# ГЕОЛОГИЧЕСКИЕ ПРИЛОЖЕНИЯ ОДНОМЕРНОЙ СТАТИСТИЧЕСКОЙ МОДЕЛИ

## 1. Точечная оценка погрешности среднего значения

Среднее значение $\bar{x}$ из $n$ независимых значений случайной величины $x$ также является случайной величиной. Дисперсия $x$ $\sigma^2$, следовательно дисперсия среднего $\bar{x}$ $\delta^2$ в $n$ раз меньше:

$$\delta^2 = \sigma^2 / n \quad \text{или} \quad \delta = \sigma / \sqrt{n}$$

$\delta$ - *абсолютная* среднеквадратичная случайная погрешность среднего значения $\bar{x}$.

$$\tau = \frac{\delta}{\bar{x}} = \frac{\sigma}{\bar{x}^{*}\sqrt{n}} = \frac{V}{\sqrt{n}}$$ *относительная* погрешность

$V$ — коэффициент вариации. $\tau$ может быть выражена в долях единицы или в %.

Формулы $\delta$ и $\tau$ показывают, что погрешность среднего значения прямо пропорциональна изменчивости случайной величины и обратно пропорциональна корню квадратному из числа измерений.

| *Содержания меди ($Ci$ , %)* | | |
|---|---|---|
| Среднее | | 0.8742 |
| Стандартная ошибка | =станд.отклонение/50= | 0.028882719 |
| Медиана | | 0.865 |
| Мода | | 0.96 |
| Стандартное отклонение | | 0.204231661 |
| Дисперсия выборки | | 0.041710571 |
| Эксцесс | | -0.02813976 |
| Асимметричность | | 0.38519861 |
| Интервал | | 0.93 |
| Минимум | | 0.47 |
| Максимум | | 1.4 |
| Сумма | | 43.71 |
| Счет | | 50 |
| Уровень надежности(95.0%) | | 0.058041995 |

Это позволяет решать 2 задачи:

1) оценивать абсолютную или относительную погрешность при известном числе наблюдений *n*;

2) находить необходимое число измерений *n* для достижения заданной погрешности среднего значения.

**Пример** В результате анализа 16 проб гранита рассчитано среднее содержание кремнезема $\bar{x}$ = 70,35 % и среднеквадратичное отклонение

$\sigma$ = 3,20 %. Определить, чему равна среднеквадратичная погрешность среднего содержания и сколько дополнительно нужно взять проб, чтобы снизить относительную погрешность до 1 %.

Абсолютная среднеквадратичная случайная погрешность

$$\delta = 3{,}2 / \sqrt{16} = 0{,}80\%$$

относительная случайная погрешность

$\tau$ = 0,80/70,35= 1,14%.

$=\sigma/\bar{x}$

Если $\tau$ =1%=0,01, то из формулы $\tau = \dfrac{\delta}{\bar{x}} = \dfrac{V}{\sqrt{n}}$ получим

$\delta = \tau * \bar{x}$ = 0,01 • 70,35 = 0,70. Из формулы $\tau = \dfrac{\delta}{\bar{x}} = \dfrac{V}{\sqrt{n}}$ имеем

