ПРИМЕР ОБРАБОТКИ ВЫБОРКИ

Измерен характерный размер X деталей, обрабатываемых на некотором станке. Замерено 60 деталей. Данные замеров приведены в таблице.

№ детали	***Размер***	***№ детали***	***Размер***	***№ детали***	***Размер***
1	72,58	21	72,50	41	72,30
2	72,35	22	72,69	42	72,28
3	72,33	23	72,54	43	72,51
4	72,54	24	72,48	44	72,37
5	72,24	25	72,36	45	72,14
6	72,42	26	72,50	46	72,42
7	72,58	27	72,43	47	72,36
8	72,47	28	72,46	48	72,28
9	72,54	29	72,56	49	72,20
10	72,35	30	72,48	50	72,48
11	72,38	31	72,43	51	72,60
12	72,70	32	72,56	52	72,64
13	72,47	33	72,34	53	72,73
14	72,49	34	72,38	54	72,43
15	72,24	35	72,56	55	72,28
16	72,28	36	72,32	56	72,64
17	72,47	37	72,41	57	72,72
18	71,95	38	72,14	58	72,35
19	72,18	39	72,29	59	72,60
20	72,66	40	72,31	60	72,46

Обработаем результаты этого опыта по следующему плану:

I. Построим статистическое распределение выборки.
II. Вычислим оценки математического ожидания и дисперсии.
III. Построим гистограмму относительных частот, установив статистический (эмпирический) закон распределения и запишем его функцию плотности. С помощью критерия χ^2 (Пирсона) проверим гипотезу о согласии эмпирического закона распределения случайной величины X с нормальным законом распределения (законом Гаусса).
IV. Построим кривую нормального распределения, приняв за параметры кривой найденные оценки математического ожидания и дисперсии (желательно на одном чертеже с гистограммой).
V. Найдём доверительный интервал для оценки математического ожидания и дисперсии.

I. Построение статистического распределения выборки

1. Данную выборку преобразуем в вариационный (интервальный) ряд. Строим его так: диапазон изменения случайной величины X в выборке объема n делим на k

интервалов. Число интервалов определяем по формуле $k = 1 + 3,322 \cdot \lg n$ с округлением до ближайшего целого. В нашем примере $k = 1 + 3,322 \cdot \lg 60 \approx 7$.

Ширину каждого интервала берем одинаковой и равной

$$h = \frac{x_{\max} - x_{\min}}{k} = \frac{72,73 - 71,95}{7} = \frac{0,78}{7} \approx 0,111 \approx 0,12 .$$

Границы интервалов вычисляем по формуле $x_0 = x_{\min}$, $x_i = x_{i-1} + h \left(i = 1, \ldots, k \right)$.

3. Подсчитываем количество элементов n_i, попавших в *i*-й интервал (частота интервала). Если элемент совпадает с границей интервала, то он относится к предыдущему интервалу.

4. Вычисляем относительные частоты интервалов $w_i = \dfrac{n_i}{n} \left(i = 1, 2, \ldots, k \right)$.

Обозначим $\tilde{x}_i = \dfrac{x_{i-1} + x_i}{2}$. Полученные данные заносим в таблицу 1.

Таблица 1

№	Границы классов	n_i	$w_i = \frac{n_i}{n}$	$\tilde{x}_i$	$n_i \tilde{x}_i$	$\tilde{x}_i^2$	$n_i \tilde{x}_i^2$	$\frac{w_i}{h}$
1	[71,95; 72,07]	1	0,02	72,01	72,01	5185,440	5185,440	0,14
2	(72,07; 72,19]	3	0,05	72,13	216,39	5202,747	15608,221	0,42
3	(72,19; 72,31]	10	0,17	72,25	722,50	5220,063	52200,625	1,39
4	(72,31; 72,43]	17	0,28	72,37	1230,29	5237,417	89036,087	2,36
5	(72,43; 72,55]	15	0,25	72,49	1087,35	5254,800	78822,002	2,08
6	(72,55; 72,67]	10	0,17	72,61	726,10	5272,212	52722,121	1,39
7	(72,67; 72,79]	4	0,07	72,73	290,92	5289,653	21158,612	0,56
	Сумма	60	1		4345,56		314733,097	

