Параметрические и непараметрические гипотезы. Проверка параметрических гипотез.

Если в гипотезе утверждается что-то о значении какого-то параметра, гипотеза называется **параметрической**. Если гипотеза предполагает что-то, количественно не измеряемое (например, «признак имеет нормальное распределение»), гипотеза называется **непараметрической**.

Параметрические критерии используются в задачах проверки параметрических гипотез и включают в свой расчет показатели распределения, например, средние, дисперсии и т.д.

Это такие известные классические критерии, как критерий Стьюдента, критерий Фишера и др. Они позволяют сравнить основные параметры генеральных совокупностей, а также оценить разности средних и различия в дисперсиях. Критерии способны выявить тенденции изменения признака, оценить взаимодействие двух и более факторов в воздействии на изменения признака

Непараметрические критерии проверки гипотез основаны на операциях с другими данными, в частности, частотами, рангами и т.п. Это - критерий Манна-Уитни, критерий Уилкоксона и многие другие. Непараметрические критерии позволяют решить некоторые важные задачи, связанные с выявлением различий исследуемого признака, с оценкой сдвига значений исследуемого признака, выявлением различий в распределениях.

Алгоритм использования любого критерия включает в себя:

1) выбор соответствующего статистического метода;
2) формулировку нулевой и альтернативной гипотез;
3) выбор значения доверительной вероятности (уровня значимости);
4) вычисление эмпирического значения критерия;
5) нахождение критического значения критерия с помощью таблиц;
6) принятие решения на основании сравнения эмпирического и критического значений критерия.

Проверка гипотез о доле признака

а) Сравнение доли признака с нормативом

Пусть доля некоторого признака p в генеральной совокупности должна быть равной a, т.е. $H_0\!:p = a$. Рассмотрим вначале альтернативную гипотезу $H_1\!:p \neq a$, т.е. двусторонний критерий проверки.

$$\Theta = \frac{m}{n}$$ Частота появления признака в выборке

$$P\left(\left|\frac{m}{n} - a\right| \leq z_{\alpha/2}\sigma\right) \approx \left(z_{\alpha/2}\right) - \left(-z_{\alpha/2}\right) = 2\left(z_{\alpha/2}\right) = 1 - \alpha,$$

где (z) – функция Лапласа. Среднеквадратическое отклонение для биномиального распределения

$\sigma = \sqrt{\frac{a(1-a)}{n}}$,

Критические точки

$$\Theta_1 = a - z_{\alpha/2}\sqrt{\frac{a\,(1-a)}{n}}, \qquad \Theta_2 = a + z_{\alpha/2}\sqrt{\frac{a\,(1-a)}{n}}.$$

$$\Theta_1 < \Theta_{\text{набл}} < \Theta_2$$

Рассмотрим односторонний критерий проверки, в качестве альтернативной гипотезы выдвинем H_1: $p > a$. В этом случае используется z_α – квантиль уровня α, определяемый из уравнения:

$$P\left(\frac{m}{n} > \Theta_2\right) = 0,5 - (z_\alpha) = \alpha, \ \text{где} \ \Theta_2 = a + z_a\sqrt{\frac{a\,(1-a)}{n}}.$$

Гипотеза H_0 отклоняется, если $\frac{m}{n} > \Theta_2$, и принимается, если $\frac{m}{n} \le \Theta_2$.

Пусть проводится проверка соответствия содержания активного вещества в продукции стандарту, который равен 10%, т.е. проверяется нулевая гипотеза $H_0: p = 0,1$, где p – доля брака в продукции. Для контроля произведена выборка из 100 проб, которая дала $\frac{m}{n} = 0,15$. Считать ли гипотезу верной или продукцию следует забраковать как не соответствующую нормативам?

