29. Проверка гипотезы о значимости коэффициента корреляции

Пусть по выборке объёма n нормально распределённой случайной величины ξ получен отличный от нуля выборочный коэффициент корреляции $r^*_{\xi\eta}$. Нужно при заданном уровне значимости α проверить основную гипотезу H_0: $r_{\xi\eta} = 0$ при альтернативной гипотезе $H_1 : r_{\xi\eta} \neq 0$. Для этой задачи в качестве критерия выбирается случайная величина

$$T = r^*_{\xi\eta} \frac{\sqrt{n-2}}{\sqrt{1 - {r^*_{\xi\eta}}^2}} \,. \tag{29.1}$$

При справедливости основной гипотезы величина T имеет распределение Стьюдента, и критическое значение находится из соответствующей таблицы для двусторонней критической области при уровне значимости α и числе степеней свободы $n-2$.

Пример 29.1. *По выборке объёма $n = 38$, извлеченной из двумерной нормальной генеральной совокупности (ξ, η), найден выборочный коэффициент корреляции $r^*_{\xi\eta} = 0{,}1$. Требуется при уровне значимости $0{,}1$ проверить нулевую гипотезу H_0: $r_{\xi\eta=0}$ о равенстве нулю теоретического коэффициента корреляции при конкурирующей гипотезе $H_1 : r_{\xi\eta} \neq 0$.*

◀ Найдем наблюдаемое значение критерия по формуле (29.1):

$$T_{\text{набл}} = 0{,}1 \cdot \frac{\sqrt{38-2}}{\sqrt{1 - 0{,}1^2}} = \frac{0{,}6}{\sqrt{0{,}99}} \approx \mathbf{0{,}603}.$$

Так как конкурирующая гипотеза $H_1 : r^*_{xy} \neq 0$, то поэтому критическая область двусторонняя. По таблице (приложения 3) критических точек распределения Стьюдента при уровне значимости $\alpha = 0{,}1$ и числу степеней свободы $k = 38 - 2 = 36$ находим $T_{\text{кр}} = t_{\text{кр}}(0{,}1; 36) = \mathbf{1{,}69}$.

Так как $T_{\text{набл}} < T_{\text{кр}}$, то нет оснований отвергнуть нулевую гипотезу о равенстве нулю теоретического коэффициента корреляции. Это означает, что выборочный коэффициент корреляции незначимо отличается от нуля, т.е. ξ и η некоррелированы.

Значение критических точек распределения Стьюдента можно вычислить в Excel. Для этого в любой ячейке таблицы вводим команду =СТЬЮДЕНТ.ОБР.2Х(α; k). ▶

Ответ: $\boxed{\xi \text{ и } \eta \text{ некоррелированы.}}$

29.1. Сравнение математических ожиданий

Пример 29.2. *По двум независимым выборкам, объёмы которых* $n_1 = 36$ *и* $n_2 = 24$*, извлеченным из нормальных генеральных совокупностей* ξ_1 *и* ξ_2*, найдены выборочные средние* $\overline{x}_1 = 20$ *и* $\overline{x}_2 = 22$*. Соответствующие генеральные дисперсии* $D(\xi_1) = 4,\ \ D(\xi_1) = 3$*. Требуется при уровне значимости* 0,05 *проверить нулевую гипотезу* $H_0 : M(\xi) = M(\eta)$ *при конкурирующей гипотезе* $H_1 : M(\xi) \neq M(\eta)$*.*

◀ В поставленной задаче в качестве критерия проверки гипотезы примем величину:

$$Z = \frac{|\overline{x}_1 - \overline{x}_1|}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}}. \tag{29.2}$$

При справедливости основной гипотезы величина Z имеет стандартное нормальное распределение $Z \sim N(0, 1)$. Критическая область двусторонняя с вероятностью попадания $\alpha/2$ в каждую половину области.

