

Кафедра медицинской и биологической физики с курсом информатики

Лекции «Физика, математика»

Раздел 1. Основы математического анализа, биомеханика, акустика

*Статистический анализ выборок.
Числовые характеристики ряда.*

Вопросы:

1. Основные понятия математической статистики.
2. Числовые характеристики выборки.
3. Нормальное распределение. Функция Гаусса.

Доцент, к.б.н. – Цокова
Татьяна Николаевна

Значение темы в системе знаний врача■

Работники здравоохранения поставляют основную массу данных, на которых базируется медицинская статистика. Поэтому им следует знать, как эти данные могут и должны использоваться, для того чтобы, с одной стороны, повысить уровень своей работы, а с другой - улучшить организацию медицинской помощи в своей стране.

1. Основные понятия математической статистики.

Выборочный метод –

изучение большой совокупности объектов относительно некоторого количественного признака X производится по сравнительно небольшому числу случайно отобранных объектов.

Основные понятия математической статистики.

Множество всех возможных значений (вариант)- x_i^* называется **генеральной совокупностью**, а число N – **объемом** генеральной совокупности.

Относительной частотой называется отношение частоты n_i к объему выборки n :

$$p_i^* = n_i / n.$$

Основные понятия математической статистики.

Выборкой (x_1, x_2, \dots, x_n) называется совокупность значений СВ X , полученной в результате n независимых экспериментов.

(x_1, x_2, \dots, x_n) - ***простой статистический ряд.***

Основные понятия математической статистики.

Объемом совокупности называется количество объектов в совокупности.

Объем выборки n , как правило, значительно меньше объема N генеральной совокупности: $n \ll N$.

Основные понятия математической статистики.

Определение 1 Закон статистического распределение

X	x_1	x_2	x_3	\dots	x_n
$P(X)$	p_1	p_2	p_3	\dots	p_n

$$p_1 + p_2 + \dots + p_k = \sum_{i=1}^k p_i = 1.$$

Основные понятия математической статистики.

Определение 2.

Математическим ожиданием дискретной случайной величины X называется сумма произведений всех возможных значений случайной величины на соответствующие вероятности появления этих значений, т.е.

$$M(X) = \sum_{i=1}^n x_i p_i = x_1 p_1 + x_2 p_2 + \dots + x_n p_n = a.$$

Математическое ожидание -

- центр распределения;
- самое ожидаемое значение ряда, т.е. вероятность его максимальная;
- среднее значение дискретной случайной величины, т.е.

$$M(X) = \bar{X}$$

Основные понятия математической статистики.

Определение 3.

Математическое ожидание квадрата отклонения случайной величины X от её математического ожидания называют **дисперсией** случайной величины X и обозначают $D(X)$, т.е.

$$D(X) = \sum_{i=1}^n (x_i - a)^2 p_i = (x_1 - a)^2 p_1 + (x_2 - a)^2 p_2 + \dots + (x_n - a)^2 p_n$$

Дисперсия-

- мера рассеивания значений случайной величины от среднего значения;
- имеет размерность квадрата случайной величины;

Основные понятия математической статистики.

Определение 4.

Среднее квадратическое отклонение случайной величины, является арифметическим корнем из дисперсии.

$$\sigma(X) = \sqrt{D(X)}.$$

- характеристика рассеяния в единицах признака X

2. Основные числовые характеристики выборки.

Существуют точечные и интервальные оценки параметров.

Определение 5.

Точечной оценкой φ^* параметра φ называют число, которое находят по функции результатов наблюдения $\varphi^* = \varphi^*(x_1, x_2, \dots, x_n)$, дает приближенное значение теоретического параметра φ .

Требования к числовым оценкам ряда.

Она должна быть:

Несмещённой – если её математическое ожидание равно оцениваемому параметру при любом объёме выборки.

Эффективной. Эффективность характеристика точности – отношение дисперсии наилучшей оценки и данной несмещённой оценки. Наилучшая оценка – оценка с наименьшей дисперсией.

Состоятельной, если сходится по вероятности к параметру φ , т.е. $\lim P(|\varphi^* - \varphi| < \Delta) = 1$ при $n \rightarrow \infty$. Это равенство означает, что при достаточно больших n , φ^* отличается от φ на величину меньшую, чем произвольное число Δ .