Решение

$H_1 : p \neq a.$ $\alpha = 0,05$ $z_\alpha = 1,65$ $\Theta_2 = 0,1 + 1,65\sqrt{\frac{0,1(1-0,1)}{100}} = 0,149.$

$\frac{m}{n} = 0,15 > \Theta_2$

$H_1 : p \neq a$ $\alpha = 0,05$ $z_{\alpha/2} = 1,96$

$$\Theta_1 = 0,1 - 1,96\sqrt{\frac{0,1\,(1-0,1)}{100}} = 0,041,$$

$$\Theta_2 = 0,1 + 1,96\sqrt{\frac{0,1\,(1-0,1)}{100}} = 0,159.$$

$\frac{m}{n} = 0,15$

б) Сравнение долей признака в двух совокупностях

Пусть $\frac{m_1}{n_1}$ и $\frac{m_2}{n_2}$ – частоты появления одного и того же признака в двух совокупностях из n_1 и n_2

1. Большие выборки.

$$M\left(\frac{m_1}{n_1}\right) = M\left(\frac{m_2}{n_2}\right) = p$$

$$\sigma^2\left(\frac{m_1}{n_1}\right) = \frac{p(1-p)}{n_1}$$

$$\sigma^2\left(\frac{m_2}{n_2}\right) = \frac{p(1-p)}{n_2}$$

$$\Theta = \frac{m_1}{n_1} - \frac{m_2}{n_2}$$

$$M\left(\Theta\right)=M\left(\frac{m_1}{n_1}-\frac{m_2}{n_2}\right)=M\left(\frac{m_1}{n_1}\right)-M\left(\frac{m_2}{n_2}\right)=p-p=0,$$

$$\sigma^2\left(\Theta\right)=\sigma^2\left(\frac{m_1}{n_1}-\frac{m_2}{n_2}\right)=\sigma^2\left(\frac{m_1}{n_1}\right)+\sigma^2\left(\frac{m_2}{n_2}\right)=p\left(1-p\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right).$$

В данном случае необходимо использовать двусторонний критерий, т.е. альтернативную гипотезу выбрать в виде $H_1 : \frac{m_1}{n_1} \neq \frac{m_2}{n_2}$. Задавшись уровнем значимости α, найдем $z_{\alpha/2}$ из уравнения:

$$(P\left(\left|\frac{m}{n} - a\right| \leq z_{\alpha/2}\sigma\right) = \left(z_{\alpha/2}\right) - \left(-z_{\alpha/2}\right) = 2\left(z_{\alpha/2}\right) = 1 - \alpha$$

$$\Theta_1 = -z_{\alpha/2}\sqrt{p\left(1-p\right)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}, \qquad \Theta_2 = z_{\alpha/2}\sqrt{p\left(1-p\right)}\sqrt{\frac{1}{n_1} + \frac{1}{n_2}},$$

$$p = \frac{m_1 + m_2}{n_1 + n_2}.$$

$$\Theta_1 < \Theta_{\text{набл}} < \Theta_2$$

ЗАДАЧА

Пусть число бракованных изделий в экспериментальной партии составило 4 из 100, а в контрольной – 12 из 500.

Оценить с уровнем значимости $\alpha = 0,01$ существенность расхождений долей брака в этих двух партиях.

РЕШЕНИЕ

По уровню значимости $\alpha = 0,01$ находим квантиль $z_{\alpha/2} = 2,58$.

Находим точечную оценку p: $p = \frac{4+12}{100+500} = 0,027$, откуда

$$\sigma = \sqrt{0,027\,(1-0,027)}\sqrt{\frac{1}{100}+\frac{1}{500}} = 0,0177,$$

$$\Theta_1 = -2,58 \cdot 0,0177 = -0,0458, \qquad \Theta_2 = 2,58 \cdot 0,0177 = 0,0458.$$

$$\Theta_{\text{набл}} = \frac{4}{100} - \frac{12}{500} = 0,016,$$

$$\alpha = 0,01$$

2. Малые выборки.