Найдем наблюдаемое значение критерия по формуле (29.2):

$$Z_{\text{набл}} = \frac{|20 - 22|}{\sqrt{4/36 + 3/24}} = \frac{12\sqrt{2}}{\sqrt{17}} \approx \mathbf{4{,}116}.$$

Так как конкурирующая гипотеза $M(\xi) \neq M(\eta)$, то критическая область двусторонняя. Правую критическую точку найдем из равенства:

$$\Phi(Z_{\text{кр}}) = (1 - \alpha)/2 = (1 - 0{,}05)/2 = 0{,}495.$$

По таблице функции Лапласа (Приложение 2) находим $Z_{\text{кр}} = \mathbf{2{,}58}$. Так как $|Z_{\text{набл}}| > Z_{\text{кр}}$, то нулевую гипотезу отвергаем. Это означает, что выборочные средние отличаются значимо. ▶

Ответ: **Выборочные средние отличаются значимо.**

Пример 29.3. *Из нормальной генеральной совокупности с известной дисперсией* $D = 9$ *извлечена выборка объёма* $n = 81$*, и по ней найдена выборочная средняя* $\overline{x} = 5{,}35$*. Требуется при уровне значимости* 0,02 *проверить нулевую гипотезу* $H_0 : a = 5$ *при конкурирующей гипотезе* $H_1 : a \neq 5$*.*

◀ Для решения данной задачи в качестве критерия принимается величина

$$U = \frac{\overline{x} - a_0}{\sigma/\sqrt{n}} = \frac{\overline{x} - a_0}{\sigma} \cdot \sqrt{n}. \tag{29.3}$$

Найдем наблюдаемое значение критерия по формуле (29.3):

$$U_{\text{набл}} = \frac{5{,}35 - 5}{3} \cdot \sqrt{81} = \mathbf{1{,}05}.$$

Так как конкурирующая гипотеза $a \neq a_0$, то критическая область двусторонняя. Критическую точку найдем из равенства:

$$\Phi(Z_{\text{кр}}) = (1 - \alpha)/2 = (1 - 0{,}02)/2 = 0{,}49.$$

По таблице функции Лапласа (Приложение 2) находим $Z_{\text{кр}} = \mathbf{2{,}33}$. Так как $U_{\text{набл}} > Z_{\text{кр}}$, то нулевую гипотезу принимаем. Другими словами, выборочная и гипотетическая генеральные средние различаются незначимо. ▶

Ответ: **Нулевая гипотеза принимается.**

29.2. Сравнение вероятности с заданными значением

Пример 29.4. *Автоматизированная линия выпускает детали сложной формы. Для контроля качества деталей выбирается партия из 400 деталей. Если процент бракованный деталей превышает 2%, то автоматизированная линия останавливается для наладки оборудования. В отобранной партии оказалось 10 бракованных. При уровне значимости $\alpha = 0{,}1$ будет ли остановлена автоматизированная линия?*

◀ Нулевая гипотеза $H_0 : p = p_0 = 0{,}02$, а относительная частота брака $m/n = 10/400 = 0{,}025$. Примем в качестве конкурирующей гипотезы $H_1 : p > 0{,}025$ и уровень значимости $\alpha = 0{,}05$. С учётом того, что $q_0 = 1 - 0{,}025 = 0{,}975$, по формуле (29.4) найдем наблюдаемое значение критерия:

$$U = \frac{\dfrac{m}{n} - p_0}{\sqrt{p_0 q_0}} \cdot \sqrt{n}. \tag{29.4}$$

Находим

$$U_{\text{набл}} = \frac{(0{,}025 - 0{,}02) \cdot \sqrt{400}}{\sqrt{0{,}02 \cdot 0{,}975}} \approx \mathbf{0{,}714}.$$

Так как конкурирующая гипотеза состоит в том, что $p > p_0$, то критическая область правосторонняя. Критическую точку найдем из равенства:

$$\Phi(U_{\text{кр}}) = (1 - 2\alpha)/2 = (1 - 2 \cdot 0{,}1)/2 = 0{,}4.$$

По таблице функции Лапласа (Приложение 2) находим $U_{\text{кр}} = \mathbf{1{,}28}$. Так как $U_{\text{набл}} < U_{\text{кр}}$, то нет оснований отвергнуть гипотезу о том, что вероятность брака в партии не превышает 0,02. Следовательно, автоматизированная линия не требует остановки. ►

Ответ: **Автоматизированная линия не требует остановки.**