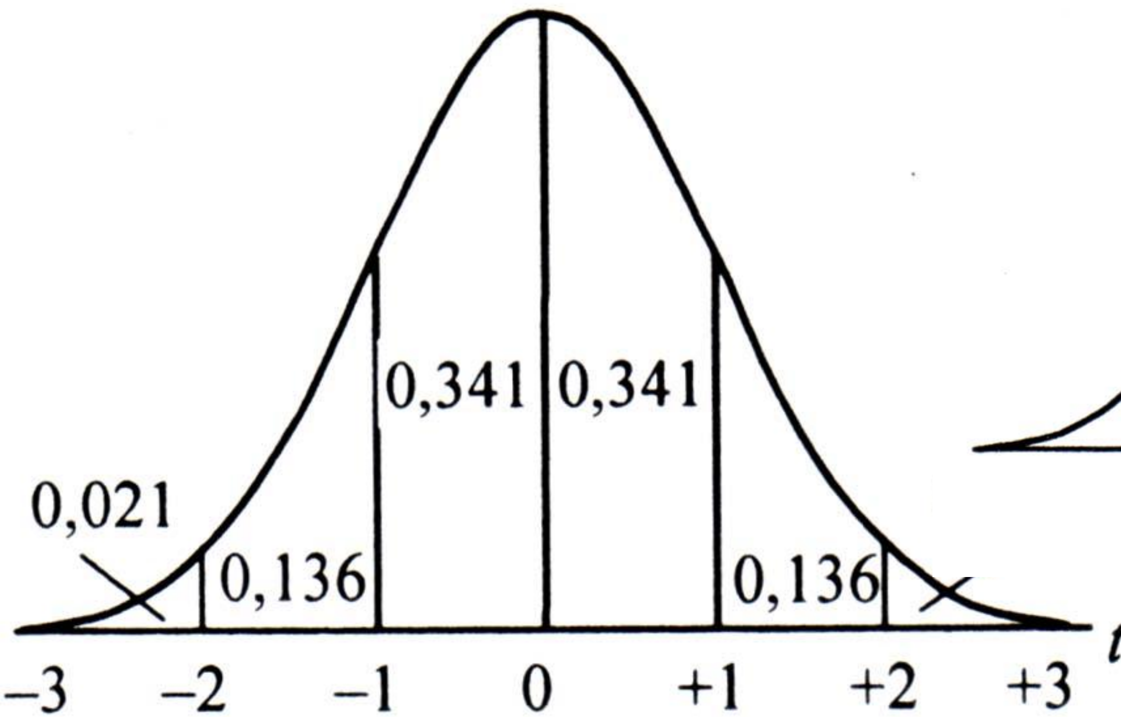
$n = \sigma^2 / \delta^2$ = $3{,}20^2$ / $0{,}70^2$ ≈ 21. Следовательно, дополнительно нужно взять и проанализировать 21 - 16 = 5 проб.

# 2. Интервальная оценка математического ожидания случайной величины

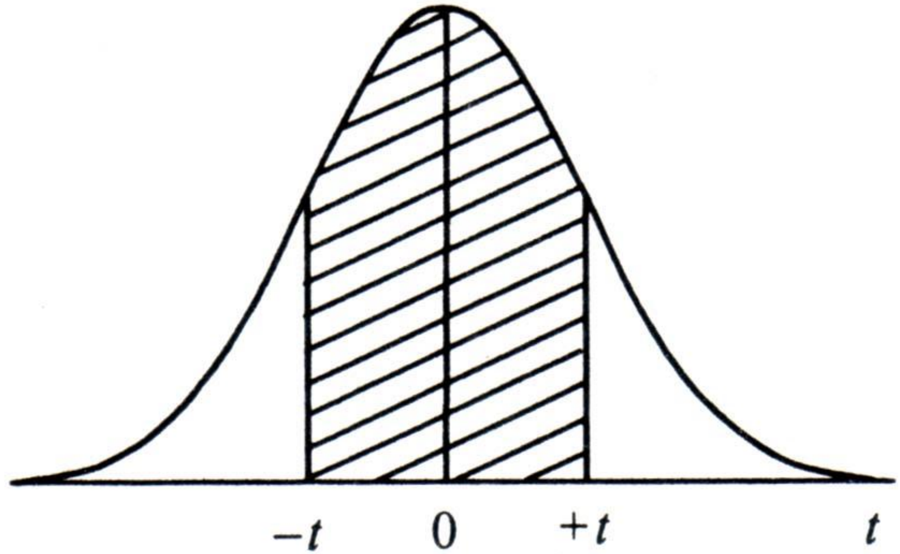
***M(x)*** в генеральной совокупности обычно неизвестно. Его можно приближенно оценить с помощью выборочного среднего значения $\bar{x}$, которое является случайной величиной и имеет дисперсию $\delta^2$. Предполагается, что случайная величина имеет распределение, близкое к нормальному. Размах значений нормально распределенной величины составляет приближенно $\pm 3\delta$. Где-то в этом интервале и заключено математическое ожидание ***M(x)***. Наиболее вероятно, что оно совпадает со средним значением $\bar{x}$, которое является ***точечной оценкой*** математического ожидания. Менее вероятно, что ***M(x)*** смещено в ту или иную сторону от среднего значения. Интервал возможных значений ***M(x)*** зависит от вероятности ***q* = Ф(t)** и выражается через коэффициент вероятности ***t*** соотношением

