Изучение вариации

Лекция 4

Вопросы лекции

- Понятие вариации признаков. Необходимость статистического изучения вариации.
- Показатели вариации: размах вариации, среднее линейное отклонение, среднее квадратическое отклонение, дисперсия, коэффициент вариации.
- Сокращенные способы расчета дисперсии. Правило сложения дисперсии.
- Изучение взаимосвязи признаков при помощи показателей вариации.

  Эмпирическое корреляционное отношение.

1 Понятие вариации признаков. Необходимость статистического изучения вариации

Вариацией называется изменяемость, колеблемость величины признака. Вариация проявляется в отклонениях от средних и зависит от множества факторов, влияющих на социально-экономическое явление.

Вариационные ряды

- При изучении совокупности интересующий нас признак у различных единиц совокупности принимает различные значения, т.е. он имеет некоторую вариацию.

Вариацией признака

- называется наличие различий в численных значениях признаков у отдельных единиц совокупности.
- Чтобы выявить характер распределения единиц совокупности по варьирующим признакам, определить закономерности в этом распределении, строят ряды распределения единиц совокупностей по какому-либо варьирующему признаку.
- Ряды распределения, построенные по количественному признаку называются **вариационными**.

При анализе вариационных рядов решают следующие задачи:

- 1) *Определение меры вариации,* т.е. количественное измерение степени колеблемости признака. Это позволяет сравнивать различные совокупности между собой по степени рассеяния и отслеживать уровень вариации признака одной и той же совокупности в различные периоды.
- 2) *Исследование закономерностей вариации* в статистических совокупностях для изучения причин, вызывающих вариацию.

Для описания статистических распределений обычно используются следующие виды характеристик (показателей):

- 1) средние величины;
- 2) характеристики вариации (рассеяния);
- 3) характеристики дифференциации и концентрации;
- 4) характеристики формы распределения.

2 Показатели вариации

Относительные показатели (коэффициент вариации, линейный коэффициент вариации, коэффициент осцилляции) строятся с учетом базы (в виде средней), выражаются в процентах и дают характеристику однородности совокупности.

- Вариация бывает случайной и систематической, существует в пространстве и во времени.

- Показатели вариации делятся на абсолютные и относительные.

Показатели вариации (абс.)

Показатель	Формула расчета показателя	
	простой	Взвешенный
Размах	$R = x_{\max} - x_{\min}$ (2.1)	
Среднее линейное отклонение	$\bar{d} = \dfrac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{n}$	$\bar{d} = \dfrac{\sum_{i=1}^{m} \lvert x_i - \bar{x} \rvert \cdot f_i}{\sum_{i=1}^{m} f_i}$ *
Дисперсия	$\sigma^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$	$\sigma^2 = \dfrac{\sum_{i=1}^{m} (x_i - \bar{x})^2 f_i}{\sum_{i=1}^{m} f_i}$
Среднее квадратическое отклонение	$\sigma = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}}$	$\sigma = \sqrt{\dfrac{\sum_{i=1}^{m} (x_i - \bar{x})^2 f_i}{\sum_{i=1}^{m} f_i}}$

- * – Здесь f_i – частота

$$\frac{f_i}{\sum f_i} = \frac{n_i}{n}$$

Показатели вариации (отн.)

Показатель	Формула расчета показателя
	простой
Коэффициент вариации	$V_{\sigma} = \frac{\sigma}{\overline{x}} \times 100\,\%$
Линейный коэффициент вариации	$V_{\overline{d}} = \frac{\overline{d}}{\overline{x}} \times 100\,\%$
Коэффициент осцилляции	$V_R = \frac{R}{\overline{x}} \times 100\%$
Коэффициент вариации	$V_{\sigma} = \frac{\sigma}{\overline{x}} \times 100\,\%$

Совокупность считается однородной, если коэффициент вариации

$$V_{\sigma} \leq 33\%$$

3 Сокращенные способы расчета дисперсии. Правило сложения дисперсии.

