5 Гипотезы и критерии согласия

1. Гипотезы и критерии согласия.
2. Критерий согласия χ^2 - Пирсона.

Пусть $x_1, x_2, ..., x_n$ – выборка из распределения теоретической случайной величины ξ с неизвестной функцией распределения $F(x)$. Проверяется гипотеза $H_0 : F(x) = F_0(x)$, где $F_0(x)$ – заданная функция распределения. Гипотезу такого вида называют *гипотезой согласия (непараметрической гипотезой)*, а критерии для ее проверки – *критериями согласия (непараметрическими критериями)*.

Для проверки гипотез согласия выбирают некоторую меру отклонения

$$\upsilon_n = \upsilon_n(F_n^*(x), F(x)) = \upsilon_n(x_1, x_2, ..., x_n)$$

эмпирической функции распределения $F_n^*(x)$ от теоретической функции распределения $F(x)$. В зависимости от вида υ_n получаются разные критерии согласия.

Критическая область имеет вид

$$K = \{x : \upsilon \geq \upsilon_{\text{крит.}}\},$$

где случайная величина υ имеет предельное распределение υ_n при $n \to \infty$.

По заданному уровню значимости α $\upsilon_{\text{крит}}$ находится из уравнения

$$P\{\upsilon_0 \geq \upsilon_{\text{крит.}}\} = \alpha,$$

где $\upsilon_0 = \upsilon|_{H_0}$ – значение υ в предположении, что верна гипотеза H_0.

Критерий проверки гипотезы H_0 строится следующим образом:

– если $\upsilon_0 \in K$, то гипотеза H_0 отвергается;

– если $\upsilon_0 \notin K$, то гипотеза H_0 согласуется с экспериментальными данными.

Критерий согласия χ^2- Пирсона. При уровне значимости α проверяется гипотеза

$$H_0 : F(x) = F_0(x),$$

где $F_0(x)$ – заданная функция распределения.

1) Числовая ось разбивается на k непересекающихся интервалов $\Delta_1 = (-\infty, z_1)$, $\Delta_2 = [z_1, z_2)$, …, $\Delta_k = [z_{k-1}, +\infty)$. Обозначим $z_0 = -\infty, z_k = +\infty$.

2) Находятся частоты n_i – число выборочных значений, попавших в интервал Δ_i, $i = 1,...,k$.

3) Вычисляются $p_i^0 = F_0(z_i) - F_0(z_{i-1})$ – теоретические вероятности попадания в интервал $\Delta_i = [z_{i-1}, z_i)$, $i = 1,...,k$.

4) Статистика критерия имеет вид

$$\chi^2_{\text{набл}} = \sum_{i=1}^{k} \frac{(n_i - np_i^0)^2}{np_i^0}.$$

5) Для выбранного уровня значимости α по таблице χ^2-распределения находим $\chi^2_{\alpha;l} = \chi^2(100\alpha\%; l)$ – 100α-процентную точку распределения χ^2 с $l = k - r - 1$ степенями свободы, где r – число неизвестных параметров теоретического распределения.

6) Схема принятия решения имеет вид:

– если $\chi^2_{\text{набл}} \geq \chi^2_{\alpha;l}$, то гипотеза H_0 отвергается;

– если $\chi^2_{\text{набл}} < \chi^2_{\alpha;l}$, то говорят, что гипотеза H_0 согласуется с экспериментальными данными.

Пример 5.1 Для выборки X из примера 1.1, используя критерий согласия χ^2-Пирсона, при уровне значимости 0,05 проверить гипотезу о том, что распределение наблюдаемой случайной величины ξ не противоречит нормальному закону с параметрами, вычисленными по выборке.

Поскольку параметры нормального распределения (математическое ожидание a и дисперсия σ^2 неизвестны), выдвигаем гипотезу $H_0 : \xi \sim N(\bar{x}, S^2)$, то есть

$$H_0 : F(x) = \frac{1}{S\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\bar{x})^2}{2S^2}} dt.$$

Числовая ось разбивается на $k = 6$ непересекающихся промежутков. Для вычисления теоретических вероятностей p_i^0 попадания случайной величины ξ в интервал Δ_i, $i = 1,...,k$, можно использовать функцию Лапласа:

$$p_i^0 = F_0(z_i) - F_0(z_{i-1}) = \Phi\left(\frac{z_i - \bar{x}}{S}\right) - \Phi\left(\frac{z_{i-1} - \bar{x}}{S}\right).$$

Вычисления приводятся в таблице 6.

