ТЕМА: Статистическое изучение вариаций

1. *Показатели вариации*

2. *Виды дисперсии*

Пример, демонстрирующий наличие совокупностей, у которых средние величины каких – либо признаков совершенно одинаковы (по уровню), а отклонения от этих средних различны.

Пример 1						**Пример 2**					
x	f	xf	$x-x_{ср}$	$(x-x_{ср})^2$	$(x-x_{ср})^2 f$	x	f	xf	$x-x_{ср}$	$(x-x_{ср})^2$	$(x-x_{ср})^2 f$
2	1	2	-3	9	9	2	30	60	-3	9	270
3	5	15	-2	4	20	3	20	60	-2	4	80
4	30	120	-1	1	30	4	10	40	-1	1	10
5	60	300	0	0	0	5	50	250	0	0	0
6	30	180	1	1	30	6	10	60	1	1	10
7	5	35	2	4	20	7	20	140	2	4	80
8	1	8	3	9	9	8	30	240	3	9	270
итого	132	660	-	-	118	итого	170	850	-	-	720

Наиболее простым является показатель *размаха признака R.* Его исчисляют как разность между наибольшими и наименьшими значениями варьирующего признака

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

В нашем примере он будет одинаковым и равняться **6** (8-2). Очевидно, что **R** улавливает только крайние отклонения, но не отражает отклонений всех вариант в ряду.

Среднее линейное отклонение определяется по следующем формулам:

Для первичного ряда

$$\bar{d} = \frac{\sum_{1}^{n} |x_i - \bar{x}|}{n} = \frac{\sum_{1}^{n} |d_i|}{n}$$

за d примем отклонение индивидуальных значений от средней арифметической.

Для вариационного ряда

$$\bar{d} = \frac{\sum_{1}^{n} |x_i - \bar{x}| f_i}{\sum_{1}^{n} f_i} = \frac{\sum_{1}^{n} |d_i| f_i}{\sum_{1}^{n} f_i}$$

Этот показатель как меру вариации признака применяют в статистике достаточно редко. В тоже время именно он демонстрирует общий методический подход к конструированию основных показателей вариации – **дисперсии и среднего квадратического отклонения.**

Если отклонения от средней возвести в квадрат и из квадратов отклонений получить среднюю величину, то этот показатель и будет называется **дисперсией**.

Дисперсия вычисляется по формуле:

$$\sigma^2 = \frac{\sum\left(x-\bar{x}\right)^2 f}{\sum f}$$

А корень квадратный из дисперсии есть **среднее квадратическое отклонение**

$$\sigma = \sqrt{\frac{\sum\left(x-\bar{x}\right)^2 f}{\sum f}}$$

Исчислим по приведенным выше данным дисперсию и среднее квадратичное отклонение.

$$\sigma_1^2 = \frac{118}{132} = 0{,}89 \qquad \sigma_2^2 = \frac{720}{170} = 4{,}2$$

$$\sigma_1 = \sqrt{0{,}89} = 0{,}94 \qquad \sigma_2 = \sqrt{4{,}2} = 2{,}05$$

Среднее квадратичное отклонение всегда выражается в тех именованных числах, в которых выражены варианты и средняя. Оно выражает абсолютную меру вариации.

По своему абсолютному значению среднее квадратичное отклонение зависит не только от степени вариации признака, но и от *абсолютных уровней вариант и средней*. Поэтому сравнивать средние квадратические отклонения вариационных рядов с разными уровнями непосредственно нельзя. Чтобы иметь возможность для такого сравнения, нужно вычислить процентное отклонение среднего квадратического отклонения к средней арифметической.
Этот относительный показатель называется **коэффициентом вариации**, и обозначается обычно буквой $\boldsymbol{v}$

$$v = \frac{\sigma \cdot 100}{\bar{x}}$$

Коэффициенты вариации дают относительную оценку вариации и позволяют сравнивать степень вариации признаков **в вариационных рядах с разным уровнем средних.**
Например, если для урожайности зерна в одном районе $\sigma = 10ц \;\; и \;\; \bar{x} = 40ц$
а в другом районе $\sigma = 9ц \;\; и \;\; \bar{x} = 30ц$
то по абсолютной величине вариация в первом районе больше, так как 10>9, а относительная мера вариации меньше, поскольку

$$v_1 = \frac{10}{40} \cdot 100 = 25\%; \;\; а \;\; v_2 = \frac{9}{30} \cdot 100 = 30\%$$

Коэффициент вариации в известной мере является критерием типичности средней. Если коэффициент вариации очень большой (обычно больше 40%), то это значит, что средняя характеризует совокупность по признаку, который существенно изменяется у отдельных ее единиц. Типичность такой средней невелика.

