МИНИСТЕРСТВО ОБРАЗОВАНИЯ
РЕСПУБЛИКИ БЕЛАРУСЬ

Белорусский национальный
технический университет

Кафедра «Высшая математика № 3»

В. И. Ерошевская
Е. Л. Ерошевская
Л. П. Минченкова

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методическое пособие

Часть 2

Минск
БНТУ
2014

МИНИСТЕРСТВО ОБРАЗОВАНИЯ РЕСПУБЛИКИ БЕЛАРУСЬ
Белорусский национальный технический университет

Кафедра «Высшая математика № 3»

В. И. Ерошевская
Е. Л. Ерошевская
Л. П. Минченкова

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методическое пособие

В 2 частях

Часть 2

Рекомендовано учебно-методическим объединением
в сфере высшего образования Республики Беларусь
по образованию в области горнодобывающей промышленности

Минск
БНТУ
2014

УДК 519.2 (075.8)
ББК 22.1я7
Е78

Р е ц е н з е н т ы:
М. В. Дубатовская, *И. В. Белько*

Ерошевская, В. И.

Е78 Математическая статистика : методическое пособие : в 2 ч. / В. И. Ерошевская, Е. Л. Ерошевская, Л. П. Минченкова. – Минск : БНТУ, 2013–2014. – Ч. 2. – 2013. – 75 с.

ISBN 978-985-550-356-0 (Ч. 2).

Методическое пособие предназначено для помощи студентам в их самостоятельной работе при изучении темы «Математическая статистика».

Вторая часть издания содержит материал, связанный с проверкой статистических гипотез и регрессионно-корреляционным анализом. Теоретический материал проиллюстрирован примерами. К практическим занятиям предложены задачи для самостоятельной работы, к которым даны ответы. Включены таблицы значений функций, необходимые для решения задач.

Часть 1 вышла в 2013 г.

УДК 519.2 (075.8)
ББК 22.1я7

ISBN 978-985-550-356-0 (Ч. 2)
ISBN 978-985-550-112-2

Занятие № 5

Понятие о статистических гипотезах. Уровень значимости и мощность критерия

Занятие № 6

Проверка статистических гипотез для непрерывных случайных величин

Занятие № 7

Проверка статистических гипотез для дискретных случайных величин

Занятие № 8

Корреляционный анализ

Занятие № 9

Линейная регрессия. Определение параметров линейной регрессии

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

Основные понятия и определения

Статистической называют гипотезу о виде неизвестного распределения или о параметрах распределений.

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

1. Гипотезы о виде закона распределения исследуемой случайной величины.
2. Гипотезы о числовых значениях параметров случайной величины.
3. Гипотезы об общем виде модели, описывающей статистическую зависимость между признаками.
4. Гипотезы о принадлежности некоторого признака к тому или иному классу.

Статистическая гипотеза H называется *простой*, если она однозначно определяет распределение случайной величины X, в противном случае гипотеза H называется *сложной*.

Проверяемая гипотеза называется *нулевой гипотезой* и обозначается H_0.

Конкурирующей (альтернативной) называют гипотезу H_0, которая противоречит основной.

Выбор альтернативной гипотезы определяется конкретной формулировкой задачи.

Правило, по которому принимается решение принять или отклонить гипотезу H_0, называется *статистическим критерием*.

Проверку статистических гипотез выполняют на основании результатов выборки. Следовательно, статистический критерий устанавливает, при каких результатах выборки проверяемая гипотеза принимается, а при каких – отвергается. Значения параметра, при которых гипотеза отвергается, образуют критическую область проверяемой гипотезы. Понятие критической области нельзя отождествлять с понятием критерия. Задача проверки гипотезы сводится к нахождению критической области для данного уровня значимости. Если значения параметров попадают в критическую область, то это указывает на несоответствие гипотезы наблюдаемым данным, и гипотеза должна быть отвергнута.

При проверке гипотезы возможны ошибки двух родов. *Ошибка первого рода* состоит в том, что нулевая гипотеза H_0 отвергается, в то время как она в действительности верна. *Ошибка второго рода* состоит в том, что нулевая гипотеза H_0 принимается, а на самом деле она неверна.

Для определения лучшего критерия проверки гипотезы H_0 необходимо среди всех критериев, имеющих одну и ту же вероятность ошибки первого рода, выбрать тот, для которого вероятность ошибки второго рода наименьшая. Допускаемую вероятность ошибки первого рода можно задать заранее.

УРОВЕНЬ ЗНАЧИМОСТИ И МОЩНОСТЬ КРИТЕРИЯ

Пусть произведена повторная выборка $x_1, x_2, \ldots, x_n$ объема n из генеральной совокупности, где, согласно гипотезе H_0, исследуемый признак X имеет функцию распределения $F(x)$. Чтобы построить статистический критерий, который позволит на основании полученных выборочных данных проверить гипотезу H_0, необходимо выполнить следующее:

1) определить статистическую характеристику гипотезы $f = f(x_1, x_2, \ldots, x_n)$, по значениям которой можно принять или отвергнуть гипотезу H_0;

2) найти допустимую вероятность α ошибки первого рода, которую называют уровнем значимости гипотезы (как правило, $\alpha = 0{,}05$; 0,01; 0,001);

3) сформулировать гипотезу H_1, альтернативную к проверяемой;

4) определить критическую область для проверяемой гипотезы, т. е. такое множество W значений статистики f, что если наблюдаемое значение попадает в W, то гипотеза H_0 отвергается в пользу альтернативной H_1. Как уже отмечалось ранее, при заданном уровне значимости α критическая область должна быть выбрана так, чтобы вероятность совершить ошибку второго рода была наименьшей. Из сказанного следует, что статистика f и критическая область W должны удовлетворять условиям

$$P_{H_0}\left(f \in W\right) = \alpha \;\text{ и }\; P_{H_1}\left(f \overline{\in} W\right) = P_{\min}. \tag{1}$$

При построении критерия желательно сократить до минимума вероятности ошибок первого и второго рода. Но это сделать невозможно. Действительно, пусть возможные значения статистики f попадают в интервал A (рис. 1), а W есть критическая область. Ошибка первого рода возникает, когда статистика f попадает в критическую область, а на самом деле нулевая гипотеза верна. Для того чтобы уменьшить эту ошибку, необходимо, насколько это возможно, сузить критическую область W.

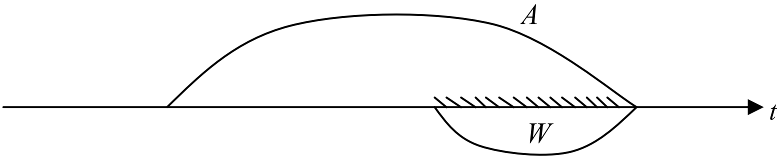

Рис. 1

Но это влияет на ошибку второго рода, так как она возникает, когда статистика, лежащая в основе критерия, попадает в область $A \backslash W$, в то время как в действительности нулевая гипотеза логична. Для того чтобы минимизировать эти ошибки, нужно стремиться к уменьшению области неотбрасывания $A \backslash W$. Таким образом, ошибки первого и второго рода конкурируют друг с другом и невозможно их сделать сколь угодно малыми.

Для определения множеств W и $A\backslash W$ вводят функцию потерь, возникающих в результате любой из ошибок. Размер областей необходимо выбирать таким образом, чтобы ущерб от каждого типа потерь был приблизительно одинаковым.

Вероятность β недопущения ошибки второго рода называют ***мощностью*** *критерия* и определяют формулой

$$\beta = P_{H_1}\left(f \overline{\in} W\right). \qquad (2)$$

Если при увеличении объема выборки вероятность β стремится к единице, то такой статистический критерий называют ***критерием*** *согласия*.

Критерий согласия проверяет, согласуется ли наблюдаемое распределение с гипотетическим. Особую роль играет проверка на нормальность распределения признака, поскольку многие критерии эту нормальность предполагают. Критерий, не использующий предположения о распределении, называют непараметрическим. Критерий согласия относится к непараметрическим критериям. Непараметрические критерии по отношению к параметрическим имеют меньшую мощность. При использовании непараметрических критериев в случаях нормального или равномерного распределения имеет место увеличение ошибок второго рода. Статистические решения становятся тогда консервативными и для отклонения нулевой гипотезы нужен большой объем выборки.

КРИТЕРИЙ СОГЛАСИЯ ПИРСОНА. ПРОВЕРКА ГИПОТЕЗЫ О РАСПРЕДЕЛЕНИИ СЛУЧАЙНОЙ ВЕЛИЧИНЫ ПО НОРМАЛЬНОМУ ЗАКОНУ

Алгоритм применения критерия Пирсона заключается в следующем:

1) из генеральной совокупности образовывается случайная выборка и на ее основании делается предположение (гипотеза) о законе распределения признака генеральной совокупности;

2) вычисляются параметры этого закона;

3) вычисляются теоретические частоты;

4) по формуле

$$\chi^2 = \sum_{i=1}^{k} \frac{\left(m_i - m_i'\right)^2}{m_i'} \tag{3}$$

находится $\chi^2_{\text{набл}}$. Теоретические частоты находятся по выражению

$$m_i' = m_i \cdot p_i\,; \tag{4}$$

5) вычисляется число степеней свободы $\nu = k - (r+1)$, где k – число интервалов из выборки; r – сумма числа параметров предполагаемого закона распределения;

6) выбирается уровень значимости α;

7) из таблицы П5 для критерия χ^2 Пирсона при известных уже значениях ν и $\chi^2_{\text{набл}}$ находится вероятность того, что случайная величина χ^2 примет значение, не меньшее $\chi^2_{\text{набл}}$ $\left(P\left(\chi^2 \geq \chi^2_{\text{набл}}\right) = \beta\right)$;

8) если $\beta > \alpha$, то следует считать несущественными расхождения между теоретическими и эмпирическими частотами. В этом случае гипотеза о распределении признака генеральной совокупности по предполагаемому закону принимается. Если $\beta \leq \alpha$, то гипотеза отвергается.

Отметим, что χ^2 – распределение можно применять только при достаточно большом объеме выборки $(n \geq 50)$ и для больших частот $m_i \geq 5$. Вместо нахождения β можно находить границу критической области из таблицы П5. Если $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то гипотеза подтверждается.

Критерий χ^2 находит применение не только в связи с проверкой нормального распределения генеральной совокупности, но его можно использовать для проверки гипотез о том, что данная выборка взята из генеральной совокупности, распределенной по биномиальному закону или закону Пуассона.

Пример 1. На предприятии провели выборочный опрос работающих об их средней заработной плате за предыдущий год. Данные опроса представлены в табл. 1.

Таблица 1

Значение признака x_i	Частота m_i	Значение признака x_i	Частота m_i
987	1	1003	11
989	2	1005	10
991	2	1007	8
993	6	1009	7
995	7	1011	3
997	12	1012	1
999	13	1016	1
1001	16		$n = 100$

С помощью критерия Пирсона проверить гипотезу о том, что средняя заработная плата по всему предприятию распределена по нормальному закону с уровнем значимости $\alpha = 0,05$.

Решение. Найдем функцию распределения признака X в генеральной совокупности, применяя формулу $F(x) = 1 + \Phi\left(\frac{x - \overline{x}_0}{\sigma_0}\right)$.

Для этого предварительно вычислим среднюю выборочную и исправленную статическую дисперсию S^2:

$$\overline{X} = \sum_{i=1}^{15} x_i \cdot \frac{m_i}{n} = 1001;$$

$$S^2 = \frac{n}{n-1}\sigma_o^2 = \frac{100}{99}\sum_{i=1}^{15}\frac{m_i}{100}(x_i - 1001)^2 = 31,82;$$

$$S = 5,64.$$

За функцию распределения признака X в генеральной совокупности принимается $F(x) = 1 + \Phi\left(\frac{x - 1001}{5,64}\right)$.

Выполним разбиение области значений случайной величины x на интервалы $S_1, S_2, \ldots, S_{11}$ таким образом, чтобы частоты были больше или равны 5. Разбиение представлено в табл. 2.

Таблица 2

Интервал	Значение признака на интервале $x_i' - x_i''$	Частота m_i	Интервал	Значение признака на интервале $x_i' - x_i''$	Частота m_i
S_1	–∞–992	5	S_7	1002–1004	11
S_2	992–994	6	S_8	1004–1006	10
S_3	994–996	7	S_9	1006–1008	8
S_4	996–998	12	S_{10}	1008–1010	7
S_5	998–1000	13	S_{11}	1010 – +∞	5
S_6	1000–1002	16			
					100

Для расчета теоретического ряда частот необходимо предварительно вычислить значения вероятностей p_i по формуле

$$p_i = P\left(x_i' < X < x_i''\right) = \Phi\left(\frac{x_i'' - \overline{X}}{S}\right) - \Phi\left(\frac{x_i' - \overline{X}}{S}\right)$$

и применить формулу для вычисления теоретических частот:

$$m_i' = nP\left(X = x_i \in S_i\right) = np_i ,$$

где $i = \overline{1,11}$.

Например,

$$p_2 = \Phi\left(\frac{994 - 1001}{5,64}\right) - \Phi\left(\frac{992 - 1001}{5,64}\right) = \Phi(-1,24) - \Phi(-1,59) =$$

$$= -0,7850 + 0,8882 = 0,0516;$$

$$m_2' = np_2 = 100 \cdot 0,0516 = 5,16 .$$

Значение функции $\Phi(x)$ вычисляется по табл. П2 (часть 1) значений функции Лапласа. Результаты вычислений представим в табл. 3.

Таблица 3

$x_i' - x_i''$	m_i	p_i	m_i'	$x_i' - x_i''$	m_i	p_i	m_i'
–∞–992	5	0,0559	5,59	1002–1004	11	0,1305	13,05
992–994	6	0,0516	5,16	1004–1006	10	0,1113	11,13
994–996	7	0,0792	7,92	1006–1008	8	0,0792	7,92
996–998	12	0,1113	11,13	1008–1010	7	0,0516	5,16
998–1000	13	0,1305	13,05	1010 – +∞	5	0,0559	5,59
1000–1002	16	0,1428	14,28				
					100	1	100

Значения χ^2 вычислим по формуле $\chi^2 = \sum_{i=1}^{11} \frac{(m_i - m_i')^2}{m_i'} = 1,74$.

Так как два параметра распределения признака в генеральной совокупности находились на основании выборки, то функцию χ^2 можно приближенно считать распределенной по закону χ^2 с $\nu = 11 - 2 - 1 = 8$ степенями свободы (здесь число интервалов $k = 11$, $r = 2$, поскольку оценивались два параметра закона распределения). При уровне значимости $\alpha = 0,05$ границей критической области будет $\chi^2_{0,05;8} = 15,5$ (по табл. П5). Так как 1,74 < 15,5, то гипотеза H_0 о нормальном распределении генеральной совокупности принимается.

Следует отметить, что на практике все шире начинают применять критерии согласия не столько для проверки согласия экспериментальных данных с некоторой гипотетической функцией, сколько для подбора наилучшей функции распределения, хотя выбор подходящего закона должен основываться прежде всего на понимании механизма изучаемого явления.

ПРОВЕРКА ГИПОТЕЗЫ О РАСПРЕДЕЛЕНИИ СЛУЧАЙНОЙ ВЕЛИЧИНЫ ПО РАВНОМЕРНОМУ ЗАКОНУ

Пусть случайная величина X распределена равномерно на заданном отрезке [a; b].

Выборочные данные сгруппируем и представим в виде последовательности интервалов $\left[x_i'; x_{i+1}'\right)$ и соответствующих им частот m_i, $i = 1, \ldots, k$, $x_1' = a$, $x_{k+1}' = b$.

