Лекция 17

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ

ЦЕЛЬ ЛЕКЦИИ: определить понятие статистических гипотез и правила их проверки; провести проверку гипотез о равенстве средних значений и дисперсий нормально распределенной случайной величины, сформулировать критерий согласия χ^2.

Определение статистической гипотезы

Статистической называется гипотеза о предполагаемом виде неизвестного распределения или утверждение относительно значений одного или нескольких параметров известного распределения. Например, совокупность наблюдаемых значений распределена по закону Пуассона, математическое ожидание случайной величины X равно a – статистические гипотезы.

Гипотеза, которая подвергается проверке, называется нулевой и обозначается H_0. Альтернативной гипотезой H_1 называется гипотеза, конкурирующая с нулевой, т. е. ей противоречащая. Простой называется гипотеза, содержащая только одно предположение. Сложная гипотеза состоит из конечного или бесконечного числа простых гипотез.

Пример. Пусть проверяется гипотеза о равенстве некоторого параметра a значению a_0, т. е. гипотеза $H_0 : a = a_0$. В этом случае альтернативной гипотезой можно рассматривать одну из следующих гипотез: $H_1 : a > a_0$; $H_1 : a < a_0$; $H_1 : a \neq a_0$; $H_1 : a > 2$. Все приведенные гипотезы простые, и только $H_1 : a > 2$ – сложная гипотеза.

Выбор альтернативной гипотезы определяется формулировкой решаемой задачи. Причина выделения нулевой гипотезы состоит в том, что чаще всего такие гипотезы рассматриваются как утверждения, которые более ценны, если они опровергаются. Это основано на общем принципе, в соответствии с которым теория должна быть отвергнута, если есть противоречащий ей факт, но не обязательно должна быть принята, если противоречащих ей фактов на текущий момент нет.

Правило, по которому выносится решение принять или отклонить гипотезу H_0, называется статистическим критерием. Проверка статистических гипотез осуществляется по результатам наблюдений (экспериментов, опытов), из которых формируют функцию результатов наблюдений, называемую проверочной статистикой. Таким образом, статистический критерий устанавливает, при каких значениях этой статистики проверяемая гипотеза принимается, а при каких она отвергается.

Пусть по n независимым наблюдениям случайной величины X получена некоторая оценка $\hat{a}$. Предположим, что есть основания считать истинное значение оцениваемого параметра a равным некоторой величине a_0. Однако даже если истинное значение параметра a и равно a_0, то оценка $\hat{a}$, скорее всего, не будет в точности равняться a_0 из-за статистической изменчивости, присущей $\hat{a}$. Поэтому возникает вопрос. Если предположить, что $a = a_0$, то при каком отклонении $\hat{a}$ от a_0 это предположение (гипотеза) должно быть опровергнуто как несостоятельное? Ответ на этот вопрос можно получить, вычислив вероятность любого значимого отклонения $\hat{a}$ от a_0, используя закон распределения случайной величины $\hat{a}$.

Если вероятность превышения разности $\hat{a}$ и заданного уровня a_0 мала, то этот уровень следует считать значимым и гипотезу $H_0 : a = a_0$ следует отвергнуть. Если вероятность превышения данной разности не является малой, то наличие этой разности можно отнести за счет обычной статистической изменчивости и гипотезу $H_0 : a = a_0$ можно считать правдоподобной. Природа статистических выводов такова, что при отклонении гипотезы можно заранее оценить вероятность возможной ошибки (отклонения истинной гипотезы); напротив, если гипотеза принята, то это не означает, что она подтверждена с заданной вероятностью. Это лишь означает, что гипотеза согласуется с опытными данными, но возможно, что для другого эксперимента гипотеза будет отвергнута.

Приведенные рассуждения представляют собой простейший вид статистической процедуры, называемой проверкой гипотез. Предполагаем, что $\hat{a}$ – несмещенная оценка параметра a – имеет плотность распределения $p(\hat{a})$. Если гипотеза $H_0 : a = a_0$ верна, то функция $p(\hat{a})$ должна иметь среднее значение a_0, как это показано на рис. 8.9.

