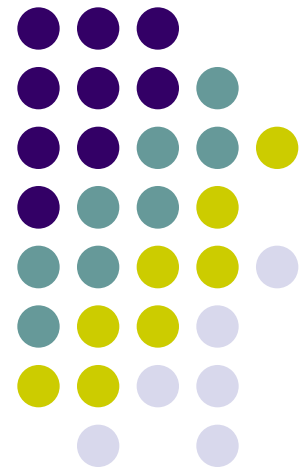
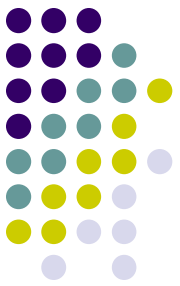


Показатели вариации

***«Всякая вещь есть форма
проявления беспредельного
разнообразия»***

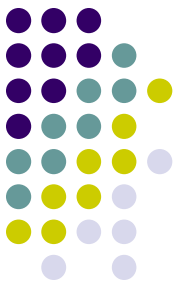
***Козьма Прутков,
Плоды раздумья***





- **План лекции**

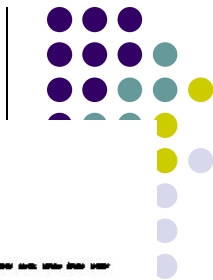
- **1. Абсолютные показатели вариации**
- **2. Относительные показатели вариации**
- **3. Меры вариации для сгруппированных данных**
- **4. Правило сложения дисперсий**
- **5. Вариация альтернативного признака**



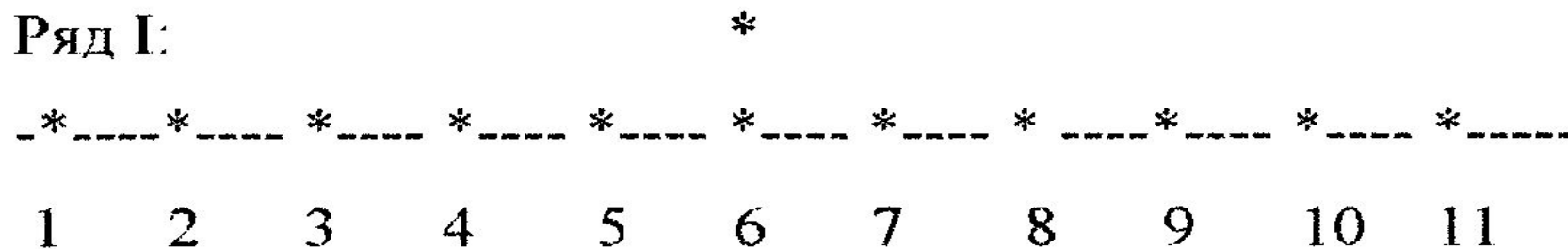
- Ряд I: 1, 2, 3, 4, 5, 6, 6, 7, 8, 9, 10, 11
- Ряд II: 4, 5, 5, 5, 6, 6, 6, 6, 7, 7, 7, 8

$$\bar{x} = 6, Me = 6, Mo = 6, n = 12$$

$$\bar{x} = 6, Me = 6, Mo = 6, n = 12$$



Ряд I:

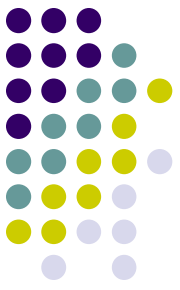


Средняя = Медиане = Моде = 6

Ряд II:



Средняя = Медиане = Моде = 6

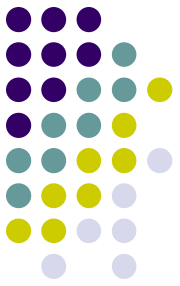


Интерквартильный размах – разница между первым и третьим квартилями. Чем больше величина интерквартильного размаха, тем больше рассеяние признака. Интерквартильный размах в ряду I равен 5.5, интерквартильный размах в ряду II равен 2.