Таблица 6

Интервал	Эмпирические частоты n_i	Теоретические вероятности попадания в интервал $p_i^0 = F_0(z_i) - F_0(z_{i-1})$	Теоретические частоты np_i^0	$\frac{(n_i - np_i^0)^2}{np_i^0}$
$(-\infty; -1,36)$	3	0,11	3,3	0,03
$[-1,36; -0,56)$	6	0,19	5,7	0,02
$[-0,56; 0,24)$	10	0,27	8,1	0,45
$[0,24; 1,04)$	3	0,24	7,2	2,45
$[1,04; 1,84)$	7	0,13	3,9	2,46
$[1,84; +\infty)$	1	0,06	1,8	0,36
	30	1,00		5,76

Таким образом, статистика критерия $\chi^2_{\text{набл}} = \sum_{i=1}^{6} \frac{(n_i - np_i^0)^2}{np_i^0} \approx 5{,}76$.

Число неизвестных параметров теоретического распределения $r = 2$ (математическое ожидание a и дисперсия σ^2).

По заданному уровню значимости $\alpha = 0{,}05$ находим значение процентной точки распределения χ^2 с $k - r - 1 = 3$ степенями свободы

$$\chi^2_{\alpha; k-r-1} = \chi^2(5\,\%; 3) \approx 7{,}8.$$

Поскольку $\chi^2_{\text{набл}} \approx 5{,}76 < \chi^2_{0,05;3} \approx 7{,}8$, делаем вывод, что гипотеза H_0 согласуется с экспериментальными данными.

Пример 5.2 Для выборки X из примера 1.1, используя критерий согласия χ^2- Пирсона, при уровне значимости 0,05 проверить гипотезу о том, что распределение наблюдаемой случайной величины ξ не противоречит равномерному закону.

Поскольку границы интервала предполагаемого равномерного распределения, на котором случайная величина ξ принимает свои значения, нам неизвестны, можно выдвинуть гипотезу о равномерном распределении случайной величины ξ на промежутке $[x_{(1)}; x_{(30)}] = [-2{,}16; 2{,}65]$. Оценки границ интервала предполагаемого равномерного распределения были получены методом максимального правдоподобия.

$$H_0: F(y) = F_0(y) = \begin{cases} 0, & x < -2{,}16, \\ \dfrac{y+2{,}16}{4{,}81}, & -2{,}16 \le x \le 2{,}65, \\ 1, & x > 2{,}65. \end{cases}$$

Вычисления приводятся в таблице 7.

Таблица 7

Интервал	Эмпири-ческие частоты n_i	Теоретические вероятности попадания в интервал $p_i^0 = F_0(z_i) - F_0(z_{i-1})$	Теоретические частоты np_i^0	$\dfrac{(n_i - np_i^0)^2}{np_i^0}$
$[-2{,}16; -1{,}36)$	3	$\frac{1}{6} \approx 0{,}16667$	5	0,8
$[-1{,}36; -0{,}56)$	6	$\frac{1}{6} \approx 0{,}16667$	5	0,2
$[-0{,}56; 0{,}24)$	10	$\frac{1}{6} \approx 0{,}16667$	5	5
$[0{,}24; 1{,}04)$	3	$\frac{1}{6} \approx 0{,}16667$	5	0,8
$[1{,}04; 1{,}84)$	7	$\frac{1}{6} \approx 0{,}16667$	5	0,8
$[1{,}84; 2{,}65]$	1	$\frac{1}{6} \approx 0{,}16667$	5	3,2
	30	1,00		10,8

Статистика критерия $\chi^2_{\text{набл}} = \sum\limits_{i=1}^{6} \dfrac{(n_i - np_i^0)^2}{np_i^0} \approx 10{,}8$.

Число неизвестных параметров теоретического распределения $r = 2$ (границы интервала предполагаемого равномерного распределения). По заданному уровню значимости $\alpha = 0{,}05$ находим значение процентной точки распределения χ^2 с $k - r - 1 = 3$ степенями свободы $\chi^2_{\alpha; k-r-1} = \chi^2(5\%; 3) \approx 7{,}8$.

Поскольку $\chi^2_{\text{набл}} \approx 10{,}8 > \chi^2_{0{,}05;3} \approx 7{,}8$, то гипотеза H_0 отвергается.

Заметим, что оценки границ интервала предполагаемого равномерного распределения могли быть получены и другими методами, например, методом моментов.

Вопросы для самоконтроля

1. Дайте определение критерия согласия.
2. Как найти статистику критерия согласия χ^2-Пирсона?
3. Как построить критическую область в критерии согласия χ^2-Пирсона?