Вероятность того, что величина $\hat{a}$ не будет превышать нижнего уровня $a_{1-\alpha/2}$, равна

$$P\left\{\hat{a} \le a_{1-\alpha/2}\right\} = \int_{-\infty}^{a_{1-\alpha/2}} p(\hat{a})\, d\hat{a} = \frac{\alpha}{2},$$

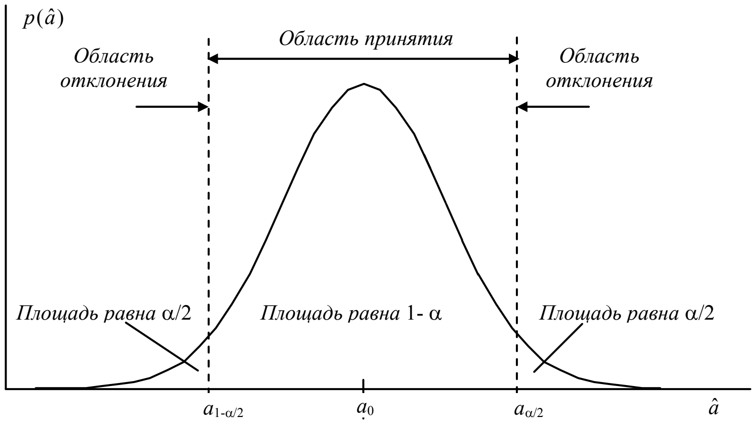

Рис. 8.9. Области принятия и отклонения при проверке гипотез

а вероятность того, что $\hat{a}$ превзойдет верхнюю границу $a_{\alpha/2}$, составит

$$P\{\hat{a} > a_{\alpha/2}\} = \int_{a_{\alpha/2}}^{\infty} p(\hat{a})d\hat{a} = \frac{\alpha}{2}.$$

Таким образом, вероятность того, что $\hat{a}$ выйдет за пределы интервала с границами $a_{1-\alpha/2}$ и $a_{\alpha/2}$, составит α. Величина α выбирается настолько малой, чтобы попадание $\hat{a}$ за пределы интервала, заключенного между $a_{1-\alpha/2}$ и $a_{\alpha/2}$, было бы практически невозможным событием. Если в результате эксперимента величина $\hat{a}$ выходит за пределы интервала $(a_{1-\alpha/2}, a_{\alpha/2})$, то в этом случае есть серьезные основания сомневаться в справедливости проверяемой гипотезы $H_0: a = a_0$. В самом деле, если гипотеза H_0 верна, то значение $\hat{a}$ будет маловероятным, и поэтому гипотезу о равенстве параметра a величине a_0 следует отвергнуть. С другой стороны, если оценка $\hat{a}$ попадает в интервал $(a_{1-\alpha/2}, a_{\alpha/2})$, то нет серьезных оснований подвергать сомнению справедливость проверяемой гипотезы H_0, и гипотезу о равенстве $a = a_0$ следует принять.

Малое значение вероятности α, используемое при проверке гипотезы, называется уровнем значимости критерия. Интервал значений $\hat{a}$, для

которых гипотезу следует отвергнуть, называется областью отклонения гипотезы, или критической областью. Интервал значений $\hat{a}$, при которых гипотезу следует принять, носит название области принятия гипотезы (см. рис. 8.9). Приведенный способ проверки гипотезы называется двусторонним критерием, так как если гипотеза H_0 верна, то величина $\hat{a}$ может быть как больше, так и меньше a_0. Необходимо проверять значимость расхождения между $\hat{a}$ и a_0 с обеих сторон. В некоторых задачах может оказаться достаточно одностороннего критерия. Например, пусть гипотеза состоит том, что $a \geq a_0$. В этом случае гипотеза будет ошибочной только тогда, когда $a < a_0$, а критерий будет использовать только нижнюю границу плотности распределения $p(\hat{a})$.

