

## Полное исследование выборки

Задание. Требуется для решения:

- Построить интервальный ряд распределения, для каждого интервала подсчитать локальные, а также накопленные частоты, построить вариационный ряд.
- Построить полигон и гистограмму
- определить выборочную среднюю, а также низшую и высшую частные средние, моду и медиану, дисперсию и среднее квадратическое отклонение, коэффициент вариации
- проверить при уровне значимости 0,05 гипотезу о нормальном законе распределения соответствующего признака с помощью критериев согласия  $\chi^2$  Пирсона, и  $\omega^2$  Смирнова
- найти точечные и интервальные оценки генеральной средней и среднего квадратического отклонения (при доверительной вероятности  $P=0,95$ )
- найти ошибки выборочных оценок
- произвести анализ всех вычисленных статистических параметров

Задание: произвести обработку данных по среднегодовому удою молока по 11-70 хозяйствам, 80 хозяйств:

23,29,39,36,32,19,33,25,30,32,29,15,14,22,28,38,31,35,23,32,42,43,22,27,27,30,38,,35,31,29,35,32,28,40,36,29,34,31,32,36,30,32,15,35,35,28,28,18,27,39,30,15,14,30,42,38,35,43,39,29,18,19,24,25,23,29,39,36,19,34,34,31,33,28,16,15,23,29,38,32,34,22

РЕШЕНИЕ.

Берем выборку объема  $n = 60$  с 11 по 70 значения:

29
15
14
22
28
38
31
35
23
32
42
43
22
27
27
30
38
35
31
29
35
32
28

Задача скачана с сайта [www.MatBuro.ru](http://www.MatBuro.ru)

©МатБюро - Решение задач по математике, статистике, экономике, программированию

Еще решения математической статистики: [www.matburo.ru/ex\\_subject.php?p=ms](http://www.matburo.ru/ex_subject.php?p=ms)

40
36
29
34
31
32
36
30
32
15
35
35
28
28
18
27
39
30
15
14
30
42
38
35
43
39
29
18
19
24
25
23
29
39
36
19
34

Упорядочиваем по возрастанию:

14
14
15
15
15
18
18
19
19
22

Задача скачана с сайта [www.MatBuro.ru](http://www.MatBuro.ru)

©МатБюро - Решение задач по математике, статистике, экономике, программированию

Еще решения математической статистики: [www.matburo.ru/ex\\_subject.php?p=ms](http://www.matburo.ru/ex_subject.php?p=ms)

22
23
23
24
25
27
27
27
28
28
28
28
29
29
29
29
29
29
30
30
30
30
31
31
31
32
32
32
32
34
34
35
35
35
35
35
35
36
36
36
38
38
38
39
39
39
40
42
42
43
43

Находим  $x_{\min} = 14$ ,  $x_{\max} = 43$ ,  $R = 43 - 14 = 29$ . Берем  $k = 6$  интервалов длины  $h = 5$  каждый. Сдвигаем начало первого интервала на 0,5 и составляем интервальный ряд, подсчитывая число значений, попадающих в каждый интервал.

начало	конец	частота
13,5	18,5	7
18,5	23,5	6
23,5	28,5	9
28,5	33,5	16
33,5	38,5	14
38,5	43,5	8
<b>Сумма</b>		<b>60</b>

Для каждого интервала подсчитаем локальные, а также накопленные частоты.

начало	конец	частота	накопл. частота
13,5	18,5	7	7
18,5	23,5	6	13
23,5	28,5	9	22
28,5	33,5	16	38
33,5	38,5	14	52
38,5	43,5	8	60

Построим вариационный ряд, выбирая в качестве вариант середины интервалов:

$x_i$	$n_i$
16	7
21	6
26	9
31	16
36	14
41	8
<b>Сумма</b>	<b>60</b>

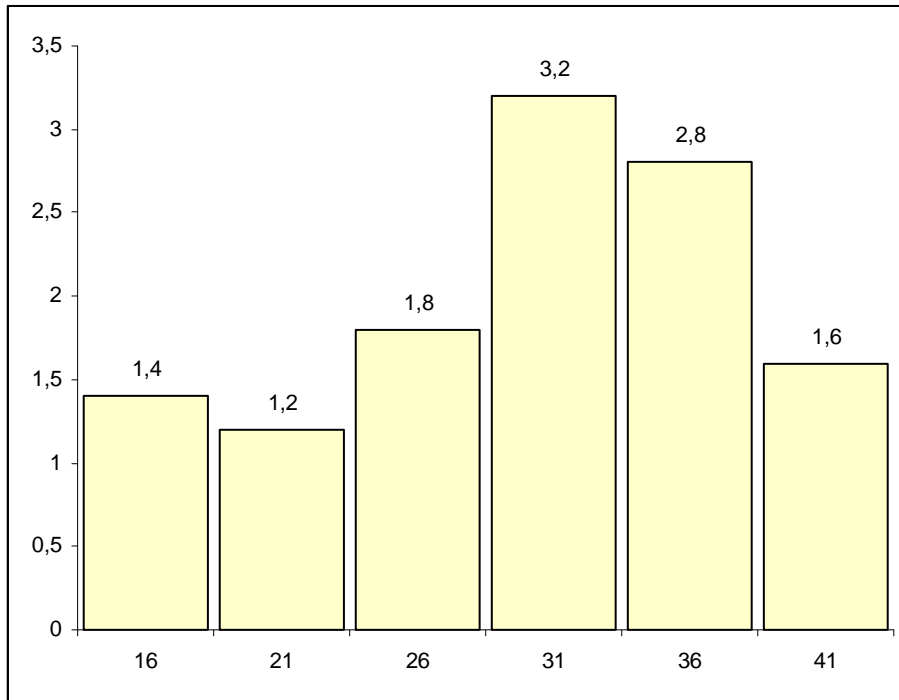
Построим полигон и гистограмму частот. Вычислим для этого плотности частот

$$w_i = \frac{n_i}{h} = \frac{n_i}{5}.$$

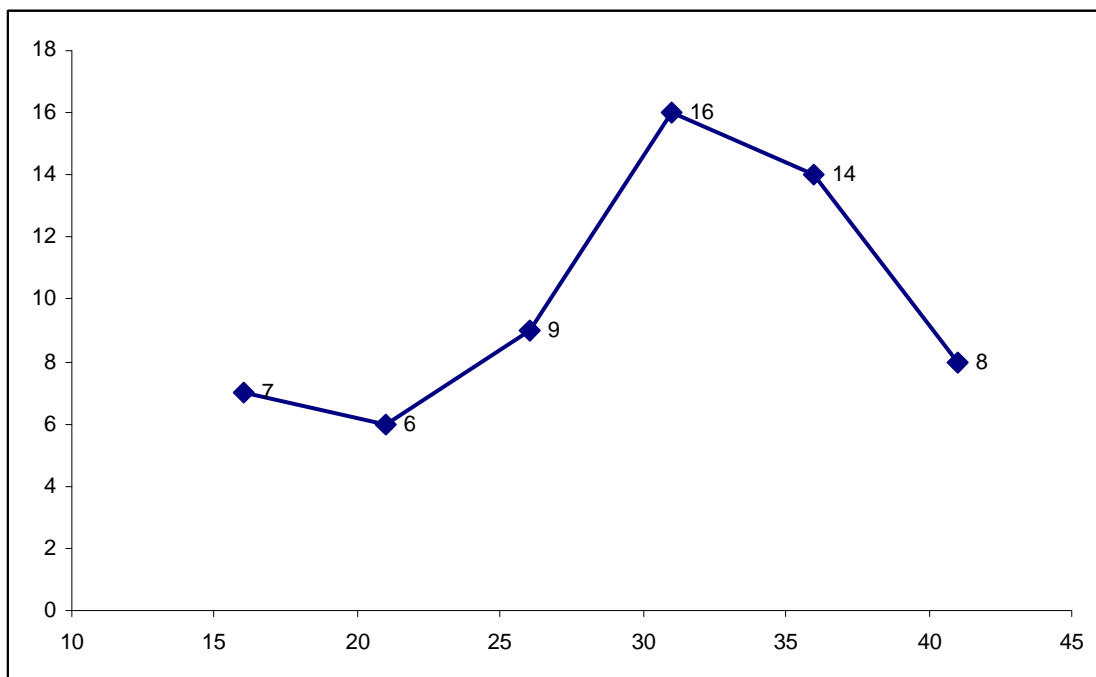
начало	конец	частота	плотность частоты
13,5	18,5	7	1,4
18,5	23,5	6	1,2
23,5	28,5	9	1,8

28,5	33,5	16	3,2
33,5	38,5	14	2,8
38,5	43,5	8	1,6

Гистограмма частот:



Полигон частот:



Определим числовые характеристики.

Выборочная средняя  $\bar{x} = \frac{1}{n} \sum x_i n_i = \frac{1}{60} 1800 = 30$ . Характеризует среднегодовой удой в среднем по выбранным 60 хозяйствам.

Выборочная дисперсия  $\bar{D} = \frac{1}{n} \sum (x_i - \bar{x})^2 n_i = \frac{1}{60} 3490 \approx 58,167$ .

