Лекция 10. Числовые характеристики вариационных рядов

На лекции рассматриваются вопросы:

1. Числовые характеристики вариационных рядов: размах вариации, средняя арифметическая (выборочная средняя), выборочная дисперсия, выборочное среднее квадратическое отклонение, коэффициент вариации, мода, медиана.

1. Числовые характеристики вариационного ряда

Пусть для изучения генеральной совокупности количественного признака X извлечена выборка объема n.

1. Размахом вариации называется число

$$R = x_{\max} - x_{\min}.$$

При этом для интервального вариационного ряда $x_{\max}$ — правый конец последнего частичного интервала, а $x_{\min}$ — левый конец первого частичного интервала.

2. Средняя арифметическая (выборочная средняя).

$$\overline{x_в} = \frac{\sum_{i=1}^{k} x_i n_i}{n},$$

где x_i — i-ая варианта признака для дискретного вариационного ряда или середина i-го интервала для интервального вариационного ряда,
n_i — частота i-го значения признака,
k — число значений признака;
или

$$\overline{x_в} = \sum_{i=1}^{k} x_i w_i ,$$

где w_i — относительная частота i-го значения признака.

Выборочная средняя характеризует среднее значение признака X генеральной совокупности.

3. Выборочная дисперсия.

$$D_в = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x_в}\right)^2 n_i}{n}$$

или

$$D_в = \sum_{i=1}^{k} \left(x_i - \overline{x_в}\right)^2 w_i .$$

Также применяют формулы:

$$D_в = \frac{\sum_{i=1}^{k} x_i^{\,2} n_i}{n} - \left(\overline{x_в}\right)^2$$

или

$$D_в = \sum_{i=1}^{k} x_i^{\,2} w_i - \left(\overline{x_в}\right)^2 .$$

Дисперсия характеризует отклонение от средней в квадратных единицах измерения признака.

4. Выборочное среднее квадратическое отклонение.

$$\sigma_в = \sqrt{D_в} .$$

5. Коэффициент вариации.

$$V = \frac{\sigma_в}{\left|\overline{x_в}\right|} \cdot 100\% \quad (\overline{x_в} \neq 0).$$

Принято считать, что если коэффициент вариации больше 35%, то изучаемая статистическая совокупность является неоднородной, и колеблемость признака высока. Следовательно, использование выборочной средней для ее характеристики неверно. В таком случае следует использовать моду и медиану для характеристики наиболее типичного значения варианты признака.

6. Модой вариационного ряда *Мо* называется то из значений x_1, x_2, $x_3, \ldots,$ x_n, которому соответствует наибольшая частота.

7. Медиана *Ме* — это значение варианты, которая является серединой вариационного ряда.

Если количество вариант четное, то медиана вычисляется как среднее двух вариант, находящихся в середине множества.

Пример:

В результате исследования построена таблица частот:

x_i	2	3	4	5
n_i	5	10	7	3

1) Вычислим размах вариации:

$$R = 5 - 2 = 3 .$$

2) Используя таблицу частот, найдем выборочную среднюю по формуле:

$$\overline{x_в} = \frac{\sum_{i=1}^{k} x_i n_i}{n}.$$

x_i	n_i	$x_i n_i$
2	5	10
3	10	30
4	7	28
5	3	15
$\sum$	25	83

Объем выборки $n = 25$.

$$\overline{x_в} = \frac{2\cdot 5 + 3\cdot 10 + 4\cdot 7 + 5\cdot 3}{25} = \frac{83}{25} = 3{,}32.$$

Итак, средний балл успеваемости равен 3,32.

3) Используя таблицу частот, найдем выборочную дисперсию по формуле:

$$D_в = \frac{\sum_{i=1}^{k} \left(x_i - \overline{x_в}\right)^2 n_i}{n}.$$

x_i	n_i	$\left(x_i - \overline{x_в}\right)^2$	$\left(x_i - \overline{x_в}\right)^2 n_i$
2	5	1,7424	8,712
3	10	0,1024	1,024
4	7	0,4624	3,2368
5	3	2,8224	8,4672
$\sum$	25		21,44

Объем выборки $n = 25$, выборочная средняя $\overline{x_в} = 3{,}32$.