При проверке статистических гипотез возможны ошибки двух типов. Во-первых, гипотеза может быть отклонена, хотя в действительности она верна. Эта возможная ошибка называется ошибкой первого рода. Во-вторых, гипотеза принимается, хотя фактически она неверна. Такая ошибка называется ошибкой второго рода. Как видно на рис. 8.9, ошибка первого рода происходит в том случае, когда при справедливости гипотезы $\hat{a}$ попадает в область ее отклонения. Таким образом, вероятность ошибки первого рода равна α, т. е. уровню значимости критерия.

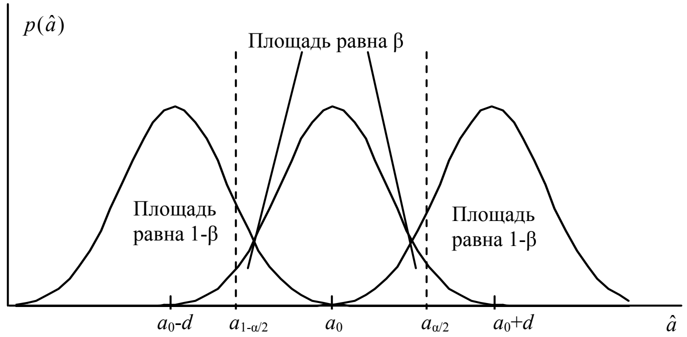

Рис. 8.10. Ошибка второго рода при проверке гипотезы

Для того чтобы найти вероятность ошибки второго рода, следует определить каким-то образом величину отклонения истинного значения параметра a от гипотетического значения параметра a_0, которое требуется определить. Предполагается, что истинное значение параметра a в действительности равно $a_0 + d$ или $a_0 - d$ (см. рис. 8.10). Если согласно гипотезе $H_0 : a = a_0$, а на самом деле $a = a_0 \pm d$, то вероятность того, что $\hat{a}$ попадет в область принятия гипотезы H_0, т. е. в интервал $(a_{1-\alpha/2}, a_{\alpha/2})$, составляет β. Таким образом, вероятность ошибки второго рода равна β при выявлении отклонения истинного значения параметра a на $\pm d$ от гипотетической величины a_0.

Вероятность $1-\beta$ называется мощностью критерия. Понятно, что при заданном значении n (объеме опытных данных) вероятность ошибки первого рода может быть сделана сколь угодно малой за счет уменьшения уровня значимости α. Однако при этом растет вероятность β – ошибка второго рода (уменьшается мощность критерия). Единственный способ уменьшить и α, и β состоит в увеличении объема выборки n, используемой при вычислении $\hat{a}$. Исходя из этих соображений определяется объем необходимых опытных данных в статистических экспериментах.

Проверка гипотезы о равенстве статистических средних значений

При проведении статистических экспериментов со случайными величинами самой различной природы большое внимание уделяется воспроизводимости результатов опытов при неоднократном повторении серии экспериментов. Часто возникает ситуация, когда среднее значение в одной серии опытов заметно отличается от величины этого параметра в другой серии. Естественно возникает вопрос, чем объяснить обнаруженное расхождение средних значений: либо случайными ошибками, либо это расхождение вызвано какими-то незамеченными или даже неизвестными ранее закономерностями.

Пусть над случайной величиной X проводятся две серии опытов. Первая серия объемом n_1 испытаний: $x_{1,1}, x_{1,2}, \ldots, x_{1,n_1}$. Вторая серия экспериментов объемом n_2 испытаний: $x_{2,1}, x_{2,2}, \ldots, x_{2,n_2}$. При этом известно, что случайная величина X распределена по нормальному закону с математическим ожиданием m_1 и дисперсией D_1 в первой серии опытов и параметрами m_2 и D_2 – во второй.

