Нормальный закон распределения

Лекция 3

План лекции:

- Закономерности нормального распределения
- Кривая нормального распределения и ее характеристики
- Интервальные оценки
- Генеральная и выборочная совокупности
- Сравнение теоретических и эмпирических распределений
- Основные этапы исследования

Нормальный закон распределения случайных величин

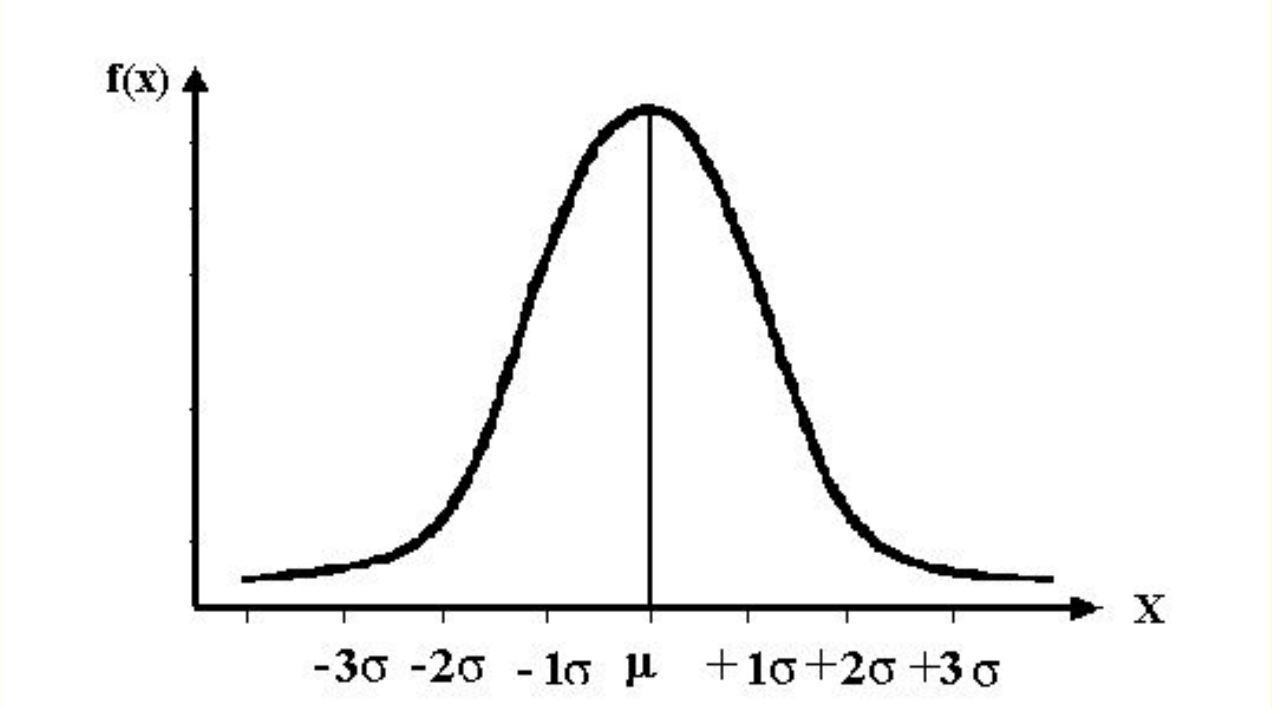
Нормальное распределение возникает тогда, когда на изменение случайной величины действует множество различных независимых факторов, каждый из которых в отдельности не имеет преобладающего значения.

ЗАКОНОМЕРНОСТИ НОРМАЛЬНОГО РАСПРЕДЕЛЕНИЯ:

- Параметр **μ** характеризует математическое ожидание (среднее арифметическое) случайной величины, являясь центром распределения и наиболее вероятным значением. Изменение математического ожидания не влияет на форму кривой, а только вызывает ее смещение вдоль оси **x**.
- Параметр **σ** характеризует изменчивость случайной величины (меру растянутости кривой вдоль оси **x**): чем больше **σ,** тем больше кривая растянута.
- График нормальной кривой симметричен относительно прямой **x=μ** (одинаковые по абсолютной величине отрицательные и положительные отклонения случайной величины от центра равновероятны).

ЗАКОНОМЕРНОСТИ РАСПРЕДЕЛЕНИЯ:

- По мере увеличения разности **(x–μ)** значение **f(x)** убывает. Это значит, что большие отклонения менее вероятны, чем малые. При **(x–μ)** $\rightarrow \infty$ значение **f(x)** стремится к нулю, но никогда его не достигает.



