.

Список литературы

1 **Герасимович, А. И.** Математическая статистика / А. И. Герасимович. – Минск : Вышэйшая школа, 1983.

2 **Гурский, Е. И.** Сборник задач по теории вероятностей и математической статистике / Е. И. Гурский. – Минск : Вышэйшая школа, 1984.

3 **Гмурман, В. Е.** Руководство к решению задач по теории вероятностей и математической статистике / В. Е. Гмурман. – Москва : Высшая школа, 2002.

4 **Булдык, Г. М.** Теория вероятностей и математическая статистика / Г. М. Булдык. – Минск : Вышэйшая школа, 1989.

5 **Калинина, В. Н.** Математическая статистика / В. Н. Калинина, В. Ф. Панкин. – Москва : Высшая школа, 1998.

6 **Копытов, А. Е.** Теория вероятностей и математическая статистика / А. Е. Копытов, Е. А. Гринглаз. – Санкт-Петербург : Питер, 2004.

7 **Гмурман, В. Е.** Теория вероятностей и математическая статистика / В. Е. Гмурман. – Москва : Высшая школа, 2004.

8 **Колемаев, В. А.** Теория вероятностей и математическая статистика / В. А. Колемаев. – Москва : Высшая школа, 2002.

9 **Белько, И. В.** Теория вероятностей и математическая статистика. Примеры и задачи / И. В. Белько. – Минск : Новое знание, 2002.

10 **Фигуркин, В. А.** Теория вероятностей и математическая статистика / В. А. Фигуркин, В. В. Оболонкин. – Минск : Новое знание, 2000.

11 Сборник задач и упражнений по высшей математике. Теория вероятностей и математическая статистика / И. П. Мацкевич [и др.]. – Минск : Вышэйшая школа, 1996.