

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ГЕОЛОГИИ

1. Математические методы моделирования в геологии: Учебник / Г.С.Поротов. Санкт-Петербургский государственный горный институт (технический университет). СПб, 2006. 223 с.
2. Каждан А. Б., Гуськов О. И. Математические методы в геологии: Учебник для вузов.— М.: Недра, 1990.— 251 с.

ОСНОВНЫЕ НАПРАВЛЕНИЯ ИСПОЛЬЗОВАНИЯ МАТЕМАТИЧЕСКИХ МЕТОДОВ В ГЕОЛОГИИ

1) накопление, хранение и систематизация (сортировка, получение выборок и пр.) геологической информации с целью более полного и быстрого ее использования;

2) обработка геологической информации преимущественно на базе методов теории вероятностей и математической статистики для описания, сравнения, классификации геологических объектов и прогнозирования их свойств;

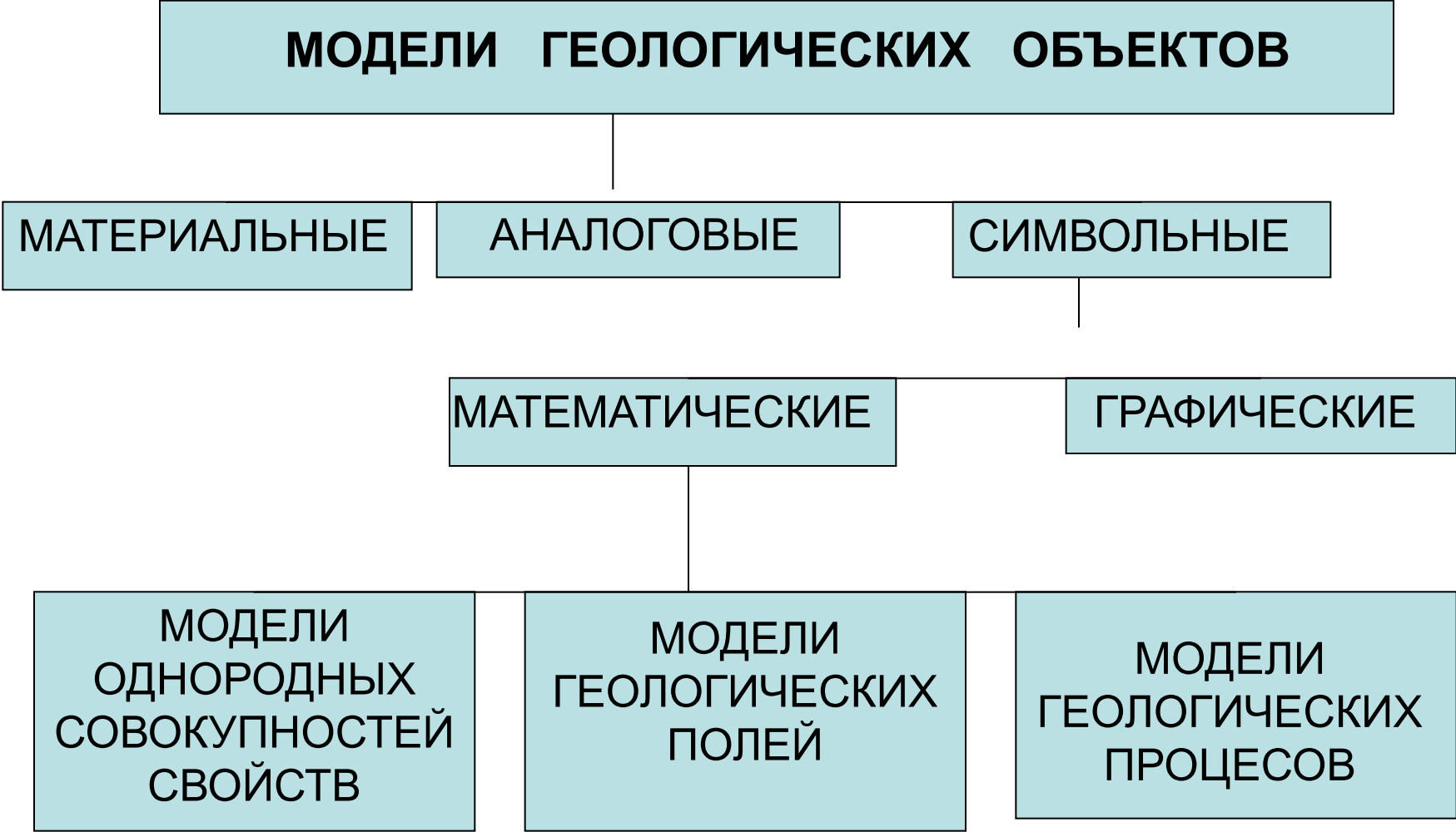
3) математическое моделирование геологических объектов и явлений для решения научных и прикладных задач;

4) автоматизация технологических операций, распространенных в геологии и горном деле (построение геологических карт и разрезов, подсчет запасов и ресурсов, проектирование разведочных и эксплуатационных работ и др).

ПОНЯТИЕ О МАТЕМАТИЧЕСКОМ МОДЕЛИРОВАНИИ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ

- *Математическая модель* - это совокупность представлений, предположений, гипотез и аксиом, отражающих существо изучаемого геологического объекта или явления.
- Модель выражается в математической форме и позволяет описывать, анализировать и прогнозировать свойства геологических объектов или последствия явлений.

МОДЕЛИ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ



```
graph TD; A[МОДЕЛИ ГЕОЛОГИЧЕСКИХ ОБЪЕКТОВ] --> B[МАТЕРИАЛЬНЫЕ]; A --> C[АНАЛОГОВЫЕ]; A --> D[СИМВОЛЬНЫЕ]; D --> E[МАТЕМАТИЧЕСКИЕ]; D --> F[ГРАФИЧЕСКИЕ]; E --> G[МОДЕЛИ ОДНОРОДНЫХ СОВОКУПНОСТЕЙ СВОЙСТВ]; E --> H[МОДЕЛИ ГЕОЛОГИЧЕСКИХ ПОЛЕЙ]; E --> I[МОДЕЛИ ГЕОЛОГИЧЕСКИХ ПРОЦЕССОВ];
```