Затем вычисляем p_i попадания X в частичные интервалы, определяем теоретические частоты $m_i' = m_i \cdot p_i$, где $n = \sum_{i=1}^{k} m_i$ – объем выборки, и находим наблюдаемое значение критерия Пирсона $\chi^2_{\text{крит}}$ по заданному уровню α и числу степеней свободы $\nu = k - 1$, принимаем одно из решений:

если $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то гипотеза H_0 о равномерном распределении X на отрезке [a; b] принимается;

если $\chi^2_{\text{набл}} \geq \chi^2_{\text{крит}}$, то гипотеза H_0 отклоняется.

Пример 2. Наблюдалось следующее распределение по минутам числа появлений на остановке автобуса, имеющего пятиминутный интервал движения.

Интервал в минутах	Число появлений автобуса
0-1	35
1-2	34
2-3	38
3-4	36
4-5	42

Проверить гипотезу о равномерном законе распределения.

Решение. 1. Вычисляем по данному вариационному ряду вероятности p_i попадания X в интервал по формуле

$$p_i = P\left(x_i' < X < x_{i+1}'\right) = \frac{x_{i+1}' - x_i'}{b - a}, \text{ где } i = 1, \ldots, m,\ x_1' = a,\ x_{m+1}' = b.$$

2. Для проверки гипотезы о том, что число появлений автобуса на остановке есть случайная величина, распределенная по равномерному закону, вычисляем критерий $\chi^2_{\text{набл}} = \sum_{i=1}^{k} \frac{\left(m_i - m_i'\right)^2}{m_i'} = \frac{m_i^2}{m_i'} - n$, для чего составим таблицу.

Таблица 4

Вычисление $\chi^2_{\text{набл}}$

	m_i'	$p_i = \frac{x_{i+1}' - x_i'}{b-a}$	$m_i' = n \cdot p_i$	$\left(m_i - m_i'\right)^2$	$\frac{\left(m_i - m_i'\right)^2}{m_i'}$	m_i^2	$\frac{m_i^2}{m_i'}$
0-1	35	$\frac{(1-0)}{5} = 0,2$	37	4	0,11	1225	33,11
1-2	34	$\frac{(2-1)}{5} = 0,2$	37	9	0,24	1156	31,24
2-3	38	$\frac{(3-2)}{5} = 0,2$	37	1	0,03	1444	39,03
3-4	36	$\frac{(4-3)}{5} = 0,2$	37	1	0,03	1296	35,03
4-5	42	$\frac{(5-4)}{5} = 0,2$	37	25	0,68	1764	47,68
Σ	185				$\chi^2_{\text{набл}} = 1,09$		186,09

Контроль: $186,09 - 185 = 1,09$. Вычисления произведены правильно.

3. Определяем $\chi^2_{\text{крит}}$ по заданному уровню значимости $\alpha = 0,05$ и числу степеней свободы $\nu = k - 1 = 5 - 1 = 4$, то есть $\chi^2_{\text{крит}(0,05;4)} = 9,5$.

4. Так как $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то нет оснований отклонить гипотезу о равномерном распределении X на отрезке [0; 5].

ПРОВЕРКА ГИПОТЕЗЫ О РАСПРЕДЕЛЕНИИ НЕПРЕРЫВНОЙ СЛУЧАЙНОЙ ВЕЛИЧИНЫ ПО ПОКАЗАТЕЛЬНОМУ (ЭКСПОНЕНЦИАЛЬНОМУ) ЗАКОНУ

Рассмотрим дифференциальную функцию показательного закона распределения:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{при} \quad x \geq 0, \\ 0 & \text{при} \quad x < 0, \end{cases}$$

где λ – постоянная положительная величина.

Как и в предыдущих проверках, исходные выборочные данные подвергаем группировке, представляем в виде последовательности частичных интервалов и соответствующих им частот, а затем вычисляем выборочную среднюю $\overline{X}$ и принимаем в качестве оценки параметра λ показательного распределения величину, обратную $\overline{X}$: $\lambda = \dfrac{1}{\overline{\overline{X}}}$.

Определение вероятности p_i попадания случайной величины X в частичные интервалы $\left[x_i'; x_{i+1}'\right)$ будем осуществлять по формулам

$$p_1 = P\left(0 < X < x_2'\right) = 1 - e^{-\lambda x_2'};$$

$$p_i = P\left(x_i' < X < x_{i+1}'\right) = e^{-\lambda x_i'} - e^{-\lambda x_{i+1}'}, \; i = 2, 3, \ldots, k-1;$$

$$p_k = P\left(x_k' < X < x' < +\infty\right) = e^{-\lambda x_k'}, \text{ где } \lambda = \frac{1}{\overline{\overline{X}}}.$$

Но если параметр λ задан, то его оценка заменяется точным значением.

При определении критического значения χ^2 по заданному уровню α число степеней свободы $\nu = k - r - 1$, где r равно 1 или 0 в зависимости от того, оценивается ли по выборке параметр λ или нет.

В результате принимаем одно из решений: если $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то считаем, что нет оснований отклонить нулевую гипотезу о распределении случайной величины X по показательному закону; если $\chi^2_{\text{набл}} > \chi^2_{\text{крит}}$, то нулевую гипотезу отклоняем.

Пример 3. Рассмотрим вариационный ряд.

Интервалы	Частоты
0–4	115
4–8	50
8–12	20
12–16	10
16–20	5
	$\sum_{i=1}^{5} m_i = n = 200$

1. Если построить гистограмму частостей, то ее вид будет напоминать экспоненциальную кривую. Поэтому произведем «выравнивание» статистических данных по показательному закону. Запишем его дифференциальную функцию:

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{при} \quad x \geq 0, \\ 0 & \text{при} \quad x < 0. \end{cases}$$

Для нахождения точечной оценки параметра $\lambda = \frac{1}{\overline{X}}$ вначале вычислим $\overline{X}$, заменив каждый интервал его серединой:

$$\overline{X} = \frac{\sum_{i=1}^{k} x_i^* m_i}{\sum_{i=1}^{k} m_i} = \frac{\sum_{i=1}^{5} x_i^* m_i}{200} = \frac{2 \cdot 115 + 6 \cdot 50 + 10 \cdot 20 + 14 \cdot 10 + 18 \cdot 5}{200} = 4{,}8\,.$$

Тогда $\lambda = \frac{1}{4{,}8} = 0{,}208$. Следовательно, дифференциальная функции предполагаемого показательного закона распределения имеет вид $f(x) = 0{,}208 e^{-0{,}208x}$.

2. Для проверки соответствия эмпирических данных с предполагаемым показательным законом распределения применим критерий согласия χ^2.

3. Вычислим вероятности попадания случайной величины X в частичные интервалы $\left(x_i'; x_{i+1}'\right)$ по формуле

$$p_i = P\left(x_i' < X < x_{i+1}'\right) = \int\limits_{x_i'}^{x_{i+1}'} f\left(x\right)\mathrm{d}x = 0,208 \int\limits_{x_i'}^{x_{i+1}'} e^{-0,208x}\mathrm{d}x =$$

$$= 0,208 \frac{e^{-0,208x}}{-0,208}\Bigg|_{x_i'}^{x_{i+1}'} = e^{-0,208x_i'} - e^{-0,208x_{i+1}'}.$$

Для нахождения $\chi^2_{\text{набл}}$ построим вспомогательную таблицу.

Таблица 5

Вычисление $\chi^2_{\text{набл}}$

Интервалы	m_i	p_i	$m_i' = n \cdot p_i$	$\left(m_i - m_i'\right)^2$	$\frac{\left(m_i - m_i'\right)^2}{m_i'}$
0–4	115	0,57	114	1	0,01
4–8	50	0,25	50	0	0
8–12	20	0,12	22	4	0,18
12–16	10	0,05	10	0	0
16–20	5	0,02	4	1	0,25
$\sum$	200	1	200	–	$\chi^2_{\text{набл}} = 0,44$

4. Найдем в таблице критических точек χ^2 распределения по уровню значимости $\alpha = 0,05$ и числу степеней свободы $\nu = k - r - 1 = 5 - 1 - 1 = 3$ критическое значение $\chi^2_{\text{крит}(0,05;3)} = 7,8$.

5. Так как $\chi^2_{\text{набл}} = 0,44 < 7,8$, то нет оснований для отклонения гипотезы об экспоненциальном законе распределения.

ПРОВЕРКА ГИПОТЕЗЫ О РАСПРЕДЕЛЕНИИ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПО БИНОМИНАЛЬНОМУ ЗАКОНУ РАСПРЕДЕЛЕНИЯ

Пусть в каждом из N одинаковых испытаний случайная величина X_i, $i = 1, 2, \ldots, N$ (i – номер испытания) может принимать только два значения x_{i1}, x_{i2}. Обозначим через X случайную величину, равную сумме x_i: $X = x_1 + x_2 + \ldots + x_N$.

Произведено n независимых наблюдений над случайной величиной X и в итоге получен вариационный ряд дискретной случайной величины X:

x_i	0	1	2	…	N
m_i	m_0	m_1	m_2	…	m_N

где $n = \sum\limits_{i=1}^{N} m_i$.

Требуется проверить гипотезу о распределении случайной величины X по биноминальному закону с параметрами (N, p), причем параметр $p\,(0 < p < 1)$ или задан, или оценивается по выборке.

Для того чтобы при заданном уровне значимости проверить нулевую гипотезу о биноминальном распределении генеральной совокупности, будем придерживаться последовательности действий:

1. Находим величину $W = \overline{p} = \dfrac{\sum\limits_{i=1}^{N} x_i m_i}{nN} = \dfrac{\sum\limits_{i=1}^{N} W_i}{N}$, принимаем ее в качестве оценки вероятности p. Если параметр p задан, то всюду в дальнейшем используем точное его значение. Затем по формуле Бернулли находим вероятности $p_i = P\left(X = x_i\right) = C_N^i W^i \left(1 - W\right)^{N-i}$, где i = 0,1,…,N, и вычисляем теоретические частоты $m_i' = n \cdot p_i$, где $n = \sum\limits_{i=0}^{N} m_i$.

2. Для проверки нулевой гипотезы применяем критерий $\chi^2 = \sum_{i=1}^{k} \frac{\left(m_i - m_i'\right)^2}{m_i'}$, где k – число групп выборки, оставшихся после объединения малочисленных групп.

3. Вычисляем $\chi^2_{\text{набл}}$.

4. Определяем значение $\chi^2_{\text{крит}}$ с помощью таблицы критических точек распределения χ^2 по заданному уровню α и числу степеней свободы $\nu = k - r - 1$, где r равно 1 или 0 в зависимости от того, оценивался ли параметр p по выборке или нет.

Пример 4. Из продукции цеха случайно отобрано 200 выборок по 5 деталей. Регистрировалось число бракованных деталей. В итоге получен вариационный ряд:

Число бракованных деталей в одной выборке (x_i)	0	1	2	3	4	5
Частота m_i (количество выборок содержащих x_i бракованных деталей)	72	77	34	14	2	1

Требуется, используя критерий Пирсона при уровне значимости $\alpha = 0{,}05$, проверить гипотезу о том, что дискретная случайная величина X (число бракованных деталей) распределена по биномиальному закону.

1. Найдем частость W и применим ее в качестве оценки вероятности того, что наудачу взятая деталь окажется бракованной.

По формуле Бернулли $p_i = C_N^i p^i q^{N-i}$ найдем вероятности p_i ($i = 0, 1, 2, 3, 4, 5$) того, что интересующее нас событие появится в $n = 5$ испытаниях ровно i раз.

$$\overline{p} = W = \frac{\sum_{i=0}^{5} W_i x_i}{nN} = \frac{0 \cdot 72 + 1 \cdot 77 + 2 \cdot 34 + 3 \cdot 14 + 4 \cdot 2 + 5 \cdot 1}{200 \cdot 5} = \frac{200}{5 \cdot 200} = 0{,}2\,.$$

Таблица 6

Вычисление $\chi^2_{\text{набл}}$

i	m_i	$W_i = \overline{p}$	$i \cdot m_i$	p_i	$m_i' = n \cdot p_i$	$(m_i - m_i')^2$	$\frac{(m_i - m_i')^2}{m_i'}$
0	72	0,36	0	0,328	65,6	40,96	0,62
1	77	0,38	77	0,410	82,0	25	0,30
2	34	0,17	68	0,205	41,0	49	1,2
3	14	0,07	42	0,051	10,2	14,44	1,4
$\begin{matrix}4\\5\end{matrix}$	$\left.\begin{matrix}2\\1\end{matrix}\right\}3$	$\left.\begin{matrix}0,02\\0,005\end{matrix}\right\}0,025$	$\left.\begin{matrix}8\\5\end{matrix}\right\}13$	$\left.\begin{matrix}0,006\\0,00\end{matrix}\right\}0,006$	$\left.\begin{matrix}1,2\\0,0\end{matrix}\right\}1,2$	$\left.\begin{matrix}0,64\\1\end{matrix}\right\}1,64$	1,37
$\sum$	200	1	200	1	200	–	4,89

2. Для проверки нулевой гипотезы выдвигаем критерий $\chi^2 = \sum_{i=1}^{k} \frac{(m_i - m_i')^2}{m_i'}$, где k – число групп выборки, оставшихся после объединения.

3. Вычисляем $\chi^2_{\text{набл}}$. При составлении расчетной таблицы для сравнения эмпирических и теоретических частот с помощью критерия Пирсона мы объединим эмпирические частоты (2+1) и (0,006 и 0). После объединения число групп выборки $k = 5$.

4. Находим $\chi^2_{\text{крит}}$ один параметр (вероятность p) оценивался по выборке, то есть $r = 1$, поэтому при определении числа степеней свободы

$$\nu = k - r - 1 = 5 - 1 - 1 = 3;$$

$$\chi^2_{\text{крит}} = \chi^2_{\text{крит}(0,05;3)} = 7,8.$$

$\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, значит, нет оснований отклонить нулевую гипотезу о биноминальном законе распределения.

ПРОВЕРКА ГИПОТЕЗЫ О РАСПРЕДЕЛЕНИИ СЛУЧАЙНОЙ ВЕЛИЧИНЫ ПО ЗАКОНУ ПУАССОНА

Пусть дискретная случайная величина X может принимать только целые неотрицательные значения x_i. Из генеральной совокупности извлечем случайную выборку объема n:

x_i	0	1	2	…	k
m_i	m_0	m_1	m_2	…	m_k

Проверим гипотезу о том, что выборка извлечена из генеральной совокупности, имеющей распределение Пуассона.

1. По заданному вариационному ряду найдем выборочную среднюю $\overline{X}$ и примем ее в качестве оценки параметра λ распределения Пуассона. Если параметр λ задан в условии задачи, то при вычислении вероятности p_i в формуле Пуассона нужно заменить его оценку точным значением:

$$p_i = P\left(X = x_i\right) = \frac{\lambda^i}{i!}e^{-\lambda}, \ldots i = 0, 1, \ldots, k-1,$$

$$p_k = P\left(X \geq k\right) = \sum_{i=k}^{\infty} \frac{\lambda^i}{i!}e^{-\lambda}.$$

2. Применим критерий Пирсона $\chi^2 = \sum_{i=1}^{k} \frac{\left(m_i - m_i'\right)^2}{m_i'}$.

3. Вычисляем $\chi^2_{\text{набл}}$ по данным выборки.

4. Определяем значение $\chi^2_{\text{крит}}$ по таблице критических точек распределения χ^2 по заданному уровню α и числу степеней свободы $\nu = k - r - 1$, где r равно 1 или 0 в зависимости от того, оценивается ли параметр λ по выборке.

5. Если $\chi^2_{\text{набл}} < \chi^2_{\text{крит}}$, то нет оснований отклонять гипотезу о распределении случайной величины по закону Пуассона. Если $\chi^2_{\text{набл}} > \chi^2_{\text{крит}}$, то гипотезу отклоняют.