По экспериментальным данным получены статистические средние значения m_1^* и m_2^*. Необходимо проверить гипотезу $H_0 : m_1 = m_2$ при альтернативной гипотезе $H_1 : m_1 \neq m_2$.

Для случая, когда дисперсии D_1 и D_2 известны, оценки m_1^* и m_2^* имеют нормальное распределение с параметрами $(m_1, D_1/n_1)$ и $(m_2, D_2/n_2)$ соответственно (см. 7.3 и 7.4). Так как случайные величины m_1^* и m_2^* независимы, то их разность тоже имеет нормальный закон распределения с параметрами $m_1 - m_2$ и $D_1/n_1 + D_2/n_2$ (нормальный закон устойчив к композиции).

Таким образом, случайная величина

$$Z = \frac{m_1^* - m_2^*}{\sqrt{D_1/n_1 + D_2/n_2}},$$

которая является нормированной разностью оценок математических ожиданий, имеет нормальное распределение с нулевым математическим ожиданием и единичной дисперсией. Эта величина выбирается в качестве проверочной статистики.

Критическим для проверяемой гипотезы являются значения

$$|Z| > z_{\alpha/2},$$

где $z_{\alpha/2}$ – квантиль порядка $1-\alpha/2$ стандартного нормального распределения, т. е. значение гауссовой случайной величины, вероятность попасть правее которой равна $\alpha/2$ (см. рис. 8.9).

Когда дисперсии D_1 и D_2 неизвестны, то сначала их необходимо оценить по экспериментальным данным. Пусть получены оценки D_1^* и D_2^*, которые незначительно отличаются друг от друга. В этом случае считается, что $D_1 = D_2 = D$, и получают оценку так называемой "объединенной" дисперсии, используя результаты обеих серий опытов:

$$\widetilde{D} = \frac{\sum_{i=1}^{n_1}\left(x_{1,i} - m_1\right)^2 + \sum_{i=1}^{n_2}\left(x_{2,i} - m_2\right)^2}{(n_1 - 1) + (n_2 - 1)} = \frac{(n_1 - 1)D_1^* + (n_2 - 1)D_2^*}{n_1 + n_2 - 2}.$$

Случайная величина $\widetilde{D}$ имеет распределение χ^2 с $k = n_1 + n_2 - 2$ степенями свободы. В качестве статистики для проверки гипотезы о равенстве математических ожиданий выбирается случайная величина

$$T = \frac{m_1^* - m_2^*}{\sqrt{\widetilde{D}(1/n_1 + 1/n_2)}},$$

имеющая распределение Стьюдента с $k = n_1 + n_2 - 2$.

Критическими для проверяемой гипотезы являются значения

$$|T| > t_{k,\alpha/2},$$

где $t_{k,\alpha/2}$ – $\alpha/2$-процентная точка распределения Стьюдента с k степенями свободы.

<u>Пример</u>. В двух сериях опытов над нормальной случайной величиной X объемом соответственно $n_1 = 20$ и $n_2 = 25$ испытаний оценены математические ожидания и получено, что в первой серии $m_1^* = 2{,}13$, а во второй – $m_2^* = 4{,}46$. Дисперсия случайной величины известна: $D = 25$. Нужно при уровне значимости $\alpha = 0{,}1$ проверить нулевую гипотезу $H_0 : m_1 = m_2$ при альтернативной гипотезе $H_1 : m_1 \neq m_2$.

Значение проверочной статистики Z равно

$$Z = \frac{m_1^* - m_2^*}{\sqrt{D(1/n_1 + 1/n_2)}} = \frac{2{,}13 - 4{,}46}{\sqrt{1/20 + 1/25}} = -1{,}55.$$

С использованием табл. 8.4 получаем, что $z_{\alpha/2} = z_{0,05} = 1{,}64$. Так как $|Z| < z_{\alpha/2}$, то нулевая гипотеза принимается, т. е. математические ожидания совпадают с уровнем значимости 0,1.