- *Размах вариации - разность между наибольшим и наименьшим значениями признака*

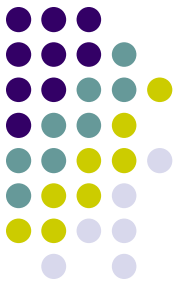
$$R = x_{\max} - x_{\min}$$

$$\frac{\sum (x - \bar{x})}{n}$$

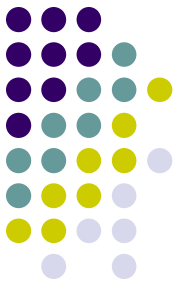


Сумма отклонений всех вариантов от их средней арифметической, согласно свойству средней арифметической, всегда равна нулю.

$$L = \frac{\sum |x - \bar{x}|}{n}$$



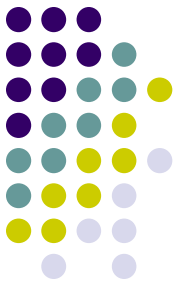
$$L = \frac{\sum |x - \bar{x}| f}{\sum f}$$



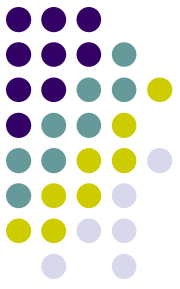
$$\sigma^2 = \frac{\sum (x - \bar{x})^2}{n}$$

$$\sigma^2 = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \cdot f_i}{\sum_{i=1}^k f_i}$$

Дисперсия вариационного ряда есть средняя арифметическая квадрата отклонения (средний квадрат отклонения) значений признаков ряда от их средней арифметической.



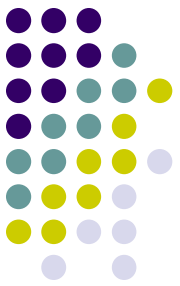
Среднее квадратическое (стандартное) отклонение вариационного ряда есть арифметическое значение корня квадратного из дисперсии.



$$\sigma = \sqrt{\sigma^2}$$

Среднее квадратическое (стандартное) отклонение вариационного ряда есть арифметическое значение корня квадратного из дисперсии.

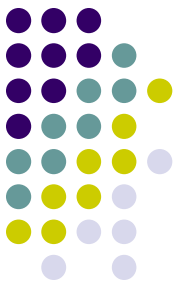
$$\sigma = \sqrt{\sigma^2}$$



Показатели относительного рассеяния

Расчет показателей меры относительного рассеяния осуществляют как отношение абсолютного показателя вариации признака к средней арифметической, умножаемое на 100%.

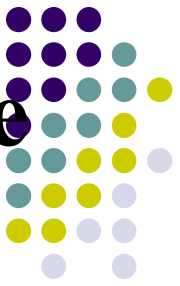
Коэффициент осцилляции отражает относительную колеблемость крайних значений признака вокруг средней:



$$K_0 = \frac{R}{\bar{x}} \cdot 100\%$$

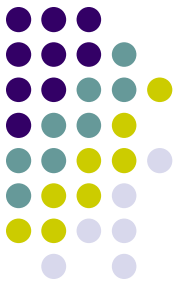
Относительное линейное отклонение характеризует долю усредненного значения признака абсолютных отклонений от средней величины:

$$K_L = \frac{L}{\bar{x}} \cdot 100$$



Коэффициент вариации является наиболее распространенным показателем колеблемости, используемым для оценки типичности средних величин:

$$V = \frac{\sigma}{\bar{x}}$$



Предположим, что стандартное отклонение в выборке валютных счетов в банке «А» и банке «В» равно \$20. Данные по банку «А» содержат информацию о счетах, сумма которых находится в пределах \$60. В банке «В» данные содержат информацию относительно счетов, сумма которых достигает \$1 миллион и больше. В первом случае стандартное отклонение в 20 единиц очень велико относительно сумм счетов. Для суммы порядка \$1 миллиона – что значит вариация плюс-минус \$20 относительно среднего?



Сравнивая эти два случая, можно сказать, что такая абсолютная мера рассеяния как стандартное отклонение не передает существенной информации при сравнении вариационных рядов. Коэффициент вариации создан специально как относительная мера вариации.



Чем меньше значение коэффициента вариации, тем однороднее совокупность по изучаемому признаку и типичнее средняя. В статистике совокупности, имеющие коэффициент вариации больше 30-35 %, принято считать неоднородными.