Выборочное среднеквадратическое отклонение  $\bar{\sigma} = \sqrt{\bar{D}} = 7,627$ .

Характеризует разброс показателя среднегодового уdoa в выборке из 60 хозяйств.

Расчетная таблица:

$x_i$	$n_i$	$x_i n_i$	$(x_i - \bar{x})^2 n_i$
16	7	112	1372
21	6	126	486
26	9	234	144
31	16	496	16
36	14	504	504
41	8	328	968
<b>Сумма</b>	<b>60</b>	<b>1800</b>	<b>3490</b>

Моду подсчитаем по формуле (модальный интервал 28,5-33,5)

$$M_o = x_{M_o} + i_{M_o} \frac{f_{M_o} - f_{M_o-1}}{2f_{M_o} - f_{M_o-1} - f_{M_o+1}} = 28,5 + 5 \frac{16 - 9}{2 \cdot 16 - 9 - 14} \approx 32,389.$$

Характеризует наиболее часто встречающийся (вероятный) среднегодовой удой.

( $x_{M_o}$  - нижняя граница модального интервала,

$f_{M_o}$  - частота модального интервала,

$f_{M_o-1}$  - частота интервала, предшествующего модальному,

$f_{M_o+1}$  - частота интервала, следующего за модальным.

$i_{M_o}$  - длина модального интервала)

Медиану подсчитаем по формуле (медианный интервал 28,5-33,5).

Характеризует показатель среднегодового уdoa, находящийся в середине вариационного ряда (выборки).

$$M_e = x_{M_e} + i_{M_e} \frac{\frac{1}{2}n - S_{M_e-1}}{f_{M_e}} = 28,5 + 5 \frac{30 - 22}{16} = 31$$

( $x_{M_e}$  - нижняя граница медианного интервала,

$S_{M_e-1}$  - накопленная частота интервала, предшествующего медианному,

$i_{M_e}$  - величина интервала,

$f_{M_e}$  - частота медианного интервала.)

Коэффициент вариации  $V = \frac{\bar{\sigma}}{\bar{x}} 100\% = \frac{7,627}{30} 100\% \approx 25,42\%$ . Так как он меньше 30%, выборку можно считать однородной.

Проверим при уровне значимости 0,05 гипотезу о нормальном законе распределения соответствующего признака с помощью критерия согласия  $\chi^2$  Пирсона. Пронормируем случайную величину  $X$ , то есть перейдем к величине  $Z = \frac{x - \bar{x}}{\bar{\sigma}}$ , вычислим концы интервалов по формулам  $z_i = \frac{x_i - \bar{x}}{\bar{\sigma}}$ ,  $z_{i+1} = \frac{x_{i+1} - \bar{x}}{\bar{\sigma}}$ . Вычислим теоретические (выравнивающие частоты)  $n_i' = nP_i$ , где  $n = 60$ ,  $P_i = \Phi(z_{i+1}) - \Phi(z_i)$  - вероятность попадания в интервал  $(z_i, z_{i+1})$ ,  $\Phi(z)$  - функция Лапласа. Для нахождения значений составим расчетную таблицу:

$x_i$	$x_{i+1}$	$n_i$	$z_i$	$z_{i+1}$	$\Phi(z_i)$	$\Phi(z_{i+1})$	$P_i$	$n_i'$	$\frac{(n_i - n_i')^2}{n_i'}$
13,5	18,5	7	$-\infty$	-1,508	-0,500	-0,434	0,066	3,948	2,360
18,5	23,5	6	-1,508	-0,852	-0,434	-0,303	0,131	7,874	0,446
23,5	28,5	9	-0,852	-0,197	-0,303	-0,078	0,225	13,500	1,500
28,5	33,5	16	-0,197	0,459	-0,078	0,177	0,255	15,289	0,033
33,5	38,5	14	0,459	1,115	0,177	0,367	0,191	11,437	0,574
38,5	43,5	8	1,115	$+\infty$	0,367	0,500	0,133	7,952	0,000
<b>Сумма</b>		<b>60</b>						<b>60,000</b>	<b>4,914</b>

Сравним эмпирические и теоретические частоты, используя критерий Пирсона:

$$\chi^2 = \sum \frac{(n_i - n_i')^2}{n_i'} = 4,914.$$

По таблице критических точек распределения  $\chi^2$  по уровню значимости  $\alpha = 0,05$  и числу степеней свободы  $k = 6 - 3 = 3$ , находим  $\chi^2_{кр.} = 7,8$ . Так как  $\chi^2_{набл.} = 4,914 < \chi^2_{кр.} = 7,8$ , то можно принять гипотезу о нормальном распределении данной величины по критерию Пирсона.