Если предположить, что дисперсия случайной величины X неизвестна, то ее оценки по результатам обеих серий испытаний оказались равными: $D_1^* = 24{,}63$ и $D_2^* = 31{,}17$. Кроме этого, предполагаем, что полученные оценки мало отличаются одна от другой, и поэтому вычисляем оценку объединенной дисперсии по формуле

$$\widetilde{D} = \frac{(n_1 - 1)D_1^* + (n_2 - 1)D_2^*}{n_1 + n_2 - 2} = \frac{19D_1^* + 24D_2^*}{43} = 28{,}28,$$

а значение проверочной статистики

$$T = \frac{m_1^* - m_2^*}{\sqrt{\widetilde{D}(1/n_1 + 1/n_2)}} = \frac{2{,}13 - 4{,}46}{\sqrt{28{,}28(1/20 + 1/25)}} = -1{,}46.$$

Используя таблицу процентных точек t-распределения Стьюдента (прил. 4), получаем

$$t_{n_1 + n_2 - 2;\, \alpha/2} = t_{43;\, 0,05} = 1{,}693.$$

Так как и в этом случае $|T| < t_{43;\, 0,05}$, то гипотеза о равенстве математических ожиданий принимается.

Проверка гипотезы о равенстве двух дисперсий

Проверка гипотезы о равенстве дисперсий особенно важна в различных экспериментах над случайными величинами, поскольку знание дисперсии позволяет оценить степень рассеивания случайной величины и соответственно судить о точности и надежности результатов. Как и в предыдущей задаче, проводятся две серии опытов по n_1 и n_2 испытаний, а полученные экспериментальные данные имеют нормальное распределение с параметрами (m_1, D_1) и (m_2, D_2) соответственно. По опытным данным найдены несмещенные оценки D_1^* и D_2^*. Ставится задача определить, лежит ли различие между этими оценками в границах возможных случайных изменений, т. е. можно ли оба значения рассматривать как оценки дисперсии одной и той же случайной величины X, имеющей нормальное распределение.

Таким образом, необходимо проверить гипотезу $H_0 : D_1 = D_2$. В качестве альтернативной гипотезы выбирается гипотеза $H_1 : D_1 \neq D_2$.

При сравнении дисперсий в качестве проверочной статистики выбирается случайная величина

$$F = \frac{D_1^*}{D_2^*}. \tag{8.26}$$

Распределение величины F находим из условия, что при справедливости гипотезы H_0 дисперсии равны, т. е. $D_1 = D_2 = D$. Поэтому перепишем выражение (8.26), учитывая (8.19) в виде

$$F = \frac{D_1^*}{D_2^*} = \frac{D_1^*/D}{D_2^*/D} = \frac{\chi^2_{n_1-1}/(n_1-1)}{\chi^2_{n_2-1}/(n_2-1)}. \qquad (8.27)$$

Соотношение (8.27) не зависит от неизвестного параметра D. Случайная величина F имеет распределение Фишера или F-распределение с $k_1 = n_1 - 1$ и $k_2 = n_2 - 1$ степенями свободы.

Критическими для проверяемой гипотезы являются значения:

$$\frac{D_1^*}{D_2^*} < f_{k_1,k_2;1-\alpha/2}\,; \quad \frac{D_1^*}{D_2^*} > f_{k_1,k_2;\alpha/2}\,.$$

Для сокращения объема таблиц процентных точек распределения Фишера за значение D_1^* принимается большая из полученных оценок дисперсии.

<u>**Пример**</u>. В двух сериях опытов над нормальной случайной величиной X объемом соответственно $n_1 = 25$ и $n_2 = 20$ испытаний получены оценки дисперсии: $D_1^* = 35{,}56$; $D_2^* = 17{,}41$. Необходимо проверить гипотезу о равенстве дисперсий при уровне значимости $\alpha = 0{,}1$.