МАТЕРИАЛЬНЫЕ

АНАЛОГОВЫЕ

СИМВОЛЬНЫЕ

МАТЕМАТИЧЕСКИЕ

ГРАФИЧЕСКИЕ

МОДЕЛИ
ОДНОРОДНЫХ
СОВОКУПНОСТЕЙ
СВОЙСТВ

МОДЕЛИ
ГЕОЛОГИЧЕСКИХ
ПОЛЕЙ

МОДЕЛИ
ГЕОЛОГИЧЕСКИХ
ПРОЦЕССОВ

МОДЕЛИ ОДНОРОДНЫХ
СОВОКУПНОСТЕЙ СВОЙСТВ

```
graph TD; A[МОДЕЛИ ОДНОРОДНЫХ СОВОКУПНОСТЕЙ СВОЙСТВ] --> B[СТАТИСТИЧЕСКИЕ МОДЕЛИ]; B --> C[ОДНОМЕРНЫЕ]; B --> D[ДВУХМЕРНЫЕ]; B --> E[ТРЕХМЕРНЫЕ]; C --> F[СТАТИСТИЧЕСКИЙ АНАЛИЗ]; D --> G[РЕГРЕССИОННЫЙ АНАЛИЗ]; E --> H[ФАКТОРНЫЙ АНАЛИЗ<br/>КЛАСТЕРНЫЙ АНАЛИЗ<br/>РАСПОЗНАВАНИЕ ОБРАЗОВ];
```

СТАТИСТИЧЕСКИЕ МОДЕЛИ

ОДНОМЕРНЫЕ

ДВУХМЕРНЫЕ

ТРЕХМЕРНЫЕ

СТАТИСТИЧЕСКИЙ
АНАЛИЗ

РЕГРЕССИОННЫЙ
АНАЛИЗ

ФАКТОРНЫЙ АНАЛИЗ
КЛАСТЕРНЫЙ АНАЛИЗ
РАСПОЗНАВАНИЕ
ОБРАЗОВ

ОСНОВНЫЕ ПОНЯТИЯ

СТАТИСТИЧЕСКОГО МОДЕЛИРОВАНИЯ

Статистическая совокупность - множество случайных, однородных, повторяющихся объектов и явлений, обладающих качественной общностью или множество значений признака, обладающих свойствами случайных величин.

Одномерная, если характеризуется *одним* признаком, двухмерная- двумя признаками, многомерной - *тремя и более признаками* (трехмерной, четырехмерной,...).

Раздел математической статистики, который занимается изучением закономерностей в одномерных статистических совокупностях –

вариационный анализ

Генеральная и выборочная совокупности

Генеральная совокупность – абстрактное понятие, это все возможные значения случайной величины изучаемого объекта

Выборочная совокупность или выборка - выборочные значения из генеральной совокупности, которые возможно или целесообразно использовать

Задача статистического анализа

- по свойствам изучаемого признака в случайной выборке определенного объема сделать с определенной вероятностью заключение о свойствах этого признака во всей генеральной совокупности.

При статистическом моделировании используются выборки, отобранные по определенным правилам.

Главным требованием является репрезентативность (или представительность) выборки, которая должна правильно представлять всю генеральную совокупность.

- Репрезентативные выборки должны удовлетворять 4 условиям: **случайности, независимости, массовости и однородности**.
- Условие **случайности** означает, что все элементы генеральной совокупности должны иметь одинаковую вероятность попадания в выборку.
- Условие **независимости** означает, что результаты каждого наблюдения в выборке не зависят от других наблюдений.
- Условие **массовости** - выборка должна быть достаточной по объему, так как в соответствии с законом больших чисел статистическая закономерность проявляется лишь в массовых явлениях.
- Условие **однородности** — выборка должна состоять из наблюдений, принадлежащих к одному объекту и выполненных одним способом.

В вариационном анализе последовательно решаются 2 задачи:

1) упорядочение исходной статистической совокупности (по возрастанию или убыванию)

вариационный ряд

2) подбор к упорядоченной статистической совокупности **теоретической модели** (вероятностной одномерной модели).

Различают **невзвешенные и взвешенные вариационные** ряды.

Невзвешенным рядом называется упорядоченная совокупность **наблюденных** значений признака.

Упорядоченная по возрастанию совокупность интервалов (или классов) значений признака и соответствующих им частот называется **взвешенным интервальным вариационным рядом.**

СОСТАВЛЕНИЕ И ИЗОБРАЖЕНИЕ ВЗВЕШЕННОГО ВАРИАЦИОННОГО РЯДА (табличный и графический)

- 1) упорядочивают значения признака по возрастанию;
- 2) определяют размах варьирования признака
$$W_u = U_{max} - U_{min};$$
- 3) определяют число классов (интервалов) группирования по эмпирической формуле:

$$k = 1 + 4 * \lg N$$

N - объем выборки

- 4) определяют ширину интервалов группирования:
$$\Delta U = W_u / k = (U_{max} - U_{min}) / (1 + 4 * \lg N);$$

СОСТАВЛЕНИЕ И ИЗОБРАЖЕНИЕ ВЗВЕШЕННОГО ВАРИАЦИОННОГО РЯДА (табличный и графический)

- 5) выбирают границы классов и определяют середины интервалов группирования.

Нижняя граница 1-го класса - ***U_{min}***.

Верхняя граница 1-го класса - ***U_{min} + ΔU***;

- 6) подсчитывают количество значений признака в каждом классе - частота класса

- 7) составляют таблицу - табличный способ изображения ***взвешенного интервального вариационного ряда распределения.***

Класс со- держаний, %(границы классов)	Середины классов C_j	Число проб (частота) n_j	Накоплен- ные частоты n_{sj}	Частоты w_j	Накоплен- ные частоты w_{sj}
30-32	31	2	2	0,013605	0.013605
32-34	33	6	8	0,040816	0.054421
34-36	35	9	17	0,061224	0.115645
36-38	37	14	31	0,095238	0.210883
38-40	39	20	51	0,136054	0.346937
40-42	41	25	76	0,170068	0.517005
42-44	43	21	97	0,142857	0.659862
44-46	45	17	114	0,115646	0.775508
46-48	47	13	127	0,088435	0.863943
48-50	49	10	137	0,068027	0.93197
50-52	51	5	142	0,034014	0.965984
52-54	53	3	145	0,020408	0.986392
54-56	55	2	147	0,013605	0.999997
Сумма		147		1	

Сумма частот всех классов вариационного ряда равна

объему выборки, т.е. $\sum_{j=1}^k n_j = N$, где k - количество классов, N - объем выборки.

Частота w_j данного класса группирования $w_j = \frac{n_j}{N}$

$$\sum w_j = 1$$

Накопленная частота n_{sj} данного класса - сумма частот с первого по данный класс

$$n_{sj} = \sum_{j=1}^s n_j$$

$$w_{sj} = \frac{n_{sj}}{N}$$

Накопленная частота W_{sj}

Различают графики вариационных рядов:

Гистограмма (или столбчатая диаграмма) - график изменения частот признака по интервалам его группирования

Интегральная гистограмма - график изменения накопленных частот признака по интервалам его группирования

Полигон распределения - график, по оси абсцисс которого откладываются середины интервалов группирования $\hat{U}_j$, а по оси ординат – соответствующие им частоты n_j .