$$\bar{x} - t\delta \angle M(x) \angle \bar{x} + t\delta$$

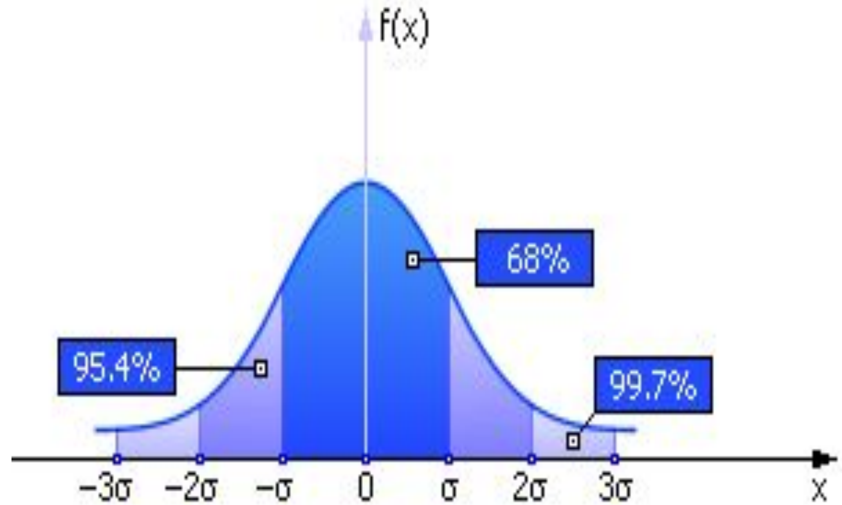
*доверительный интервал* или *интервальная оценка* математического ожидания



0,021
0,136
0,341
0,341
0,136
t
–3
–2
–1
0
+1
+2
+3

–t
0
+t
t

f(x)
68%
95,4%
99,7%
–3σ
–2σ
–σ
0
σ
2σ
3σ
x

Каждому значению вероятности $q$ соответствует определенный коэффициент вероятности $t$ и размер доверительного интервала:

| Вероятность $q = \Phi(t)$ | Коэффициент Вероятности $t$ | Доверительный интервал |
|---|---|---|
| 0.683 | 1 | $\bar{x} - \delta < M(x) < \bar{x} + \delta$ |
| 0.954 | 2 | $\bar{x} - 2\delta < M(x) < \bar{x} + 2\delta$ |
| 0.997 | 3 | $\bar{x} - 3\delta < M(x) < \bar{x} + 3\delta$ |

Используя данные примера, где среднее содержание кремнезема в граните $\bar{x}$ = 70,35 %, и $\delta$ = 0,80 %, получаем доверительные интервалы:

| Вероятность $q$ | Доверительный интервал |
|---|---|
| 0.683 | $69{,}55 < M(x) < 71{,}15$ |
| 0.954 | $68{,}75 < M(x) < 71{,}95$ |
| 0.997 | $67{,}95 < M(x) < 72{,}75$ |

Если $\bar{x}$ или другая оцениваемая величина подчиняются не нормальному закону распределения, то вероятность $q$ будет иная.

# 3. Выделение аномальных значений

- о разделении неоднородной совокупности на однородные,
- о выделении из неоднородных совокупностей аномальных значений.

Наиболее распространенный способ выделения аномальных значений (если известен или задан закон распределения случайной величины) - правило ***«трех сигм»*** основан на том, что случайная величина при нормальном законе распределения практически полностью (на 99,7 %) заключена в пределах от $\bar{x} - 3\delta$ до $\bar{x} + 3\delta$.