Пример 5. Проведено наблюдение за числом вызовов телефонной станции. С этой целью в течение 100 случайно выбранных 5-секундных интервалов времени регистрировалось число вызовов. Получен следующий вариационный ряд:

Число вызовов x_i	0	1	2	3	4	5
m_i	8	28	31	18	9	6

$$\sum_{i=1}^{6} m_i = n = 100 .$$

Проверить, используя критерий χ^2, гипотезу о том, что распределение числа вызовов согласуется с законом Пуассона. Уровень значимости принять $\alpha = 0,05$.

Вероятность ровно x вызовов в течение n случайно выбранных отрезков времени вычисляется по формуле Пуассона:

$$P_{n,x} = \begin{cases} \dfrac{\lambda^x}{x!} e^{-\lambda}, & \text{если} \quad x = 0, 1, 2, \ldots, n, \\ 0 & \text{в остальных случаях.} \end{cases}$$

1. Найдем точечную оценку параметра λ генеральной совокупности:

$$\overline{X} = \frac{\sum\limits_{i=1}^{6} x_i m_i}{\sum\limits_{i=1}^{6} m_i} = \frac{8 \cdot 0 + 28 \cdot 1 + 31 \cdot 2 + 18 \cdot 3 + 9 \cdot 4 + 6 \cdot 5}{100} = \frac{210}{100} = 2{,}1.$$

Таким образом, функция вероятностей предполагаемого закона Пуассона имеет вид $P_n\left(x\right)=\dfrac{2{,}1^x e^{-2{,}1}}{x!}$.

2. Применим критерий $\chi^2=\sum\limits_{i=1}^{k}\dfrac{\left(m_i-m_i'\right)^2}{m_i'}=\sum\limits_{i=1}^{k}\dfrac{m_i{}^2}{m_i'}-n$.

3. Находим $\chi^2_{\text{набл}}$. Для этого все необходимые вычисления приводим в табл. 7.

Таблица 7

Вычисление $\chi^2_{\text{набл}}$

Число вызовов x_i	m_i	$p_i=\dfrac{(2{,}1)^{x_i}e^{-2{,}1}}{x_i!}$	$n\cdot p_i=m_i'$	$\left(m_i-m_i'\right)^2$	$\dfrac{\left(m_i-m_i'\right)^2}{m_i'}$	m_i^2	$\dfrac{m_i^2}{m_i'}$
0	8	0,122	12,2	17,64	1,45	64	5,25
1	28	0,257	25,7	5,29	0,21	784	30,51
2	31	0,270	27,0	16,00	0,59	961	35,59
3	18	0,189	18,9	0,81	0,04	324	17,14
4	9	0,099	9,9	0,81	0,08	81	8,18
5	6	0,063	6,3	0,03	0,01	36	5,17
$\sum$	100	1,000	100	–	$\chi^2_{\text{набл}}=2{,}38$	–	102,38

Контроль: $\chi^2_{\text{набл}}=2{,}38$; $\sum\limits_{i=1}^{6}\dfrac{m_i{}^2}{m_i'}-n=102{,}38-100=2{,}38$. Вычисления произведены верно.

4. По таблице П5 по заданному уровню значимости и числу степеней свободы $\nu=k-r-1=6-1-1=4$ найдем $\chi^2_{\text{крит}\left(0{,}05;4\right)}=9{,}488$.

5. Так как $2{,}38<9{,}488$, то нет оснований для отклонения гипотезы о том, что закон распределения числа вызовов на телефонной станции является законом Пуассона.

Итак, мы рассмотрели критерий χ^2, при помощи которого проверяли гипотезу о согласии данных выборки с конкретным теоретическим законом распределения для любой случайной величины как непрерывной, так и дискретной.

ПРОВЕРКА ГИПОТЕЗЫ О ЗНАЧЕНИИ ГЕНЕРАЛЬНОЙ СРЕДНЕЙ НОРМАЛЬНО РАСПРЕДЕЛЕННОЙ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПРИ ИЗВЕСТНОЙ ГЕНЕРАЛЬНОЙ ДИСПЕРСИИ

Если из значений нормально распределенной случайной величины вычесть ее среднюю арифметическую и результат разделить на среднее квадратическое отклонение, то получим нормированную случайную величину $Z = \dfrac{X - a}{\sigma}$, $Z \to N\left(0;\, 1^2\right)$.

Переход к стандартной форме случайной величины позволит нам формализовать процедуру проверки гипотез.

Предположим, что верна нулевая гипотеза $H_0 : \bar{X} = a_0$. Преобразуем $Z = \dfrac{\bar{X} - a_0}{\sigma / \sqrt{n}}$, так как $\sigma\left(\bar{X}\right) = \dfrac{\sigma}{\sqrt{n}}$. При большом объеме $Z = \dfrac{\bar{X} - a_0}{\sigma / \sqrt{n}}$, подчиняется стандартному нормальному закону $Z \to \left(0;1^2\right)$.

Критическая область строится в зависимости от вида альтернативных гипотез:

$$H_1 : \bar{X} \neq a_0 ,$$

$$H_1 : \bar{X} = a_1 > a_0 ,$$

$$H_1 : \bar{X} = a_1 < a_0 .$$

Алгоритм проверки гипотезы о значении генеральной средней нормально распределенной генеральной совокупности при известной генеральной дисперсии приведен в табл. 8.

Таблица 8

Нулевая гипотеза	$H_0 : \overline{X} = a_0$
Альтернативная гипотеза	а) $H_1 : \overline{X} \neq a_0$ б) $H_1 : \overline{X} = a_1, a_1 > a_0$ в) $H_1 : \overline{X} = a_1, a_1 < a_0$
Уровень значимости критерия	α (часто $\alpha = 0,05$ или $\alpha = 0,01$)
Критерий	$Z = \dfrac{\overline{X} - a_0}{\sigma/\sqrt{n}}$ (предполагается, что $\sigma_{\text{ген}}$ известно)
Критические точки	Зависят от α . Это: а) границы $\pm Z_{\alpha/2}$, разделяющие критические области принятия H_0 (когда $\alpha = 0,05$, критические точки ± 1,96 когда $\alpha = 0,01$, критические точки $\pm 2,575$). Для других значений α критические точки могут быть получены из таблицы стандартного нормального распределения, т. е. $Z_{\alpha/2} = \Phi_0^{-1}\left(0,5 - \alpha/2\right)$; б) граница Z_α, разделяющая критическую область от области принятия H_0, находится как $Z_\alpha = \Phi_0^{-1}\left(0,5 - \alpha\right)$ (когда $\alpha = 0,05$, критическая точка Z_α = 1,645; когда $\alpha = 0,01$, критическая точка Z_α = 2,325); в) граница $-Z_\alpha$, разделяющая критическую область от области принятия H_0, находится как $-Z_\alpha = -\Phi_0^{-1}\left(0,5 - \alpha\right)$ (когда $\alpha = 0,05$, критическая точка $-Z_\alpha$ = –1,645; когда $\alpha = 0,01$, критическая точка $-Z_\alpha$ = –2,325)

Окончание табл. 8

Правило принятия решения	H_0 отклоняется: а) если $\lvert Z_{\text{набл}} \rvert \geq Z_{\text{кр.};\alpha/2}$ б) если $Z_{\text{набл}} \geq Z^{\text{прав}}_{\text{кр.};\alpha}$ в) если $Z_{\text{набл}} \leq -Z^{\text{прав}}_{\text{кр.};\alpha}$

ПРОВЕРКА ГИПОТЕЗЫ О ЗНАЧЕНИИ ГЕНЕРАЛЬНОЙ СРЕДНЕЙ НОРМАЛЬНО РАСПРЕДЕЛЕННОЙ ГЕНЕРАЛЬНОЙ СОВОКУПНОСТИ ПРИ НЕИЗВЕСТНОЙ ГЕНЕРАЛЬНОЙ ДИСПЕРСИИ

Пусть некоторая генеральная совокупность имеет нормальное распределение $\overline{X} \to N\left(\overline{X};\sigma^2\right)$, а параметры $\overline{X}$ и σ^2 неизвестны. Тогда по результатам случайной выборки объема n найдем их точечные оценки $\overline{X}$ и S.

Требуется проверить гипотезу $H_0 : \overline{X} = a_0$. В качестве критерия проверки нулевой гипотезы принимают случайную величину $T = \dfrac{\overline{X} - a_0}{S} \cdot \sqrt{n-1}$.

Примем без доказательства, что если гипотеза H_0 верна, то случайная величина T имеет распределение Стьюдента с числом степеней свободы $k = n - 1$.

Правило проверки описанной гипотезы H_0 приведем в табл. 9.

Таблица 9

Нулевая гипотеза	$H_0 : \overline{X} = a_0$
Альтернативная гипотеза	а) $H_1 : \overline{X} \neq a_0$ б) $H_1 : \overline{X} = a_1,\, a_1 > a_0$ в) $H_1 : \overline{X} = a_1,\, a_1 < a_0$

Продолжение табл. 9

Уровень значимости критерия	α (часто $\alpha = 0{,}05$ или $\alpha = 0{,}01$)
Критерий	$T = \dfrac{\bar{X} - a}{S/\sqrt{n-1}}$ (предполагается, что $\sigma_{\text{ген}}$ неизвестно) Если объем выборки n достаточно большой, то можно применить критерий $T = \dfrac{\bar{X} - a}{\sigma} \cdot \sqrt{n}$, в котором следует положить $\sigma = S = \sqrt{\dfrac{1}{n}\left(x_i - \bar{X}\right)^2}$
Критические точки	Зависят от α. Это: а) границы $\pm t_{\text{двуст.кр}\left(\alpha;\, k_{\text{св}} = n-1\right)}$, разделяющие критические области от области принятия H_0, определяются по таблице приложения критических точек распределения Стьюдента по заданному уровню значимости α, помещенному в верхней строке таблицы, и числу степеней свободы $k_{\text{св}} = n - 1$; б) граница $-t_{\text{кр}}^{\text{прав}}{}_{\left(\alpha;\, k_{\text{ср}}\right)}$ разделяющая правостороннюю критическую область от области принятия гипотезы H_0, определяется по уровню значимости α, помещенному в нижней строке таблицы критических точек Стьюдента, и числу степеней свободы $k_{\text{св}} = n - 1$; в) граница $-t_{\text{кр}}^{\text{прав}}{}_{\left(\alpha;\, k_{\text{ср}}\right)}$, разделяющая левостороннюю критическую область от области принятия гипотезы H_0, определяется вначале $t_{\text{кр}}^{\text{прав}}{}_{\left(\alpha;\, k_{\text{ср}}\right)}$ и затем полагается $t_{\text{кр}}^{\text{прав}} = -t_{\text{кр}}^{\text{прав}}$

Окончание табл. 9

Правило принятия решения	H_0 отклоняется: а) если $\lvert t_{\text{набл}} \rvert \geq t_{\text{двуст. кр}}$ б) если $t_{\text{набл}} \geq t_{\text{кр}}^{\text{прав}}$ в) если $t_{\text{набл}} \leq -t_{\text{кр}}^{\text{прав}}$

Пример 6. Менеджер кредитного отдела нефтяной компании выясняет, является ли среднемесячный баланс владельцев кредитных карточек, равным 75 у.е. Аудитор случайным образом отобрал 100 счетов и нашел, что среднемесячный баланс владельцев составил 83,4 у.е. с выборочным стандартным отклонением, равным 23,65 у.е. Определить на 5%-м уровне значимости, может ли этот аудитор утверждать, что средний баланс отличен от 75 у.е.

Решение. 1. Исходя из условия задачи, сформулируем гипотезы:

$H_0 : \overline{X} = 75$ у.е.;

$H_1 : \overline{X} \neq 75$ у.е,

$\sigma_{\text{ген}}$ – неизвестно. Уровень значимости $\alpha = 0{,}05$.

2. Для проверки гипотезы H_0 применим критерий

$$T = \frac{\overline{X} - a_0}{\sigma} \cdot \sqrt{n-1}$$

с двусторонней критической областью.

3. Вычислим $t_{\text{набл}} = \dfrac{83{,}4 - 75}{23{,}65}\sqrt{100-1} \approx 3{,}53$.

4. $t_{\text{двуст.кр}(\alpha=0{,}05;\, k_{\text{св}}=99)} \approx 2$.

5. Так как $t_{\text{набл}} \geq t_{\text{двуст.кр}}$, то нулевая гипотеза отклоняется. Значит, среднемесячный баланс владельцев карточек отличен от 75 у.е.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ДВУХ СРЕДНИХ НОРМАЛЬНО РАСПРЕДЕЛЕННЫХ ГЕНЕРАЛЬНЫХ СОВОКУПНОСТЕЙ С ИЗВЕСТНЫМИ ДИСПЕРСИЯМИ

Пусть X_1 и X_2 – нормально распределенные генеральные совокупности с известными дисперсиями σ_1^2 и σ_2^2 и неизвестными генеральными средними $\bar{X}_1$ и $\bar{X}_2$ (математическими ожиданиями), причем $X_1 \to \left(\bar{X}_1 : \sigma_1^2\right)$ и $X_2 \to \left(\bar{X}_2 : \sigma_2^2\right)$. Из генеральных совокупностей извлечены две независимые выборки с объемами n_1 и n_2. В результате получены $\bar{X}_1$, $\bar{X}_2$ и $\sigma_{1\,\text{выб}}^2$, $\sigma_{2\text{выб}}^2$ – средние арифметические и дисперсии выборочных совокупностей. Известно, $X_1 \to \left(\bar{X}_1; \dfrac{\sigma_1^2}{n_1}\right)$ и $X_2 \to \left(\bar{X}_2; \dfrac{\sigma_2^2}{n_2}\right)$.

Требуется на уровне значимости α проверить нулевую гипотезу: $\bar{X}_1$ и $\bar{X}_2$.

Так как выборки независимы, то независимы и выборочные средние $\bar{X}_1$ и $\bar{X}_2$. Поэтому

$$D\left(\bar{X}_1 - \bar{X}_2\right) = D\left(\bar{X}_1\right) + D\left(\bar{X}_2\right) = \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}.$$

Если справедлива нулевая гипотеза H_0,

$$M\left(\bar{X}_1 - \bar{X}_2\right) = M\left(\bar{X}_1\right) - M\left(\bar{X}_2\right) = 0.$$

Тогда случайная величина (статистика) $Z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \to N\left(0;1^2\right)$ и применяется в качестве критерия для проверки нулевой гипотезы $H_0 : \bar{X}_1 = \bar{X}_2$.

Алгоритм построения критических областей запишем в виде табл. 10.