В нашем примере в 1-м случае $$v=\frac{0{,}94}{5}\cdot 100=0{,}188\,(18{,}8\%)$$

Во 2-м случае $$v=\frac{2{,}05}{5}\cdot 100=0{,}41\,(41\%)$$

Внутри – и межгрупповая дисперсия. Правило сложение дисперсии

Если статистическая совокупность разбита на группы по какому – либо признаку, то для оценки влияния различных факторов, определяющих колеблемость индивидуальных значений признака, можно применить разложение дисперсии на составляющие: меж – и внутри групповую дисперсию.

Если рассчитать дисперсию по всей совокупности (общую σ_o^2), то полученный показатель будет характеризовать вариацию признака имеющую место в результате действия различных факторов.

Чтобы выполнить задачу выделения в составе общей дисперсии той ее части, которая обусловлена влиянием какой – либо определенного фактора, следует разбить изучаемую совокупность на группы, положив в основу интересующий нас фактор. Затем изучить раздельно вариацию признака внутри однородных в отношении данного фактора групп и изменения в величине признака от группы к группе. Это позволит разложить общую дисперсию на две, одна будет характеризовать часть вариации обусловленной влиянием фактора, положенного в основу группировки, а вторая - вариацию, происходящую под влиянием прочих факторов.

Вариацию, обусловленную влиянием фактора, положенного в основу группировки характеризует межгрупповая дисперсия δ^2 (которая является мерой колеблемости частных средних по группам ($\bar{x}_j$) вокруг общей средней ($\bar{x}_o$), и исчисляется по формуле:

$$\delta^2 = \frac{\sum_{j=1}^{k}\left(\bar{x}_j - \bar{x}_o\right)^2 \cdot n_j}{\sum_{j=1}^{k} n_j}$$

где к – число групп;

n_j – число единиц в j-ой группе ;

$\bar{x}_j$ - частная средняя по j – ой группе;

$\bar{x}_o$ - общая средняя по совокупности единиц.

Вариацию, обусловленную влиянием прочих факторов, характеризует внутригрупповая дисперсия:

$$\sigma_j^2 = \frac{\sum_{1}^{nj}\left(x_j - \bar{x}_j\right)^2}{n_j}$$

а, по совокупности в целом – средняя из внутригрупповых дисперсий

$$\bar{\sigma}^2 = \frac{\sum_{1}^{k}\sum_{1}^{nj}\left(x_{ij} - \bar{x}_j\right)^2}{\sum_{1}^{k} n_j} \quad или \quad \bar{\sigma}^2 = \frac{\sum_{1}^{k}\sigma_j^2 \cdot n_j}{\sum_{1}^{k} n_j};$$

$$\sigma_o^2 = \delta^2 + \bar{\sigma}^2$$

Например, при изучении влияния стажа работы рабочих на производительность труда в цехе были получены такие данные:

Группы рабочих по стажу, (лет)	**Число рабочих в группе, (чел.) nj**	**Средняя выработка рабочих в данной группе, (в у.е.)** $\bar{x}_j$	**Дисперсия выработки в группе** σ_j^2
До 5	**30**	**20**	**3**
5 и более	**40**	**23**	**2**
Всего	**70**		

Вариация признака за счет изменения стажа работы будет определяться величиной межгрупповой дисперсии. Средняя часовая выработка всех рабочих

$$\bar{x}_o = \frac{20{,}0 \cdot 30 + 23{,}0 \cdot 40}{70} = \frac{600 + 920}{70} = \frac{1520}{70} = 21{,}7$$

Межгрупповая дисперсия

$$\delta^2 = \frac{(20{,}0 - 21{,}7)^2 \cdot 30 + (23{,}0 - 21{,}7)^2 \cdot 40}{70} = 2{,}20$$

Вариация признака за счет изменения стажа работы будет определяться величиной межгрупповой дисперсии. Средняя часовая выработка всех рабочих

$$\bar{x}_o = \frac{20{,}0 \cdot 30 + 23{,}0 \cdot 40}{70} = \frac{600 + 920}{70} = \frac{1520}{70} = 21{,}7$$

Межгрупповая дисперсия

$$\delta^2 = \frac{(20{,}0 - 21{,}7)^2 \cdot 30 + (23{,}0 - 21{,}7)^2 \cdot 40}{70} = 2{,}20$$

Вариация выработки под действием всех прочих факторов будет характеризоваться величиной средней из внутригрупповой дисперсий

$$\overline{\sigma}^2 = \frac{3{,}0 \cdot 30 + 2{,}0 \cdot 40}{70} = 2{,}43$$

Вариация выработки под действием всех факторов будет характеризоваться величиной общей дисперсии:

$$\sigma_o^2 = 2{,}43 + 2{,}20 = 4{,}63$$

Это означает, что на 47,6% (2,20/4,63×100%) дисперсия выработки обусловлена различиями в стаже работы, а на 52,4% (2,43/4,63×100%) влиянием прочих факторов