Если значение случайной величины отличается от $\bar{x}$ больше чем на $3\delta$, то оно является аномальным и не должно участвовать в расчете $\bar{x}$ и среднеквадратичного отклонения.

Для удобства случайную величину нормируют по формуле $$t = \frac{x - \bar{x}}{\sigma}$$

И правило «трех сигм» преобразуется:

если нормированное значение $|t| > 3$, то оно является аномальным.

**Пример.**

Средняя зольность угля $\overline{x}$ = 6,5 %, среднеквадратичное отклонение

$\sigma$ = 2,1%. Определить, не является ли аномальной проба угля с зольностью

15 %.

Найдем нормированное значение $\boldsymbol{t}$ =(15 -6,5)/2,1=4,05.

Поскольку $\boldsymbol{t}$ >3, проба является аномальной и относится к другой совокупности.

На основе приведенных данных можно определить, какие вообще значения зольности являются аномальными.

Так как $\overline{x} - 3\delta$ = 6,5 -3*2,1 = 0,2 %;

$\overline{x} + 3\delta$ = 6,5 + 3*2,1 = 12,8 %,

то аномальными являются значения зольности < 0,2 и > 12,8 %.

Аномальные значения можно выявить на графике пробит-функции, с помощью критерия Титьена-Мура.

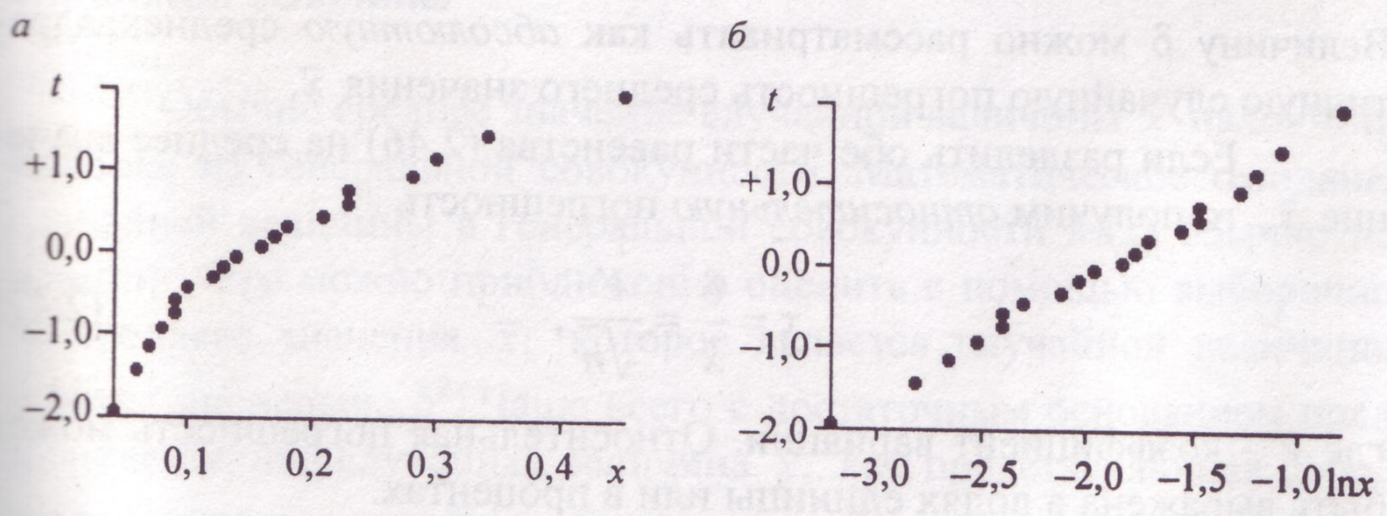



Рис.2.14. Графики пробит-функции для исходных (*а*) и преобразованных (*б*) данных

По оси ординат квантили соответствуют вероятностям p=(2n-1)/2N,   где n – порядковый № в упорядоченном ряду

# Выявление по критерию Титьена-Мура

$$L = \frac{N-n}{N}\frac{\sigma_{N-n}^2}{\sigma_N^2},$$

где $\sigma_N^2$ – дисперсия исходной совокупности; $\sigma_{N-n}^2$ – дисперсия после исключения $n$ предполагаемых аномальных значений.