Однако у такого способа оценки вариации есть и существенный недостаток. Например, исходная совокупность рабочих, имеющих средний стаж 15 лет, со средним квадратическим отклонением $\sigma = 10$ лет, «состарилась» еще на 15 лет. Теперь $\bar{x} = 30$ лет, а среднеквадратическое отклонение по-прежнему равно 10. Совокупность, ранее бывшая неоднородной ($10/15 \cdot 100 = 66,7\%$), со временем оказывается, таким образом, вполне однородной ($10/30 \cdot 100 = 33,3\%$).

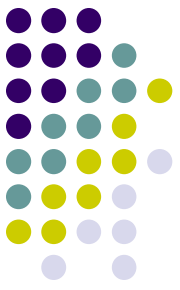


Меры вариации для сгруппированных данных.



Правило сложения дисперсий

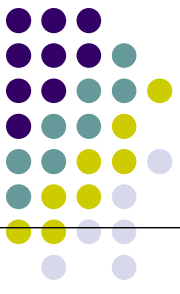
Вариация признаков, как правило, обусловлена влиянием различных факторов. Если совокупность разбить на группы по факторному признаку, то это окажет определенное влияние на значение вариации признака в группах. Выявить долю вариации, определяемую теми или иными факторами, можно разделяя всю совокупность на группы по фактору, влияние которого исследуется.



Для этих целей используются показатели вариации для сгруппированных данных. В этом случае выделяют три вида дисперсий:

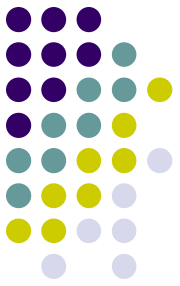
Общую дисперсию; внутригрупповую дисперсию, межгрупповую дисперсию.

Общая дисперсия измеряет вариацию признака во всей совокупности под влиянием всех факторов. Внутригрупповая дисперсия измеряет вариацию признака внутри группы, а межгрупповая дисперсия измеряет вариацию групповых средних относительно общей средней.

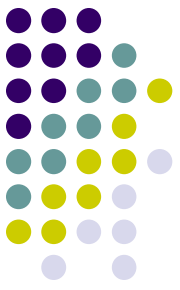


Значение признака x	Число единиц в j -й группе				Итого
	1	2	...	m	
x_1	f_1	s_1	...	t_1	$f_1 + s_1 + \dots + t_1 = n_1$
x_2	f_2	s_2	...	t_2	$f_2 + s_2 + \dots + t_2 = n_2$
...
x_k	f_k	s_k	...	t_k	$f_k + s_k + \dots + t_k = n_k$
Итого	N_1	N_2	...	N_m	N

Вычислим m частных средних



$$\bar{x}_1 = \frac{\sum_{i=1}^k x_i f_i}{N_1}$$



$$\bar{x}_2 = \frac{\sum_{i=1}^k x_i s_i}{N_2}$$

...

$$\bar{x}_m = \frac{\sum_{i=1}^k x_i t_i}{N_m}.$$

На основе частных средних

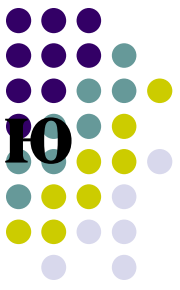


$\bar{x}_1, \bar{x}_2, \dots, \bar{x}_m$
определяем общую среднюю

$$\bar{x}_{\text{общ.}} = \frac{\sum_{j=1}^m \bar{x}_j N_j}{N}, \text{ где}$$

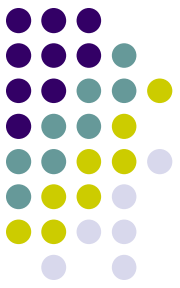
$$N = \sum_{j=1}^m N_j = \sum_{i=1}^k n_i$$

Общая дисперсия отражает вариацию признака за счет всех условий, действующих в совокупности



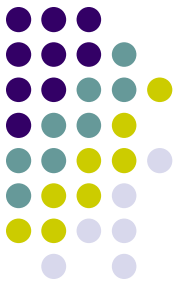
$$\sigma_{\text{общ.}}^2 = \frac{\sum_{i=1}^k (x_i - \bar{x}_{\text{общ.}})^2 n_i}{N}$$