Таблица 10

Нулевая гипотеза	$H_0 : \bar{X}_1 = \bar{X}_2$
Альтернативная гипотеза	а) $H_1 : \bar{X}_1 \neq \bar{X}_2$ б) $H_1 : \bar{X}_1 > \bar{X}_2$ в) $H_1 : \bar{X}_1 < \bar{X}_2$
Уровень значимости критерия	α (часто $\alpha = 0,05$ или $\alpha = 0,01$)
Критерий	$Z = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ (предполагается, что генеральные дисперсии σ_1^2 и σ_2^2 известны)
Критические точки	Зависят от α. Это: а) границы $\pm Z_{\text{кр}(\alpha/2)}$, разделяющие двустороннюю область от области принятия H_0, находятся как $\pm Z_{\text{кр}(\alpha/2)} = \pm\Phi^{-1}_{0\left(0,5-\frac{\alpha}{2}\right)}$ по таблице П2, часть 1, функции Лапласа $\Phi_0(Z) = \dfrac{1}{\sqrt{2\pi}}\int\limits_0^Z e^{-\frac{x^2}{2}}\mathrm{d}x$ (когда $\alpha = 0,05$, критические точки $\pm 1,96$; когда $\alpha = 0,01$, критические точки $\pm 2,575$); б) граница $Z^{\text{прав}}_{\text{кр } \alpha}$, разделяющая правостороннюю критическую область от области принятия гипотезы H_0, находится как $\Phi^{-1}_{0(0,5-\alpha)}$ (когда α = 0,05, критическая точка Z_α = 1,645; когда α = 0,01, критическая точка Z_α = 2,325);

Критические точки	Зависят от α . Это: в) граница $-Z_{\text{кр};\alpha}^{\text{лев}}$, разделяющая левосторооннюю критическую область от области принятия гипотезы H_0 , находится как $-Z_\alpha = -\Phi_{0(0,5-\alpha)}^{-1}$ (когда $\alpha = 0,05$, критическая точка $-Z_\alpha = -1,645$; когда $\alpha = 0,01$, критическая точка $-Z_\alpha = -1,645$)
Правило принятия решения	H_0 отклоняется, если: а) $\lvert Z_{\text{набл}} \rvert \geq Z_{\text{кр}}$ б) $Z_{\text{набл}} \geq Z_{\text{кр}}$ в) $Z_{\text{набл}} \leq -Z_{\text{кр}}$

Пример 7. По двум независимым выборкам, объемы которых $n_1 = 40$ и $n_2 = 50$, извлеченным из нормальных генеральных совокупностей, найдены выборочные средние: $\overline{X}_1 = 130$ и $\overline{X}_2 = 140$, генеральные дисперсии известны: $\sigma_1^2 = 80$, $\sigma_2^2 = 100$. Требуется при уровне значимости $\alpha = 0,01$ проверить нулевую гипотезу $H_0 : \overline{X}_1 = \overline{X}_2$ при конкурирующей гипотезе: $\overline{X}_1 \neq \overline{X}_2$.

Решение. Найдем наблюдаемое значение критерия:

$$Z_{\text{набл}} = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} = \frac{130 - 140}{\sqrt{80/440 + 100/50}} = \frac{-10}{\sqrt{4}} = \frac{-10}{2} = -5 .$$

По условию конкурирующая гипотеза имеет вид $\overline{X}_1 \neq \overline{X}_2$, поэтому критическая область – двусторонняя.

Найдем правую критическую точку из равенства $\Phi\left(Z_{\text{кр}}\right) = \left(1 - \alpha\right) / 2 = \left(1 - 0,01\right) / 2 = 0,495$.

По таблице П2, часть 1, функции Лапласа находим $Z_{кр} = 2{,}58$.

Так как $\left|Z_{набл}\right| \geq Z_{кр}$, то нулевую гипотезу отвергаем, т. е. выборочные средние различаются значимо.

ПРОВЕРКА ГИПОТЕЗЫ О ДИСПЕРСИЯХ ДВУХ НОРМАЛЬНО РАСПРЕДЕЛЕННЫХ СОВОКУПНОСТЕЙ

Рассмотрим две генеральные совокупности X_1 и X_2, распределенные нормально, т.е. $X_1 \to N\left(a_1;\sigma_1\right)$ и $X_2 \to N\left(a_2;\sigma_2\right)$. Из них извлекаются две независимые случайные выборки с объемами n_1 и n_2, находятся исправленные дисперсии $S_1^2 = \dfrac{n_1}{n_1 - 1}\sigma^2_{выб}$, $S_2^2 = \dfrac{n_2}{n_2 - 1}\sigma^2_{выб}$.

Требуется проверить нулевую гипотезу $H_0 : \sigma_1^2 = \sigma_2^2$. В качестве критерия проверки этой гипотезы применяется статистика $F = \dfrac{S_1^2}{S_2^2}$, которая при условии справедливости нулевой гипотезы имеет распределение Фишера-Снедекора со степенями свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$. Предполагается, что $S_1^2 > S_2^2$.

Алгоритм построения критических областей запишем в виде табл. 11.

Таблица 11

Нулевая гипотеза	$H_0 : \sigma_1^2 = \sigma_2^2$
Альтернативная гипотеза	а) $H_1 : \sigma_1^2 > \sigma_2^2$ б) $H_1 : \sigma_1^2 \neq \sigma_2^2$ Замечание: гипотеза $H_1 : \sigma_1^2 < \sigma_2^2$ заменяется гипотезой $H_1 : \sigma_1^2 > \sigma_2^2$

Уровень значимости критерия	α (часто $\alpha = 0{,}05$ или $\alpha = 0{,}01$). Для промежуточных значений уровней значимости критические точки можно получить из таблицы критических точек распределения Фишера-Снедекора (k_1 – число степеней свободы большей дисперсии, k_2 – число степеней свободы меньшей дисперсии), применяя интерполирование
Критерий	$F = \dfrac{S_1^2}{S_2^2}$, где S_1^2 и S_2^2 – исправленные выборочные дисперсии, $S_1^2 > S_2^2$
Критические точки	Зависят от α. Это: а) граница $F^{\text{прав}}_{\text{кр}(\alpha,\, k_1, k_2)}$, разделяющая правостороннюю критическую область от области принятия H_0, где числа степеней свободы $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$, находится по таблице критических точек распределения Фишера-Снедекора; б) граница $F^{\text{прав}}_{\text{кр}\left(\frac{\alpha}{2},\, k_1, k_2\right)}$ находится по таблице критических точек распределения Фишера-Снедекора, а граница $F^{\text{прав}}_{1\,\text{кр}}$ находится как правосторонняя по уровню значимости $\dfrac{\alpha}{2}$
Правило принятия решения	Нулевая гипотеза H_0 отклоняется, если а) $F_{\text{набл}} \geq F_{\text{кр}(\alpha,\, k_1,\, k_2)}$; б) $F_{\text{набл}} \geq F_{\text{кр}\left(\frac{\alpha}{2},\, k_1,\, k_2\right)}$

Пример 8. Менеджер предприятия решил выяснить, существует ли разница в производительности труда рабочих дневной и вечерней смены. Случайно организованная выборка 10 рабочих дневной смены показала, что средний выпуск продукции составил 74,3 ед./ч, а выборочная дисперсия оказалась равной $\sigma_1^2 = 16$ ед./ч. Выборка же 10 рабочих вечерней смены выявила, что средний выпуск продукции равнялся $69,7$ ед./ч, а $\sigma_2^2 = 18$ ед./ч. На 1%-м уровне значимости $(\alpha = 0,01)$ определить, существует ли разница в производительности труда рабочих вечерней и дневной смены.

Решение. Так как выборочные дисперсии различны, проверим предварительно нулевую гипотезу $H_0 : \sigma_1^2 = \sigma_2^2$ о равенстве генеральных дисперсий, пользуясь критерием Фишера-Снедекора $S_1^2 = \frac{10}{10-1} \cdot 18 \approx 20$, $S_2^2 = \frac{10}{10-1} \cdot 16 \approx 17,8$.

Найдем отношение большей исправленной дисперсии к меньшей: $F_{\text{набл}} = \frac{20}{17,8} = 1,12$. Дисперсия 20 больше дисперсии 17,8; дисперсия 18 больше дисперсии 16. Поэтому в качестве альтернативной примем гипотезу $H_1 : \sigma_1^2 > \sigma_2^2$. В этом случае критическая область правосторонняя. По таблице П7 критических точек распределения Фишера-Снедекора по уровню значимости $\alpha = 0,01$ и числам степеней свободы $k_1 = 9$ и $k_2 = 9$ находим $F_{\text{кр}(0,01,\, 9,9)}^{\text{прав}} = 5,35$.

Так как $F_{\text{набл}} < F_{\text{кр}}$ (1,12 < 5,35) – нет оснований отклонить нулевую гипотезу о равенстве генеральных дисперсий, поскольку предположение о равенстве генеральных дисперсий выполняется, сравним средние.

Итак:

1. $H_0 : \overline{X}_1 = \overline{X}_2$.

$H_1 : \overline{X} > \overline{X}_2$ (так как 74,3 > 69,7 – имеем правостороннюю критическую область.

2. В качестве критерия проверки нулевой гипотезы примем случайную величину $T = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}} \cdot \sqrt{\dfrac{n_1 n_2}{n_1+n_2}}$.

3. $t_{\text{набл}} = \dfrac{74{,}3 - 69{,}7}{\sqrt{\dfrac{9 \cdot 17{,}8 + 9 \cdot 20}{10+10-2}}} \cdot \sqrt{\dfrac{10 \cdot 10}{10+10}} \approx 2{,}366$.

4. Находим $t_{\text{кр}}$ по таблице П6 критических точек распределения Стьюдента $t_{\text{кр}(0{,}01;\ k_{\text{св}}=18)} = 2{,}55$.

Так как $t_{\text{набл}} < t_{\text{кр}}$, то нет оснований отклонить гипотезу H_0, т. е. не существует разницы в производительности труда рабочих дневной и вечерней смены, а имеющие место различия случайны, незначимы.

ПРОВЕРКА ГИПОТЕЗЫ О РАВЕНСТВЕ ГЕНЕРАЛЬНЫХ СРЕДНИХ ДВУХ НОРМАЛЬНО РАСПРЕДЕЛЕННЫХ СОВОКУПНОСТЕЙ ПРИ НЕИЗВЕСТНЫХ РАВНЫХ ГЕНЕРАЛЬНЫХ ДИСПЕРСИЯХ (МАЛЫЕ НЕЗАВИСИМЫЕ ВЫБОРКИ)

Имеются две генеральные совокупности $X_1 \to N\left(\overline{X}_1; \sigma_1^2\right)$ и $X_2 \to N\left(\overline{X}_2; \sigma_2^2\right)$, из которых известны выборки с объемами n_1 и n_2 с параметрами $\overline{X}_1$, σ_1^2 и $\overline{X}_2$, σ_2^2. Пусть генеральные дисперсии σ_1^2 и σ_2^2 неизвестны, например, по той причине, что нельзя получить хорошие оценки генеральных дисперсий по выборкам малого объема. Но если предположить, что генеральные дисперсии равны, то примем без доказательства, что можно построить критерий проверки нулевой гипотезы, представляющий собой статистику $T = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}} \cdot \sqrt{\dfrac{n_1 n_2}{n_1+n_2}}$, которая при справедливости нулевой гипотезы имеет *t*-распределение Стьюдента с числом степеней свободы $k_{\text{св}} = n_1 + n_2 - 2$.

Критическая область строится в зависимости от вида альтернативной гипотезы.

Правила построения критических областей и принятия решений приведены в табл. 12.

Таблица 12

Нулевая гипотеза	$H_0 : \overline{X}_1 = \overline{X}_2$
Альтернативная гипотеза	а) $H_1 : \overline{X}_1 \neq \overline{X}_2$ б) $H_1 : \overline{X} > \overline{X}_2$ в) $H_1 : \overline{X} < \overline{X}_2$
Уровень значимости критерия	α (часто $\alpha = 0{,}05$ или $\alpha = 0{,}01$)
Критерий	$T = \dfrac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}} \cdot \sqrt{\dfrac{n_1 n_2}{n_1 + n_2}}$ (предполагается, что генеральные дисперсии σ_1^2 и σ_2^2 неизвестны, но равны)
Критические точки	Зависят от α . Это: а) границы $\pm t_{\text{двуст. кр}\,(\alpha;\, k_{\text{св}} = n_1 + n_2 - 2)}$, разделяющие двустороннюю область от области принятия H_0, находятся по таблице точек распределения Стьюдента по уровню значимости α, помещенному в верхней строке таблицы и числу степеней свободы $k_{\text{св}} = n_1 + n_2 - 2$;

Окончание табл. 12

Критические точки	б) граница $t^{\text{прав}}_{\text{кр}\,(\alpha;\, k_{\text{св}}=n_1+n_2-2)}$, разделяющая правостороннюю критическую область от области принятия гипотезы H_0, определяется по уровню α, помещенному в нижней строке таблицы критических точек Стьюдента и числу степеней свободы $k_{\text{св}} = n_1 + n_2 - 2$; в) граница $t^{\text{лев}}_{\text{кр}(\alpha;\, k_{\text{св}}=n_1+n_2-2)}$, разделяющая левостороннюю критическую область от области принятия гипотезы H_0, определяется вначале $t^{\text{прав}}_{\text{кр}(\alpha;\, k_{\text{св}}=n_1+n_2-2)}$ и затем полагается $t^{\text{лев}}_{\text{кр}} = -t^{\text{прав}}_{\text{кр}}$
Правило принятия решения	H_0 отклоняется, если: а) $\lvert t_{\text{набл}} \rvert \geq t_{\text{двуст. кр}}$; б) $t_{\text{набл}} \geq t^{\text{прав}}_{\text{кр}}$; в) $t_{\text{набл}} \leq -t^{\text{лев}}_{\text{кр}}$

Пример 9. Из двух партий изделий, изготовленных на двух одинаково настроенных станках, извлечены малые выборки, объемы которых $n_1 = 10$ и $n_2 = 12$. Получены следующие результаты: $\bar{X}_1 = 3{,}6$, $\bar{X}_2 = 3{,}5$, $S^2_{X_1} = 0{,}0267$, $S^2_{X_2} = 0{,}0255$. Требуется при уровне значимости $\alpha = 0{,}01$ проверить гипотезу $H_0 : \bar{X}_1 = \bar{X}_2$ при альтернативной гипотезе $H_0 : \bar{X}_1 \neq \bar{X}_2$. Предполагается, что случайные величины X_1 и X_2 распределены нормально.

Решение. Рассматриваемый в этом параграфе критерий предполагает, что генеральные дисперсии одинаковы, но исправленные дисперсии различны, поэтому вначале нужно сравнить дисперсии, используя критерий Фишера-Снедекора. Сделаем это, приняв в качестве альтернативной гипотезы $H_1 : D(X_1) \neq D(X_2)$. Найдем наблюдаемое значение критерия: $F_{\text{набл}} = \dfrac{S_{X_1}^2}{S_{X_2}^2} = \dfrac{0,0267}{0,0255} = 1,05$. По таблице П7 критических точек распределения Фишера-Снедекора находим $F_{\text{кр}(0,01,\ 9,11)} = 4,63$. Так как $F_{\text{набл}} < F_{\text{кр}}$ – дисперсии различаются незначимо и, следовательно, можно считать, что допущение о равенстве генеральных дисперсий выполняется.

Сравним средние, для чего вычислим наблюдаемое значение критерия Стьюдента:

$$T_{\text{набл}} = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\dfrac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}}} \cdot \sqrt{\frac{n_1 n_2}{n_1 + n_2}} = 1,45 .$$

По условию, альтернативная гипотеза имеет вид $\bar{X}_1 \neq \bar{X}_2$, поэтому критическая область двусторонняя. По уровню значимости $\alpha = 0,01$ и числу степеней свободы $k_{\text{св}} = n_1 + n_2 - 2 = 10 + 12 - 2 = 20$ находим по таблице критических точек распределения Стьюдента критическую точку $t_{\text{двуст. кр}(0,01;\ 20)} = 2,53$.

Так как $t_{\text{набл}} < t_{\text{двуст.кр}}$ – нет оснований отвергнуть гипотезу о равенстве средних.