Аналогичный график для частостей –

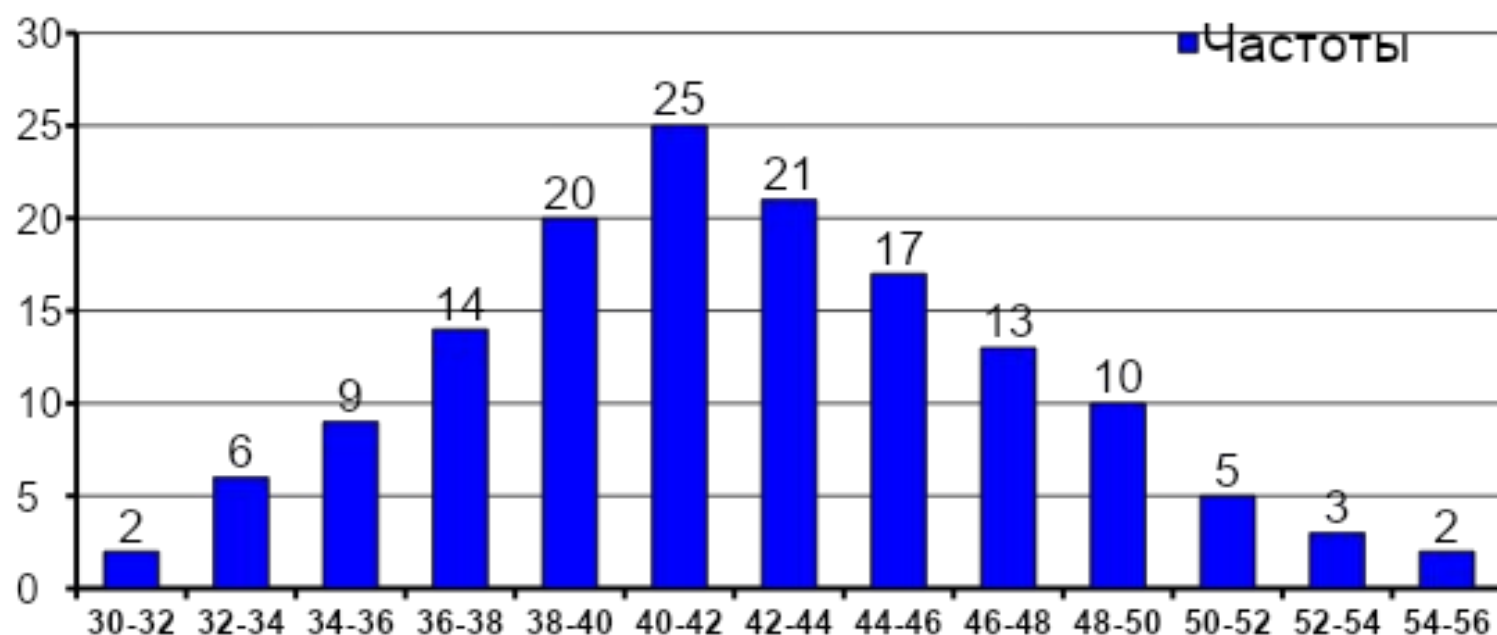
вариационная кривая, а для накопленных частостей - кумулятивная кривая (кумулята).

Построение гистограммы

- По оси **y** – функция плотности распределения **частота**
- По оси **x** - **интервалы** (классы) **значений** **случайной величины** (признака)

Гистограмма

n_j

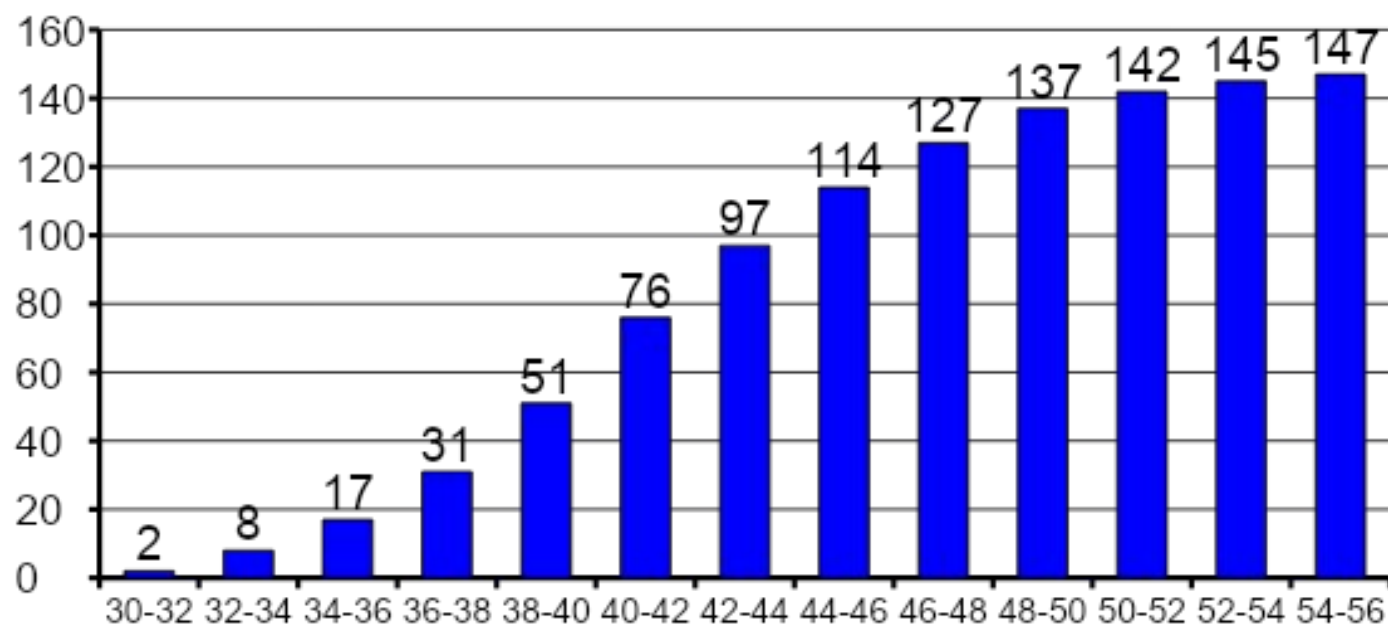


Границы классов, %
(интервалы группирования)

