СРС: Проверка гипотезы о нормальном распределении генеральной совокупности по критерию Пирсона

Литература: Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высшая школа, 2004. – С.251-253.

Краткая справка

Пусть эмпирическое распределение задано в виде последовательности равноотстоящих вариант и соответствующих им частот:

x_i	x_1	x_2	…	x_k
n_i	n_1	n_2	…	n_k

Требуется, используя критерий Пирсона, проверить гипотезу о том, что генеральная совокупность Х распределена нормально при заданном уровне значимости α.

Алгоритм проверки гипотезы:

1. Вычислить выборочную среднюю $\bar{x}_\text{в}$ и выборочное среднее квадратическое отклонение $\sigma_\text{в}$.

2. Вычислить теоретические частоты:

$$n'_i = \frac{nh}{\sigma_\text{в}} \cdot \varphi(u_i),$$

где n – объем выборки (сумма всех частот), h – шаг (разность между двумя соседними вариантами),

$$u_i = \frac{x_i - \bar{x}_\text{в}}{\sigma_\text{в}}, \ \ \varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}.$$

3. Сравнить эмпирические и теоретические частоты с помощью критерия Пирсона. Найти наблюдаемое значение критерия:

$$\chi^2_\text{набл} = \sum \frac{(n_i - n'_i)^2}{n'_i}.$$

По таблице критических точек распределения χ^2, по заданному уровню значимости α и числу степеней свободы $df = k - 3$ (k – число групп выборки) находят критическую точку $\chi^2_\text{кр}(\alpha; df)$ правосторонней критической области[1].

Если $\chi^2_\text{набл} < \chi^2_\text{кр}$, то нет оснований отвергнуть гипотезу о нормальном распределении генеральной совокупности, т.е. эмпирические и теоретические частоты различаются незначимо (случайно).

Если $\chi^2_\text{набл} > \chi^2_\text{кр}$, то гипотезу о нормальном распределении генеральной совокупности отвергают.

[1] При использовании критерия Пирсона число степеней свободы $df = k - 1 - r$, где r – число параметров, оцениваемых по выборке. Нормальное распределение определяется двумя параметрами: математическим ожиданием a и средним квадратическим отклонением σ. Оба эти параметра оцениваются по выборке –выборочным средним и выборочным средним квадратическим отклонением, то $r = 2$.

Решение типового примера

Используя критерий Пирсона, при уровне значимости $\alpha = 0{,}05$ проверить, согласуется ли гипотеза о нормальном распределении генеральной совокупности X с эмпирическим распределением выборки объема $n = 200$:

x_i	4	9	14	19	24	29	34	39	44
n_i	8	29	30	35	23	22	23	21	9

Составим расчетную таблицу:

n_i	x_i	x_i^2	$u_i = \frac{x_i - \bar{x}}{\sigma}$	$\varphi(u_i)$	$n_i' = \frac{nh}{\sigma_в} \cdot \varphi(u_i)$	$n_i - n_i'$	$(n_i - n_i')^2$	$\frac{(n_i - n_i')^2}{n_i'}$
8	4	16	-1,70	0,09	8,48	-0,48	0,23	0,03
29	9	81	-1,25	0,18	16,52	12,48	155,73	5,37
30	14	196	-0,80	0,29	26,23	3,77	14,22	0,47
35	19	361	-0,35	0,38	33,96	1,04	1,09	0,03
23	24	576	0,11	0,40	35,84	-12,84	164,94	7,17
22	29	841	0,56	0,34	30,85	-8,85	78,33	3,56
23	34	1156	1,01	0,24	21,65	1,35	1,82	0,08
21	39	1521	1,46	0,14	12,39	8,61	74,14	3,53
9	44	1936	1,91	0,06	5,78	3,22	10,36	1,15
Сумма								**21,40**
Среднее	**22,83**	**643,48**						

Найдем выборочное среднее

$$\overline{x_в} = \frac{1}{n} \cdot \sum x_i n_i = 22{,}83.$$

Найдем выборочное среднее квадратическое отклонение:

$$\overline{x^2} = \frac{1}{n} \cdot \sum x_i^2 n_i = 643{,}48.$$

$$\sigma_в = \sqrt{\overline{x^2} - \bar{x}^2} = \sqrt{643{,}48 - 22{,}83^2} = 11{,}07.$$

Стандартизируем переменную x_i по формуле $u_i = \frac{x_i - \bar{x}}{\sigma}$ (см. соответствующий столбец расчетной таблицы) и рассчитаем значения дифференциальной функции нормального стандартного распределения по формуле

$$\varphi(u) = \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}}.$$

Замечание: в MS Excel значения функции $\varphi(u)$ можно рассчитать по формуле: =НОРМРАСП(u;0;1;0), где:

u – ссылка на ячейку со значением аргумента функции;

0 – значение математического ожидания для стандартной нормальной случайной величины;

1 – значение среднего квадратического отклонения для стандартной нормальной случайной величины;

0 – логическое значение (ложь) – отказ от вычисления интегральной функции распределения.

Для расчета теоретических частот n_i' определим шаг: $h = x_2 - x_1 = 9 - 4 = 5$. Также учитываем, что объем выборки $n = 200$.

Определяем наблюдаемое значение критерия:

$$\chi^2_{\text{набл}} = \sum \frac{(n_i - n_i')^2}{n_i'} = 21{,}4.$$

По таблице критических точек распределения χ^2, по заданному уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $df = k - 3 = 9 - 3 = 6$ (k – число групп выборки) находит критическую точку $\chi^2_{\text{кр}}(\alpha; df) = \chi^2_{\text{кр}}(0{,}05; 6) = 12{,}6$.

Замечание: в MS Excel критическое значение $\chi^2_{\text{кр}}(0{,}05; 6)$ можно найти по формуле: =ХИ2ОБР(0,05;6).

Т.к. $21{,}4 = \chi^2_{\text{набл}} > \chi^2_{\text{кр}}(0{,}05; 6) = 12{,}6$, то гипотезу о нормальном распределении генеральной совокупности отвергаем.

Ответ: *гипотезу о нормальном распределении генеральной совокупности отвергаем*.