Основные числовые характеристики выборки

1. **Средняя выборочная** (среднее взвешенное значение признака в выборке):

$$\bar{X}_B = \frac{\sum_{i=1}^K x_i n_i}{n} = \frac{1}{n} \cdot (x_1 \cdot n_1 + x_2 n_2 + \dots + x_k n_k). \quad (1)$$

Основные числовые характеристики выборки

1. **Дисперсия выборочная.** Характеризует разброс (рассеяние) значений вариант x_i от выборочного среднего значения \bar{x}_B и измеряется в квадратных единицах признака X :

$$D_B = \frac{1}{n} \cdot \sum_{i=1}^k (x_i - \bar{x}_B)^2 \cdot n_i = \frac{1}{n} \left[(x_1 - \bar{x}_B)^2 n_1 + (x_2 - \bar{x}_B)^2 n_2 + \dots + (x_k - \bar{x}_B)^2 n_k \right] \quad (2)$$

Основные числовые характеристики выборки

На практике используют ещё одну формулу для расчёта дисперсии:

$$D_{\theta} = \overline{x^2} - (\overline{x})^2, \quad (3)$$

где

$$\overline{x} = \frac{1}{n} \cdot \sum_{i=1}^k x_i n_i; \quad \overline{x^2} = \frac{1}{n} \cdot \sum_{i=1}^k x_i^2 n_i.$$

Основные числовые характеристики выборки

1. Среднее квадратичное отклонение выборки – характеристика рассеяния значений признака в выборке от среднего выборочного в единицах признака X :

$$\sigma_v = \sqrt{D_v} . \quad (4)$$

Оценки имеют следующий вид (если $n < 30$)

$$\bar{x} \approx \bar{x}_v ;$$

$$D \approx \frac{n}{n-1} \cdot D_v = S_v^2 ; \quad (5)$$

$$\sigma \approx S_B = \sqrt{\frac{n}{n-1} D_B} , \quad (6)$$

где S_B^2 - исправленная выборочная дисперсия.

Исправленная выборочная дисперсия:

$$S^2(X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1} \quad (7) - (9)$$

$$S^2(X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2 m_i}{n-1} \quad \text{где } m_i \text{ - частота.}$$

$$S^2(X) = \frac{n\sigma^2(x)}{n-1}$$

Основные числовые характеристики выборки

Определение 6.

Доверительным интервалом φ^* основе точечной оценки называется интервал, который покрывает параметр φ с заданной надежностью P ,

$$P(|\varphi^* - \varphi| < \Delta) = p$$

*Доверительный интервал для нормально
распределённой случайной величины*

при $n < 30$:

$$M(X) = \bar{X} \pm \frac{t_{n,p} S}{\sqrt{n}}, \quad (10)$$

где $t_{n-1,p}$ -коэффициент Стьюдента (таблица),
 p - доверительная вероятность, равная 0,95.

при $n > 30$:

$$M(X) = \bar{X} \pm 1,96 \frac{\sigma}{\sqrt{n}}; \quad (11)$$

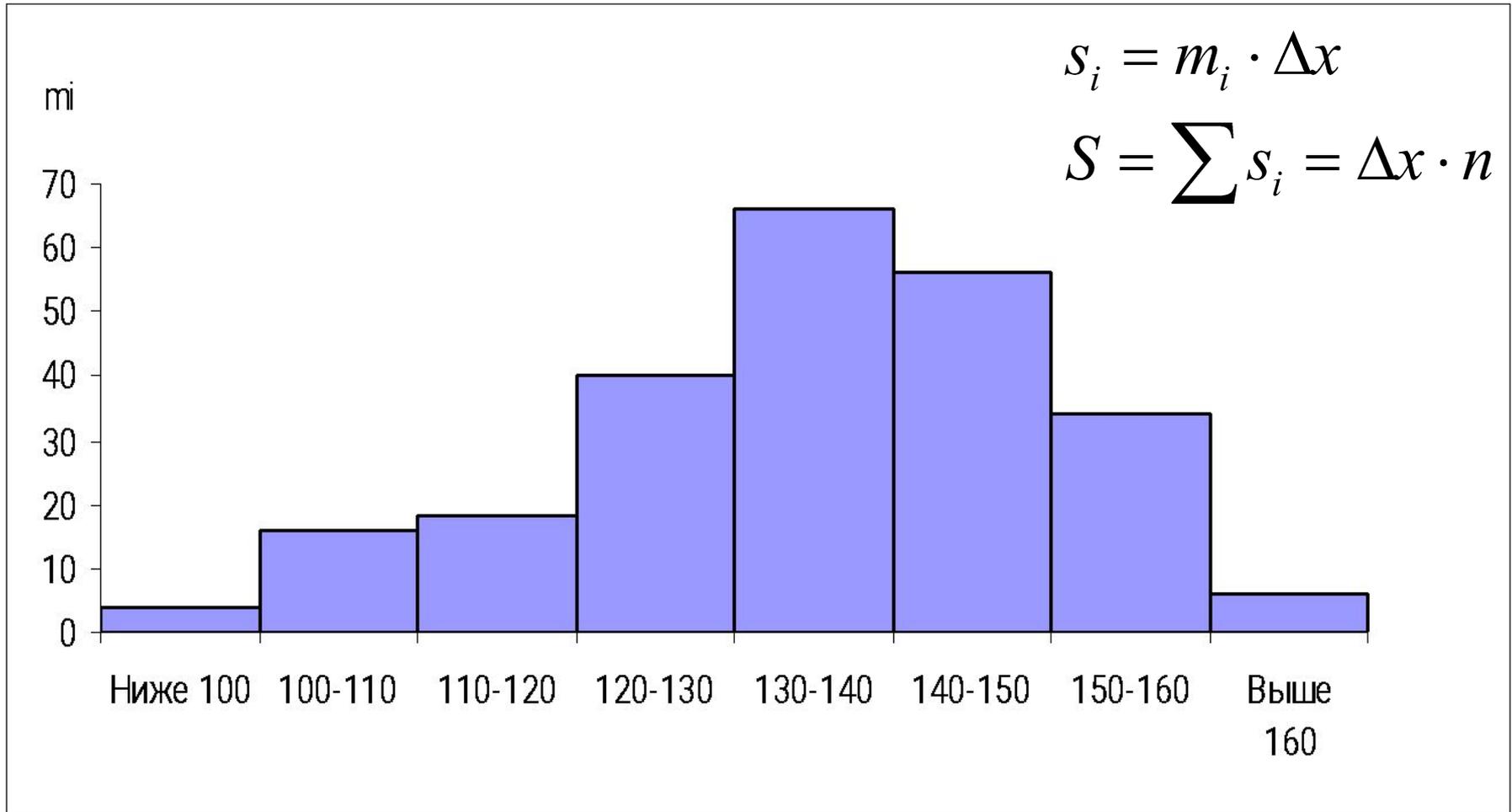
3. Нормальное распределение. Функция Гаусса. Гистограмма

Определение 7.