Интегральная гистограмма

n Sj

■ Накопленные частоты

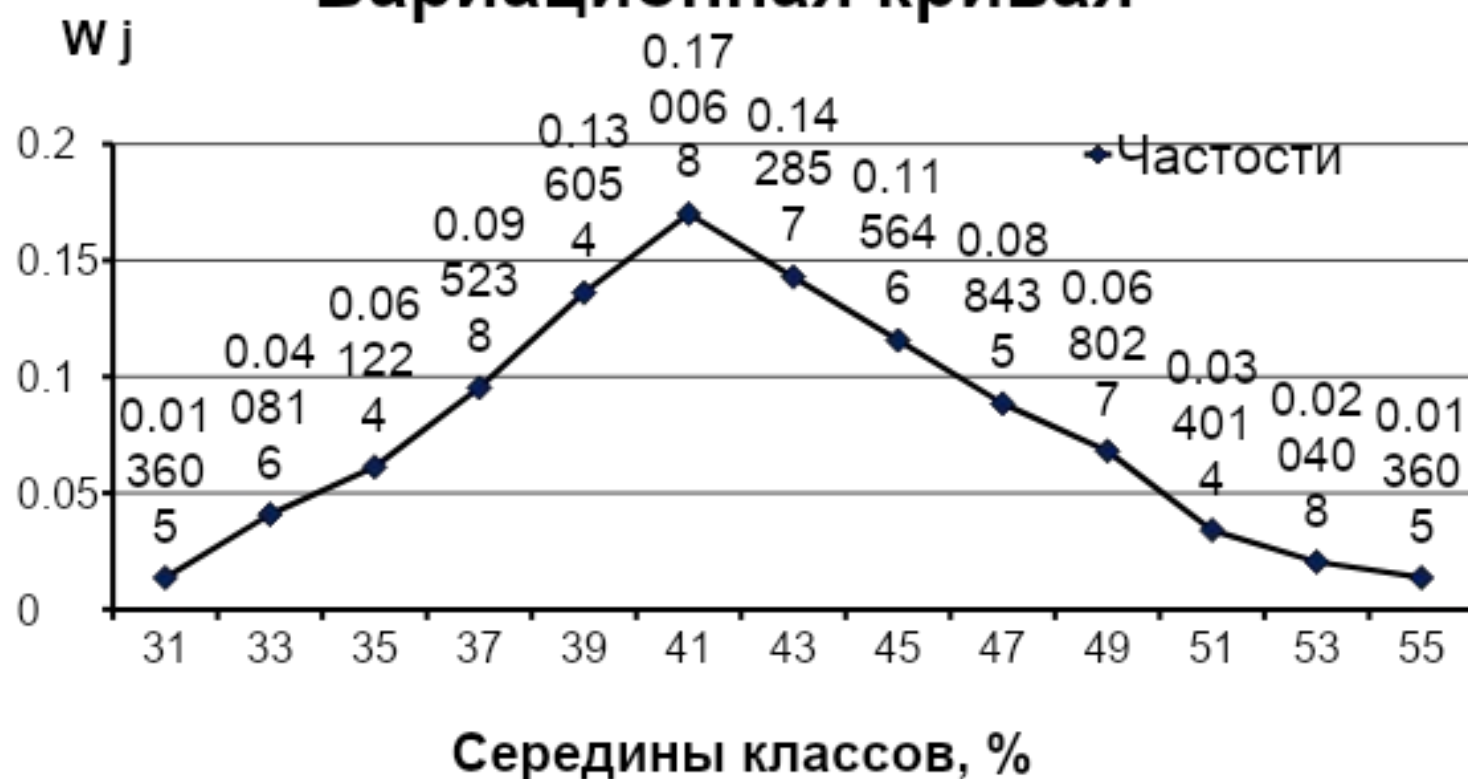


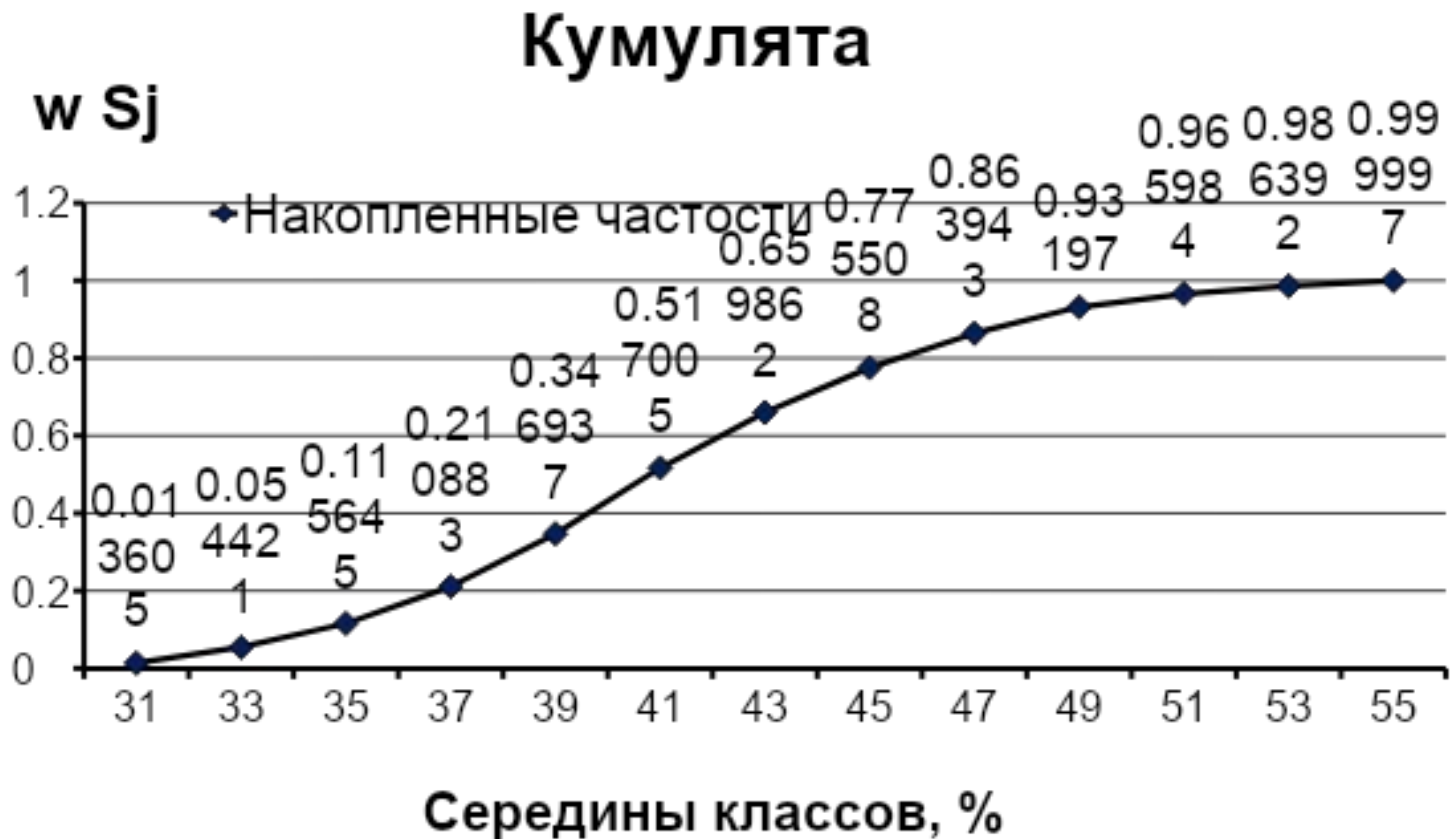
Границы классов, %

Полигон распределения



Вариационная кривая





Графическое изображение рядов - наглядно, но не полно.