Межгрупповая дисперсия отражает вариацию между группами за счет фактора, положенного в основу группировки



$$\delta^2 = \frac{\sum_{j=1}^m (\bar{x}_j - \bar{x})^2 N_j}{N}$$

**Частная групповая дисперсия
отражает вариацию внутри
каждой группы**

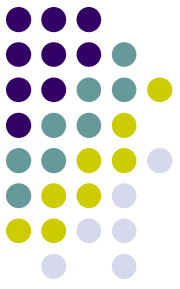


$$\sigma_1^2 = \frac{\sum_{i=1}^k (x_i - \bar{x}_1)^2 f_i}{N_1}$$

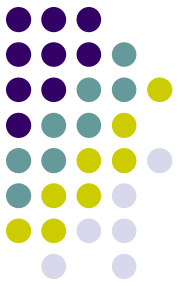
$$\sigma_2^2 = \frac{\sum_{i=1}^k (x_i - \bar{x}_2)^2 s_i}{N_2}$$

.....

$$\sigma_m^2 = \frac{\sum_{i=1}^k (x_i - \bar{x}_m)^2 t_i}{Nm}$$



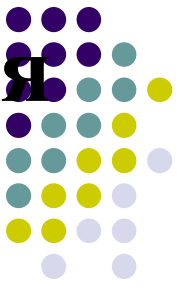
В общем виде частная дисперсия:



$$\sigma_j^2 = \frac{\sum_{i=1}^k (x_i - \bar{x}_j)^2 N_{ij}}{N_j}$$

где N_{ij} - частоты для значений x_i , в каждой j -группе, $j=1, \dots, m$.

Внутригрупповая дисперсия - средняя арифметическая из групповых дисперсий



$$\overline{\sigma^2} = \frac{\sum_{j=1}^m \sigma_j^2 N_j}{N}$$

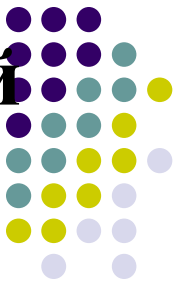
Общая дисперсия равна сумме межгрупповой дисперсии и средней из внутригрупповых дисперсий



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

Суть правила: общая дисперсия, возникающая под воздействием всех факторов, формируется из дисперсии, возникающей за счет фактора группировки и дисперсии, возникающей под воздействием всех прочих факторов.

С помощью правила сложения дисперсий оценивается удельный вес факторов, лежащих в основе группировки, во всей совокупности факторов, воздействующих на результативный признак. Для этого применяется коэффициент детерминации



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

Эмпирическое корреляционное отношение



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

Показатель принимает значения в интервале $[0, 1]$ и характеризует взаимосвязь между изучаемым признаком и признаком, положенным в основу группировки. Если связь между признаками отсутствует, то $\eta = 0$. В этом случае дисперсия групповых средних равна нулю ($\delta^2 = 0$), то есть все групповые средние равны между собой и межгрупповой вариации нет. Если связь функциональная, то $\eta = 1$. В этом случае дисперсия групповых средних равна общей дисперсии ($\delta^2 = \sigma^2$), то есть внутригрупповая дисперсия отсутствует. Это означает, что группировочный признак полностью определяет вариацию изучаемого признака, а влияние прочих факторных признаков равно нулю.



Промежуточные значения η оцениваются по степени их близости к граничным



**Качественная оценка связи между признаками
(шкала Чеддока)**

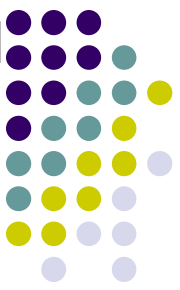
η	Связь	η	Связь
0	Отсутствует	0,5-0,7	Заметная
0-0,2	Очень слабая	0,7-0,9	Тесная
0,2-0,3	Слабая	0,9-0,99	Весьма тесная
0,3-0,5	Умеренная	1	Функциональная



