

## 28. Проверка статистических гипотез

**Определение 28.1.** *Статистической называется гипотеза о виде распределения или о значениях его параметров.*

Гипотезы будем обозначать  $H_0, H_1, H_2, \dots$ .

Различают проверяемую или основную гипотезу  $H_0$  и альтернативную или конкурирующую  $H_1$ , которая должна противоречить основной.

Для проверки статистической гипотезы на основании выборки  $x_1, x_2, \dots, x_n$  вычисляют значение критерия, зависящего от наблюдений:

$$T = T(x_1, x_2, \dots, x_n).$$

Для каждого типа статической гипотезы существует своя функция  $T = T(x_1, x_2, \dots, x_n)$ .

Всё множество значений критерия делится на так называемую критическую область, при попадании в которую критерия проверяемая гипотеза отвергается, и область принятия гипотезы.

Критические области бывают двусторонние и односторонние: левосторонняя и правосторонняя.

Если критическая область правосторонняя, т.е.  $(t_{кр2}; +\infty)$ , при выполнении условия  $T_{набл} > t_{кр2}$  делают вывод: проверяемая гипотеза  $H_0$  отвергается с уровнем значимости  $\alpha$  в пользу гипотезы  $H_1$ ; если это условие не выполняется, т.е.  $T_{набл} \leq t_{кр2}$ , делают более осторожный вывод: нет оснований для того, чтобы отвергнуть гипотезу  $H_0$  в пользу гипотезы  $H_1$  с уровнем значимости  $\alpha$ .

Если критическая область левосторонняя, т.е.  $(-\infty; t_{кр1})$ , гипотеза  $H_0$  отвергается при выполнении условия  $T_{набл} < t_{кр1}$ .

В случае двусторонней критической области вида  $(-\infty; t_{кр1}) \cup (t_{кр2}; +\infty)$  гипотеза  $H_0$  отвергается при выполнении условия  $T_{набл} < t_{кр1}$  или  $T_{набл} > t_{кр2}$ .

**Определение 28.2.** *Вероятность того что гипотеза отвергается, хотя на самом деле она верна, называется **уровнем значимости критерия** и обычно обозначается  $\alpha$ .*

**Определение 28.3.** ***Критериями согласия** называют критерии для проверки гипотез о виде закона распределения случайной величины.*

## 28.1. Критерий Пирсона о проверки гипотезы о законе распределения

Для простоты изложения рассмотрим алгоритм только для доказательства гипотезы о нормальном распределении н.с.в.  $\xi$ . Схема расчётов с помощью критерия Пирсона (критерия  $\chi^2$ ) следующая.

- (1) Всё множество наблюдений разбиваем на  $s$  интервалов вида  $(a_{j-1}; a_j]$  с размахом  $h$  и подсчитываем эмпирические частоты — количество наблюдений  $m_j$ , попавших в  $j$ -ый интервал. Относительная частота наблюдений, попавших в  $j$ -ый интервал, равна  $P_j^* = \frac{m_j}{n}$ .
- (2) Строим гистограмму относительных частот  $P_j^*/h$ . На основании вида гистограммы выбираем в качестве предполагаемого какой-то закон распределения изучаемой случайной величины.
- (3) Находим выборочное среднее  $\bar{x}$  и несмещённое среднеквадратичное отклонение  $S^*$ . Определяем теоретические частоты  $m'_j$  для  $j$ -го интервала  $(a_{j-1}; a_j]$ . В нашем случае применяем формулы определения вероятности попадания нормально распределённой н.с.в. в интервал (18.5):

$$m'_j = (F(t_j) - F(t_{j-1})) \cdot n, \quad t_j = \frac{a_j - \bar{x}}{S^*},$$

где  $F(x)$  — теоретическая функция распределения, найденная на этапе 1.

- (4) Вычисляем критерий  $\chi^2_{\text{набл}}$  (критерий Пирсона):

$$\chi^2_{\text{набл}} = \sum_{j=1}^s \frac{(m_j - m'_j)^2}{m'_j}. \quad (28.1)$$

- (5) Определяем число степеней свободы  $k$  случайной величины  $\chi^2$ :

$$k = s - 1 - r, \quad (28.2)$$

где  $r$  — число параметров закона распределения (для нормального закона распределения  $r = 2$ ),  $s$  — число интервалов.