II. Вычисление оценок математического ожидания и дисперсии

Известно, что оценки математического ожидания и дисперсии вычисляются по формулам $\bar{x} = \dfrac{1}{n} \sum_{i=1}^{n} n_i x_i$, $D = \dfrac{1}{n} \sum_{i=1}^{n} n_i x_i^2 - \left(\bar{x} \right)^2$, $S^2 = \dfrac{n}{n-1} D$, где n_i – частота элемента x_i, n – объём выборки. Для сокращения вычислений элементам выборки, попавшим в *i*-й интервал, припишем значения равные серединам интервалов $\tilde{x}_i = \dfrac{x_{i-1} + x_i}{2} \ \left(i = 1, 2, \ldots, k \right)$.

Для вычисления оценки $\bar{x}$ подсчитаем произведения $n_i \tilde{x}_i$ и внесем в таблицу 1. В нашем примере $\bar{x} = \dfrac{1}{60} \cdot 4345,56 = 72,426 \approx 72,43$ – оценка математического ожидания.

Для вычисления оценки S^2 подсчитаем произведения $n_i \tilde{x}_i^2$ и внесем в таблицу 1.

В нашем примере

$$D=\frac{1}{n}\sum_{i=1}^{k}n_i{x_i}^2-\left(\overline{x}\right)^2=\frac{1}{60}\cdot 314733,097-72,426^2\approx 5245,552-5245,526=0,026.$$

$$S^2=\frac{n}{n-1}\cdot D=\frac{60}{59}\cdot 0,026\approx 0,026.$$

Далее вычисляем оценку среднего квадратического отклонения

$$S=\sqrt{S^2}=\sqrt{0,026}=0,161\approx 0,16.$$

Для сравнения подсчитаем S по «правилу 3σ». Так как для случайной величины, имеющей нормальное распределение, почти все рассеивания укладывается на участке 3σ, то с помощью этого правила можно ориентировочно определить оценку среднего квадратического отклонения случайной величины. Берем максимальное практически возможное отклонение от среднего значения и делим его на три.

В нашем примере $\overline{x}=72,43$, $x_{\max}=72,73$, $x_{\min}=71,95$, $x_{\max}-\overline{x}=72,73-72,43=0,3$, $\overline{x}-x_{\min}=72,43-71,95=0,48$, $\overline{\sigma}=\frac{0,48}{3}=0,16$.

III. Построение гистограммы относительных частот

Для построения гистограммы заполним последний столбец таблицы 1. Строим точки с координатами $\left(\tilde{x}_i,\frac{w_i}{h}\right)$.

Если построенные точки гистограммы соединим плавной линией (на рис. 1 – пунктирная линия), то эта линия будет аналогом плотности распределения случайной величины и, следовательно, по виду гистограммы можно выдвинуть предположение о виде закона распределения случайной величины.

В нашем примере по виду гистограммы (рис. 1) выдвигаем гипотезу о нормальном распределении (или о распределении, близком к нормальному) случайной величины с плотностью

$$f\left(x\right)=\frac{1}{S\sqrt{2\pi}}\cdot e^{-\frac{\left(x-\overline{x}\right)^2}{2S^2}}=\frac{1}{0,16\sqrt{2\pi}}\cdot e^{-\frac{\left(x-72,43\right)^2}{2\cdot 0,16^2}}.$$