Таким образом, средние размеры изделий существенно не различаются.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

1. Используя критерий Пирсона, при уровне значимости 0,05 проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема $n = 200$:

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

2. Используя критерий Пирсона, при уровне значимости 0,01 установить, случайно или значимо расхождение между эмпирическими частотами m_i и теоретическими частотами m_i', которые вычислены, исходя из гипотезы о нормальном распределении генеральной совокупности X :

m_i	8	16	40	72	36	18	10
m_i'	6	18	36	76	39	18	7

3. В результате испытания 200 элементов на длительность работы получено эмпирическое распределение (в первом столбце указаны интервалы времени в часах, во втором – частота, т. е. количество элементов, проработавших время в пределах соответствующего интервала)

$x_i - x_{i+1}$	m_i	$x_i - x_{i+1}$	m_i
0–5	133	15–20	4
5–10	45	20–25	2
10–15	15	25–30	1

Требуется, при уровне значимости 0,05, проверить гипотезу о том, что время работы элементов распределено по показательному закону.

4. В итоге испытаний 1000 элементов на время безотказной работы получено эмпирическое распределение (в первом столбце указаны интервалы времени в часах; во втором – частота m_i , т. е. количество отказавших элементов в i -м интервале).

$x_i - x_{i+1}$	m_i	$x_i - x_{i+1}$	m_i
0–10	365	40–50	70
10–20	245	50–60	45
20–30	150	60–70	25
30–40	100		$n = 1000$

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

5. В течение 10 часов регистрировали прибытие автомашин к бензоколонке и получили эмпирическое распределение (в первом столбце указан интервал времени в часах, во втором – частота, т. е. количество машин, прибывших в этом интервале). Всего было зарегистрировано 200 машин.

$x_{i-1} - x_i$	m_i	$x_{i-1} - x_i$	m_i
8–9	12	13–14	6
9–10	40	14–15	11
10–11	22	15–16	33
11–12	16	16–17	18
12–13	28	7–18	14

Требуется при уровне значимости 0,01 проверить гипотезу о том, что время прибытия машин распределено равномерно.

6. В результате взвешивания 800 стальных шариков получено эмпирическое распределение (в первом столбце указан интервал веса в граммах, во втором столбце – частота, т. е. количество шариков, вес которых принадлежит этому интервалу).

$x_{i-1} - x_i$	m_i	$x_{i-1} - x_i$	m_i
20,0–20,5	91	23,0–23,5	79
20,5–21,0	76	23,5–24,0	73
21,0–21,5	75	24,0–24,5	80
21,5–22,0	74	24,5–25,0	77
22,0–23,0	92		$n = 800$
22,5–23,0	83		

Требуется при уровне значимости 0,01 проверить гипотезу о том, что вес шариков X распределен равномерно.

7. Отдел технического контроля проверил $n = 100$ партий изделий по $N = 10$ изделий в каждой партии и получил следующее эмпирическое распределение дискретной случайной величины X – числа не-

стандартных изделий (в первой строке указано число x_i нестандартных изделий в одной партии; во второй строке – частота m_i, т. е. количество партий, содержащих x_i нестандартных изделий):

x_i	0	1	2	3	4	5	6	7
m_i	2	3	10	22	26	20	12	5

Требуется при уровне значимости 0,01 проверить гипотезу о том, что случайная величина X распределена по биноминальному закону.

8. В библиотеке случайно отобрано 200 выборок по 5 книг. Регистрировалось число поврежденных книг (подчеркивания, помарки и т. д.). В итоге получено следующее эмпирическое распределение (в первой строке указано число x_i поврежденных книг в одной выборке; во второй – частота m_i, т. е. количество выборок, содержащих x_i поврежденных книг):

x_i	0	1	2	3	4	5
m_i	72	77	34	14	2	1

Требуется при уровне значимости 0,05 проверить гипотезу о том, что дискретная случайная величина X (число поврежденных книг) распределена по биномиальному закону.

9. Отдел технического контроля проверил n = 200 партий одинаковых изделий и получил следующее эмпирическое распределение (в первой строке указано количество x_i нестандартных изделий в одной партии; во второй строке – частота m_i, т. е., количество партий, содержащих x_i нестандартных изделий):

x_i	0	1	2	3	4
m_i	116	56	22	4	2

Требуется при уровне значимости 0,05 проверить гипотезу о том, что количество нестандартных изделий X распределено по закону Пуассона.

10. В результате проверки 500 контейнеров со стеклянными изделиями установлено, что число поврежденных изделий X имеет следующее эмпирическое распределение (в первой строке указано количество x_i поврежденных изделий в одном контейнере, во второй строке – частота m_i, т.е. число контейнеров, содержащих x_i поврежденных изделий):

x_i	0	1	2	3	4	5	6	7
m_i	199	169	87	31	9	3	1	1

Требуется при уровне значимости 0,01 проверить гипотезу о том, что случайная величина X – число поврежденных изделий – распределена по закону Пуассона.

Указание. Объединить частоты трех последних групп.

11. Дисперсия генеральной совокупности равна 100. Выборка 25 единиц из этой совокупности дала среднюю арифметическую, равную 17. Можно ли отклонить $H_0 : a = 21$ при конкурирующей гипотезе $H : a \neq 21$? Принять $\alpha = 0,05$.

12. Производители нового вида аспирина утверждают, что он снимает головную боль за 30 мин. Случайная выборка 121 человека, страдающего головными болями, показала, что новый тип аспирина снимает головную боль за 28,6 минуты при среднем квадратическом отклонении 4,2 минуты. Проверить на уровне значимости $\alpha = 0,05$ справедливость утверждения производителей аспирина о том, что это лекарство излечивает головную боль за 30 минут.

13. Компания, производящая средства для потери веса, утверждает, что прием таблеток в сочетании со специальной диетой позволяет сбросить в среднем в неделю 400 граммов веса. Случайным образом отобраны 25 человек, использующих эту терапию, и обнаружено, что в среднем они сбросили 430 граммов при исправленном среднем квадратическим отклонении 110 граммов. Проверить гипотезу о том, что средняя потеря в весе составляет 400 граммов. Уровень значимости $\alpha = 0,05$.

14. Из нормальной генеральной совокупности с известным средним квадратическим отклонением $\sigma = 30$ извлечена выборка объема $n = 49$ и по ней найдена выборочная средняя $\overline{X} = 126{,}5$. Требуется, при уровне значимости $\alpha = 0{,}05$, проверить нулевую гипотезу $H_0 : a = a_0 = 120$ при конкурирующей гипотезе $H_0 : a \neq a_0$.

15. Инженер по контролю качества проверяет среднее время горения нового вида электролампы. Для проверки в порядке случайной выборки было отобрано 100 ламп, среднее время горения которых оказалось 1075 часов. Предположим, что среднее квадратическое отклонение времени горения для генеральной совокупности известно и равно 100 часам. Используя уровень значимости $\alpha = 0{,}05$, проверить гипотезу о том, что время горения ламп более 1000 часов.

16. Производители стирального порошка утверждают, что средний вес коробки с порошком составляет 325 граммов. Случайная выборка 25 коробок с порошком обнаружила, что средний вес коробки составляет 323,8 грамм, а среднее квадратическое отклонение равно 11,7 граммов. При уровне значимости $\alpha = 0{,}01$ определить отличается ли средний вес коробок от 325 грамм.

17. Имеются данные о результатах проверки качества деталей:

Партия		Основные характеристики	
		средняя прочность, кг/см^2	σ^2
До изменения технологии изготовления	100	40	250
После изменения технологии изготовления	100	44	150

Определить, является ли повышение прочности деталей с 40 до 44 кг/см^2 существенным настолько, что его можно считать следствием изменения технологии, или же это результат случайной колеблемости показателей, и поэтому изменение технологии нельзя считать эффективным? Принять уровень значимости $\alpha = 0{,}05$.

18. Ежедневная заработная плата в определенной отрасли есть случайная величина, распределенная по нормальному закону, со

средней $\bar{X} = 13,2$ у.е. и $\sigma = 2,5$ у.е. Если компания в этой отрасли нанимает 40 рабочих и платит им в среднем 12,2 у.е., можно ли эту компанию обвинить в том, что она платит слишком низкую зарплату? Уровень значимости принять равным $\alpha = 0,05$.

19. Инвестиционный фонд объявил, что средний годовой доход по акциям предприятий некоторой отрасли промышленности составил 11,5 %. Инвестор, желая проверить, является ли это заявление правильным, произвел случайную выборку из 41 акции этой отрасли. Средний годовой доход по ним составил 10,8 % и выборочное среднее квадратическое отклонение 3,4 %. Имеет ли инвестор достаточную информацию для того, чтобы опровергнуть заявление инвестиционного фонда? Принять уровень значимости $\alpha = 0,05$.

20. Эксперт страховой компании полагает, что в течение последних нескольких лет средний размер страховых обязательств по импортным автомобилям был равен 1000 у.е. Для проверки этого предположения была организована выборка 25 владельцев автомобилей. Проведенное обследование показало, что средний размер страхового обязательства равен 1050 у.е. с исправленным стандартным отклонением $S = 47$ у.е. Могут ли результаты выборки опровергнуть утверждение эксперта страховой компании, что средний размер страхового обязательства равен 1000 у.е.? Уровень значимости принять равным $\alpha = 0,01$.

21. Для проверки эффективности нового лекарства были отобраны две случайные группы по 15 человек, болеющих гриппом. При применении старого лекарства средний срок выздоровления составлял 11 дней с выборочной дисперсной $S_1^2 = 3$, при применении нового – срок выздоровления составил 8 дней с $S_2^2 = 4$. Проверить на уровне значимости $\alpha = 0,01$ гипотезу о преимуществе нового лекарства.

22. Месячная заработная плата для выборки из 50 рабочих определенной фирмы составляет 110 при среднем квадратическом отклонении 1,44 у.е., а заработная плата для выборки из 40 рабочих другой фирмы равна 105 у.е. при среднем квадратическом отклонении 1,50 у.е. Выше ли зарплата в первой фирме, чем во второй? Уровень значимости принять $\alpha = 0,01$.

ОТВЕТЫ

1. $\chi^2_{\text{набл}} = 22,2;\quad \chi^2_{\text{кр}}(0,05;6) = 12,6$.

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

2. $\chi^2_{\text{набл}} = 3,061;\quad \chi^2_{\text{кр}}(0,01;4) = 13,3$.

$\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$ – гипотезу о нормальном распределении генеральной совокупности не отвергаем.

3. $\chi^2_{\text{набл}} = 1,29;\quad \chi^2_{\text{кр}}(0,05;2) = 6,0.$

$\chi^2_{\text{набл}} < \chi^2_{\text{кр}}$ – гипотеза о распределении X по показательному закону не отвергается.

4. $\chi^2_{\text{набл}} = 15,88;\quad \chi^2_{\text{кр}}(0,01;5) = 15,1.$

$\chi^2_{\text{набл}} > \chi^2_{\text{кр}}$ – гипотеза о показательном распределении времени безотказной работы элементов отвергается.

5. $\bar{x}_{\text{в}} = 12,71;\quad \sigma_{\text{в}} = 2,86;\quad a* = 7,76;\quad b* = 17,66;\quad f(x) = 0,101$, $k = 7;\quad \chi^2_{\text{набл}} = 53,43;\quad \chi^2_{\text{кр}} = 18,5.$ Гипотеза о равномерном распределении времени отвергается. Данные наблюдений не согласуются с этой гипотезой.

6. $\overline{X}_{\text{в}} = 22,47;\quad \sigma_{\text{в}} = 1,44;\quad a* = 19,98;\quad b* = 24,96;\quad f(x) = 0,2$; $k = 7;\quad \chi^2_{\text{набл}} = 4,39;\quad \chi^2_{\text{кр}}(0,01;7) = 18,5.$ Нет оснований отвергнуть гипотезу о равномерном распределении X.

7. $p* = 0,4;\quad k = 5;$ теоретические частоты: 0,60; 4,04; 12,24; 25,05; 20,05; 11,15; 4,25; $\chi^2_{\text{набл}} = 0,68;\quad \chi^2_{\text{кр}}(0,01;5) = 15,1.$ Нет оснований отвергнуть гипотезу о распределении X по биномиальному закону.

8. $p* = 0,2;\quad k = 2;$ теоретические частоты: 65,54; 81,92; 40,96; 10,24; 1,28; 0,06; $\chi^2_{\text{набл}} = 4,65;\quad \chi^2_{\text{кр}}(0,05;2) = 6,0.$ Нет оснований отвергнуть гипотезу о распределении X по биномиальному закону.

9. $\overline{X}_в = 0{,}6$; $\lambda = 0{,}6$; теоретические частоты: 65,86; 19,76; 3,96; 0,60; $\chi^2_{набл} = 2{,}54$; $\chi^2_{кр}\left(0{,}05;2\right) = 6{,}0$. Так как $\chi^2_{набл} < \chi^2_{кр}$ – нет оснований отвергнуть гипотезу о распределении случайной величины X по закону Пуассона.

10. $k = 4$; $\lambda = \overline{X}_в = 1$; теоретические частоты: 183,95; 183,95; 92,00; 30,65; 7,65; 1,55; 0,25; 0,05; $\chi^2_{набл} = 8{,}32$; $\chi^2_{кр}\left(0{,}01;4\right) = 13{,}3$. Нет оснований отвергнуть гипотезу о распределении X по закону Пуассона.

11. $H_0 : \overline{X} = 21$ отклоняется в пользу $H_1 : \overline{X} < 21$.

12. $H_0 : \overline{X} = 30$ отклоняется в пользу $H_1 : \overline{X} < 30$.

13. Нет оснований отклонить $H_0 : \overline{X} = 400$ в пользу конкурирующей гипотезы $H_1 : \overline{X} > 400$.

14. Нет оснований отклонить H_0.

15. $H_0 : \overline{X} = 1000$ отклоняется в пользу $H_1 : \overline{X} > 1000$, т. е. имеющиеся различия не случайны, значимы.

16. Нет оснований отклонить гипотезу $H_0 : \overline{X} = 300$ в пользу конкурирующей гипотезы $H_1 : \overline{X} \neq 300$.

17. Гипотеза $\overline{H}_0 : \overline{X}_1 = \overline{X}_2$ отклоняется в пользу $H_1 : \overline{X}_1 < \overline{X}_2$, т. е. изменение технологии эффективно.

18. Да. $H_0 : \overline{X} = 13{,}2$ у.е. отклоняется в пользу $\overline{H}_1 : \overline{X} < 13{,}2$ (т. к. 12,2 < 13,2).

19. Не имеет. Нет оснований отклонить гипотезу $H_0 : \overline{X} = 11{,}5$ в пользу конкурирующей гипотезы $\overline{H}_1 : \overline{X} < 11{,}5$.

20. Гипотеза $H_0 : \overline{X} = 1000$ у.е. отклоняется в пользу альтернативной гипотезы $H_0 : \overline{X} = 1050$ у.е.

21. Гипотеза $H_0 : M\left(X\right) > M\left(Y\right)$ принимается, т. е. новое лекарство более эффективно старого.

22. Гипотеза $H_0 : \overline{X}_1 = \overline{X}_2$ не отвергается в пользу альтернативной гипотезы $H_0 : \overline{X}_1 > \overline{X}_2$, т. е. зарплата в двух фирмах одинаковая на уровне значимости $\alpha = 0{,}01$.

КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

Обозначим через X независимую переменную, а через Y зависимую от нее переменную. Зависимость Y от X называется ***функциональной***, если каждому значению величины X соответствует единственное значение величины Y. Однако гораздо чаще в окружающем нас мире каждому значению переменной X соответствует не одно, а множество значений переменной Y. Это объясняется тем, что на результирующую переменную действует не только контролируемый фактор X, но и множество других неконтролируемых случайных факторов.

Если каждому значению переменной X соответствует определенное (условное) распределение переменной Y, то такая зависимость называется **стохастической** или **вероятностной**.