Наиболее полным является аналитический способ исследования, при котором определяют **числовые характеристики вариационного ряда**.

СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ ВАРИАЦИОННОГО РЯДА

- 1) характеристики (меры) положения;
- 2) характеристики (меры) рассеяния.

Мерами положения называются характерные точки на оси абсцисс графика распределения, около которых группируется подавляющее число наблюдений (различные виды **средних, мода и медиана**).

Среднее арифметическое для невзвешенного ряда:

$$\bar{U} = \frac{1}{N} \cdot \sum_{i=1}^N U_i$$

среднее взвешенное - для взвешенного интервального

ряда:

$$\bar{U}_{взв} = \frac{1}{N} \cdot \sum_{j=1}^K U_j \cdot n_j$$

Для приближенной оценки среднего при первичной обработке данных используются непараметрические показатели:

мода и медиана.

Модой (M_o) называется то значение признака, которому соответствует максимальная частота или частость. Графически мода - это то значение признака U_i , которому соответствует максимум на вариационной кривой или полигоне распределения.

Вариационный ряд и вариационная кривая с 1 модой называются одномодальными, с 2 модами - бимодальными, с 3 модами и более - полимодальными.

Медианой (Me) – называется то значение признака, которое соответствует середине упорядоченного вариационного ряда, т.е. которое делит упорядоченный ряд на две равные по численности части. Графически медиану определяют по кумулятивной кривой - это то значение признака, которому соответствует накопленная частота на кумуляте $w_{sj}=0,5$.

Соотношение среднего, моды и медианы: у симметричных распределений

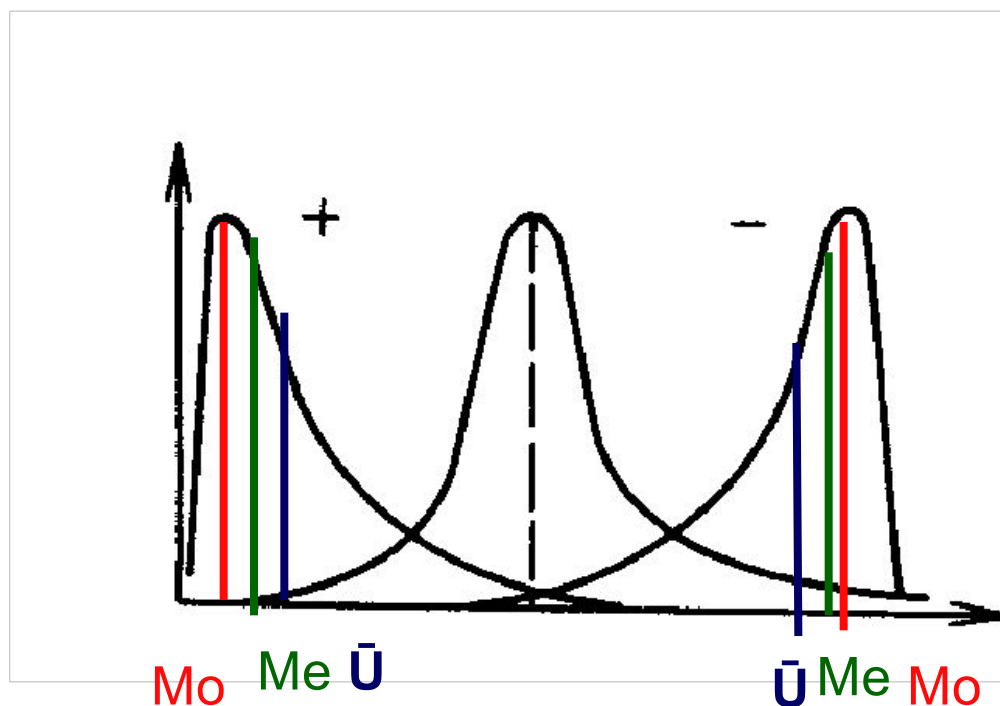
$$Mo = Me.$$

У правоасимметричных (положительно асимметричных)

рядов распределения $Mo < Me < \bar{U}$,

у левоасимметричных (отрицательно асимметричных):