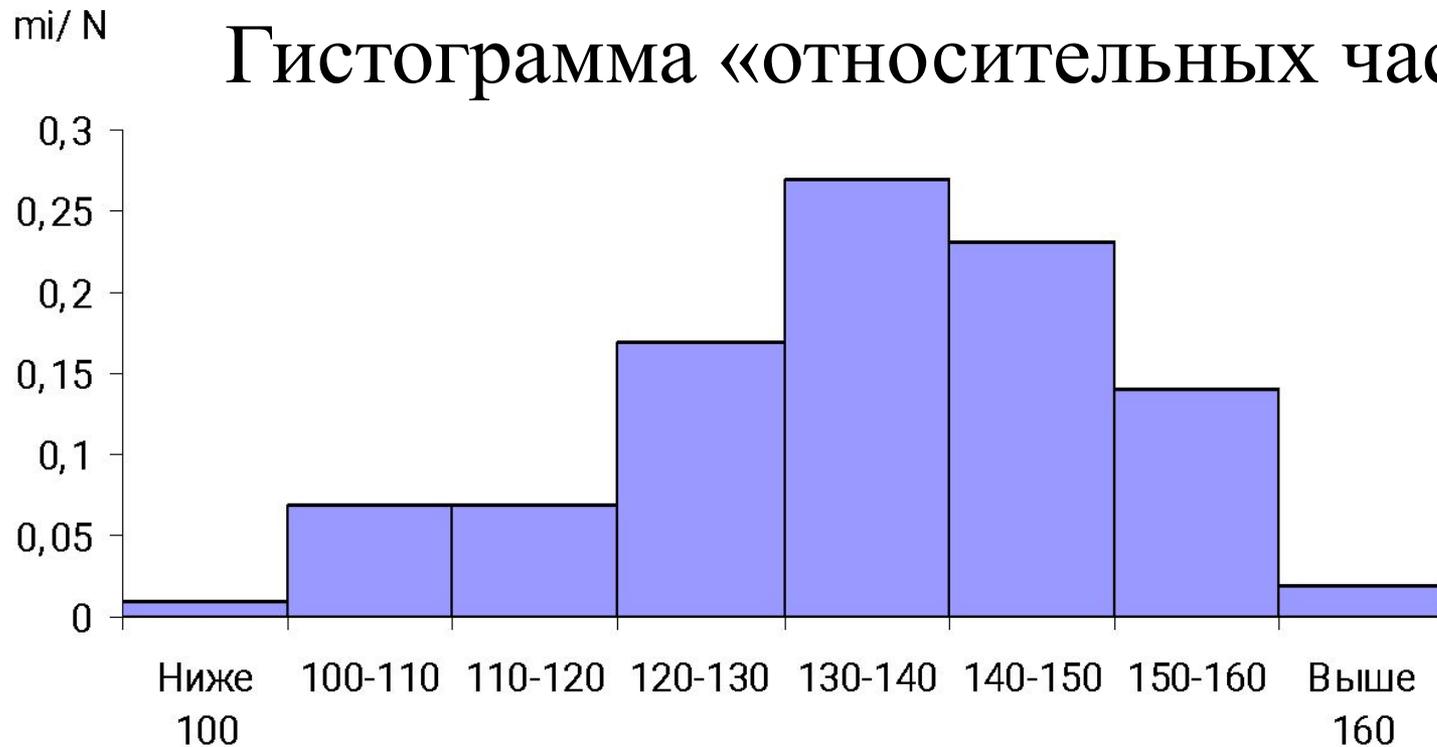
Гистограмма – графическое представление частотного распределения количественной случайной величины, сгруппированной в классы равной ширины площадями прямоугольников.