Величина проверочной статистики равна

$$F = \frac{D_1^*}{D_2^*} = \frac{35{,}56}{17{,}42} = 2{,}04\,.$$

По таблице процентных точек F-распределения Фишера находим критическое значение

$$f_{k_1,k_2;\,\alpha/2} = f_{24,19;\,0,05} = 2{,}11\,.$$

Получили, что $F < f_{24,19;\,0,05}$. Поэтому нулевая гипотеза о равенстве дисперсий принимается с уровнем значимости 0,1.

Критерий согласия χ^2

На практике часто возникает задача аппроксимации построенной гистограммы аналитическим выражением, представляющим собой некоторый теоретический закон распределения (плотности вероятности $f(x)$). При этом стремятся к тому, чтобы такая аппроксимация была в опреде-

ленном смысле наилучшей. Заметим, что любая аналитическая функция $f(x)$, с помощью которой аппроксимируется статистическое распределение, должна обладать основными свойствами плотности распределения:

$$\left.\begin{array}{l} f(x) \ge 0 \\ \int\limits_{-\infty}^{\infty} f(x)dx = 1 \end{array}\right\}.$$

Чтобы оценить, насколько хорошо выбранный теоретический закон распределения согласуется с экспериментальными данными, используются так называемые критерии согласия. Таких критериев существует несколько, но наиболее часто применяется критерий согласия χ^2, предложенный Пирсоном.

Критерий согласия χ^2 является непараметрическим критерием проверки статистических гипотез в отличие от ранее рассмотренных критериев, которые являются параметрическими.

Пусть проведено n независимых опытов, в каждом из которых случайная величина X приняла определенное значение. Результаты опытов сведены в k интервалов, и построены статистический ряд, выборочная функция распределения и гистограмма, т. е. экспериментальные данные описываются выборочным законом распределения $P^*(x)$. Необходимо проверить, согласуются ли экспериментальные данные с гипотезой $H_0 : P(x) = P^*(x)$ о том, что случайная величина X имеет выбранный теоретический закон распределения $P(x)$, который может быть задан функцией распределения $F(x)$ или плотностью $f(x)$. Альтернативная гипотеза в этом случае – $H_1 : P(x) \ne P^*(x)$.

Знание теоретического закона распределения позволяет найти теоретические вероятности попадания случайной величины в каждый интервал (x_i, x_{i+1}), $i = \overline{1,k}$: $p_1, p_2, \ldots, p_k$.

Проверка согласованности теоретического и статистического распределений сводится к оценке расхождений между теоретическими вероятностями p_i и полученными частотами p_i^*. В качестве меры расхождения удобно выбрать сумму квадратов отклонений $(p_i^* - p_i)$, взятых с некоторыми "весами" c_i:

$$\Delta = \sum_{i=1}^{k} c_i (p_i^* - p_i)^2 .$$

Смысл коэффициентов c_i ("весов" интервалов) состоит в том, что отклонения, относящиеся к разным интервалам, нельзя считать одинаковыми по значимости. То есть одно и то же по абсолютной величине отклонение $p_i^* - p_i$ может быть мало значимым, если сама вероятность p_i велика, и, наоборот, быть заметным, если эта вероятность мала. Естественно, веса c_i выбирать по величине обратно пропорционально вероятностям p_i.

Пирсон доказал, что если выбрать $c_i = n/p_i$, то при больших n закон распределения случайной величины Δ практически не зависит от функции распределения $F(x)$ и числа испытаний n, а зависит только от числа разрядов k и стремится к распределению χ^2.

Обозначив через χ^2 меру расхождения Δ, получаем

$$\chi^2 = n\sum_{i=1}^{k} \frac{(p_i^* - p_i)^2}{p_i} = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i} . \qquad (8.28)$$

Распределение χ^2 зависит от параметра r, называемого числом "степеней свободы". Для критерия согласия Пирсона $r = k - s$, где k – число интервалов, s – число независимых условий ("связей"), накладываемых на частоты p_i^* и параметры распределения. Так, при аппроксимации нормального распределения $s = 3$, а при исследовании распределения Пуассона $s = 2$.