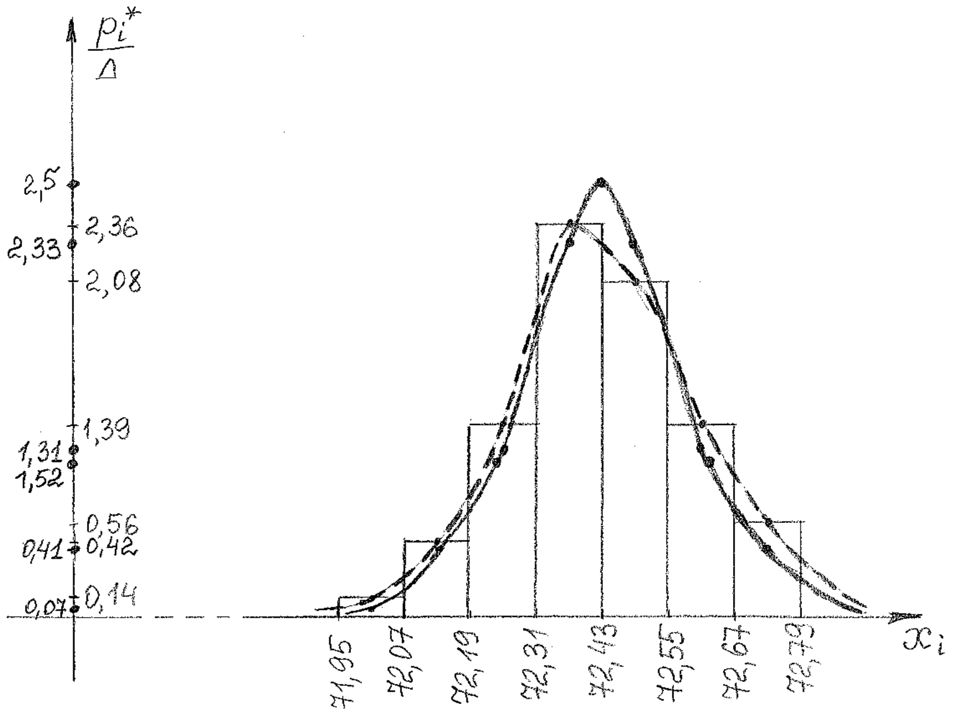

Рис. 1

IV. Проверка гипотезы о нормальном распределении случайной величины

Ввиду ограниченного числа наблюдений статистический закон распределения обычно в какой-то мере отличается от теоретического, даже если предположение о законе распределения сделано правильно. В связи с этим возникает необходимость решить следующую задачу: является ли расхождение между статистическим и теоретическим законами распределения следствием ограниченного числа наблюдений или оно является существенным и связано с тем, что действительное распределение случайной величины не соответствует выдвинутой гипотезе.

Для проверки гипотезы о нормальном распределении рассматриваемой величины заполняем таблицу 2.

1. Производим новую классификацию выборки: объединяем интервалы, для которых $n_i < 5$ в один. В нашем примере после объединения $k = 4$.
2. Левую границу первого интервала примем равной $-\infty$, правую границу последнего примем равной $+\infty$.
3. Вычисляем вероятности p_i попадания варианты в каждый интервале по формуле $p_i = \Phi\left(\dfrac{x_{i+1} - \overline{x}}{\overline{\sigma}}\right) - \Phi\left(\dfrac{x_i - \overline{x}}{\overline{\sigma}}\right)$, где $i = 1, 2, 3, 4$ – номер интервала, $\Phi(x)$ – функция Лапласа.
4. Вычисляем n_i и p_i с учетом объединения интервалов.
5. В нашем примере $\chi^2_{эмп} = \sum\limits_{i=1}^{k} \dfrac{\left(n_i - np_i\right)^2}{np_i} = 0{,}166$.

6. По заданному уровню значимости $\alpha = 0{,}05$ и числу степеней свободы $r = 4 - 2 - 1 = 1$ по таблице находим $\chi^2_{кр} = 3{,}84$.

7. Так как $\chi^2_{эмп} = 0{,}166 < \chi^2_{кр} = 3{,}84$, то нет оснований отвергнуть выдвинутую гипотезу, т.е. гипотезу о том, что случайная величина X распределена по нормальному закону, можно с надежностью $\gamma = 1 - \alpha = 0{,}95$ считать правдоподобной, не противоречащей опытным данным.