Высоты каждого прямоугольника пропорциональны частотам классов, а ширина интервала, одинаковая для всех.

Гистограмма «частот»



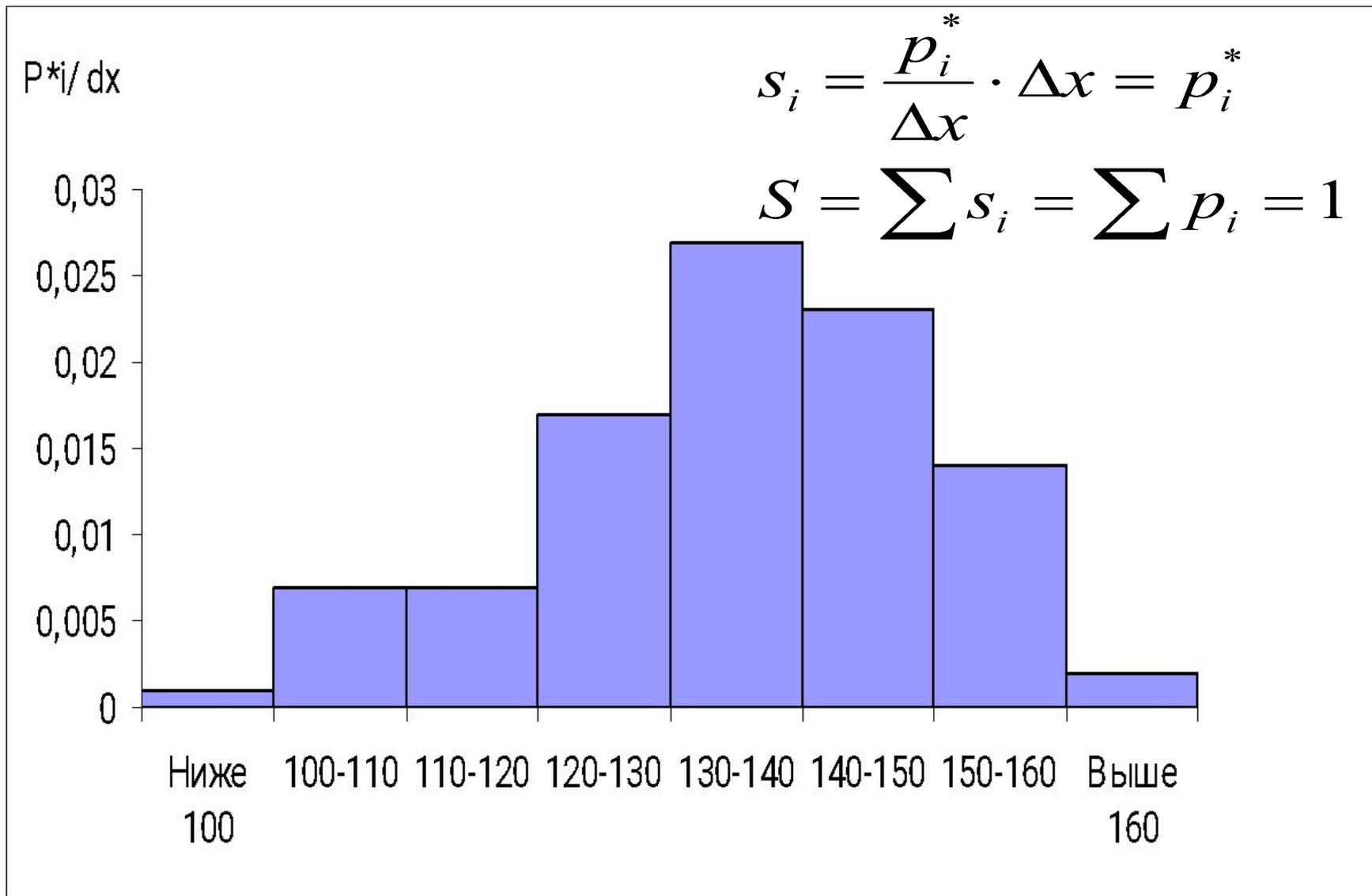
Гистограмма «относительных частот»