Кривая нормального распределения

Функция плотности распределения вероятностей:

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

Функция распределения вероятностей:

$$F(x)=\frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx$$

$$t=\frac{x-\mu}{\sigma}\quad\Rightarrow\quad F(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x}e^{-\frac{t^2}{2}}dt$$

Вероятность попадания значения случайной величины в интервал от a до b:

$$P(a\langle x\langle b)=\Phi\left(\frac{b-\mu}{\sigma}\right)-\Phi\left(\frac{a-\mu}{\sigma}\right)$$

причем $\Phi(-t) = 1 - \Phi(t)$

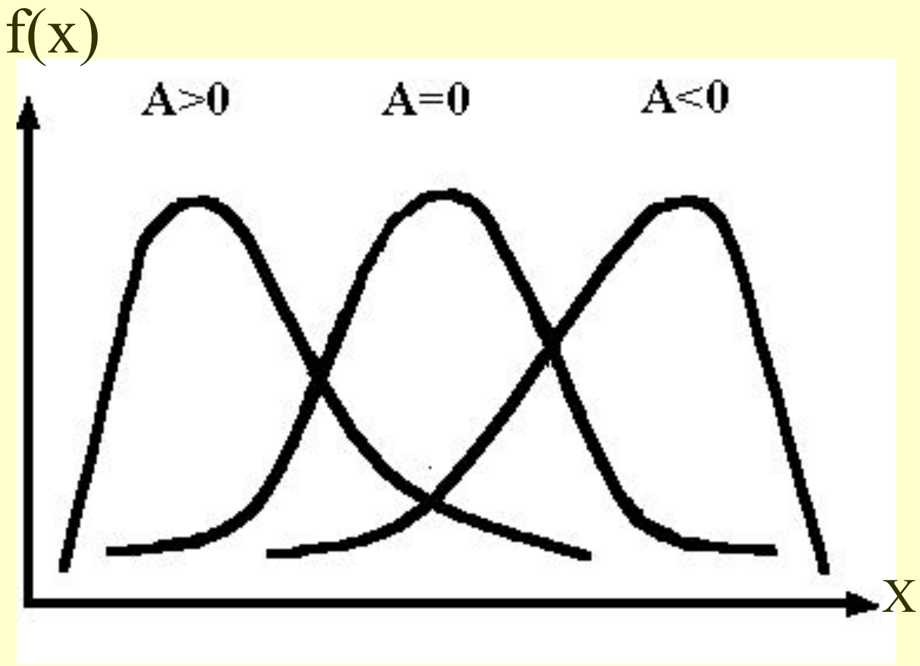
Характеристики кривой:

- Коэффициент асимметрии
- Показатель эксцесса

КОЭФФИЦИЕНТ **АСИММЕТРИИ**

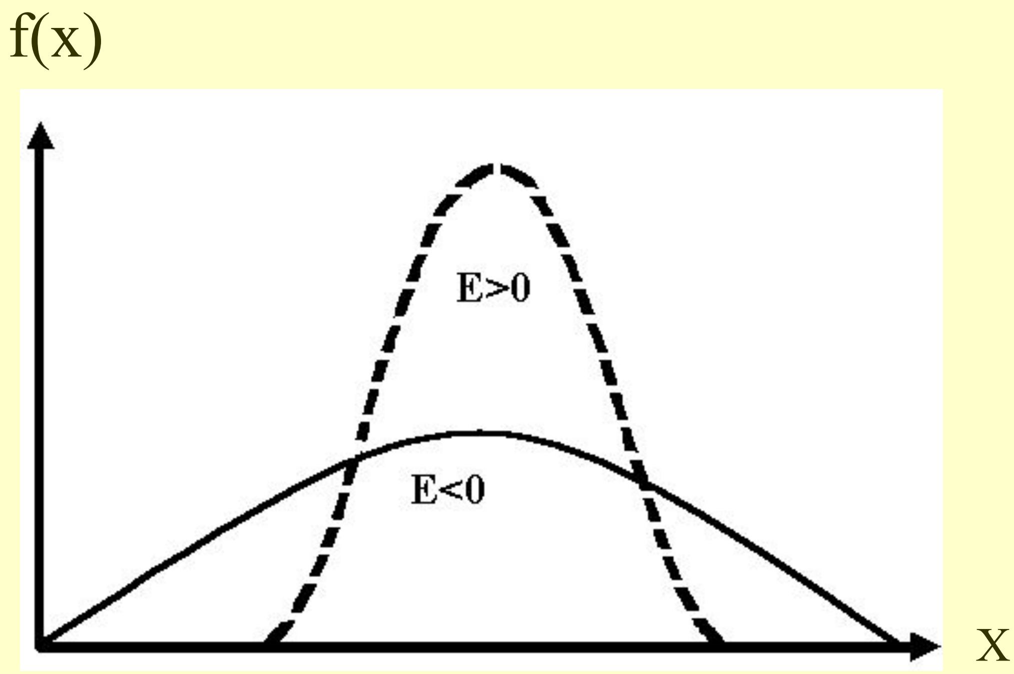
$$A = \frac{M(x - M(x))^3}{\sigma^3}$$

A>0 - правоасимметричные,
A<0 - левоасимметричные



ПОКАЗАТЕЛЬ **ЭКСЦЕССА**

$$E = \frac{M(x - M(x))^4}{\sigma^4} - 3$$



Интервальные оценки

$$t = \frac{x - \mu}{\sigma} \longleftarrow \text{нормированное отклонение}$$

x – μ=σt

$\pm$ **1σ – 68,3%;**

$\pm$ **2σ – 95,5%;**

$\pm$ **3σ – 99,7% всех вариант**

Доверительные вероятности и доверительные интервалы

- Вероятности 0,95 и 0,99 (95% и 99%) – доверительные вероятности
- $\Delta x = \pm\sigma t$ – доверительный интервал

Вероятности	Интервалы
0,95	$\pm 1{,}96\sigma$
0,99	$\pm 2{,}58\sigma$
0,999	$\pm 3{,}03\sigma$

$\alpha = 1 - P \longleftarrow$ уровень значимости

Генеральная и выборочные совокупности

- Наиболее общую совокупность, подлежащих изучению объектов называют **генеральной.**
- Выборка считается **репрезентативной**, если каждый объект выборки отобран случайно из генеральной совокупности, то есть все объекты имеют одинаковую вероятность попасть в выборку.

Сравнительная характеристика

Характеристики	Совокупность	
	Генеральная	**Выборочная**
Математическое ожидание	**μ**	$\overline{X}$
Среднее квадратическое отклонение	**σ**	**s**
Средняя квадратическая ошибка (стандартная ошибка)		$s_{\overline{x}} = \frac{s}{\sqrt{n}}$

$$\mu = \overline{X} \pm t s_{\overline{x}} \longleftarrow$$ значение генеральной средней с доверительным интервалом

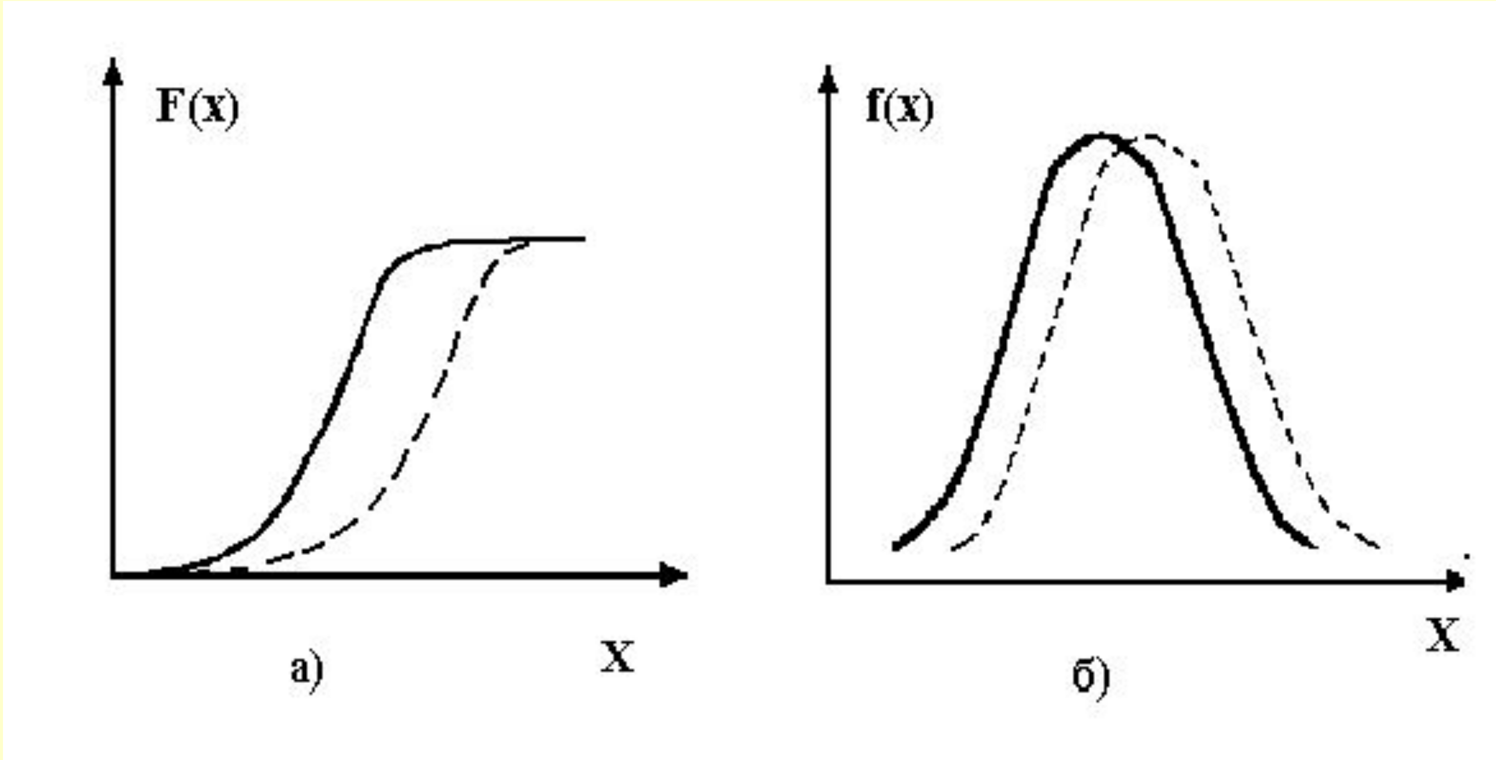
Сравнение теоретических и эмпирических распределений