Если значение $L$ окажется меньше критерия $L_{доп}$ при заданной вероятности $\alpha$, то исключенные значения являются аномальными. Для примера приведена табл.2.18 с вероятностью $\alpha = 0{,}05$ [14].

**Критерий Титьена – Мура при $\alpha = 0,05$**

| $N$ | Количество исключенных значений $n$ | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| 3 | 0,003 | | | | | | | | | |
| 4 | 0,051 | 0,001 | | | | | | | | |
| 5 | 0,125 | 0,018 | | | | | | | | |
| 6 | 0,203 | 0,055 | 0,010 | | | | | | | |
| 7 | 0,273 | 0,106 | 0,032 | | | | | | | |
| 8 | 0,326 | 0,146 | 0,064 | 0,022 | | | | | | |
| 9 | 0,372 | 0,194 | 0,099 | 0,045 | | | | | | |
| 10 | 0,418 | 0,233 | 0,129 | 0,070 | 0,034 | | | | | |
| 11 | 0,454 | 0,270 | 0,162 | 0,098 | 0,054 | | | | | |
| 12 | 0,489 | 0,305 | 0,196 | 0,125 | 0,076 | 0,042 | | | | |
| 13 | 0,517 | 0,337 | 0,224 | 0,150 | 0,098 | 0,060 | | | | |
| 14 | 0,540 | 0,363 | 0,250 | 0,174 | 0,122 | 0,079 | 0,050 | | | |
| 15 | 0,556 | 0,387 | 0,276 | 0,197 | 0,140 | 0,097 | 0,066 | | | |
| 16 | 0,575 | 0,410 | 0,300 | 0,219 | 0,159 | 0,115 | 0,082 | 0,055 | | |
| 17 | 0,594 | 0,427 | 0,322 | 0,240 | 0,181 | 0,136 | 0,100 | 0,072 | | |
| 18 | 0,608 | 0,447 | 0,337 | 0,259 | 0,200 | 0,154 | 0,116 | 0,086 | 0,062 | |
| 19 | 0,624 | 0,462 | 0,354 | 0,277 | 0,209 | 0,168 | 0,130 | 0,099 | 0,074 | |
| 20 | 0,639 | 0,484 | 0,377 | 0,299 | 0,238 | 0,188 | 0,150 | 0,115 | 0,088 | 0,066 |
| 25 | 0,696 | 0,550 | 0,450 | 0,374 | 0,312 | 0,262 | 0,222 | 0,184 | 0,154 | 0,126 |

# 4. Выделение однородных совокупностей

Заключение о неоднородности совокупности лучше всего делать по гистограмме частот. Геологическая причина появления двух совокупностей заключается в том, что бедные руды возникли путем замещения алюмосиликатных пород, а богатые - карбонатных пород.