Корреляционной зависимостью между двумя переменными называется функциональная зависимость между значениями одной из них и условным математическим ожиданием другой. Корреляционная зависимость может быть представлена в виде:

$$\varphi(x) = M_x(Y), \tag{5}$$

$$\psi(y) = M_y(X). \tag{6}$$

Уравнения (5) и (6) называются ***уравнениями регрессии*** *соответственно Y на X и X на Y, а их графики –* ***линиями регрессии****. Если $\varphi(x) = \text{const}$, $\psi(y) = \text{const}$, то говорят, что корреляционная связь отсутствует.*

Переменная X в корреляционном анализе может быть как детерминированной, так и случайной.

Основной задачей корреляционного анализа является выявление тесноты связи между переменными X и Y и количественная оценка тесноты этой связи.

Корреляционный анализ следует применять только в том случае, когда данные наблюдений или эксперимента, можно считать случайными и выбранными из нормальной совокупности.

В корреляционном анализе экспериментальные данные задаются корреляционной таблицей. Вариант такого задания рассмотрим на следующем примере.

Для исследования зависимости годового объема производства Y от основных фондов X получены статистические данные по 20 предприятиям (табл. 13).

Таблица 13

y_j \ x_i		12,5	17,5	22,5	27,5	n_j
20–21	20,5	1	–	–	–	1
21–22	21,5	–	2	–	–	2
22–23	22,5	–	1	2	–	3
23–24	23,5	–	–	3	3	6
24–25	24,5	–	–	–	8	8
n_i		1	3	5	11	$n=20$

В первой строке таблицы записаны значения переменной X, в первом столбце – интервалы изменения переменной Y, а во втором – середины этих интервалов. Для дальнейших расчетов необходимы только середины интервалов, поэтому сами интервалы могут отсутствовать. Центральную часть таблицы занимают частоты n_{ij} (число предприятий), соответствующие значениям переменных $X = x_i$ и $Y = y_j$. В последней строке записаны частоты $n_i = \sum_{j=1}^{m} n_{ij}$, а в последнем столбце – частоты $n_j = \sum_{i=1}^{l} n_{ij}$.

Здесь $l = 4$ – число средних значений величины X, $m = 5$ – число средних значений величины Y, $n = \sum n_i = \sum n_j = 20$ – число всех значений. Таблица такого вида называется **корреляционной таблицей**.

Основной оценкой для тесноты связи между переменными X и Y служит **выборочный коэффициент** r, который определяется формулой

$$r = \frac{\overline{XY} - \overline{X} \cdot \overline{Y}}{S_x S_y}. \qquad (7)$$

Свойства выборочного коэффициента корреляции аналогичны свойствам коэффициента корреляции между случайными величинами X и Y.

Перечислим свойства коэффициента корреляции:

1. Коэффициент корреляции принимает значение на отрезке $\left[-1,\ 1\right]$, т. е.

$$-1 \le r \le 1.$$

В зависимости от того, насколько $|r|$ приближается к 1, различают слабую, умеренную и сильную связь, т. е. чем ближе $|r|$ к 1, тем теснее связь.

2. Если переменные X и Y умножить на одно и тоже число, то коэффициент корреляции не изменится.

3. Если $r = \pm 1$, то корреляционная связь между X и Y представляет собой линейную зависимость. Запишем более подробные формулы для вычисления коэффициента корреляции:

$$r = \frac{n\sum\limits_{i=1}^{l}\sum\limits_{j=1}^{m} x_i y_j n_{ij} - \left(\sum\limits_{i=1}^{l} x_i n_i\right)\left(\sum\limits_{j=1}^{m} y_j n_j\right)}{\sqrt{n\sum\limits_{i=1}^{l} x_i^2 n_i - \left(\sum\limits_{i=1}^{l} x_i n_i\right)^2} \cdot \sqrt{n\sum\limits_{j=1}^{m} y_j^2 n_j - \left(\sum\limits_{j=1}^{m} y_j n_j\right)^2}} \qquad (8)$$

$$r = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\left(x_i - \overline{X}\right)\left(y_j - \overline{Y}\right)n_{ij}}{n\,S_x S_y} \tag{9}$$

Если данные не сгруппированы, то формулы (8), (9) упрощаются:

$$r = \frac{n\sum_{i=1}^{n} x_i y_i - \left(\sum_{i=1}^{n} x_i\right)\cdot\left(\sum_{j=1}^{n} y_j\right)}{\sqrt{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}\cdot\sqrt{n\sum_{j=1}^{n} y_j^2 - \left(\sum_{j=1}^{n} y_j\right)^2}} \tag{10}$$

$$r = \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)\left(y_i - \overline{Y}\right)}{n\,S_x S_y} \tag{11}$$

Так как r вычисляется по данным выборки, то в отличие от генерального коэффициента корреляции ρ, r является величиной случайной. Если $r \neq 0$, то возникает вопрос, объяснятся ли это действительной линейной связью между X и Y или вызвано случайными факторами. Для выяснения значимости коэффициента корреляции проверяется гипотеза H_0 об отсутствии линейной корреляционной связи между переменными X и Y, т. е. $H_0 : \rho = 0$.

При справедливости этой гипотезы статистика

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \tag{12}$$

имеет распределение Стьюдента с $k = n - 2$ степенями свободы. Поэтому гипотеза H_0 отвергается, если $|t| > t_{\text{к, кр}}$, где $t_{\text{к, кр}}$ – табличное значение распределения Стьюдента, определенное на уровне значимости α при числе степеней свободы $k = n - 2$.

Пример 10. Для данных табл. 13 найти выборочный коэффициент корреляции, проверить его значимость на уровне $\alpha = 0{,}05$.

Решение. Для вычислений составим таблицу. Находим суммы $\sum_i x_i n_i = 480$; $\sum_j y_j n_j = 468$; $\sum_i x_i^2 n_i = 11925$; $\sum_j y_j^2 n_j = 10979$ и заносим их в таблицу. Вычислим

$$\sum_i \sum_j x_i y_j n_{ij} = 20{,}5 \cdot 12{,}5 + 21{,}5 \cdot 17{,}5 \cdot 2 + 22{,}5 \cdot 17{,}56 +$$

$$+22{,}5 \cdot 22{,}5 \cdot 2 + 23{,}5 \cdot 22{,}5 \cdot 3 +$$

$$+23{,}5 \cdot 27{,}5 \cdot 3 + 24{,}5 \cdot 27{,}5 \cdot 8 = 11330.$$

j,i	x_i	n_i	y_j	n_j	$x_i n_i$	$y_j n_j$	x_i^2	$x_i^2 n_i$	y_j^2	$y_j^2 n_j$
1	12,5	1	20,5	1	12,5	20,5	156,25	156,25	420,25	420,25
2	17,5	3	21,5	2	52,5	43,0	306,25	918,75	462,25	924,5
3	22,5	5	22,5	3	112,5	67,5	506,25	2531,25	506,25	1518,75
4	27,5	11	23,5	6	302,5	141,0	756,25	8318,75	552,25	3313,5
5			24,5	8		196,0			600,25	4802
$\sum$					480	468		11925		10979

Подставляя полученные значения сумм в (8), найдем выборочный коэффициент корреляции

$$r = \frac{20 \cdot 11330 - 480 \cdot 468}{\sqrt{20 \cdot 11925 - 480^2} \cdot \sqrt{20 \cdot 10979 - 468^2}} = \frac{1960}{90 \cdot 23{,}58} = 0{,}934\,.$$

Проверим значимость r на уровне $\alpha = 0{,}05$. Для этого вычислим статистику

$$t = \frac{r(n-2)}{\sqrt{1-r^2}} = \frac{0{,}924\sqrt{18}}{\sqrt{1-0{,}924^2}} = 10{,}25.$$

По таблице распределения П6 Стьюдента $k = n - 2 = 18$ находим критическое значение $t_{кр} = 1{,}734$ Так как $t > t_{кр}$, то считаем r значимым.

КОРРЕЛЯЦИОННОЕ ОТНОШЕНИЕ

Введенный ранее коэффициент корреляции является показателем тесноты связи лишь в случае линейной зависимости. Для получения показателя тесноты связи при зависимости любого вида запишем правило сложения дисперсий в форме $S_y^2 = S_{фy}^2 + S_0^2$,

где $S_y^2 = \frac{1}{n}\sum_{j=1}^{m} y_j^2 n_j - \left(\frac{1}{n}\sum_{j=1}^{m} y_j n_j\right)^2$ – общая дисперсия переменной *Y*,

$S_{фy}^2 = \frac{\sum_{i=1}^{l}\left(\overline{Y}_i - \overline{Y}^2\right)n_i}{n}$ – межгрупповая дисперсия, $S_0^2 = \frac{\sum_{i=1}^{l}\sum_{j=1}^{m}\left(y_j - \overline{Y}_i\right)^2 n_i}{n}$ – остаточная дисперсия.

Здесь $\overline{Y}_i = \frac{1}{n_i}\sum_{j=1}^{m} y_j n_{ij}$ – групповые средние.

Остаточная дисперсия представляет собой ту часть дисперсии *Y*, которая обусловлена неучтенными случайными факторами. Межгрупповая дисперсия выражает ту часть дисперсии *Y*, которая вызвана изменчивостью *X*.

Величина $\eta_{yx} = \frac{S_{фy}}{S_y}$ называется **эмпирическим корреляционным отношением.** Чем теснее связь, тем большее влияние на вариацию *Y* оказывает изменчивость *X*, тем больше η_{yx}. Величина $\overline{d} = \eta_{yx}^2$ является эмпирическим коэффициентом детерминации и показывает, какая часть общей вариации *Y* обусловлена вариацией *X*. Чем ближе $\overline{d}$ к 1, тем теснее наблюдения примыкают к линии регрессии, тем лучше регрессия описывает зависимость переменных. Для линейной модели $\overline{d} = r^2$.

Пример 11. Для данных табл. 13 найти корреляционное отношение η_{yx}.

Для вычисления эмпирического корреляционного отношения найдем групповые средние $\overline{Y}_i$ и $\overline{Y}$:

$$\overline{Y_1} = 20,5\ ;\ \overline{Y_2} = \frac{1}{3}\left(21,5 \cdot 2 + 22,5\right) = 21,83;$$

$$\overline{Y_3} = \frac{1}{5}\left(22,5 \cdot 2 + 23,5 \cdot 3\right) = 23,1;\ \ \overline{Y_4} = \frac{1}{11}\left(23,5 \cdot 3 + 25,5 \cdot 8\right) = 24,23;$$

$$\overline{Y_i} = \frac{1}{n}\sum_{i=1}^{4} Y_j n_i = \frac{1}{20}\left(20,5 + 21,83 \cdot 3 + 23,1 \cdot 5 + 24,23 \cdot 11\right) = 23,4.$$

Тогда

$$S_{\text{ф}y}^2 = \frac{1}{n}\sum_{i=1}^{4}\left(\overline{Y_i} - \overline{Y}\right)^2 n_i = \frac{1}{20}\Big[\left(20,5 - 23,4\right)^2 + \left(21,83 - 23,4\right)^2 \cdot 3 +$$

$$+\left(23,1 - 23,4\right)^2 \cdot 5 + \left(24,33 - 23,4\right)^2 \cdot 11\Big] = 1,29;$$

$$S_y^2 = \frac{1}{n}\sum_{j=1} y_j^2 n_j - \left(\frac{1}{n}\sum_{j=1} y_j^2 n_j\right)^2 = \frac{1}{20}10979 - \left(\frac{1}{20} - 468\right)^2 = 1,39.$$

Вычисляем корреляционное отношение

$$\eta_{yx} = \frac{S_{\text{ф}y}}{S_y} = \sqrt{\frac{1,29}{1,39}} = 0,963\ .$$

РЕГРЕССИОННЫЙ АНАЛИЗ

При изучении стохастических зависимостей в задачах техники, экономики и т. д. одним из главных моментов является установление вида зависимости Y от $X_1, X_2, \ldots, X_n$, т. е. вида уравнения регрессии. Это связано в первую очередь с необходимостью прогнозирования исследуемых процессов.

Установление формы зависимости, оценка функции регрессии и ее параметров являются задачами регрессионного анализа.

В регрессионном анализе изучаются модели вида

$$Y = \varphi\left(X\right) + \varepsilon, \tag{13}$$

где X – неслучайная независимая переменная, называемая фактором; Y – случайная зависимая переменная (результирующий признак); ε – случайная переменная, характеризующая отклонение от линии регрессии (остаточная переменная). Основными предпосылками регрессионного анализа являются:

1) для каждого из наблюдений с номером i , $i = \overline{1;n}$, $\varepsilon_i \in N\left(a, \sigma\right)$, где

$$\sigma^2 = DY_i = D\varepsilon_i = \mathrm{const}\,; \tag{14}$$

2) для разных наблюдений переменные ε_i и ε_j некоррелированы, $M\left(\varepsilon_i\,\varepsilon_j\right) = 0$, $i \neq j$. Оценкой функции регрессии $\varphi\left(x\right) = M_x\left(Y\right)$ является функция

$$Y_x = \varphi\left(x, b_0, b_1, \ldots, b_p\right), \tag{15}$$

где x – значения величины X; $Y_x = M_x Y;\, b_0 b_1, \ldots, b_p$ – параметры функции регрессии.

Таким образом, задача регрессионного анализа состоит в определении функции φ, ее параметров $b_0, b_1, \ldots, b_p$ и дальнейшего статистического исследования уравнения (15).

КОРРЕЛЯЦИОННОЕ ПОЛЕ

Рассмотрим понятие корреляционного поля на примере табл. 13, где через x_i и y_j обозначены соответственно середины интервалов изменения переменных, а через n_i, n_j – их частоты.

Зависимость между X и Y изобразим точками координатной плоскости (x, y). На рис. 2 точки (x_i, y_j) откладываются соответственно значениям переменных X и Y.

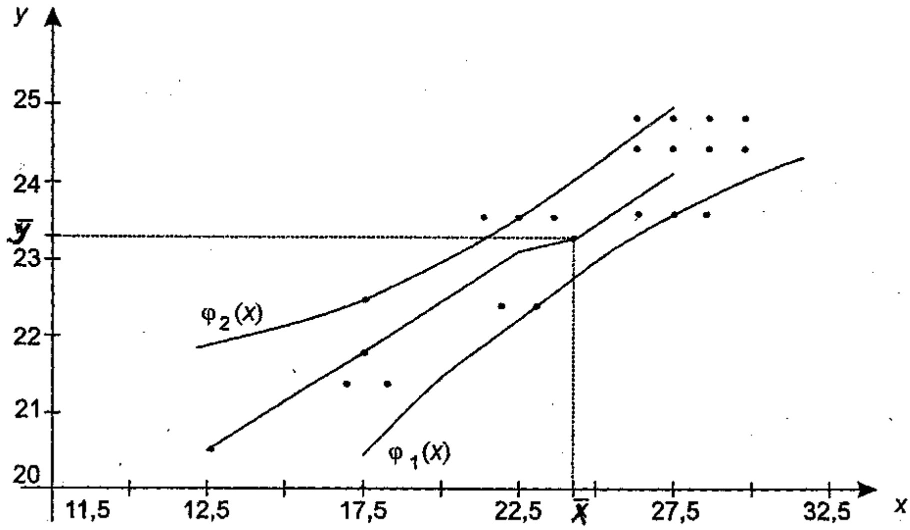

Рис. 2

Графическое изображение (рис. 2) называется полем корреляции.

Для каждого значения $x_i, i = \overline{1, l}$, вычислим групповые средние $\overline{Y}_i$. Соединив полученные точки ломаной линией, получим график эмпирической линии регрессии Y по X.

По виду эмпирической линии регрессии можно предположить, что связь между величинами X и Y является линейной.

ЛИНЕЙНАЯ РЕГРЕССИЯ

Если функция φ в (15) линейна по x, т. е.

$$Y_x = a + bx, \tag{16}$$

то говорят, что имеет место линейная регрессия Y по X.