Тогда выборочная дисперсия:

$$D_в = \frac{(2-3{,}32)^2\cdot 5 + (3-3{,}32)^2\cdot 10 + (4-3{,}32)^2\cdot 7 + (5-3{,}32)^2\cdot 3}{25} =$$

$$= \frac{8{,}712 + 1{,}024 + 3{,}2368 + 8{,}4672}{25} =$$

$$= \frac{21{,}44}{25} = 0{,}8576.$$

Вычислим выборочную дисперсию 2-м способом по формуле:

$$D_в = \frac{\sum_{i=1}^{k} x_i^{\,2} n_i}{n} - \left(\overline{x_в}\right)^2.$$

x_i	n_i	x_i^2	$x_i^2 n_i$
2	5	4	20
3	10	9	90
4	7	16	112
5	3	25	75
$\sum$	25		297

$$D_в = \frac{2^2 \cdot 5 + 3^2 \cdot 10 + 4^2 \cdot 7 + 5^2 \cdot 3}{25} - 3{,}32^2 =$$
$$= \frac{297}{25} - 11{,}0224 = 0{,}8576 .$$

4) Вычислим выборочное среднее квадратическое отклонение:

$$\sigma_в = \sqrt{D_в} = \sqrt{0{,}8576} \approx 0{,}9 .$$

5) Вычислим коэффициент вариации:

$$V = \frac{\sigma_в}{\left|\overline{x_в}\right|} \cdot 100\% \approx \frac{0{,}9}{3{,}32} \cdot 100\% \approx 27{,}11\% .$$

6) Так как варианта $x_2 = 3$ имеет наибольшую частоту $n_2 = 10$, то мода $Mo = 3$.

7) Количество вариант четное, следовательно, медиана:

$$Me = \frac{3+4}{2} = 3{,}5 .$$

Пример:

При исследовании получили интервальный вариационный ряд. Составили таблицу частот и относительных частот:

$[x_{i-1}; x_i)$	$[22;24)$	$[24;26)$	$[26;28)$	$[28;30)$	$[30;32)$	$[32;34]$
n_i	2	12	34	40	10	2
w_i	0,02	0,12	0,34	0,40	0,10	0,20

1) Вычислим размах вариации:

$$R = 34 - 22 = 12 .$$

2) Используя интервальный вариационный ряд, найдем выборочную среднюю по формуле:

$$\overline{x_в} = \sum_{i=1}^{k} x_i w_i .$$

$[x_{i-1};x_i)$	w_i	x_i	$x_i w_i$
$[22;24)$	0,02	23	0,46
$[24;26)$	0,12	25	3,00
$[26;28)$	0,34	27	9,18
$[28;30)$	0,40	29	11,60
$[30;32)$	0,10	31	3,10
$[32;34]$	0,02	33	0,66
$\sum$			28

$$\overline{x_в} = \sum_{i=1}^{k} x_i w_i = 23\cdot 0{,}02 + 25\cdot 0{,}12 + 27\cdot 0{,}34 + 29\cdot 0{,}4 + 31\cdot 0{,}1 + 33\cdot 0{,}2 =$$

$$= 0{,}46 + 3 + 9{,}18 + 11{,}6 + 3{,}1 + 0{,}66 = 28.$$

Итак, среднее время обработки детали равно 28 мин.

3) Используя интервальный вариационный ряд, найдем выборочную дисперсию по формуле:

$$D_в = \sum_{i=1}^{k} \left(x_i - \overline{x_в}\right)^2 w_i .$$

$[x_{i-1};x_i)$	w_i	x_i	$\left(x_i - \overline{x_в}\right)^2$	$\left(x_i - \overline{x_в}\right)^2 w_i$
$[22;24)$	0,02	23	25	0,50
$[24;26)$	0,12	25	9	1,08
$[26;28)$	0,34	27	1	0,34
$[28;30)$	0,40	29	1	0,40
$[30;32)$	0,10	31	9	0,90
$[32;34]$	0,02	33	25	0,50
$\sum$				3,72

Выборочная средняя $\overline{x_в} = 28$.

$$D_в = (23-28)^2\cdot 0{,}02 + (25-28)^2\cdot 0{,}12 + (27-28)^2\cdot 0{,}34 + (29-28)^2\cdot 0{,}4 +$$

$$+ (31-28)^2\cdot 0{,}1 + (33-28)^2\cdot 0{,}02 =$$

$$= 0{,}5 + 1{,}08 + 0{,}34 + 0{,}4 + 0{,}9 + 0{,}5 = 3{,}72.$$

5) Вычислим выборочное среднее квадратическое отклонение:

$$\sigma_в = \sqrt{D_в} = \sqrt{3{,}72} \approx 1{,}93.$$

6) Вычислим коэффициент вариации:

$$V = \frac{\sigma_в}{\left|\overline{x_в}\right|}\cdot 100\% \approx \frac{1{,}93}{28}\cdot 100\% \approx 6{,}89\%.$$

7) Так как частота попадания признака в интервал $[28;30)$ наибольшая, равная 40, то мода равна середине соответствующего интервала, то есть $Mo = 29$.

8) Количество частичных интервалов четное, следовательно, медиана:

$$Me = \frac{\frac{26+28}{2} + \frac{28+30}{2}}{2} = 28.$$