Схема применения критерия χ^2 для оценки согласованности теоретического и статистического распределения сводится к следующим процедурам (этапам):

1. На основании полученных экспериментальных данных $x_1, x_2, \ldots, x_n$ рассчитываются значения частот p_i^* в каждом из k интервалов.
2. Вычисляются, исходя из теоретического распределения, вероятности p_i попадания значений случайной величины в интервалы (x_i, x_{i+1}).
3. По формуле (8.28) рассчитывается значение χ^2.

4. Определяется число степеней свободы $r = k - s$.
5. По таблице процентных значений распределения χ^2 (прил. 3) определяется вероятность α того, что случайная величина, имеющая распределение χ^2 с r степенями свободы превзойдет полученное на этапе 3 значение χ^2. Если эта вероятность мала, то гипотеза $H_0 : P(x) = P^*(x)$ отбрасывается как неправдоподобная. Если же эта вероятность относительно велика, то гипотезу $H_0 : P(x) = P^*(x)$ можно признать не противоречащей опытным данным.

Пример. Пусть случайная величина X – значения напряжения на выходе генератора шума. Проверим, согласуются ли полученные данные с нормальным законом распределения.

Получено $n = 500$ значений, при этом оценки математического ожидания и среднего квадратичного значения соответственно равны: $m = 0{,}344; \sigma = 3{,}2605$. Для теоретического нормального распределения с полученными параметрами m и σ вычисляем вероятности попадания в каждый из 10 интервалов по формуле

$$p_i = \Phi\left(\frac{x_{i+1} - m}{\sigma}\right) - \Phi\left(\frac{x_i - m}{\sigma}\right),$$

где x_{i+1}, x_i – границы i-го интервала, $\Phi(x)$ – функция Лапласа, таблица значений которой приведена в прил. 2.

Затем создается таблица, содержащая число попаданий m_i в каждый разряд и соответствующие значения np_i для $n = 500$.

Интервалы	-10÷-8	-8÷-6	-6÷-4	-4÷-2	-2÷0	0÷2	2÷4	4÷6	6÷8	8÷10
m_i	5	9	47	85	112	122	69	34	12	5
np_i	2,25	10,32	32,75	72,36	110,94	118,12	88,34	44,84	15,98	3,95

По формуле (8.28) получаем

$$\chi^2 = \sum_{i=1}^{10} \frac{(m_i - np_i)^2}{np_i} = 20{,}20 .$$

Так как число степеней свободы $r = 10 - 3 = 7$, то по таблице процентных точек χ^2 распределения (прил. 3) находим, что $\chi^2_{7;0,01} = 18{,}48$.

Поскольку $\chi^2 > \chi^2_{7;0,01}$ для малой вероятность $\alpha = 0,01$, следует признать: полученные экспериментальные данные противоречат проверяемой гипотезе о том, что случайная величина X распределена по нормальному закону.

При использовании критерия согласия (χ^2 или любого другого) положительный ответ нельзя рассматривать как утвердительный о правильности выбранной гипотезы. Определенным является лишь отрицательный ответ, т. е. если полученная вероятность α мала, то можно отвергнуть выбранную гипотезу $H_0 : P(x) = P^*(x)$ и отбросить ее как явно не согласующуюся с экспериментальными данными. Если же вероятность α велика, то это не может считаться доказательством справедливости гипотезы $H_0 : P(x) = P^*(x)$, а указывает только на то, что гипотеза не противоречит экспериментальным данным.

При использовании критерия согласия χ^2 достаточно большими должны быть не только общее число опытов n (несколько сотен), но и значения m_i в отдельных интервалах. Для всех интервалов должно выполняться условие $m_i \geq 5$. Если для некоторых интервалов это условие нарушается, то соседние интервалы объединяются в один.