Для расчета дисперсии можно использовать модифицированную формулу:

$$\sigma^2 = \overline{x^2} - (\bar{x})^2$$

Выведем эту формулу

$$\sigma^2 = \frac{\sum\limits_i (x_i - \bar{x})^2 f_i}{\sum\limits_i f_i} = \frac{\sum\limits_i x_i^2 f_i}{\sum\limits_i f_i} - \frac{2\sum\limits_i x_i \bar{x} f_i}{\sum\limits_i f_i} + \frac{\sum\limits_i \bar{x}^2 f_i}{\sum\limits_i f_i} =$$

$$= \frac{\sum\limits_i x_i^2 f_i}{\sum\limits_i f_i} - 2\bar{x}\frac{\sum\limits_i x_i f_i}{\sum\limits_i f_i} + \frac{\sum\limits_i \bar{x}^2 f_i}{\sum\limits_i f_i} = \overline{x^2} - 2(\bar{x})^2 + (\bar{x})^2 = \overline{x^2} - (\bar{x})^2$$

- Размах вариации, среднее линейное и среднее квадратичное отклонение – это именованные величины.
- Единицей измерения у них и у исходных значений признака совпадают. Дисперсия может быть задана в ед.^{2} признака или в % отклонений.

Общая дисперсия совокупности:

$$\sigma_{\text{общ}}^{2} = \frac{\sum_{i=1}^{k}(x_i - \widetilde{X})^2 m_i}{N}$$

- Общая дисперсия отражает вариацию признака за счет всех факторов, действующих в данной совокупности.

Вариацию между группами за счет признака-фактора, положенного в основу группировки, отражает **межгрупповая дисперсия**, которая исчисляется как средний квадрат отклонений групповой средней от общей средней:

Межгрупповая дисперсия:

$$\delta^2 = \frac{\sum_{j=1}^{l} (\bar{x}_j - \widetilde{X})^2 n_j}{N}$$

- **Межгрупповая дисперсия** характеризует систематическую вариацию результативного признака, т.е. вариацию между группами за счет признака-фактора, положенного в основу группировки.

- Вариацию внутри каждой группы изучаемой совокупности отражает **внутригрупповая дисперсия**, которая исчисляется как средний квадрат отклонений значений признака *x* от частной средней $\bar{x}_j$:

Внутригрупповая дисперсия

$$\sigma_j^2 = \frac{\sum_{i=1}^{k}(x_i - \bar{x}_j)^2 f_{ij}}{n_j} \text{ или } \sigma_j^2 = \frac{\sum_{i=1}^{k} x_i^2 f_{ij}}{n_j} - (\bar{x}_j)^2$$

Для всей совокупности внутригрупповую вариацию будет выражать **средняя из внутригрупповых дисперсий**, которая рассчитывается как средняя арифметическая из внутригрупповых дисперсий:

Средняя из внутригрупповых дисперсий

$$\overline{\sigma^2} = \frac{\sum_{j=1}^{l} \sigma_j^2 n_j}{N}$$

Внутригрупповая дисперсия

- отражает случайную вариацию, т.е. часть вариации обусловленную влиянием неучтенных факторов и не зависящую от признака-фактора, положенного в основу группировки.

Между представленными видами дисперсий существует определенное соотношение, которое известно как *правило сложения дисперсий*:

$$\sigma^2_{общ} = \overline{\sigma^2} + \delta^2$$

Таким образом

общая дисперсия складывается из двух слагаемых: первое – средняя из внутригрупповых дисперсий – измеряет вариацию внутри частей совокупности, второе – межгрупповая дисперсия – вариацию между средними этих частей.

Правило сложения дисперсий

- позволяет выявить зависимость результатов от определяющих факторов с помощью соотношения межгрупповой и общей дисперсий.
- Это соотношение называется **эмпирическим коэффициентом детерминации** (η^2) и показывает долю вариации результативного признака под влиянием факторного.