- (6) По заданному уровню значимости  $\alpha$  и числу степеней свободы  $k$  по таблице критических точек распределения  $\chi^2$  (таблица приложения 4) находим критическую точку  $\chi_{\text{кр}}^2(\alpha; k)$ . Если  $\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2(\alpha; k)$  — нет оснований отвергнуть гипотезу о принятом (нормальном) законе распределения. Если  $\chi_{\text{набл}}^2 > \chi_{\text{кр}}^2(\alpha; k)$  — гипотезу отвергают с уровнем значимости  $\alpha$ .

Вернёмся к примеру 27.3.

**Пример 28.1.** Из генеральной совокупности, являющейся нормальной случайной величиной  $\xi$ , извлечена выборка объёма  $n = 50$ , результаты которой сгруппированы с постоянным размахом интервала  $h=4$  и помещены в таблицу.

$i$	1	2	3	4	5	6
$(x_i; x_{i+1}]$	(10; 14]	(14; 18]	(18; 22]	(22; 26]	(26; 30]	(30; 34]
$m_i$	4	11	15	12	6	2

Найти закон распределения случайной величины  $\xi$ .

На рисунках 58, 59 изображены графики гистограммы и эмпирической функции распределения данной случайной величины. Если гистограмму аппроксимировать непрерывной функцией, то можно заметить относительную схожесть полученного графика с графиком функции Гаусса, а график эмпирической функции распределения похож на график функции распределения нормальной случайной величины. При решении примера 18.1 были приведены графики функции плотности и распределения нормально распределённой непрерывной случайной величины, рис. 41 и 42.

Поэтому выдвигаем основную гипотезу  $H_0$ , состоящую в том, что случайная величина  $\xi$  подчиняется нормальному закону распределения.

Для доказательства выдвинутой статистической гипотезы о нормальном распределении случайной величины  $\xi$  применяем критерий Пирсона.

При решении примера 27.3 найдены выборочная средняя  $\bar{x} \approx 20,88$  и несмещённая выборочная дисперсия  $S^{*2} \approx 25,659$ .  $S^* \approx 5,066$ .

В предположении, что случайная величина  $\xi$  подчиняется нормальному закону распределения, по формулам (18.5) найдём вероятности попадания случайной величины  $\xi$  в интервалы  $[x_i; x_{i+1}]$ .

$$P(x_1 \leq \xi < x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right).$$

Умножая полученные вероятности на объём выборки, найдём теоретические частоты  $m_i'$  попадания случайной величины  $\xi$  в те же интервалы.

$$m_i' = n \cdot \left( \Phi\left(\frac{x_{i+1} - \bar{x}}{S^*}\right) - \Phi\left(\frac{x_i - \bar{x}}{S^*}\right) \right).$$

Заполним таблицу, в которой  $t_i = \frac{x_i - \bar{x}}{S^*}$ . Значения функции Лапласа  $\Phi(t_i)$  берем из таблицы приложения 2.

$i$	1	2	3	4	5	6	7
$x_i$	10	14	18	22	26	30	34
$t_i$	-2,15	-1,36	-0,57	0,22	1,01	1,8	2,59
$\Phi(t_i)$	-0,5	-0,3729	-0,2157	0,0871	0,3438	0,4641	0,5
$m_i'$	5,576	7,86	15,14	12,83	6,015	1,555	

Для того чтобы сумма теоретических частот была равна 1, значения функции Лапласа на левой и правой границах выбираем  $\Phi(t_1) = -0,5$  и  $\Phi(t_7) = 0,5$ , соответственно. Таким образом, мы расширяем до бесконечности крайние интервалы.

По формуле  $\chi_{\text{набл}}^2 = \sum_{i=1}^6 \frac{(m_i - m_i')^2}{m_i'}$ , находим  $\chi_{\text{набл}}^2 = 2,206$ .

По таблице приложения 4 для  $\alpha = 0,05$  и  $k = 6 - 1 - 2 = 3$ , находим критическую точку  $\chi_{\text{кр}}^2(0,05; 3) = 7,81$ . Поскольку по заданной выборке и уровне значимости  $\alpha = 0,05$   $\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2(0,05; 3)$ , поэтому нет оснований отвергать гипотезу  $H_0$  о нормальном распределении.

Значение функции  $\chi_{\text{кр}}^2(\alpha; k)$  можно вычислить в Excel. Для этого в любой ячейке таблицы вводим команду =ХИ2.ОБР.ПХ( $\alpha$ ; k).

**Ответ:** На базе полученной выборки делаем вывод, что исследуемая непрерывная случайная величина подчиняется нормальному закону распределения.

Рассмотрим ещё один пример демонстрирующий весь алгоритм Пирсона.