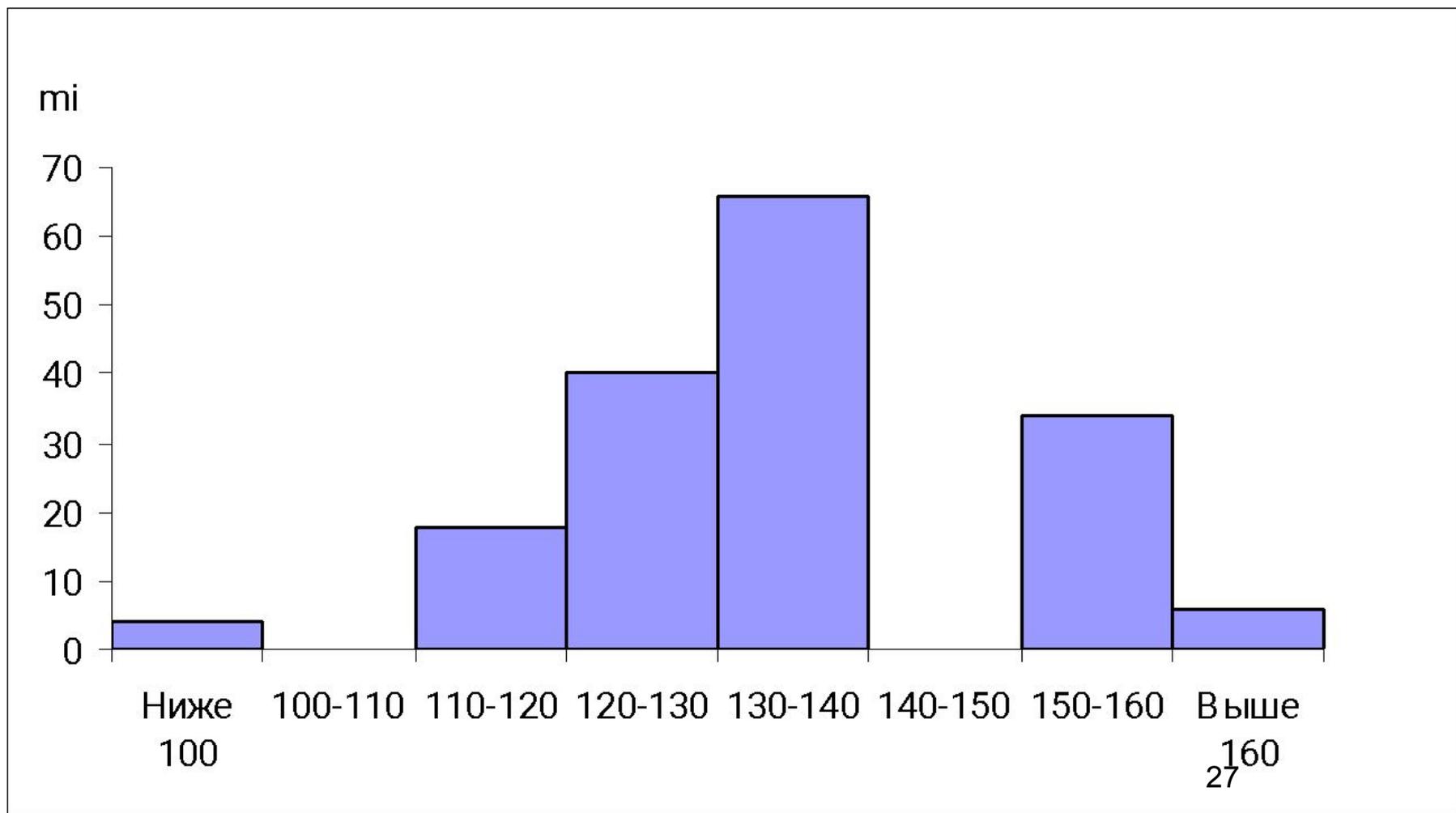
$$s_i = \frac{m_i}{n} \cdot \Delta x$$

$$S = \sum s_i = \Delta x \cdot \frac{\sum m_i}{n} = \Delta x \cdot \frac{n}{n} = \Delta \tilde{\sigma}$$

Гистограмма «приведённых частот»



Замечание. Вид гистограммы зависит от Δx , т.е. от числа интервалов m . Если m велико, гистограмма имеет «гребенчатый» вид. Если m – мало, потеря особенностей, т.е. имеем равномерное распределение.



Согласно центральной предельной теореме закон распределения суммы большого числа независимых СВ, влияние каждого из которых на всю сумму ничтожно мало, близок к нормальному.

А.М.Ляпунов (1857 – 1918)

Нормальное распределение (распределение Гаусса).

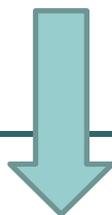
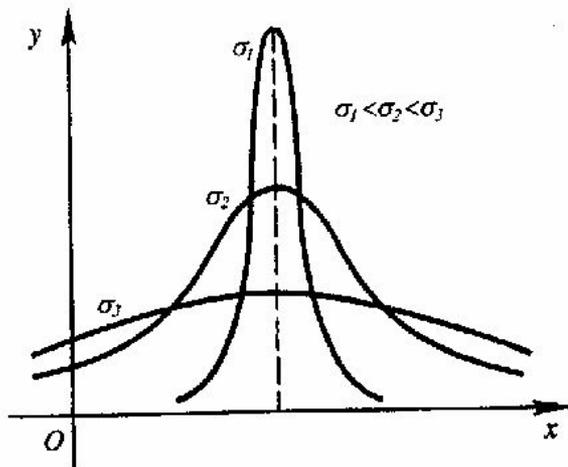
$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\bar{x})^2}{2 \cdot \sigma^2}} \quad \text{функция Гаусса.}$$



Это симметричная, быстро затухающая функция.

Имеет куполообразную форму с вершиной при $x = \bar{x}$,

где \bar{x} – определяет точку максимума и ось симметрии, σ - расстояние от этой оси до точки перегиба ($M(X) = \bar{x}$, $D(X) = \sigma$).



функция Гаусса.

$$f(x) = \frac{1}{\sigma \cdot \sqrt{2\pi}} \cdot e^{-\frac{(x-\bar{x})^2}{2 \cdot \sigma^2}}$$



Свойства нормального распределения:

- а) Наиболее вероятны значения x , близкие к ожидаемому среднему значению
- б) Отклонения от среднего значения в обе стороны равновероятны.
- в) Большие отклонения x от среднего значения маловероятны.

Свойства нормального распределения

- г) При уменьшении σ увеличивается вероятность значений, близких к μ , рассеяние уменьшается, кривая сжимается.
- д) При увеличении σ график кривой Гаусса становится более расплывчатым, что говорит об увеличении рассеяния.