Распределение $\Theta = \frac{m_1}{n_1} - \frac{m_2}{n_2}$

Совокупность	Фактические частоты			Теоретические частоты	
	A	$\overline{A}$	Всего	A	$\overline{A}$
Выборка 1	m_1	m_1	m_1	m_1	m_1
Выборка 2	m_2	m_2	m_2	m_2	m_2
Всего	$m_1 + m_2$	$m_1 + m_2$	$m_1 + m_2$	–	–

Вид соответствующей плотности вероятности

$$p = \frac{m_1 + m_2}{n_1 + n_2}$$

$$\chi^2 = \frac{(m_1 - pn_1)^2}{pn_1} + \frac{[\overline{m_1} - (1-p)\,n_1]^2}{(1-p)\,n_1} + \frac{(m_1 - pn_1)^2}{pn_1} + \frac{[\overline{m_2} - (1-p)\,n_2]^2}{(1-p)\,n_2}.$$

Нулевую гипотезу формулируем в виде «обе совокупности есть выборки из одной генеральной совокупности». В данном случае естественно применить односторонний критерий: определив для данного уровня значимости $alpha$ критическое значение χ_0^2, при $\chi^2 > \chi_0^2$ отклоняем нулевую гипотезу, при $\chi^2 \le \chi_0^2$ считаем расхождения между выборками незначимыми.

ЗАДАЧА

Пусть число бракованных изделий в экспериментальной партии составило 9 из 50, а в контрольной – 7 из 30. Оценить с уровнем значимости $\alpha = 0,05$ существенность расхождений долей брака в этих двух партиях.

РЕШЕНИЕ

Расчет теоретических частот производим по оценке:

$$p = \frac{m_1 + m_2}{n_1 + n_2} = \frac{9 + 7}{16 + 64} = \frac{16}{80} = 0,2.$$

Совокупность	A	$\overline{A}$	Всего	A	$\overline{A}$
Экспериментальная партия	9	41	50	10	40
Контрольная партия	7	23	30	6	24
Всего	16	64	80	–	–

$$\chi^2_{\text{эксп}} = \frac{(9-10)^2}{10} + \frac{(41-40)^2}{40} + \frac{(7-6)^2}{6} + \frac{(23-24)^2}{24} = 0,333.$$

При $\alpha = 0,05$ и $\nu = 1$ $\chi^2_0 = 3,8$ и $\chi^2_{\text{эксп}} < \chi^2_0$, поэтому нет оснований отвергать нулевую гипотезу.

Проверка гипотез о среднем значении

$$z = \frac{\overline{X_B} - a}{\sigma\left(\overline{X_B}\right)}.$$

а) Сравнение среднего значения с нормативом

Так как $M(z) = 0$ и $\sigma(z) = 1$, случайная величина z распределена по стандартному нормальному закону. При справедливости гипотезы H_0 математическое ожидание $M\left(\overline{X_B}\right) = \overline{X} = a$ и отклонения z нуля – следствие случайных погрешностей выборки.

Если в качестве альтернативной гипотезы выдвигается $H_1 : \overline{X} \neq a$, задавшись уровнем значимости α, из уравнения: $P\left(|z| \leq z_{\alpha/2}\right) = 2\left(z_{\alpha/2}\right) = 1 - \alpha$

найдем значение $z_{\alpha/2}$ и критические точки для двусторонней проверки: $z_1 = -z_{\alpha/2}$, $z_2 = z_{\alpha/2}$. Если найденное на основании выборки значение $z_{\text{эксп}}$ будет удовлетворять неравенству $|z_{\text{эксп}}| < z_{\alpha/2}$, то гипотеза H_0 не отклоняется, если $|z_{\text{эксп}}| \geq z_{\alpha/2}$, то H_0 отклоняется.

