ПРАКТИЧЕСКОЕ ЗАНЯТИЕ 12.

ПРОВЕРКА СТАТИСТИЧЕСКИХ ГИПОТЕЗ.

1. ПОСТРОЕНИЕ ГИСТОГРАММЫ РАСПРЕДЕЛЕНИЯ.

Для наглядного представления о выборке часто используют график, называемый *гистограммой*. Для построения гистограммы интервал, содержащий все элементы выборки, разбивают на l непересекающихся интервалов (как правило, равной длины). Подсчитывают числа n_i попаданий результатов экспериментов в каждый i-й интервал и строят столбиковую диаграмму, откладывая по оси ординат значения средней плотности $n_i/(nh_i)$, где h_i – длина i-го интервала. Площадь каждого столбика равна n_i/n, что соответствует относительной частоте попадания элементов выборки в i-й интервал. Площадь под всей ступенчатой фигурой равна единице.

При увеличении объема выборки и уменьшении интервалов группировки гистограмма приближается к функции плотности генеральной совокупности. Гистограмма является эмпирической функцией плотности, она дает приближенную функцию плотности генеральной совокупности (ее оценку) по случайной выборке.

Задача 1.5. Построить гистограмму для выборки, представленной первыми двумя строками таблицы 1.6.

Таблица 1.6

Интервалы	10–20	20–30	30–40	40–50	50–60	60–70	70–80	80–90
Частоты, n_i	1	4	6	13	10	8	6	2
$n_i/(nh_i)$	0,002	0,008	0,012	0,026	0,020	0,016	0,012	0,004

Решение.

Согласно условию задачи: $n=\sum_{i=1}^{8} n_i = 50$; $h_i = 10$; $i = 1,\dots,8$. Значения средней плотности $n_i/(nh_i)$, необходимые для построения гистограммы, рассчитаны в последней строке таблицы 1.6. Гистограмма представлена на рис. 1.1.

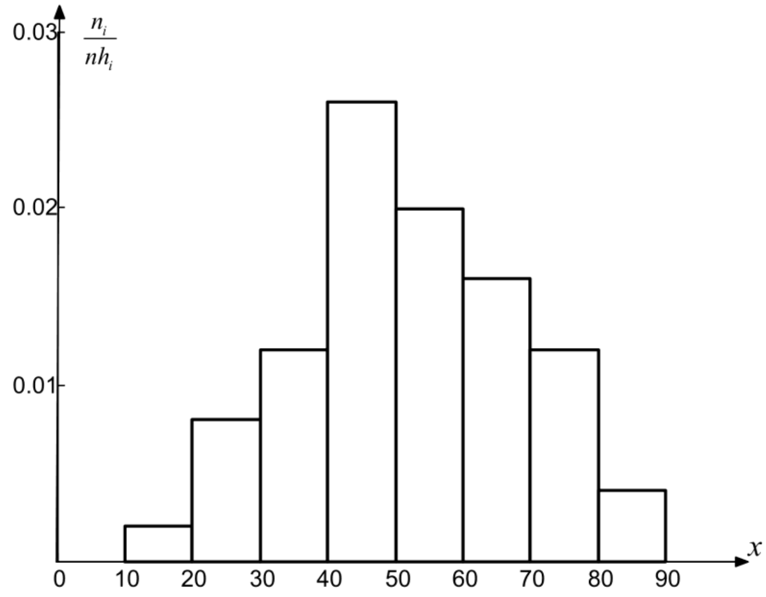

Рис. 1.1. Гистограмма.

2. Проверка гипотезы о виде распределения генеральной совокупности.

Если распределение случайной величины X не известно, можно рассмотреть гипотезу о том, что X имеет функцию распределения $F(x)$. Критерии значимости для проверки таких гипотез называют *критериями согласия*.

Пусть X_1, X_2, ..., X_n – выборка наблюдений случайной величины X. Проверяется гипотеза H_0, утверждающая, что X имеет функцию распределения $F(x)$.

Проверку гипотезы H_0 при помощи критерия χ^2 проводят следующим образом. По выборке находят оценки неизвестных параметров предполагаемого закона распределения случайной величины X. Область возможных значений случайной величины X разбивают на l интервалов. Подсчитывают числа n_i попаданий результатов экспериментов в каждый i-й интервал. Используя предполагаемый закон распределения случайной величины X, находят вероятности p_i того, что значение X принадлежит i-тому интервалу. Затем сравнивают полученные частоты n_i/n с вероятностями p_i. Критерий согласия Пирсона требует принять гипотезу о пригодности проверяемого распределения с уровнем значимости α, если значение *взвешенной суммы квадратов отклонений*:

$$\chi^2 = n\sum_{i=1}^{l}\frac{(n_i/n - p_i)^2}{p_i} = \sum_{i=1}^{l}\frac{(n_i - np_i)^2}{np_i} \tag{1.55}$$

меньше квантиля распределения χ^2-распределения с $k = l - 1$ степенями свободы, то есть

$$\chi^2 < \chi^2_{1-\alpha}(k),$$

в противном случае эту гипотезу отвергают, как противоречащую результатам эксперимента. Если при этом некоторые параметры распределения оценивают по результатам той же выборки, то квантиль χ^2-распределения следует брать для $k = l - 1 - m$ степеней свободы, где m – число оцениваемых параметров.