В случае $\bar{x} = 0$, $\sigma = 1$ функция Гаусса называется плотностью нормированного и центрированного распределения (локальная функция Лапласа)

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$$

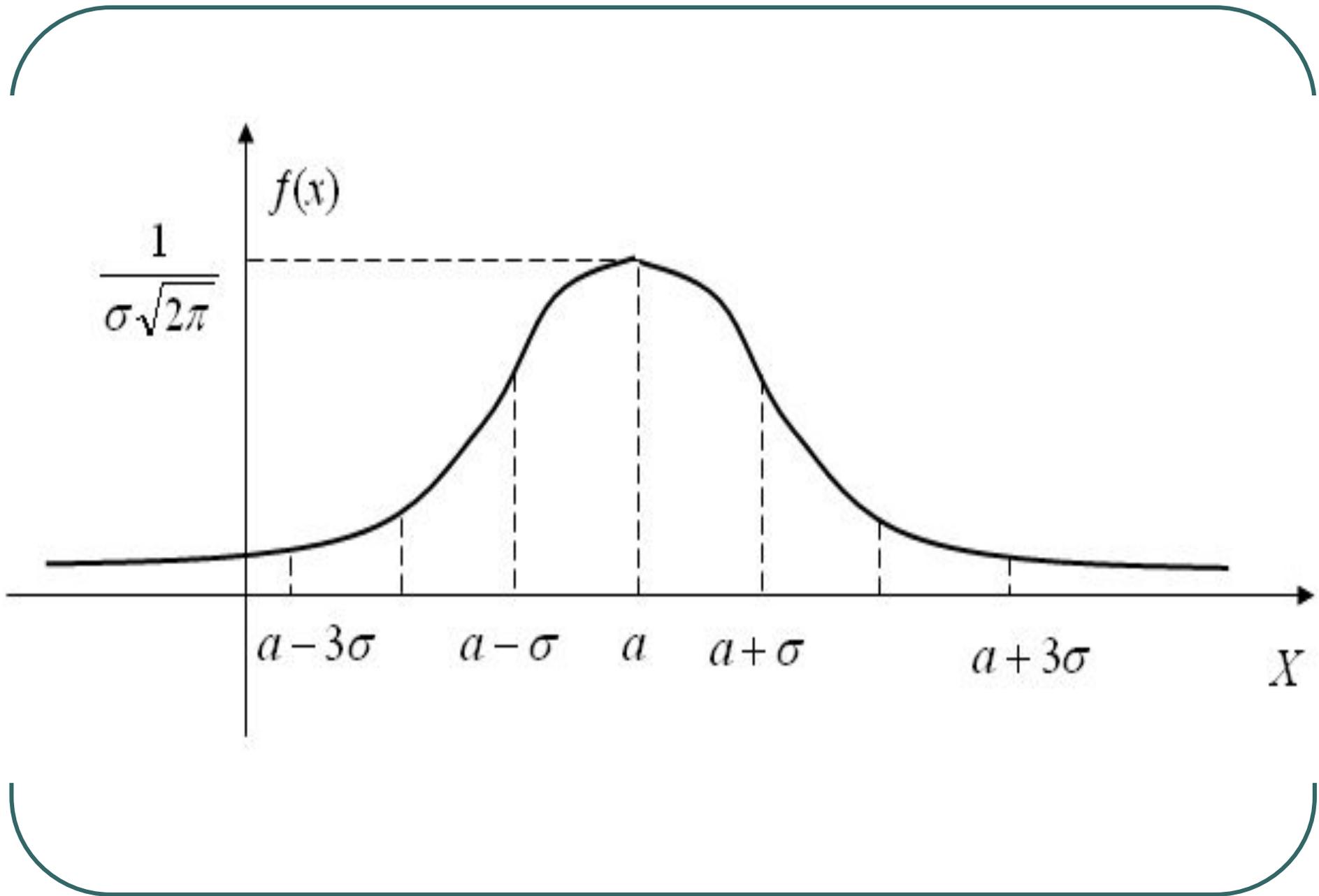
и вычисляется с помощью таблиц.

Правило «трех сигм».

$$P(|X - \bar{X}| < 3\sigma) \approx 0,9973 \text{ (99,7\%)}.$$

почти достоверно (на 99,7%), значения нормально распределенной случайной величины отличаются от своего среднего значения \bar{x} менее, чем на 3σ , то есть практически все значения нормально распределенной случайной величины X попадают в интервал:

$$(\bar{x} - 3\sigma, \bar{x} + 3\sigma).$$



Если случайная величина X_{meas} распределена по нормальному закону, то, обычно, за абсолютную погрешность принимают её **среднеквадратичное отклонение**.

Абсолютная погрешность измеряется в тех же единицах измерения, что и сама величина.

Литература:

- Морозов, Ю.В.
ОСНОВЫ высшей математики и
статистики: учебник / Ю.В. Морозов.
– М.: Медицина, 2001