ЗАДАЧА

Проверяется гипотеза о среднем размере детали H_0:$\overline{X} = 4,8$ мм относительно альтернативной H_1: $\overline{X} \neq 4,8$ мм. Из предыдущих измерений известно, что $\sigma = 0,4$ мм. По результатам выборки объемом $n = 100$ получено $\overline{X_B} = 5,2$ мм.

Примем уровень значимости $\alpha = 0,05$ и найдем соответствующий квантиль $z_{\alpha/2} = 1,96$. Допустимая область значений параметра $z \in (-1,96; 1,96)$. Вычисляя выборочное значение параметра $z_{\text{эксп}} = \frac{5,2-4,8}{0,4} = 1 < 1,96$, видим, что оно попало в допустимую область. Гипотеза H_0 не отклоняется.

Если альтернативная гипотеза задается неравенствами H_1:$\overline{X} > a$ или H_1:$\overline{X} < a$, то производится односторонняя проверка при помощи значения z_α, вычисляемого из уравнения:

$$P\left(|z| < z_\alpha\right) = 0,5 - (z_\alpha) = \alpha$$

и критическая область задается неравенством $z > z_\alpha$ или $z < z_\alpha$.

$$s = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \overline{X_B}\right)^2}.$$

Выборочная несмещенная оценка

$$t = \frac{\overline{X_B} - a}{s}\sqrt{n-1},$$

б) Сравнение средних значений двух совокупностей

Пусть имеются две совокупности, характеризующиеся средними $\overline{X}, \overline{Y}$ и дисперсиями σ_x^2, σ_y^2. Выдвигается гипотеза, что эти средние равны, т.е. $H_0: \overline{X} = \overline{Y}$. Для проверки этой гипотезы из каждой совокупности производится выборка: из первой – объемом n_1, в результате которой получаются $\overline{X_B}$ и s_x^2, из второй – объемом n_2, в результате которой получаются $\overline{Y_B}$ и s_y^2. s_x^2 и s_y^2 – дисперсии выборок.

Для проверки основной гипотезы используем критерий:

$$\Theta = \frac{\overline{X_B} - \overline{Y_B}}{\sqrt{D\left(\overline{X_B} - \overline{Y_B}\right)}}.$$

Так как $M\left(\overline{X_B}\right) = \overline{X}, M\left(\overline{Y_B}\right) = \overline{Y}$, при справедливости нулевой гипотезы H_0 будем иметь $M\left(\Theta\right) = 0$. Используя свойства дисперсии и предполагая выборки (а следовательно, и выборочные средние) независимыми, получим:

$$\sigma^2\left(\overline{X_B} - \overline{Y_B}\right) = \sigma^2\left(\overline{X_B}\right) + \sigma^2\left(\overline{Y_B}\right) = \frac{\sigma_x^2}{n_1} + \frac{\sigma_y^2}{n_2}.$$

Теперь сделаем дополнительное предположение, что дисперсии обеих совокупностей равны, т.е. $\sigma_x^2 = \sigma_y^2 = \sigma^2$. Это предположение нуждается в специальной проверке, о которой речь пойдет в следующем разделе. Если принять это предположение, то

$$\sigma^2\left(\overline{X_B} - \overline{Y_B}\right) = \sigma^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right).$$

Подставляя это выражение в формулу для критерия, получаем:

$$\Theta = \frac{\overline{X_B} - \overline{Y_B}}{\sigma \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}.$$

$$s^2 \approx \frac{\sum_{i=1}^{n_1}\left(x_i - \overline{X_B}\right)^2 + \sum_{i=1}^{n_2}\left(y_i - \overline{Y_B}\right)^2}{n_1 + n_2 - 2} = \frac{n_1 s_x^2 + n_2 s_y^2}{n_1 + n_2 - 2},$$

$$z = \frac{\overline{X_B} - \overline{Y_B}}{\sqrt{n_1 s_x^2 + n_2 s_y^2}} \cdot \frac{\sqrt{n_1 + n_2 - 2}}{\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$