Таблица 2

i	Границы классов	$\frac{x_i - \bar{x}}{S}$	$\Phi\left(\frac{x_i - \bar{x}}{S}\right)$	$\frac{x_{i+1} - \bar{x}}{S}$	$\Phi\left(\frac{x_{i+1} - \bar{x}}{S}\right)$	p_i	n_i	np_i	$\frac{(n_i - np_i)^2}{np_i}$
1	(–∞;72,31]	–∞	–0,5000	–0,75	–0,2734	0,2266	14	13,596	0,012
2	(72,31;72,43]	–0,75	–0,2734	0,00	0,0000	0,2734	17	16,404	0,022
3	(72,43;72,55]	0,00	0,0000	0,75	0,2734	0,2734	15	16,404	0,120
4	(72,55; +∞)	0,75	0,2734	+∞	0,5000	0,2266	14	13,596	0,012
	Сумма					1	60		0,166

Итак, теоретическая плотность распределения имеет вид:

$$f(x) = \frac{1}{0{,}16\sqrt{2\pi}} \cdot e^{-\frac{(x-72{,}43)^2}{2\cdot 0{,}16^2}} .$$

Построим график этой функции. Для этого возьмем 7 точек с абсциссами $\tilde{x}_i$ из таблицы 1 и вычислим ординаты этих точек. Составим таблицу 3.

Таблица 3

i	$\tilde{x}_i$	$\tilde{x}_i - 72{,}43$	$(\tilde{x}_i - 72{,}43)^2$	$\frac{(\tilde{x}_i - 72{,}43)^2}{B}$	$e^{-\frac{(\tilde{x}_i - 72{,}43)^2}{B}}$	$A \cdot e^{-\frac{(\tilde{x}_i - 72{,}43)^2}{B}}$
1	72,01	–0,42	0,1760	3,52	0,0296	0,07
2	72,13	–0,30	0,0900	1,80	0,1650	0,41
3	72,25	–0,18	0,0324	0,65	0,5220	1,31
4	72,37	–0,06	0,0036	0,07	0,9320	2,33
5	72,49	0,06	0,0036	0,07	0,9320	2,33
6	72,61	0,18	0,0324	0,65	0,5220	1,31
7	72,73	0,30	0,0900	1,80	0,1650	0,41

Примечания: $A = \frac{1}{S\sqrt{2\pi}} = \frac{1}{0{,}16 \cdot 2{,}51} \approx 2{,}5$, $B = 2S^2 = 2 \cdot 0{,}16^2 \approx 0{,}05$.

Для более точного построения графика найдём максимум и точки перегиба функции $f(x)$: максимум находится в точке $(\overline{x};\, A)$, т.е. (72,43; 2,5), точки перегиба имеют абсциссы $x = \overline{x} \pm S$, т.е. это точки (72,27; 1,52) и (72,59; 1,52).

Строим график $f(x)$ на рис. 1 (сплошная линия).

Сравним теоретическую и эмпирическую плотности распределения случайной величины:

Таблица 4

№	1	2	3	4	5	6	7
$\tilde{x}_i$	72,01	72,13	72,25	72,37	72,49	72,61	72,73
$f(x)$	0,07	0,41	1,31	2,33	2,33	1,31	0,41
$\frac{w_i}{h}$	0,14	0,42	1,39	2,36	2,08	1,39	0,56

Сравнивая значения ординат плотности распределения случайной величины и плотности относительных частот, мы наблюдаем незначительное отклонение этих величин друг от друга, что также свидетельствует о правильности выбора закона распределения.

V. Нахождение доверительных интервалов для математического ожидания и дисперсии

По доверительной вероятности $\gamma = 0{,}95$ и числу степеней свободы $k = n - 1 = 59$ величину $t_\gamma = 2{,}001$, $\varepsilon_\gamma = \dfrac{S \cdot t_\gamma}{\sqrt{n}} = \dfrac{0{,}16 \cdot 2{,}001}{\sqrt{60}} \approx 0{,}04$.

Доверительный интервал для математического ожидания имеет вид:

$$72{,}43 - 0{,}04 < a < 72{,}43 + 0{,}04 \text{ или } 72{,}39 < a < 72{,}47.$$

В нашем примере для $\gamma = 0{,}95$ и $n = 60$ по таблице находим $q = 0{,}188$. Тогда

$$0{,}16 \cdot 0{,}812 < \sigma < 0{,}16 \cdot 1{,}188 \text{ или } 0{,}12 < \sigma < 0{,}19.$$

Аналогично, для других доверительных вероятностей.