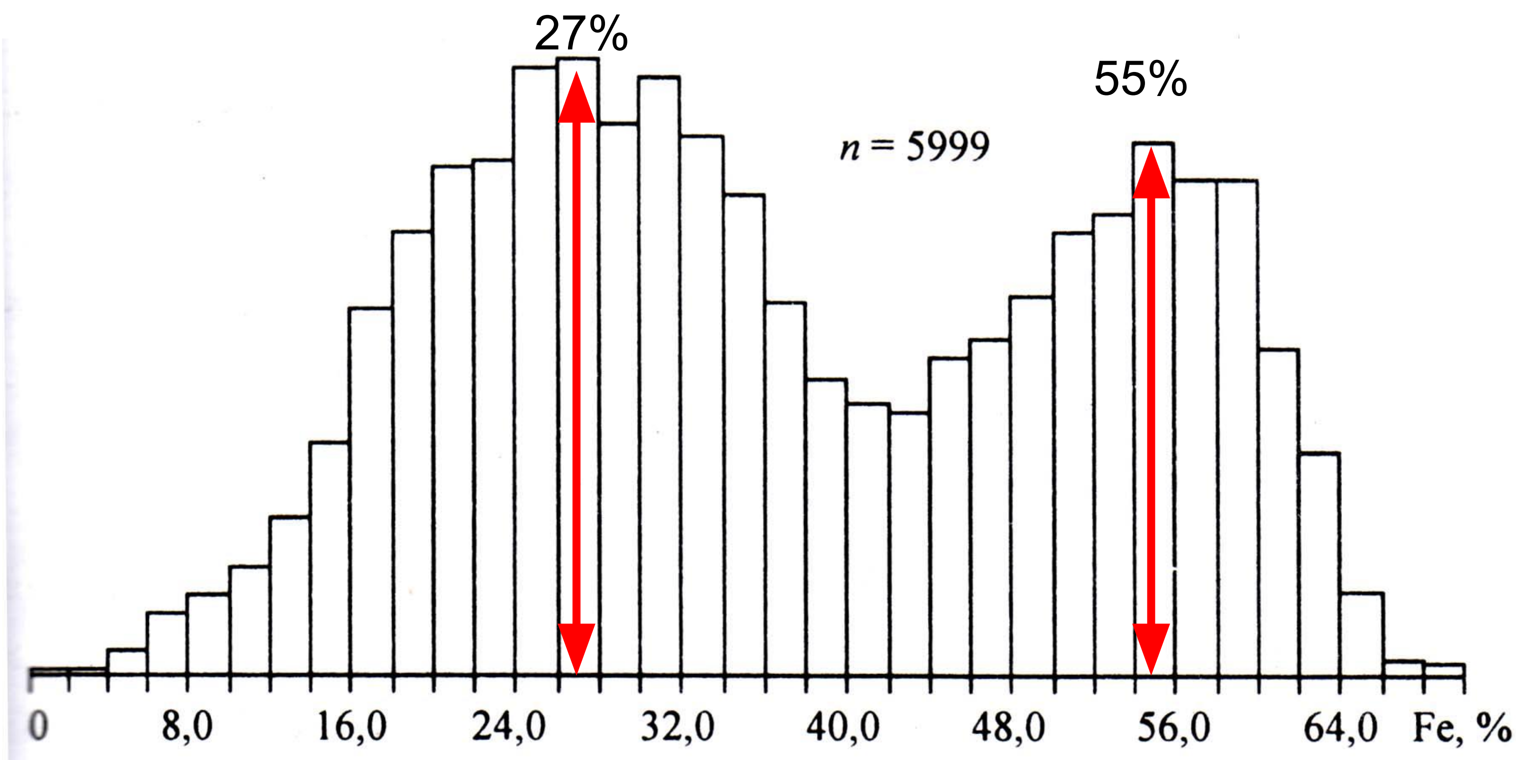



Рис.2.17. Гистограмма содержаний железа в рудах Качарского месторождения

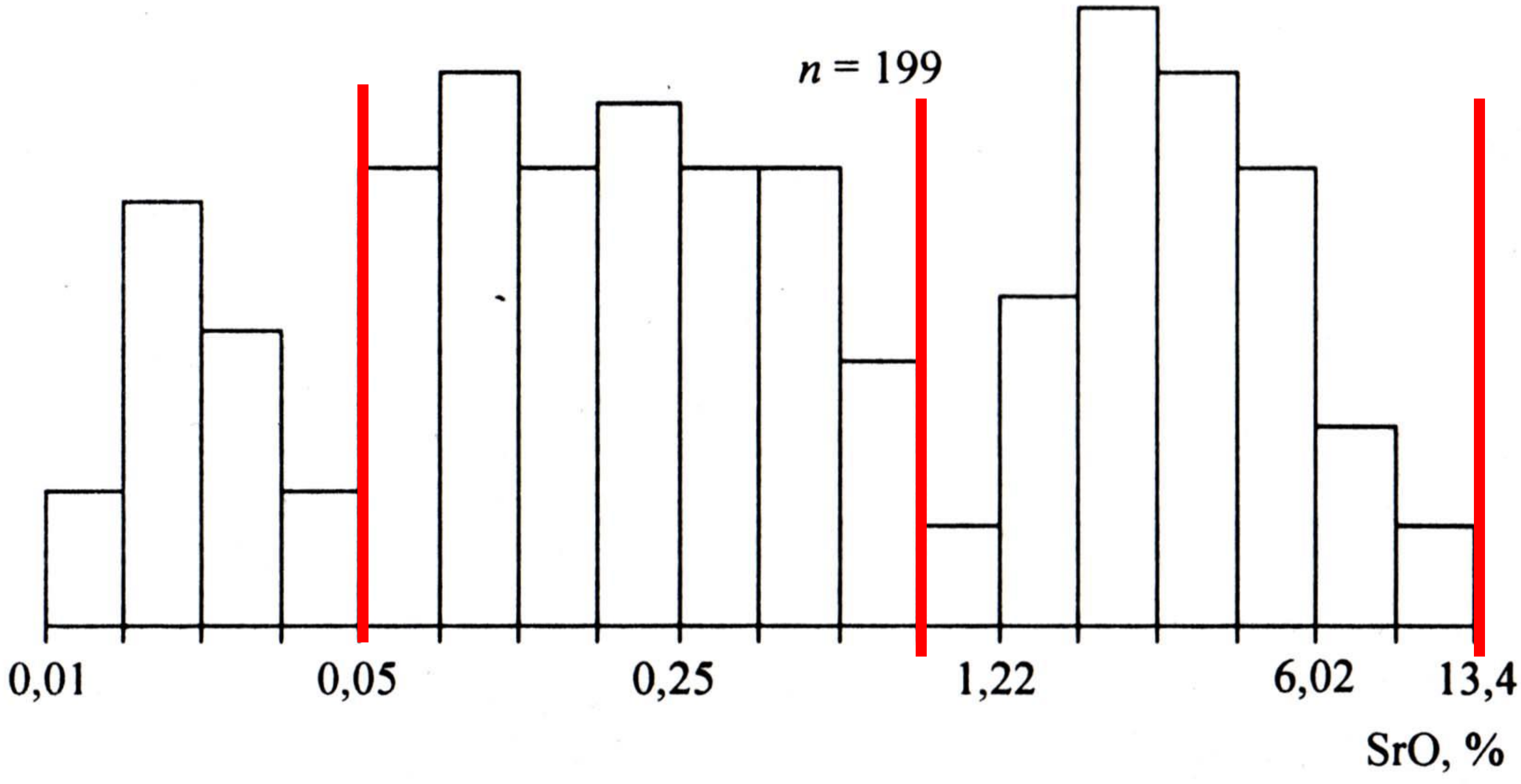
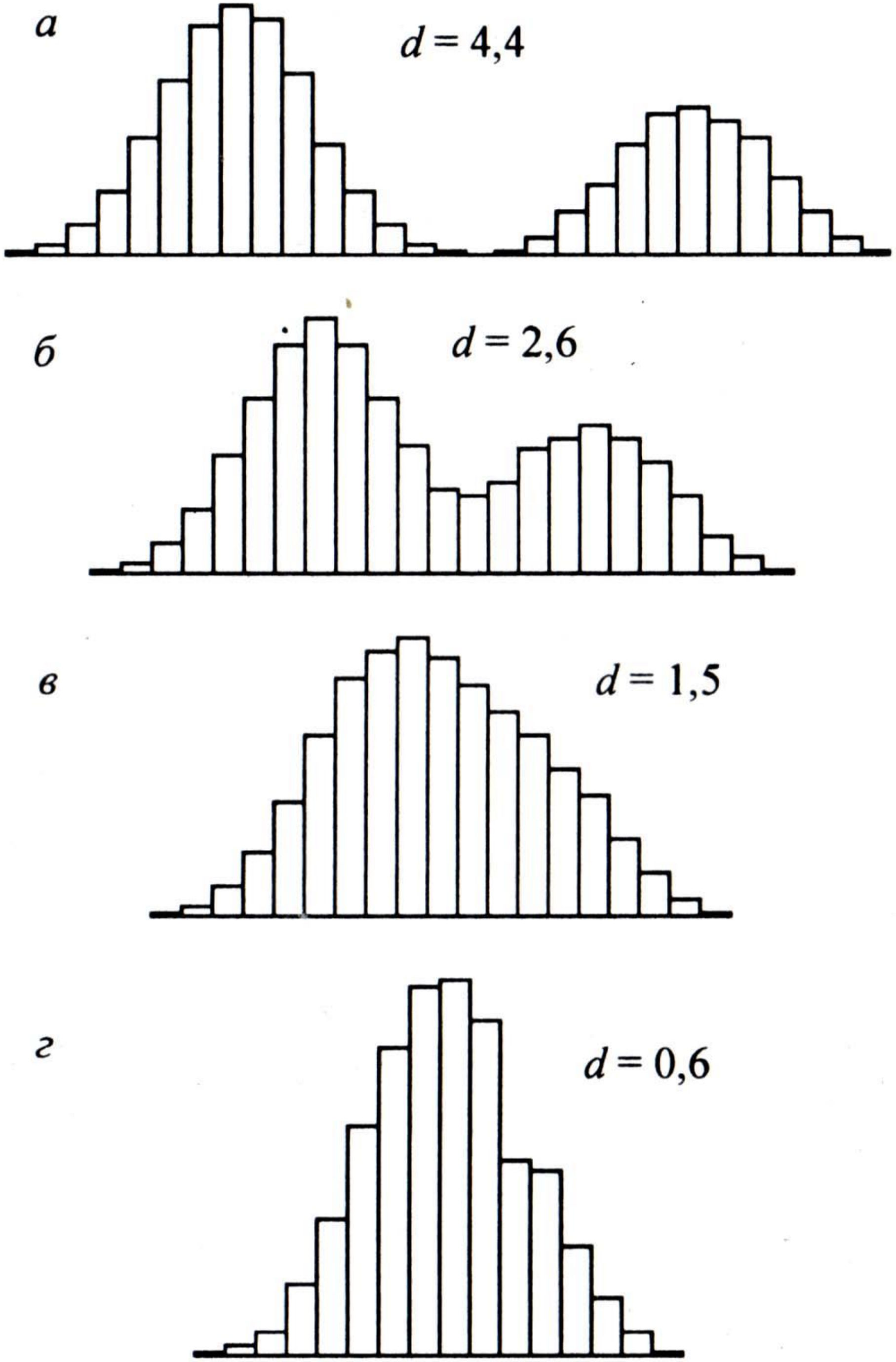



Рис.2.18. Гистограмма содержания стронция в апатитах различных природных образований. Масштаб по горизонтальной оси логарифмический

Наиболее чистыми по содержанию стронция являются апатиты из гранитоидов, ультрабазитов и метаморфических пород. Средние по содержанию стронция - апатиты скарновых месторождений и некоторых массивов щелочных пород. Наиболее высокие содержания стронция наблюдаются в апатитах Хибинской группы месторождений.




Однородные совокупности, входящие в смешанную совокупность, различаются средними значениями $\bar{x}$, $\bar{y}$ и дисперсиями $\sigma_x^2$, $\sigma_y^2$.

Показателем, определяющим возможность аналитического разделения смешанных совокупностей при условии нормального их распределения, является ***раздвиг*** распределений:

$$d = \frac{\left|\bar{x} - \bar{y}\right|}{\sqrt{\sigma_x^2 + \sigma_y^2}}$$

1. очень большой (d>4)
2. большой ($d$ = 2÷4)
3. малый ($d$= 0,7÷2)
4. незначительный (d<0,7)