**Пример 28.2.** Лаборатория электролампового завода провела испытания 100 ламп на продолжительность горения и получила следующие результаты (в часах):

812, 817, 828, 833, 841, 820, 822, 825, 826, 824, 826, 825, 817, 826, 834, 818, 842, 826, 837, 827, 821, 835, 823, 824, 815, 833, 830, 824, 816, 828, 822, 826, 827, 822, 837, 816, 825, 810, 823, 831, 826, 814, 838, 831, 824, 812, 827, 839, 828, 836, 815, 836, 817, 828, 823, 832, 819, 826, 818, 820, 811, 828, 810, 822, 836, 816, 829, 821, 833, 821, 829, 823, 832, 823, 831, 826, 832, 827, 829, 826,

836, 821, 838, 818, 822, 819, 823, 828, 826, 820, 825, 828, 822, 835, 824, 825, 820, 829, 825, 824/.

В предположении, что случайная величина  $\xi$  — продолжительность горения лампы является нормальной случайной величиной, провести исследования по схеме примера 27.3. Диапазон наблюдений  $\xi \in [810; 842]$  разбить на 8 интервалов с размахом  $h = 4$ .

◀ Применяем алгоритм метода Пирсона 27.6, используемый в предыдущем примере 27.3. Для решения данной задачи предлагается использовать компьютерные программы.

(1) По заданной выборке заполняем таблицу, в которой  $P_i^* = \frac{m_i}{h \cdot n}$ .

$i$	1	2	3	4	5	6	7	8	9
$x_i$	810	814	818	822	826	830	834	838	842
$m_i$	5	9	13	24	25	10	9	5	
$P_i^*$	0,0125	0,0225	0,0325	0,06	0,0625	0,025	0,0225	0,0125	

Далее строим гистограмму относительных частот, рис.28.2. По графику аппроксимации гистограммы можно отметить, что исследуемая н.с.в. может иметь нормальное распределение.

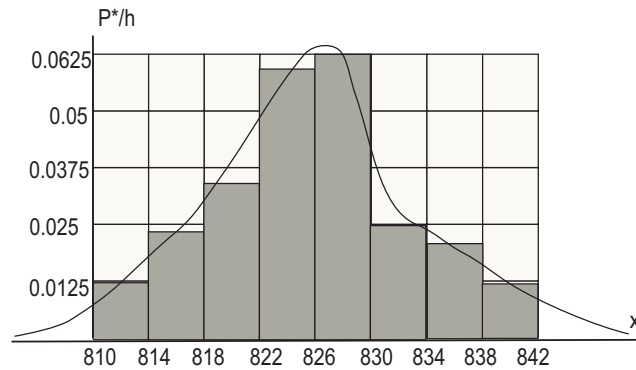


Рисунок 60. Гистограмма примера 28.2

Находим выборочное среднее  $\bar{x}$  и несмещённое выборочное среднеквадратическое отклонение  $S^*$ .

$$\bar{x} = \frac{1}{n} \sum_{i=1}^s m_i \cdot U_i, \quad S^2 = \frac{n}{n-1} \left( \sum_{i=1}^s m_i \cdot U_i^2 - (\bar{x})^2 \right), \quad S^* = \sqrt{S^2},$$

где  $U_i = (x_{i+1} + x_i)/2$  — координаты центра интервалов  $[x_i; x_{i+1})$ .

$$\bar{x} = 825,8, \quad S^2 = 48,3, \quad S^* = 6,95.$$

Теперь находим теоретические частоты.

В предположении, что случайная величина  $\xi$  подчиняется нормальному закону распределения, по формулам (18.5) найдём вероятности попадания случайной величины  $\xi$  в интервалы  $[x_i; x_{i+1}]$ .

$$P(x_1 \leq \xi < x_2) = \Phi\left(\frac{x_2 - a}{\sigma}\right) - \Phi\left(\frac{x_1 - a}{\sigma}\right).$$

Умножая полученные вероятности на объём выборки, найдём теоретические частоты  $m_i'$  попадания случайной величины  $\xi$  в те же интервалы.

$$m_i' = n \cdot \left( \Phi\left(\frac{x_{i+1} - \bar{x}}{S^*}\right) - \Phi\left(\frac{x_i - \bar{x}}{S^*}\right) \right).$$

Заполним таблицу, в которой  $t_i = \frac{x_i - \bar{x}}{S^*}$ . Значения функции Лапласа  $\Phi(t_i)$  берем из таблицы приложения 2. При этом левую и правую границы расширяем до  $\mp 10$ .