Группировка населения отдельных областей России по среднему размеру ежемесячных денежных льгот пенсионеров

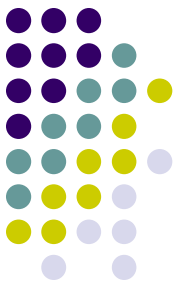
Место проживания	Средний размер денежных льгот, руб. \bar{x}_j	Численность пенсионеров, тыс.чел. N_j	Дисперсия льгот в области (группе) σ_j^2
Курская обл.	264,3	341,4	9025
Курганская обл.	310,4	235,5	2704
Камчатская обл.	490,4	38,9	2116
Итого	296,2	615,8	7171,2

Найдем средний размер льгот (руб.) по трем областям в целом:



$$\bar{x} = \frac{\sum_j \bar{x}_j N_j}{\sum_j N_j} = \frac{264,3 \cdot 341,4 + 310,4 \cdot 235,5 + 490,4 \cdot 38,9}{615,8} = 296,2$$

Вариация льгот по отдельным областям, обусловленная различием в месте проживания пенсионеров, характеризуется межгрупповой дисперсией:



$$\delta^2 = \frac{\sum_{j=1}^m (\bar{x}_j - \bar{x})^2 N_j}{N} = \frac{(264,3 - 296,2)^2 \cdot 341,4 + (310,4 - 296,2)^2 \cdot 235,5}{615,8} + \frac{(490,4 - 296,2)^2 \cdot 38,9}{615,8} = 3023,6 \text{ руб.}^2$$

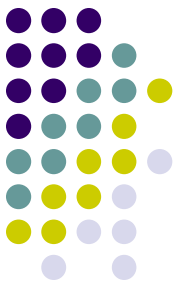
Средняя из групповых дисперсий дает обобщающую характеристику случайной вариации, обусловленную отдельными факторами, кроме места проживания пенсионеров (например, характером занятости, стажем работы и т.п.):

$$\overline{\sigma^2} = \frac{9025 \cdot 341,4 + 2704 \cdot 235,5 + 2116 \cdot 38,9}{615,8} = 6171,2 \text{ руб.}^2$$

Вариация льгот в изучаемых областях России, обусловленная влиянием всех факторов, вместе взятых, определяется общей дисперсией:



$$\sigma^2 = \delta^2 + \bar{\sigma}^2 = 2976,1 + 6171,2 = 9147,3$$



$$\eta^2 = \frac{\delta^2}{\sigma_{\text{общ.}}^2} = \frac{2976,1}{9147,3} = 0,325$$



Полученный коэффициент детерминации показывает, что дисперсия льгот зависит от места проживания пенсионеров на 32,5 %.
Остальные 67,5% определяются другими неучтенными факторами.

$$\eta = \sqrt{\eta^2} = \sqrt{0,325} = 0,57$$


Полученное значение эмпирического корреляционного отношения позволяет утверждать, что существует заметная связь между местом проживания пенсионеров и размером льгот.

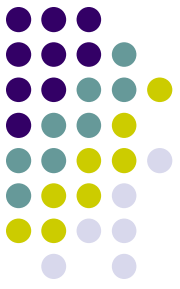
Вариация альтернативного признака



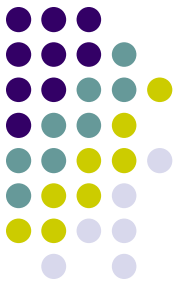
При статистическом выражении колеблемости альтернативных признаков наличие изучаемого признака обозначается 1, а его отсутствие – 0. Доля вариантов, обладающих изучаемым признаком обозначается p , а доля вариантов, не обладающих признаком q . Следовательно, $p + q = 1$.



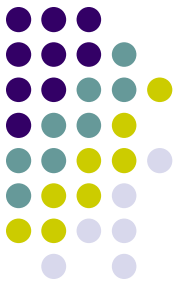


$$\bar{x} = \frac{\sum_i x_i f_i}{\sum_i f_i} = \frac{1 \cdot p + 0 \cdot q}{p + q} = p$$



$$\sigma_p^2 = \frac{\sum_i (x_i - \bar{x})^2 f_i}{\sum_i f_i} = \frac{(1-p)^2 \cdot p + (0-p)^2 \cdot q}{p+q} = q^2 p + p^2 q = pq(p+q) = pq$$

Если значения 1 и 0 встречаются одинаково часто, то дисперсия достигает своего максимума



Пример. На 10000 населения приходится 4000 мужчин и 6000 женщин. Определить среднее квадратическое отклонение по полу.