Задача 1.24. Проведено $n = 100$ экспериментов, их результаты разнесены по восьми интервалам, как указано в табл. 1.9. Проверить пригодность нормального или логарифмически нормального закона распределения ($\alpha = 0{,}05$). Случайная величина Y имеет логарифмически нормальный закон распределения, если логарифм ее значений имеет нормальный закон распределения.

Таблица 1.9

Исходные данные задачи 1.24 и результаты расчета.

$y_{i-1} - y_i$	n_i	t_i	$\Phi(t_i)$	np_i	$n_i - np_i$	$(n_i - np_i)^2/np_i$
1,0÷1,4	12	–1,131	–0,3710	12,90	–0,90	0,06
1,4÷1,8	11	–0,848	–0,3018	6,92	4,08	2,41
1,8÷2,6	24	–0,283	–0,1114	19,04	4,96	1,29
2,6÷3,4	23	0,283	0,1114	22,28	0,72	0,02
3,4÷4,2	12	0,848	0,3018	19,04	–7,04	2,60
4,2÷5,0	6	1,414	0,4213	11,95	–5,95	2,96
5,0÷5,8	7	1,980	0,4761	5,48	1,52	0,42
5,8÷7,4	5	∞	0,5000	2,39	2,61	2,85
Σ	100	–	–	100	0	12,61

Решение.

Проведя первичную обработку результатов эксперимента, определили среднее значение $\overline{Y} = 3$ и эмпирическую дисперсию $S_y^2 = 2$. Вероятности для нормального распределения вычисляем по формуле (1.17): $p_i = \Phi(t_i) - \Phi(t_{i-1})$, где $t_i = (y_i - 3)/\sqrt{2}$, $\Phi(t)$ – интеграл вероятностей (см. табл. Приложения [2]). Результаты расчета приведены в правой части таблицы 1.9. Крайние интервалы для расчета нормального распределения следует считать бесконечными, поэтому полагаем $t_0 = -\infty$, $\Phi(t_0) = -0{,}5$, $t_8 = +\infty$, $\Phi(t_8) = 0{,}5$. Вычисленные значения p_i умножаем на 100. Сумма чисел последнего столбца, равная $\chi_n^2 = 12{,}61$, превосходит квантиль 11,07 для $\mathcal{P} = 0{,}95$ и $k = 8 - 3 = 5$. Поэтому гипотезу о пригодности нормального закона отвергаем .

Для проверки пригодности логарифмически нормального распределения, то есть нормального закона распределения логарифмов величины y, прологарифмируем полученные в эксперименте значения y, обозначив их $\ln y = x$, и проведем весь расчет со значениями величины X. Первичная обработка этих значений дает среднее значение $\overline{X} = 1$ и эмпирическую дисперсию $S_X^2 = 0{,}20$. Здесь полагаем $t_i = (x_i - 1)/0{,}45$, результаты расчета приведены в табл. 1.10.

Таблица 1.10

Результаты расчета (к задаче 1.24)

$x_{i-1} - x_i$	n_i	t_i	$\Phi(t_i)$	np_i	$n_i - np_i$	$(n_i - np_i)^2/np_i$
0÷0,336	12	–1,476	–0,4300	7,00	5,00	3,57
0,336÷0,588	11	–0,916	–0,3202	10,98	0,02	0,00
0,588÷0,956	24	–0,098	–0,0390	28,12	–4,12	0,60
0,956÷1,224	23	0,498	0,1908	22,98	0,02	0,00
1,224÷1,435	12	0,967	0,3332	14,24	–2,24	0,35
1,435÷1,609	6	1,353	0,4120	7,88	–1,88	0,45
1,609÷1,758	7	1,684	0,4539	4,19	2,81	1,88
1,758÷2,001	5	∞	0,5000	4,61	0,39	0,03
Σ	100	–	–	100	0	6,88

Здесь $\chi_n^2 = 6{,}88$ оказывается меньше соответствующей квантили 11,07, и поэтому нет оснований отвергать гипотезу о пригодности логарифмически нормального распределения для описания результатов эксперимента.