$i$	1	2	3	4	5	6	7	8	9
$x_i$	810	814	818	822	826	830	834	838	842
$t_i$	-10	-1,7	-1,13	-0,553	0,023	0,599	1,17	1,75	10
$\Phi(t_i)$	-0,5	-0,456	-0,37	-0,21	0,0092	0,225	0,38	0,46	0,5
$P_i$	0,0442	0,0854	0,161	0,219	0,216	0,155	0,0801	0,0401	
$m_i'$	4,422	8,542	16,07	21,89	21,61	15,46	8,008	4,008	

По формуле  $\chi_{\text{набл}}^2 = \sum_{i=1}^8 \frac{(m_i - m_i')^2}{m_i'}$ , находим  $\chi_{\text{набл}}^2 = 3,714$ .

По таблице приложения 4 для  $\alpha = 0,05$  и  $k = 8 - 1 - 2 = 5$ , находим критическую точку  $\chi_{\text{кр}}^2(0,05; 5) = 11,1$ . Поскольку по заданной выборке и уровне значимости  $\chi_{\text{набл}}^2 < \chi_{\text{кр}}^2(0,05; 5)$ , следовательно нет оснований отвергать гипотезу  $H_0$  о нормальном распределении случайной величины  $\xi$ . ►

**Ответ:** На базе полученной выборки делаем вывод, что исследуемая непрерывная случайная величина подчиняется нормальному закону распределения.

Такие объёмные задачи решать вручную долго и неинтересно, поэтому прилагаем `matha`-программу с краткими комментариями, по которой проводились вычисления.

```
kill(all)$fpprintprec:3$ n:100$ numer:true$ load(distrib)$
x1:[ 812, 817, 828, 833, 841, 820, 822, 825, 826, 824, 826, 825,
    817, 826, 834, 818, 842, 826, 837, 827, 821, 835, 823, 824,
    815, 833, 830, 824, 816, 828, 822, 826, 827, 822, 837, 816,
    825, 810, 823, 831, 826, 814, 838, 831, 824, 812, 827, 839,
    828, 836, 815, 836, 817, 828, 823, 832, 819, 826, 818, 820,
```

```

811, 828, 810, 822, 836, 816, 829, 821, 833, 821, 829, 823,
832, 823, 831, 826, 832, 827, 829, 826, 836, 821, 838, 818 ,
822, 819, 823, 828, 826, 820, 825, 828, 822, 835, 824, 825,
820, 829, 825, 824];
/*{\it Сортируем список x1 в порядке возрастания значений}.*
x2:sort(x1);
/* {\it Разбиваем выборку на s интервалов постоянной длины h (размах)}.*
s:8; h:(x2[n] -x2[1])/s;
/*{\it Координаты границ элементов}.*
a:makelist(x2[1]+(i-1)*h, i, 1, s+1);
/* {\it Координаты середин элементов}.*
U:makelist((a[j]+a[j+1])/2, j, 1, s);
/*{\it Определение частоты наблюдений по интервалам}.*
m:makelist(0, i, 1, s)$ for j:1 while j<=n do(
    k:fix((x2[j] -x2[1])/h)+1,if k>s then k:s,m[k]:m[k]+1);
/* {\it Вывод значений эмпирической частоты наблюдений по
интервалам}.*
m;
/* {\it Контроль объёма выборки}.*
sum(m[i], i, 1, s);
PZ:makelist(m[i]/(n*h),i,1,s);sum(PZ[i],i,1,s);
/* {\it Строим график}.*
wxplot2d(['discrete,makelist([U[j], m[j]], j, 1, s)]],
    [style,[lines, 3, 5]], [gnuplot_preamble,"set grid"],
    [ylabel,""])$
/* Находим выборочное среднее Mx и среднеквадратическое S */
Mx:sum(m[j]*U[j], j, 1, s)/n;
S2:n/(n-1)*(sum(m[j]*U[j]^2, j, 1, s)/n-Mx^2);
S:sqrt(S2);
F(x):=cdf_normal (x, 0, 1) -0.5;
ti:makelist((a[j] -Mx)/S , j, 1, s+1)$ ti[1]:-10$ ti[s+1]:10$ ti;
Pj:makelist(F(ti[j]) , j, 1, s+1) ;
P:makelist(Pj[j+1]-Pj[j],j,1,s);
/*{\it Теоретические частоты}.*
m1:makelist(n*P[j], j, 1, s);
nn:sum(m1[j], j, 1, s);
/* {\it Вычисление}  $\chi^2_{\text{набл}}$ . */
x2nabl:sum((m[j] -m1[j])^2/m1[j], j, 1, s);

```