Для определения параметров a и b используем нормальную систему метода наименьших квадратов

$$\begin{cases} na + b\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i, \\ a\sum_{i=1}^{n} x_i + b\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i. \end{cases} \tag{17}$$

Это линейная система двух уравнений с двумя неизвестными a и b, из которой следует, что

$$a = \frac{\sum_{i=1}^{n} y_i \cdot \sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} x_i y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}, \tag{18}$$

$$b = \frac{n\sum_{i=1}^{n} x_i y_i - \sum_{i=1}^{n} x_i \cdot \sum_{i=1}^{n} y_i}{n\sum_{i=1}^{n} x_i^2 - \left(\sum_{i=1}^{n} x_i\right)^2}. \tag{19}$$

Учитывая, что

$\sum_{i=1}^{n} x_i = n\overline{X}$, $\sum_{i=1}^{n} y_i = n\overline{Y}, \sum_{i=1}^{n} x_i^2 = n\overline{X}^2, \sum_{i=1}^{n} x_i y_i = n\overline{XY}$, из (18), (19) получим

$$a=\frac{\overline{Y}\cdot\overline{X}^2-\overline{X}\cdot\overline{XY}}{\overline{X}^2-\left(\overline{X}\right)^2}, \tag{20}$$

$$b=\frac{\overline{XY}-\overline{X}\cdot\overline{Y}}{\overline{X}^2-\left(\overline{X}\right)^2}. \tag{21}$$

Коэффициент b в уравнении регрессии (16) называется коэффициентом регрессии Y по X и обозначается b_{yx}.

Этот коэффициент находится по формуле

$$b_{yx}=r\frac{S_y}{S_x}. \tag{22}$$

Линейное уравнение регрессии можно записать в обычно принятой в математической статистике форме:

$$y_x-\overline{Y}=b_{yx}\left(x-\overline{X}\right) \tag{23}$$

или

$$y_x-\overline{Y}=r\frac{S_y}{S_x}\left(x-\overline{X}\right). \tag{24}$$

Аналогично, уравнение регрессии X на Y имеет вид

$$x_y-\overline{X}=r\frac{S_x}{S_y}\left(y-\overline{Y}\right). \tag{25}$$

СТАТИСТИЧЕСКИЙ АНАЛИЗ УРАВНЕНИЯ РЕГРЕССИИ

Для того чтобы установить, соответствует ли выбранная регрессионная модель (13) экспериментальным данным, используют основное уравнение дисперсионного анализа

$$\theta = \theta_{ф} + \theta_0 ,$$

где

$$\theta = \sum_{j=1}^{m} \left(y_j - \overline{Y} \right)^2 n_j \qquad (26)$$

– общая сумма квадратов отклонений y от средней,

$$\theta_{ф} = \sum_{i=1}^{l} \left(y_{xi} - \overline{Y} \right)^2 n_i \qquad (27)$$

– сумма квадратов, обусловленная регрессией,

$$\theta_0 = \sum_{i=1}^{l} \sum_{j=1}^{m} \left(y_j - y_{x_i} \right)^2 n_{ij} \qquad (28)$$

– остаточная сумма квадратов.

Для несгруппированной выборки формулы (26)–(28) несколько упрощаются:

$$\theta = \sum_{i=1}^{n} \left(y_i - \overline{Y} \right)^2 , \qquad (29)$$

$$\theta_{ф} = \sum_{i=1}^{n} \left(y_{x_i} - \overline{Y} \right)^2 , \qquad (30)$$

$$\theta_0 = \sum_{i=1}^{n} \left(y_i - y_{x_i} \right)^2 . \qquad (31)$$

Для заданного уровня α находим критическое значение $F_{\text{кр}}$ распределения Фишера при $k_1 = l-1$, $k_2 = n-l$ степенях свободы, где n – число наблюдений, l – число групп в корреляционной таблице или число оцениваемых параметров в несгруппированной выборке. Если статистика

$$t = \frac{\theta_{\text{ф}}(n-l)}{\theta_0(l-1)} > F_{\text{кр}}, \tag{32}$$

то уравнение регрессии считается значимым, т. е. соответствующим экспериментальным данным на уровне значимости α.

В случае линейной регрессии при $l=2$ уравнение регрессии значимо на уровне α, если

$$t = \frac{\theta_{\text{ф}}(n-2)}{\theta_0} > F_{\text{кр}}. \tag{33}$$

Воздействие неучтенных случайных факторов в линейной модели определяется остаточной дисперсией σ_0^2. Оценкой этой дисперсии является выборочная остаточная дисперсия

$$S_0^2 = \frac{1}{n-l}\theta_0. \tag{34}$$

Пример 12. Для зависимости Y от X, заданной корреляционной табл. 13, найти оценки параметров a и b уравнения линейной регрессии $Y_x = a + bx$, остаточную дисперсию; выяснить значимость уравнения регрессии при $\alpha = 0{,}005$.

Решение. Воспользуемся предыдущими результатами

$$\bar{X} = \frac{1}{n}\sum_i x_i n_i = \frac{1}{20}\cdot 480 = 24;\ \bar{Y} = 23{,}4;\ r = 0{,}934;$$

$$\overline{X^2} = \frac{1}{n}\sum_i {x_i}^2 n_i = \frac{1}{20}\cdot 11925 = 596{,}25;\ S_x^2 = 20{,}25;$$

$$S_x = 4,5;$$

$$\overline{Y^2} = \frac{1}{n}\sum_j {y_j}^2 n_j = \frac{1}{20}\cdot 10979 = 548,95;\ S_y^2 = 1,39;$$

$$S_y = 1,179.$$

Согласно формуле (24), уравнение регрессии будет иметь вид $Y_x - 23,4 = 0,934\cdot\frac{1,179}{4,5}\left(x-24\right)$ или $Y_x = 17,53 + 0,245\cdot x$, тогда $a = 17,53;\ b = 0,245$.

Для выяснения значимости уравнения регрессии вычислим суммы $\theta_ф$ и θ_0. Составим расчетную таблицу:

x_i	$y_{x_i} = a + b_{x_i}$	n_i	$\left(Y_{x_i} - \overline{Y}\right)^2 n_i$	$\sum\limits_{j=1}^{m}\left(Y_j - Y_{x_i}\right)^2 n_{ij}$
12,5	20,59	1	7,90	0,0081
17,5	21,81	3	7,58	0,6683
22,5	23,03	5	0,68	1,2245
27,5	24,26	11	8,14	1,8128
$\sum$		20	24,30	3,7137

Из (27) и (28) по данным таблицы получим $\theta_ф = 24,30$; $\theta_0 = 3,71$.

$k_1 = l - 1 = 4 - 1 = 3,\ k_2 = n - l = 20 - 4 = 16$ и $\alpha = 0,05$, по табл. П7 находим $F_{кр} = 3,24$.

Вычислим статистику

$$t = \frac{\theta_ф\left(n-l\right)}{\theta_0\left(l-1\right)} = \frac{24,30\cdot\left(20-4\right)}{3,71\cdot\left(4-1\right)} = \frac{24,30\cdot 16}{3,71\cdot 3} = 35.$$

Так как $t > F_{\text{кр}}$, то уравнение регрессии значимо. Остаточная дисперсия равна $S_0^2 = \frac{1}{n-l}\theta_0 = \frac{3,71}{16} = 0,232$.

ЗАДАЧИ ДЛЯ САМОСТОЯТЕЛЬНОГО РЕШЕНИЯ

1. В таблице представлены результаты случайной выборки измерения роста (X) и веса (Y) 50 мужчин.

Y, кг / X, см	60	70	80	90
160	2	5	4	1
170	2	8	9	4
180	0	4	6	5

Найти выборочное уравнение прямой регрессии Y на X, коэффициент корреляции и коэффициенты регрессии $\rho_{y/x}$, $\rho_{x/y}$.

Ответ: $\overline{Y}_x = 0,3691 \cdot x + 14,0274$, $\overline{X}_y = 0,2572 \cdot y + 150,7992$, $r =$ $= 0,3081$.

2. Найти коэффициент линейной корреляции между признаками X и Y и написать уравнения прямых регрессий Y на X и X на Y, если распределение признаков X и Y приводится в таблице.

Y / X	11	16	21	26	31	36
25	2	4				
35		6	3			
45			6	45	4	
55			2	8	6	
65				4	7	3

Ответ: $\overline{Y}_x = 15873 \cdot x + 7,1425$, $\overline{X}_y = 0,3822 \cdot y + 7,2215$, $r = 0,7789$.

3. Для использования зависимости объема производства Y от основных фондов X получены статистические данные по 52 предприятиям за год.

y_i	x_j, млн руб.								
	12,5	17,5	22,5	27,5	32,5	37,5	42,5	47,5	52,5
200–210	1								
210–220		2							
220–230		1	2						
230–240			3	3	1				
240–250				8	9				
250–260				1	7	5			
260–270						2			
270–280						1	3		
280–290								2	
290–300									1

Найти уравнение прямой регрессии Y на X, отражающее зависимость объема производства от основных фондов, и, наоборот, уравнение регрессии X на Y, отражающее величину основных фондов предприятия при определении объема производства. Вычислить коэффициент корреляции.

Ответ: $\overline{Y}_x = 1{,}9132 \cdot x + 187{,}7306$, $\overline{X}_y = 0{,}4369 \cdot y - 76{,}86$, $r =$ $= 0{,}9143$.

ЗАДАЧИ ДЛЯ ВНЕУДИТОРНОЙ РАБОТЫ

При исследовании статистической зависимости между случайными величинами по данной корреляционной таблице требуется:

1. Найти числовые характеристики выборки

– среднее выборочное ;

– выборочное среднее квадратическое отклонение;

– выборочный коэффициент корреляции;

– выборочный коэффициент детерминации.

2. Построить корреляционное поле; по характеру расположения точек на корреляционном поле подобрать общий вид функции регрессии.

3. Найти эмпирические функции регрессии Y на X и X на Y.

4. Построить графики эмпирических функций регрессий.

5. Проверить значимость коэффициента корреляции. (Уровень значимости принять равным $\alpha = 0{,}05$.)

Вариант 1

$y \backslash x$	–30	–20	–10	0	n_y
–8	1	–	–	–	1
–4	4	1	–	–	5
0	1	15	1	–	17
4	–	2	13	–	15
8	–	–	2	1	3
12	–	–	–	9	9
n_x	6	18	16	10	$n = 50$

Вариант 2

y \ x	0,7	0,9	1,1	1,3	1,5	n_y
0,5	3	2	–	–	–	5
0,6	6	14	3	1	–	24
0,7	1	6	27	12	2	48
0,8	–	–	5	4	8	17
0,9	–	–	–	2	4	6
n_x	10	22	35	19	14	n = 100

Вариант 3

y \ x	0,75	0,80	0,85	0,90	n_y
10	6	–	–	–	6
20	4	2	–	–	6
30	–	12	–	–	12
40	–	6	4	1	11
50	–	–	6	9	15
n_x	10	20	10	10	n = 50

Вариант 4

y \ x	30	50	70	90	110	130	n_y
40	4	1	–	–	–	–	5
60	1	3	3	–	–	–	7
80	–	2	7	4	–	–	13
100	–	–	4	8	1	–	13
120	–	–	–	3	5	–	8
140	–	–	–	–	1	3	4
n_x	5	6	14	15	7	3	$n = 50$

Вариант 5

y \ x	75	85	90	95	n_y
2	3	–	–	–	3
3	2	4	1	–	7
4	1	18	9	–	28
5	–	1	8	1	10
6	–	–	–	2	2
n_x	6	23	18	3	$n = 50$

Вариант 6

y \ x	2	4	6	8	10	n_y
0,06	–	–	–	–	4	4
0,07	–	–	–	3	3	6
0,08	–	5	10	12	–	27
0,09	–	2	4	4	–	10
0,10	1	1	1	–	–	3
n_x	1	8	15	19	7	$n = 50$

Вариант 7

y \ x	1,35	1,45	1,55	1,65	1,75	n_y
6	–	–	–	3	1	4
7	–	1	4	1	–	6
8	1	14	12	1	–	28
9	1	5	1	–	–	7
10	5	–	–	–	–	5
n_x	7	20	17	5	1	$n = 50$

Вариант 8

$y \backslash x$	30	35	40	45	50	55	n_y
12	3	–	–	–	–	–	3
14	3	5	–	–	–	–	8
16	–	2	10	12	–	–	24
18	–	–	–	2	3	5	10
20	–	–	–	–	1	4	5
n_x	6	7	10	14	4	9	$n = 50$

Вариант 9

$y \backslash x$	0,02	0,06	0,10	0,14	0,18	n_y
0,05	2	4	2	–	–	8
0,10	–	2	10	4	–	16
0,15	–	–	3	10	5	18
0,20	–	–	–	2	5	7
0,25	–	–	–	–	1	1
n_x	2	6	15	16	11	$n = 50$

Вариант 10

$y \backslash x$	0,65	0,95	1,25	1,55	1,85	2,15	n_y
0,3	5	–	–	–	–	–	5
0,4	2	6	6	1	–	–	15
0,5	–	7	5	4	–	–	16
0,6	–	1	4	5	2	–	12
0,7	–	–	–	–	1	1	2
n_x	7	14	15	10	3	1	$n = 50$

Вариант 11

$x \backslash y$	6	8	10	12	14	n_x
30	2	17	9	3	–	31
35	–	10	17	9	–	36
40	–	3	24	16	13	56
45	–	–	6	24	12	42
50	–	–	2	11	22	35
n_y	2	30	58	63	47	$n = 200$

Вариант 12

x \ y	20	30	40	50	60	n_x
15	7	5	–	–	–	12
25	20	23	–	–	–	43
35	–	30	47	2	–	79
45	–	10	11	20	6	47
55	–	–	9	7	3	19
n_y	27	68	67	29	9	$n = 200$

Вариант 13

x \ y	3	9	15	21	27	33	n_x
25	–	–	–	1	–	1	2
35	–	–	1	5	4	5	15
45	–	–	2	18	10	2	32
55	–	6	14	2	2	–	24
65	–	6	3	–	–	–	9
75	4	8	–	–	–	–	12
85	6	–	–	–	–	–	6
n_y	10	20	20	26	16	8	$n = 100$

Вариант 14

x \ y	2	2,5	3	3,5	4	n_x
1000	–	–	–	2	3	5
2000	–	–	3	6	2	11
3000	–	4	6	3	–	13
4000	1	6	4	1	–	12
5000	6	3	–	–	–	9
n_y	7	13	13	12	5	$n = 50$

Вариант 15

x \ y	1,25	1,5	1,75	2	2,25	n_x
8	–	–	1	2	3	6
13	–	–	1	4	3	8
18	–	4	7	1	–	12
23	2	7	5	–	–	14
28	6	4	–	–	–	10
n_y	8	15	14	7	6	$n = 50$

ПРИЛОЖЕНИЕ

Таблица П5

Значения функции $\chi^2_{\alpha;\upsilon};\ P\left(\chi^2 \geq \chi^2_{\alpha;\upsilon}\right) = \alpha$