4 Изучение взаимосвязи признаков при помощи показателей вариации. Эмпирическое корреляционное отношение.

Эмпирический коэффициент детерминации

$$\eta^2 = \frac{\delta^2}{\sigma^2_{общ}}$$

Эмпирическое корреляционное отношение

η^2 и $\eta \in [0, 1]$

(η) показывает тесноту связи между исследуемым явлением и группировочным признаком:

$$\eta = \sqrt{\frac{\delta^2}{\sigma^2_{\text{общ}}}}$$

- Если связь отсутствует, то η = 0. В этом случае межгрупповая дисперсия равна нулю ($\delta^2=0$), т.е. все групповые средние равны между собой и межгрупповой вариации нет. Это означает, что группировочный признак не влияет на вариацию исследуемого признака *x*.
- Если связь функциональная, то η = 1. В этом случае дисперсия групповых средних равна общей дисперсии ($\delta^2=\sigma^2_{\text{общ}}$). Это означает, что группировочный признак полностью определяет характер изменения изучаемого признака.

- Чем больше значение корреляционного отношения приближается к единице, тем полнее (сильнее) корреляционная связь между признаками (см.таблица ниже).

Качественная оценка связи между признаками (шкала Чэддока)

Значение $\eta_{теор}$	Характер связи		Значение $\eta_{теор}$	Характер связи
$\eta = 0$	Отсутствует		$0{,}5 \le \eta < 0{,}7$	Заметная
$0 < \eta < 0{,}2$	Очень слабая		$0{,}7 \le \eta < 0{,}9$	Сильная
$0{,}2 \le \eta < 0{,}3$	Слабая		$0{,}9 \le \eta < 1$	Весьма сильная
$0{,}3 \le \eta < 0{,}5$	Умеренная		$\eta = 1$	Функциональная

Пример решения задачи

Определим групповые дисперсии, среднюю из групповых дисперсий, межгрупповую дисперсию, общую дисперсию по данным о производительности труда в двух бригадах.

Изготовлено деталей за час, шт. (производительность труда)	Количество рабочих, имеющих соответствующую производительность труда	
	в бригаде 1	в бригаде 2
x_i	f_{i1}	f_{i2}
10	1	0
12	3	0
14	3	1
16	2	3
18	1	2
20	0	4

Промежуточные расчеты занесем в таблицы:

x_i	Бр. 1	Бр. 2	m_i	Промежуточные расчеты для определения средних величин		
	f_{i1}	f_{i2}		$x_i \cdot f_{i1}$	$x_i \cdot f_{i2}$	$x_i \cdot m_i$
10	1	0	1	10	0	10
12	3	0	3	36	0	36
14	3	1	4	42	14	56
16	2	3	5	32	48	80
18	1	2	3	18	36	54
20	0	4	4	0	80	80
Σ	n_1=10	n_2=10	N=20	$\Sigma x_i \cdot f_{i1}$=138	$\Sigma x_i \cdot f_{i2}$=178	$\Sigma x_i \cdot m_i$ =316

x_i	Промежуточные расчеты для определения дисперсий					
	$(x_i - \overline{x}_1)$	$(x_i - \overline{x}_2)$	$(x_i - \widetilde{X})$	$(x_i - \overline{x}_1)^2 \cdot f_{i1}$	$(x_i - \overline{x}_2)^2 \cdot f_{i2}$	$(x_i - \widetilde{X})^2 \cdot m_i$
10	-3,8	-7,8	-5,8	14,44	0,00	33,64
12	-1,8	-5,8	-3,8	9,72	0,00	43,32
14	0,2	-3,8	-1,8	0,12	14,44	12,96
16	2,2	-1,8	0,2	9,68	9,72	0,20
18	4,2	0,2	2,2	17,64	0,08	14,52
20	6,2	2,2	4,2	0,00	19,36	70,56
Σ	–	–	–	51,60	43,60	175,20