$$Mo > Me > \bar{U}.$$

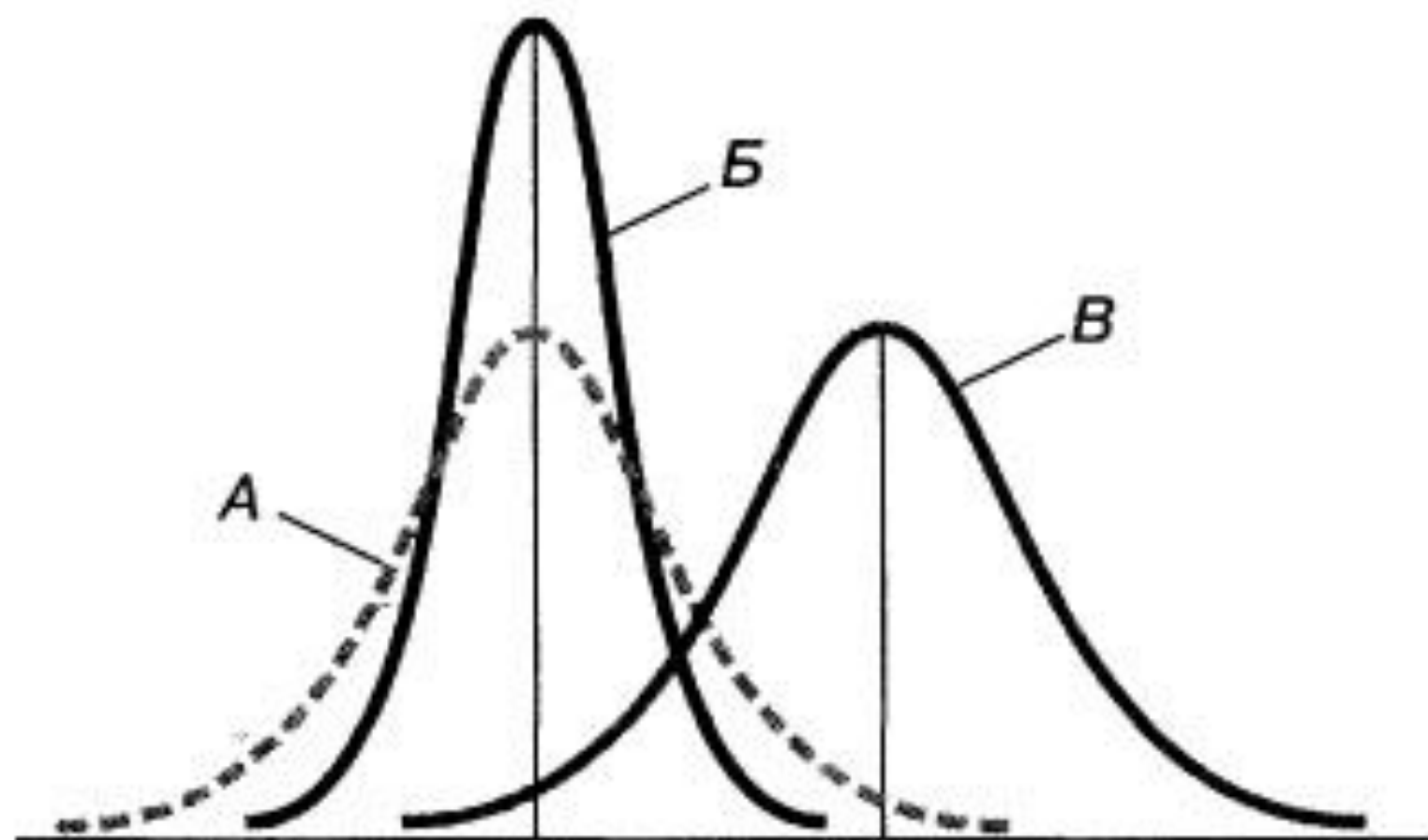


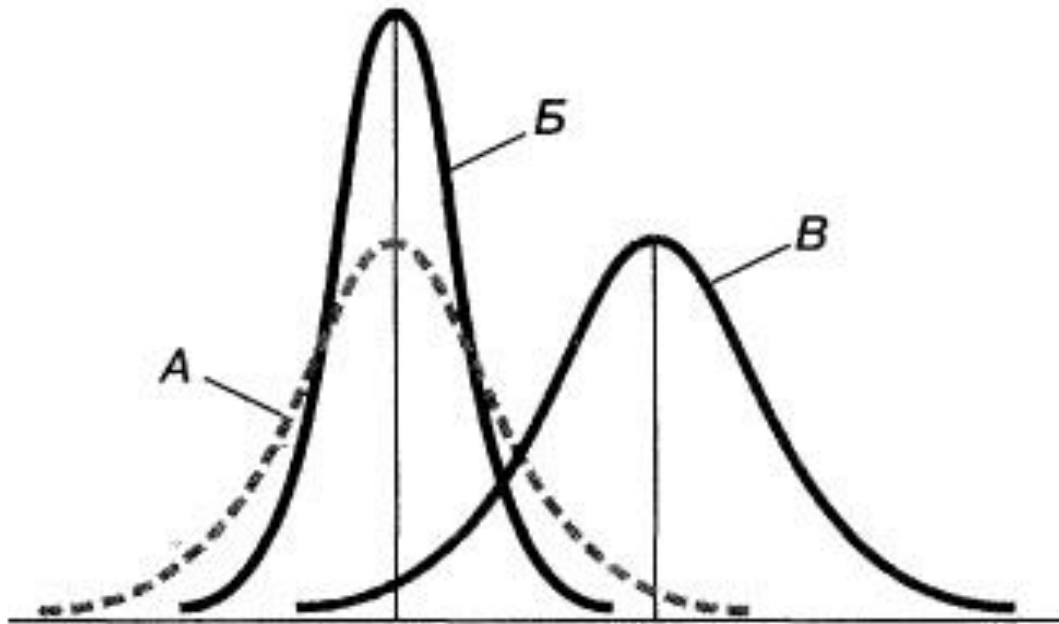
Мерами рассеяния называются статистические характеристики, которые указывают на степень и характер концентрации или рассеяния значений признака относительно мер положения.

**выборочная дисперсия,
стандартное отклонение,
коэффициент вариации,
выборочные показатели асимметрии и эксцесса.**

Выборочная дисперсия - средний квадрат отклонений значений признака от средней величины.

Графически дисперсия отражает сжатость или растянутость вариационной кривой или полигона распределения вдоль оси абсцисс.





- 1) Распределения A и Б имеют одинаковое математическое ожидание μ , но разные стандартные отклонения.
- 2) Распределения A и В имеют одинаковое стандартное отклонение σ , но разные математические ожидания.
- 3) Распределения Б и В имеют разные математические ожидания и стандартные отклонения.

для невзвешенного ряда и неупорядоченного

признака:
$$S^2 = \frac{1}{N} \cdot \sum_{i=1}^N (U_i - \bar{U})^2$$

для взвешенного интервального ряда:

$$S^2 = \frac{1}{N} \cdot \sum_{j=1}^k (\hat{U}_j - \bar{U})^2 \cdot n_j$$

S^2 - заниженная оценка истинной дисперсии
несмещенные оценки дисперсии:

$$S^2_{\text{несм.}} = \frac{1}{N-1} \cdot \sum_{i=1}^N (U_i - \bar{U})^2$$

$$S^2_{\text{несм.}} = \frac{1}{N-1} \cdot \sum_{j=1}^k (\hat{U}_j - \bar{U})^2 \cdot n_j$$

Для сравнения дисперсий разных признаков рассчитывают безразмерный параметр - коэффициент вариации V - относительное стандартное отклонение арифметическое

$$V = \frac{S}{\bar{U}} \cdot 100\%$$

По значениям коэффициента вариации V в геологии выделяют **5 типов распределений**:

- 1) весьма равномерное $V < 5\%$,
 - 2) равномерное $V = 5-40\%$,
 - 3) неравномерное $V = 40-100\%$
 - 4) весьма неравномерное $V = 100-150\%$,
 - 5) крайне неравномерное $V > 150\%$.
-

Выборочный **показатель асимметрии** указывает на характер и степень симметрии или асимметрии вариационной кривой.
для невзвешенного ряда и неупорядоченного признака:

$$A = \frac{\sum_{i=1}^N (U_i - \bar{U})^3}{N \cdot S^3}$$

для взвешенного интервального ряда:

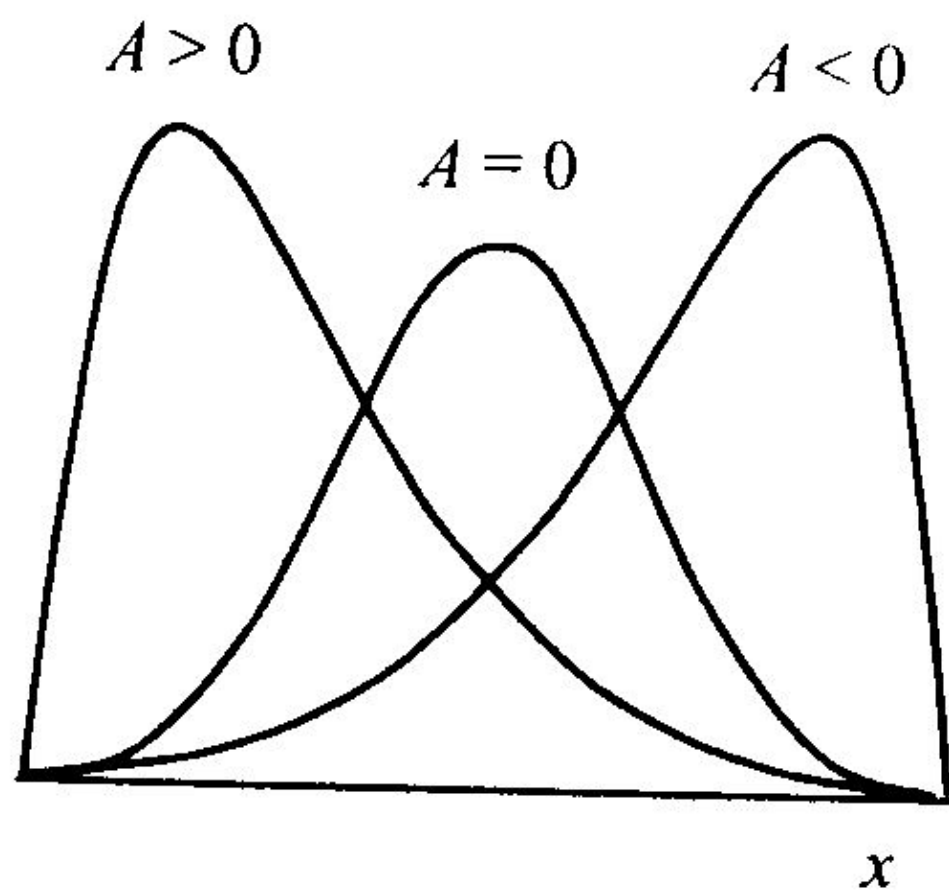
$$A = \frac{\sum_{j=1}^k (U_i - \bar{U})^3 \cdot n_j}{N \cdot S^3}$$

По степени асимметрии выделяют 4 типа вариационных кривых:

- 1) симметричные $A = 0$;
- 2) практически симметричные $|A| \leq 0,1$
- 3) слабо асимметричные $0,1 < |A| \leq 0,5$
- 4) сильно асимметричные $|A| > 0,5$

По характеру асимметрии различают 2 типа кривых:

- 1) правоасимметричные (положительно асимметричные) при $A > 0$;
 - 2) левоасимметричные (отрицательно асимметричные) при $A < 0$.
-



Выборочный **показатель эксцесса** характеризует степень крутовершинности или плосковершинности вариационной кривой по сравнению с кривой нормального распределения. Рассчитывается по формулам: для невзвешенного ряда:

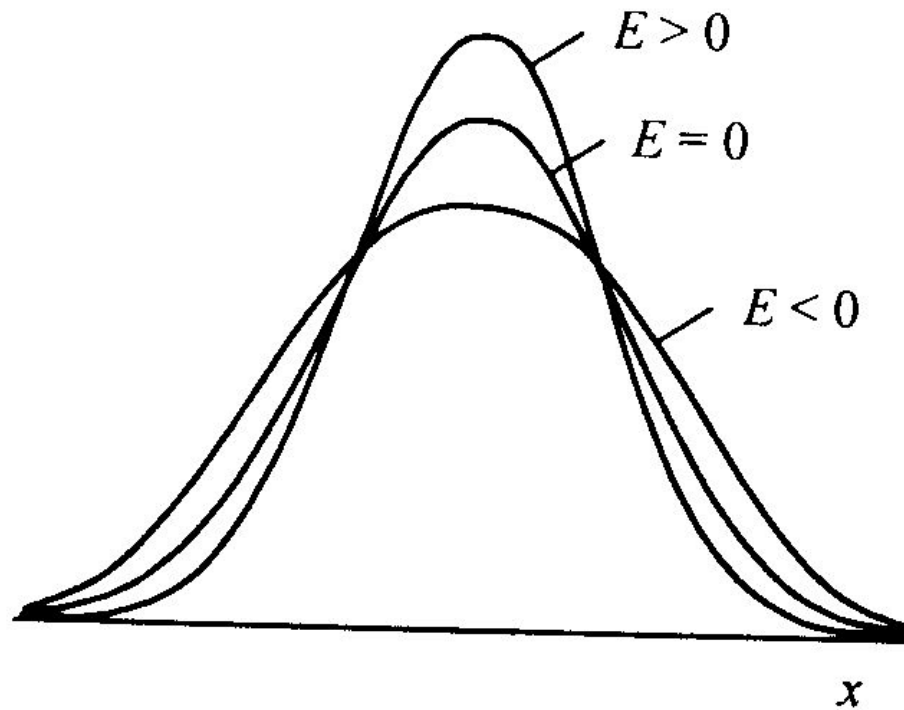
$$E = \frac{\sum_{i=1}^N (U_i - \bar{U})^4}{N \cdot S^4} - 3,$$

для взвешенного ряда

$$E = \frac{\sum_{j=1}^k (\hat{U}_j - \bar{U})^4 \cdot n_j}{N \cdot S^4} - 3$$

Коэффициент эксцесса (безразмерный параметр)

Знак эксцесса указывает на положение вершины вариационной кривой относительно вершины кривой нормального распределения.



ВИДЫ СТАТИСТИЧЕСКИХ МОМЕНТОВ

Статистическим моментом q-го порядка называется среднее из q-тых степеней отклонений значений признака от некоторой постоянной величины U_0 .

Для взвешенного интервального ряда рассчитывается по формуле:

$$M_q = \frac{1}{N} \sum_{j=1}^k (\hat{U}_j - U_0)^q \cdot n_j$$

где q - порядок момента (M_q), N - объем выборки, k - количество интервалов вариационного ряда, $\hat{U}_j$ - середина j -го интервала вариационного ряда, n_j - частоты.

Различают 5 видов моментов:

- 1) начальные m_q ;
- 2) центральные μ_q ;
- 3) основные v_q ;
- 4) смешанные;
- 5) условные $^K m_q$.

Начальными m_q называются моменты, в которых $U_0=0$, т.е. начало отсчета совмещается с нулем (началом координат).