То есть можно считать, что показатель среднегодового уdoa распределен нормально с параметрами  $a = 30, \sigma = 7,627$ .

Проверим при уровне значимости 0,05 гипотезу о нормальном законе распределения соответствующего признака с помощью критерия согласия  $\omega^2$  Смирнова. Вычислим значение статистики критерия по формуле:

$n\omega^2 = \frac{1}{12n} + \sum_{i=1}^n \left\{ F(x_i) - \frac{2i-1}{2n} \right\}^2$ , где будем считать что  $F(x)$  - функция распределения для нормального закона с параметрами  $a = 30, \sigma = 7,627$ .

Расчетная таблица:

$i$	$x_i$	$\left\{ F(x_i) - \frac{2i-1}{2n} \right\}^2$
1	14	9,26E-05
2	14	4,96E-05
3	15	0,000291
4	15	0,001138
5	15	0,00254
6	18	0,001146
7	18	0,002553
8	19	0,002539
9	19	0,004497
10	22	0,000126
11	22	0,000778
12	23	0,000152
13	23	0,00084
14	24	8,6E-05
15	25	0,000207
16	27	0,007867
17	27	0,005188
18	27	0,003065
19	28	0,007786
20	28	0,005122
21	28	0,003014
22	28	0,001462
23	29	0,005306
24	29	0,003156
25	29	0,001561
26	29	0,000522
27	29	3,81E-05
28	30	0,001736
29	30	0,000625
30	30	6,94E-05
31	30	6,94E-05
32	31	0,000738
33	31	0,00011
34	31	3,81E-05
35	32	0,000808
36	32	0,000138
37	32	2,4E-05
38	32	0,000465
39	34	0,003406
40	34	0,001738
41	35	0,004755
42	35	0,002734
43	35	0,001269
44	35	0,000359
45	35	5,24E-06



46	35	0,000207
47	36	8,6E-05
48	36	5,47E-05
49	36	0,000579
50	38	0,000778
51	38	0,000126
52	38	2,95E-05
53	39	3,62E-05
54	39	0,000114
55	39	0,000746
56	40	0,000396
57	42	2,73E-07
58	42	0,000261
59	43	0,000366
60	43	0,001282

**Сумма 0,08527**

Получаем:  $n\omega^2 = \frac{1}{12 \cdot 60} + 0,08527 \approx 0,087$ . Критическое значение для вероятности 0,95 будет  $n\omega_{\text{крит}}^2 = 0,126$ . Так как наблюдаемое значение меньше критического, можно принять гипотезу о нормальном распределении.

Найдем точечные и интервальные оценки генеральной средней при доверительной вероятности  $P = 0,95$ .

Точечная оценка: выборочная средняя  $\bar{x} = \frac{1}{n} \sum x_i n_i = \frac{1}{60} 1800 = 30$ .

Используем формулу для интервальной оценки:  $\bar{x} - t_\gamma \frac{s}{\sqrt{n}} < a < \bar{x} + t_\gamma \frac{s}{\sqrt{n}}$ , где  $\bar{x} = 30$ ,  $s = 7,691$  (см. ниже),  $n = 60$ ,  $t_\gamma$  определяется из таблицы Стьюдента  $t_\gamma(60; 0,95) = 2,001$ .

Получаем:

$$30 - 2,001 \frac{7,691}{\sqrt{60}} < a < 30 + 2,001 \frac{7,691}{\sqrt{60}},$$

$$28,013 < a < 31,987.$$

Найдем точечные и интервальные оценки среднего квадратичного отклонения при доверительной вероятности  $P = 0,95$ .

Точечная оценка (смещенная): выборочное среднеквадратическое отклонение  $\bar{\sigma} = \sqrt{\bar{D}} = 7,627$ .

Точечная оценка (несмещенная): исправленное выборочное среднеквадратическое отклонение  $s = \sqrt{\frac{n}{n-1} \bar{D}} = 7,691$ .

Используем формулу для интервальной оценки:  $s(1-q) < \sigma < s(1+q)$ , где  $q$  определяется из таблицы по заданным  $n = 60$  и  $\gamma = 0,95$ ,  $q = 0,188$ .

Получаем после подстановки известных данных:

$$7,691(1-0,188) < \sigma < 7,691(1+0,188)$$

$$6,245 < \sigma < 9,137$$

Найдем ошибки выборочных оценок:

$$\text{Для математического ожидания } \Delta_a = 2,001 \frac{7,691}{\sqrt{60}} \approx 1,987.$$

$$\text{Для среднеквадратического отклонения } \Delta_\sigma = 0,188 \cdot 7,691 \approx 1,446.$$