Решение. Доля мужчин в населении $p=4000/10000=0,4$; доля женщин $q=6000/10000=0,6$. Тогда дисперсия

$$\sigma_p^2 = pq = 0,4 \cdot 0,6 = 0,24$$

а среднее квадратическое отклонение

$$\sigma_p = \sqrt{\sigma_p^2} = \sqrt{0,24} = 0,49$$

Внутригрупповая дисперсия доли



$$\sigma_{p_i}^2 = p_i(1 - p_i)$$

Средняя из внутригрупповых дисперсий

$$\overline{\sigma_{p_i}^2} = \frac{\sum p_i(1 - p_i)n_i}{\sum n_i} = p_i(1 - p_i)$$

межгрупповая дисперсия

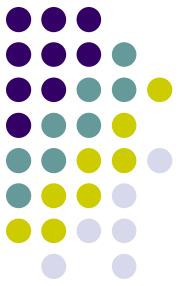


$$\delta_{pi}^2 = \frac{\sum (p_i - \bar{p})^2 n_i}{\sum n_i}$$

где n_i – численность единиц в отдельных группах;

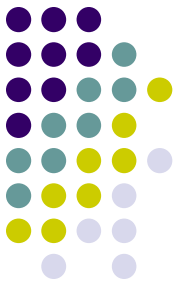
\bar{p} – доля изучаемого признака во всей совокупности, которая определяется по формуле:

$$\bar{p} = \frac{\sum p_i n_i}{\sum n_i}$$



- **Общая дисперсия**

$$\sigma_{\bar{p}}^2 = \bar{p}(1 - \bar{p})$$

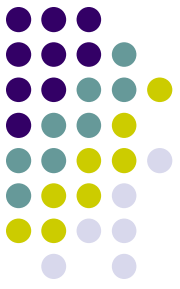


Правило сложения дисперсии доли признака

$$\sigma_{\bar{p}}^2 = \overline{\sigma_{p_i}^2} + \delta_{p_i}^2$$



Цех	Удельный вес основных рабочих в % (p_i)	Численность всех рабочих (чел.)
1	80	100
2	75	200
3	90	150
Итого	-	450



Определим долю рабочих в целом по фирме:

$$\bar{p} = \frac{0,80 \cdot 100 + 0,75 \cdot 200 + 0,90 \cdot 150}{450} = \frac{365}{450} = 0,81$$

Общая дисперсия доли
основных рабочих по всей
фирме в целом будет равна:



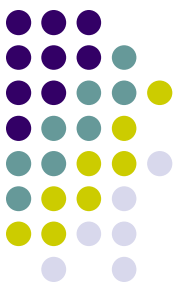
$$\sigma_{\bar{p}}^2 = 0,81 \cdot (1 - 0,81) = 0,154$$

- Внутрицеховые дисперсии:

$$\sigma_{p1}^2 = 0,8 \cdot 0,2 = 0,16;$$

$$\sigma_{p2}^2 = 0,75 \cdot 0,25 = 0,19;$$

$$\sigma_{p3}^2 = 0,9 \cdot 0,1 = 0,09.$$



Средняя из внутрицеховых дисперсий:

$$\overline{\sigma_{pi}^2} = \frac{0,16 \cdot 100 + 0,19 \cdot 200 + 0,09 \cdot 150}{450} = \frac{675}{450} = 0,15$$



Межгрупповая дисперсия:

$$\delta_{pi}^2 = \frac{(0,8 - 0,81)^2 \cdot 100 + (0,75 - 0,81)^2 \cdot 200 + (0,9 - 0,81)^2 \cdot 150}{450} =$$

$$= \frac{365}{450} = 0,004$$

$$0,154 = 0,15 + 0,004$$