1) 0 порядка ($q=0$): $m_0 = \frac{\sum \hat{U}_j^0 \cdot n_j}{N} = \frac{\sum n_j}{N} = \frac{N}{N} = 1;$

2) 1 порядка ($q=1$): $m_1 = \frac{\sum \hat{U}_j^1 \cdot n_j}{N} = \bar{U}$ - т.е. среднее взвешенное;

3) 2 порядка ($q=2$): $m_2 = \frac{\sum \hat{U}_j^2 \cdot n_j}{N} = \bar{U}^2$ - т.е. средневзвешенный квадрат;

4) 3 порядка ($q=3$): $m_3 = \frac{\sum \hat{U}_j^3 \cdot n_j}{N} = \bar{U}^3$ - т.е. средневзвешенный куб;

5) 4 порядка ($q=4$): $m_4 = \frac{\sum \hat{U}_j^4 \cdot n_j}{N} = \bar{U}^4$ - т.е. средневзвешенная четвертая степень.

Центральными называются **моменты**, в которых $U_0 = \bar{U}$, т.е. отклонения рассчитываются относительно средней величины ряда.

$$1) \text{ 0 порядка: } \mu_0 = \frac{\sum_{j=1}^k (\hat{U}_j - \bar{U})^0 \cdot n_j}{N} = \frac{\sum n_j}{N} = \frac{N}{N} = 1;$$

$$2) \text{ 1 порядка: } \mu_1 = \frac{\sum_{j=1}^k (\hat{U}_j - \bar{U})^1 \cdot n_j}{N} = \frac{\sum \hat{U}_j \cdot n_j}{N} - \frac{\bar{U} \sum n_j}{N} = \bar{U} - \bar{U} = 0$$

$$3) \text{ 2 порядка: } \mu_2 = \frac{\sum_{j=1}^k (\hat{U}_j - \bar{U})^2 \cdot n_j}{N} = S^2 \text{ - выборочная дисперсия;}$$

$$4) \text{ 3 порядка: } \mu_3 = \frac{\sum_{j=1}^k (\hat{U}_j - \bar{U})^3 \cdot n_j}{N} \text{ для расчетов коэфф. асимметрии;}$$

$$5) \text{ 4 порядка: } \mu_4 = \frac{\sum_{j=1}^k (\hat{U}_j - \bar{U})^4 \cdot n_j}{N} \text{ для расчетов коэфф. эксцесса.}$$

При расчетах центральных моментов чаще используют более простые формулы св начальных моментов с центральными:

$$\mu_2 = m_2 - (m_1)^2 ;$$

$$\mu_3 = m_3 - 3m_1m_2 + 2m_1^3 ;$$

$$\mu_4 = m_4 - 4m_1m_3 + 6m_1^2m_2 - 3m_1^4$$

Основные моменты представляют собой нормированные, т.е. приведенные к одному масштабу безразмерные величины. Они рассчитываются путем деления центральных моментов q -го порядка на стандартные отклонения в q -ой степени, соответствующей порядку момента.

1) основной момент 0 порядка:

$$v_0 = \frac{\mu_0}{S^0} = \frac{1}{1} = 1$$

2) основ, момент 1 порядка:

$$v_1 = \frac{\mu_1}{S^1} = \frac{0}{S^1} = 0$$

3) основной момент 2 порядка:

$$v_2 = \frac{\mu_2}{S^2} = \frac{S^2}{S^2} = 1 ;$$

4) 3 порядка:

$$v_3 = \frac{\mu_3}{S^3} = A$$

— выборочный коэффициент асимметрии;

5) 4 порядка:

$$v_4 = \frac{\mu_4}{S^4}$$

- для расчета эксцесса: $E = v_4 - 3$.

Условные моменты ${}^k m_q$ рассчитывают, когда начало отсчета совпадает с серединой какого-либо произвольно взятого интервала ($U_0 = \hat{U}_j$).

При большом объеме выборки статистики положения и рассеяния интервального вариационного ряда рассчитывают с помощью условных начальных моментов первых 4 порядков. Расчеты проводятся методом замены переменных в следующей последовательности:

Интервалы со- держаний желе- за, %	Середины интервалов $\hat{U}_j$	Часто- ты n_j	Услов- ные классы k_j	$n_j \cdot k_j$	$n_j \cdot k_j^2$	$n_j \cdot k_j^3$	$n_j \cdot k_j^4$
1	2	3	4	5	6	7	8
48-50	49	6	-2	-12	24	-48	96
50-52	51	10	-1	-10	10	-10	10
52-54	Me=53	15	0	0	0	0	0
54-56	55	13	1	13	13	13	13
56-58	57	6	2	12	24	48	96
Сумма		50		3	71	3	215

2) совмещают место нуля U_0 с серединой медианного интервала ($U_0 = \hat{U}_{Me} = U_3 = 53\%$),

определяют номера условных классов по формуле: $K_j = \frac{\hat{U}_j - U_0}{\Delta U}$ с учетом знака; при этом номер медианного класса будет равен 0, так как $K_{Me} = \frac{U_{Me} - U_{Me}}{\Delta U} = \frac{(53 - 53)}{2} = 0$, номера домедианных классов будут уменьшаться на 1, а послемедианных классов - увеличиваться на 1;

3) находят произведения частот на номера условных классов K_j в степени от 1 до 4 для каждого класса (столбцы 5-8), затем их суммы по столбцам 5-8, после чего - начальные условные моменты 4-х порядков по формулам:

$${}^k m_1 = \frac{j = \sum_{j=1}^k n_j k_j}{N} = \frac{3}{50} = 0,06; \quad {}^k m_2 = \frac{j = \sum_{j=1}^k n_j k_j^2}{N} = \frac{71}{50} = 1,42;$$
$${}^k m_3 = \frac{j = \sum_{j=1}^k n_j k_j^3}{N} = \frac{3}{50} = 0,06 \quad {}^k m_4 = \frac{j = \sum_{j=1}^k n_j k_j^4}{N} = \frac{215}{50} = 4,30$$

4) находят центральные моменты 2, 3 и 4 порядков, используя те же, что и раньше, формулы связи условных начальных моментов с центральными:

Класс со- держаний, %(границы классов)	Середины классов C_j	Число проб (частота) n_j	Накоплен- ные частоты n_{sj}	Частоты w_j	Накоплен- ные частоты w_{sj}	
30-32	31	2	2	0,013605	0.013605	
32-34	33	6	8	0,040816	0.054421	
34-36	35	9	17	0,061224	0.115645	
36-38	37	14	31	0,095238	0.210883	
38-40	39	20	51	0,136054	0.346937	
40-42	41	25	76	0,170068	0.517005	
42-44	43	21	97	0,142857	0.659862	
44-46	45	17	114	0,115646	0.775508	
46-48	47	13	127	0,088435	0.863943	
48-50	49	10	137	0,068027	0.93197	
50-52	51	5	142	0,034014	0.965984	
52-54	53	3	145	0,020408	0.986392	
54-56	55	2	147	0,013605	0.999997	
Сумма		147		1		