- **Нулевая гипотеза.** Согласно этой гипотезе первоначально принимается, что между эмпирическим и теоретическим распределением признака в генеральной совокупности достоверного различия нет.

Средние квадратические ошибки s_A (асимметрии) и s_E (эксцесса)

$$s_A = \sqrt{\frac{6(n-1)}{(n+1)(n+3)}} \quad s_E = \sqrt{\frac{24n(n-2)(n-3)}{(n-1)^2(n+3)(n+5)}}$$

Для достаточно большой выборки ($n>30$), если показатели асимметрии (A) и эксцесса (E) в два и более раза превышают показатели их средних квадратических ошибок, гипотезу о нормальности распределения нужно **отвергнуть**.



Сравнение теоретических и экспериментальных распределений по:

а) критерию Колмогорова – Смирнова,

б) критерию Пирсона.

Пунктирная линия – эмпирическое распределение, сплошная – теоретическое распределение.

Критерий Пирсона

$$\chi^2_{\text{эмп.}} = \sum_{i=1}^{k} \frac{(m_i - np_i)^2}{np_i}$$

где $\mathbf{m_i}$ – экспериментальные частоты попадания значения случайной величины в интервал,

$\mathbf{np_i}$ – теоретические частоты.

- **Число степеней свободы** – это общее число величин, по которым вычисляются соответствующие статистические показатели, минус число тех условий, которые связывают эти величины, то есть уменьшают возможности вариации между ними. Число степеней свободы определяется по следующей **формуле:**

  df=k–r–1, где k – число интервалов, r – число параметров предполагаемого распределения. Для нашего случая **r=2**, следовательно, **df=k–3.**
- По заданному уровню значимости **(α)** и числу степеней свободы **df,** находим критическое значение $\chi^2_{кр}$ **(α,df).**
- **Если** $\chi^2_{эмп} < \chi^2_{кр}$ гипотеза о согласии эмпирического и теоретического распределения **не отвергается.**

Основные этапы исследования:

- **Сгруппировать исследуемый ряд по классам. Подсчитать середины интервалов и частоты попадания в интервал.**
- **Построить гистограмму и полигон распределения.**
- **Найти эмпирическую функцию распределения и построить ее график.**
- **Вычислить числовые (точечные) характеристики распределения.**
- **Найти интервальные оценки для генеральной средней.**
- **Проверить гипотезу о том, что генеральная совокупность, из которой извлечена выборка, распределена по нормальному закону, используя критерий Пирсона χ^2.**

БЛАГОДАРЮ ЗА ВНИМАНИЕ