υ\α	0,20	0,10	0,05	0,02	0,01	0,001
1	1,642	2,706	3,841	5,412	6,635	10,827
2	3,219	4,605	5,991	7,824	9,210	13,815
3	4,642	6,251	7,815	9,837	11,345	16,266
4	5,989	7,779	9.488	11,668	13,277	18,467
5	7,289	9,236	11,070	13,388	15,086	20,515
6	8,558	10,645	12,592	15,033	16,812	22,457
7	9,803	12,017	14,067	16,622	18,475	24,322
8	11,030	13,362	15,507	18,168	20,090	26,125
9	12,242	14,684	16,919	19,679	21,666	27,877
10	13,442	15,987	18,307	21,161	23,209	29,588
11	14,631	17,275	19,675	22,618	24,725	31,264
12	15,812	18,549	21,026	24,054	26,217	32,909
13	16,985	19,812	22,362	25,472	27,688	34,528 1
14	18,151	21,064	23,685	26,683	29,141	36,123
15	19,311	22,307	24,996	28,259	30,578	37,697
16	20,465	23,542	26,296	29,633	32,000	39,252
17	21,615	24,769	27,587	30,995	33,409	40,790
18	22,760	25,989	28,869	32,346	34,805	42,312
19	23,900	27,204	30,144	33,687	36,191	43,820
20	25,038	28,412	31,410	35,020	37,566	45,315
21	26,171	29,615	32,671	36,343	38,932	46,797
22	27,301	30,813	33,924	37,659	40,289	48,268
23	28,429	32,007	35,172	38,968	41,638	49,728
24	29,553	33,196	36,415	40,270	42,980	51,179
25	30,675	34,382	37,652	41,566	44,312	52,620
26	31,795	35,563	38,885	42,856	45,642	54,052
27	32,912	36,741	40,113	44,140	46,963	55,476
28	34,027	37,916	41,337	45,419	48,278	56,893
29	35,139	39,087	42,557	46,693	49,588	58,302
30	36,250	40,256	43,773	47,962	50,892	59,703

Таблица П6

Распределение Стьюдента

Значения $t_{\alpha;\,\nu}$ удовлетворяют условию

$$P\left(t \geq t_{\alpha;\upsilon}\right) = \int_{t_{\alpha;\upsilon}}^{\infty} S\left(t,\upsilon\right)dt = \alpha$$

υ\α	0,40	0,30	0,20	0,10	0,05	0,025	0,010	0,005	0,001	0,0005
1	2	3	4	5	6	7	8	9	10	11
1	0,325	0,727	1,376	3,078	6,314	12,71	31,82	63,66	318,3	636,6
2	0,289	0,617	1,061	1,886	2,920	4,303	6,965	9,925	22,33	32,60
3	0,277	0,584	0,978	1,638	2,353	3,182	4,541	5,841	10,22	12,94
4	0,271	0,569	0,941	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	0,267	0,559	0,920	1,476	2,015	2,571	3,365	5,032	5,893	6,859
6	0,265	0,553	0,906	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	0,263	0,549	0,896	1,415	1,895	2,365	2,998	3,499	4,785	5,405
8	0,262	0,546	0,889	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	0,261	0,543	0,883	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	0,260	0,542	0,879	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	0,260	0,540	0,876	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	0,259	0,539	0,873	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	0,259	0,538	0,870	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	0,258	0,537	0,868	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	0,258	0,536	0,866	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	0,258	0,535	0,865	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	0,257	0,534	0,863	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	0,257	0,534	0,862	1,330	1,734	2,101	2,552	2,878	3,611	3,922
19	0,257	0,533	0,861	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	0,257	0,533	0,860	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	0,257	0,532	0,859	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	0,256	0,532	0,858	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	0,256	0,532	0,858	1,319	1,714	2,069	2,500	2,807	3,485	3,767
24	0,256	0,531	0,857	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	0,256	0,531	0,856	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	0,256	0,531	0,856	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	0,256	0,531	0,855	1,314	1,703	2,052	2,473	2,771	3,421	3,690

Окончание табл. П6

1	2	3	4	5	6	7	8	9	10	11
28	0,256	0,530	0,855	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	0,256	0,530	0,854	1,311	1,699	2,045	2,462	7,756	3,396	3,659
30	0,256	0,530	0,854	1,310	1,697	2,042	2,457	2,750	3,385	3,646
40	0,255	0,529	0,851	1,303	1,684	2,021	2,423	2,704	3,307	3,551
50	0,255	0,528	0,849	1,298	1,676	2,009	2,403	2,678	3,262	3,495
60	0,254	0,527	0,848	1,296	1,671	2,000	2,390	2,660	3,232	3,460
80	0,254	0,527	0,846	1,292	1,664	1,990	2,374	2,639	3,195	3,415
100	0,254	0,526	0,845	1,290	1,660	1,984	2,365	2,626	3,174	3,389
200	0,254	0,525	0,843	1,286	1,653	1,972	2,345	2,601	3,131	3,339
500	0,253	0,525	0,842	1,283	1,648	1,965	2,334	2,586	3,106	3,310
∞	0,253	0,524	0,842	1,282	1,645	1,960	2,326	2,576	3,090	3,291

Таблица П7

Значения $F_{\alpha,} k_1, k_2$ – критерия Фишера-Снедекора

α = 0,05																			
k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	240	242	244	246	248	249	250	251	252	253	254
2	18,5	19,0	19,2	19,2	19,3	19,3	19,3	19,4	19,4	19,4	19,4	19,4	19,4	19,4	19,5	19,5	19,5	19,5	19,5
3	10,1	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79	8,74	8,70	8,66	8,64	8,62	8,59	8,57	8,55	8,53
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96	5,91	5,86	5,80	5,77	5,75	5,72	5,69	5,66	5,63
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74	4,68	4,62	4,56	4,53	4,50	4,46	4,43	4,40	4,36
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06	4,00	3,94	3,87	3,84	3,81	3,77	3,74	3,70	3,67
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64	3,57	3,51	3,44	3,41	3,38	3,34	3,30	3,27	3,23
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35	3,28	3,22	3,15	3,12	3,08	3,04	3,01	2,97	2,93
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14	3,07	3,01	2,94	2,90	2,86	2,83	2,79	2,75	2,71
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98	2,91	2,85	2,77	2,74	2,70	2,66	2,62	2,58	2,54
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85	2,79	2,72	2,65	2,61	2,57	2,53	2,49	2,45	2,40
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75	2,69	2,62	2,54	2,51	2,47	2,43	2,38	2,34	2,30
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67	2,60	2,53	2,46	2,42	2,38	2,34	2,30	2,25	2,21
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60	2,53	2,46	2,39	2,35	2,31	2,27	2,22	2,18	2,13
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54	2,48	2,40	2,33	2,29	2,25	2,20	2,16	2,11	2,07
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49	2,42	2,35	2,28	2,24	2,19	2,15	2,11	2,06	2,01
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45	2,38	2,31	2,23	2,19	2,15	2,10	2,06	2,01	1,96
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41	2,34	2,27	2,19	2,15	2,11	2,06	2,02	1,97	1,92
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38	2,31	2,23	2,16	2,11	2,07	2,03	1,98	1,93	1,88
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35	2,28	2,20	2,12	2,08	2,04	1,99	1,95	1,90	1,84

Продолжение табл. П7

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32	2,25	2,18	2,10	2,05	2,01	1,96	1,92	1,87	1,81
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30	2,23	2,15	2,07	2,03	1,98	1,94	1,89	1,84	1,78
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27	2,20	2,13	2,05	2,01	1,96	1,91	1,86	1,81	1,76
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25	2,18	2,11	2,03	1,98	1,94	1,89	1,84	1,79	1,73
25	4,24	3,39	2,99	2,76	2,60	4,49	2,40	2,34	2,28	2,24	2,16	2,09	2,01	1,96	1,92	1,87	1,82	1,77	1,71
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22	2,15	2,07	1,99	1,95	1,90	1,85	1,80	1,75	1,69
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20	2,13	2,06	1,97	1,93	1,88	1,84	1,79	1,73	1,67
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19	2,12	2,04	1,96	1,91	1,87	1,82	1,77	1,71	1,65
29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22	2,18	2,10	2,03	1,94	1,90	1,85	1,81	1,75	1,70	1,64
30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21	2,16	2,09	2,01	1,93	1,89	1,84	1,79	1,74	1,68	1,62
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	2,08	2,00	1,92	1,84	1,79	1,74	1,69	1,64	1,58	1,51
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	1,99	1,92	1,84	1,75	1,70	1,65	1,59	1,53	1,47	1,39
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	1,91	1,83	1,75	0,66	1,61	1,55	1,50	1,43	1,35	1,25
∞	3,84	3,00	3,60	2,37	2,21	2,10	2,01	1,94	1,83	1,83	1,75	1,67	1,57	1,52	1,46	1,39	1,32	1,22	1,00
$\alpha = 0{,}01$																			
1	4052	4999,5	5403	5625	5764	5859	5928	5982	6022	6056	6106	6157	6209	6235	6261	6287	6313	6339	6366
2	98,50	99,00	99,17	99,25	99,30	99,33	99,36	99,37	99,39	99,40	99,42	99,43	99,45	99,46	99,47	99,47	99,48	99,49	99,50
3	34,12	30,82	29,46	28,71	28,24	27,91	27,67	27,49	27,35	27,23	27,05	26,87	26,69	26,60	26,50	26,41	26,32	26,22	26,13
4	21,20	18,00	16,69	15,98	15,52	15,21	14,98	14,80	14,66	14,55	14,37	14,20	14,02	13,93	13,84	13,75	13,65	13,56	13,46
5	16,26	13,27	12,06	11,39	10,97	10,67	10,46	10,29	10,16	10,05	9,89	9,72	9,55	9,47	9,38	9,29	9,20	9,11	9,02
6	13,75	10,92	9,78	9,15	8,75	8,47	8,26	8,10	7,98	7,87	7,72	7,56	7,40	7,31	7,23	7,14	7,06	6,97	6,88
7	12,25	9,55	8,45	7,85	7,46	7,19	6,99	6,84	6,72	6,62	6,47	6,31	6,16	6,07	5,99	5,91	5,82	5,74	5,65
8	11,26	8,65	7,59	7,01	6,63	6,37	6,18	6,03	5,91	5,81	5,67	5,52	5,36	5,28	5,20	5,12	5,03	4,95	4,86
9	10,56	8,02	6,99	6,42	6,06	5,80	5,61	5,47	5,35	5,26	5,11	4,96	4,81	4,73	4,65	4,57	4,48	4,40	4,31
10	10,04	7,56	6,55	5,99	5,64	5,39	5,20	5,06	4,94	4,85	4,71	4,56	4,41	4,33	4,25	4,17	4,08	4,00	3,91
11	9,65	7,21	6,22	5,67	5,32	5,07	4,89	4,74	4,63	4,54	4,40	4,25	4,10	4,02	3,94	3,86	3,18	3,69	3,60

Окончание табл. П7

k_2 \ k_1	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
12	9,33	6,93	5,95	5,41	5,06	4,82	4,64	4,50	4,39	4,30	4,16	4,01	3,86	3,78	3,70	3,62	3,54	3,45	3,36
13	9,07	6,70	5,74	5,21	4,86	4,62	4,44	4,30	4,19	4,10	3,96	3,82	3,66	3,59	3,51	3,43	3,34	3,25	3,17
14	8,86	6,51	5,56	5,04	4,69	4,46	4,28	4,14	4,03	3,94	3,80	3,66	3,51	3,43	3,35	3,27	3,18	3,09	3,00
15	8,68	6,36	5,42	4,89	4,56	4,32	4,14	4,00	3,89	3,80	3,67	3,52	3,37	3,29	3,21	3,13	3,05	2,96	2,87
16	8,53	6,23	5,29	4,77	4,44	4,20	4,03	3,89	3,78	3,69	3,55	3,41	3,26	3,18	3,10	3,02	2,93	2,84	2,75
17	8,40	6,11	5,18	4,67	4,34	4,10	3,93	3,79	3,68	3,59	3,46	3,31	3,16	3,08	3,00	2,92	2,83	2,75	2,65
18	8,29	6,01	5,09	4,58	4,25	4,01	3,84	3,71	3,60	3,51	3,37	3,23	3,08	3,00	2,92	2,84	2,75	2,66	2,57
19	8,18	5,93	5,01	4,50	4,17	3,94	3,77	3,63	3,52	3,43	3,30	3,15	3,00	2,92	2,84	2,76	2,67	2,58	2,49
20	8,10	5,85	4,94	4,43	4,10	3,87	3,70	3,56	3,46	3,37	3,23	3,09	2,94	2,86	2,78	2,69	2,61	2,52	2,42
21	8,02	5,78	4,87	4,37	4,04	3,81	3,64	3,51	3,40	3,31	3,17	3,03	2,88	2,80	2,72	2,64	2,55	2,46	2,36
22	7,95	5,72	4,82	4,31	3,99	3,76	3,59	3,45	3,35	3,26	3,12	2,98	2,83	2,75	2,67	2,58	2,50	2,40	2,31
23	7,88	5,66	4,76	4,26	3,94	3,71	3,54	3,41	3,30	3,21	3,07	2,93	2,78	2,70	2,62	2,54	2,45	2,35	2,26
24	7,82	5,61	4,72	4,22	3,90	3,67	3,50	3,36	3,26	3,17	3,03	2,89	2,74	2,66	2,58	2,49	2,40	2,31	2,21
25	7,77	5,57	4,68	4,18	3,85	3,63	3,46	3,32	3,22	3,13	2,99	2,85	2,70	2,62	2,54	2,45	2,36	2,27	2,17
26	7,72	5,53	4,64	4,14	3,82	3,59	3,42	3,29	3,18	3,09	2,96	2,81	2,66	2,58	2,50	2,42	2,33	2,23	2,13
27	7,68	5,49	4,60	4,11	3,78	3,56	3,39	3,26	3,15	3,06	2,93	2,78	2,63	2,55	2,47	2,38	2,29	2,20	2,10
28	7,64	5,45	4,57	4,07	3,75	3,53	3,36	3,23	3,12	3,03	2,90	2,75	2,60	2,52	2,44	2,35	2,26	2,17	2,06
29	7,60	5,42	4,54	4,04	3,73	3,50	3,33	3,20	3,09	3,00	2,87	2,73	2,57	2,49	2,41	2,33	2,23	2,14	2,03
30	7,56	5,39	4,51	4,02	3,70	3,47	3,30	3,17	3,07	2,98	2,84	2,70	2,55	2,47	2,39	2,30	2,21	2,11	2,01
40	7,31	5,18	4,31	3,83	3,51	3,29	3,12	2,99	2,89	2,80	2,66	2,52	2,37	2,29	2,20	2,11	2,02	1,92	1,80
60	7,08	4,98	4,13	3,65	3,34	3,12	2,95	2,82	2,72	2,63	2,50	2,35	2,20	2,12	2,03	1,94	1,84	1,73	1,60
120	6,85	4,79	3,95	3,48	3,17	2,96	2,79	2,66	2,56	2,47	2,34	2,19	2,03	1.95	1,86	1,76	1,66	1,53	1,38
∞	6,63	4,61	3,78	3,32	3,02	2,80	2,64	2,51	2,41	2,32	2,18	2,04	1,88	1,79	1,70	1,59	1,47	1,32	1,00

k_1 – число степеней свободы для большой дисперсии, k_2 – для меньшей дисперсии.

Содержание

Учебное издание

ЕРОШЕВСКАЯ Вера Ивановна
ЕРОШЕВСКАЯ Елена Леонидовна
МИНЧЕНКОВА Лариса Павловна

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Методическое пособие

В 2 частях

Часть 2

Редактор *Л. Н. Шалаева*
Компьютерная верстка *А. Г. Занкевич*

Подписано в печать 08.05.2014. Формат 60×84 $^{1}/_{16}$. Бумага офсетная. Ризография.
Усл. печ. л. 4,42. Уч.-изд. л. 3,45. Тираж 50. Заказ 1018.

Издатель и полиграфическое исполнение: Белорусский национальный технический университет.
Свидетельство о государственной регистрации издателя, изготовителя, распространителя печатных изданий № 1/173 от 12.02.2014. Пр. Независимости, 65. 220013, г. Минск.