

Основы измерения и количественного анализа данных

1. Описательная статистика.
2. Сущность распределения признака Нормальное распределение.
3. Статистические гипотезы.
4. Уровни статистической значимости





Определение

- ❖ **Описательная (дескриптивная) статистика** – комплекс математических процедур, целью которых является описание характера распределения переменных посредством основных статистических показателей

Расчет статистических показателей

- ❖ **Среднее арифметическое** – это сумма всех чисел в конкретном массиве данных, делённая на их количество. Отражает среднюю тенденцию для данной переменной в указанной выборке

Расчет статистических показателей

- ❖ **Мода** – такое числовое значение, которое встречается в выборке наиболее часто. В одной выборке может быть несколько мод.
- ❖ Мода отражает наиболее часто встречаемое значение (число), а не частоту его встречаемости (число раз повторений числа)

	<i>Mo</i>				
Значение	1	2	3	4	5
Частота	2	7	18	9	4

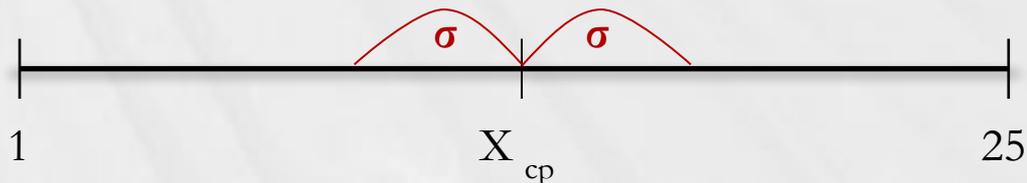
Расчет статистических показателей

- ❖ **Медиана** – это величина (число в числовом ряду), по отношению к которой по крайней мере 50 % выборочных значений меньше ее и по крайней мере 50 % больше.
- ❖ Медиана – значение, которое делит упорядоченное множество данных пополам.



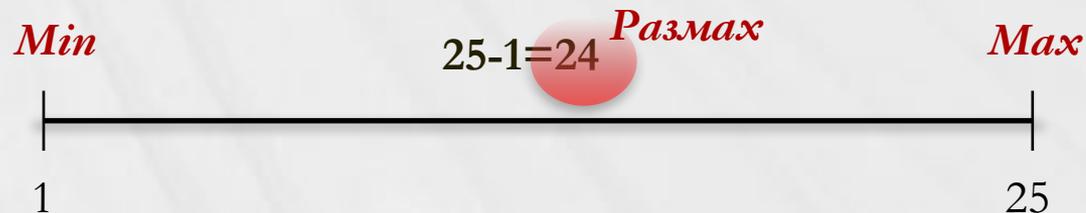
Расчет статистических показателей

- ❖ **Дисперсия (σ^2)** – мера рассеяния, которая характеризует вариацию признака всей совокупности под влиянием всех тех факторов, которые обусловили данную вариацию. Дисперсию трудно интерпретировать содержательно. Однако, квадратный корень из этого значения является стандартным отклонением и хорошо поддается интерпретации.
- ❖ **Стандартное отклонение (σ)** – показатель рассеяния. Стандартное отклонение показывает, насколько хорошо среднее значение описывает всю выборку. При нормальном распределении в пределах одного стандартного отклонения находится около 65 % значений ряда переменной



Расчет статистических показателей

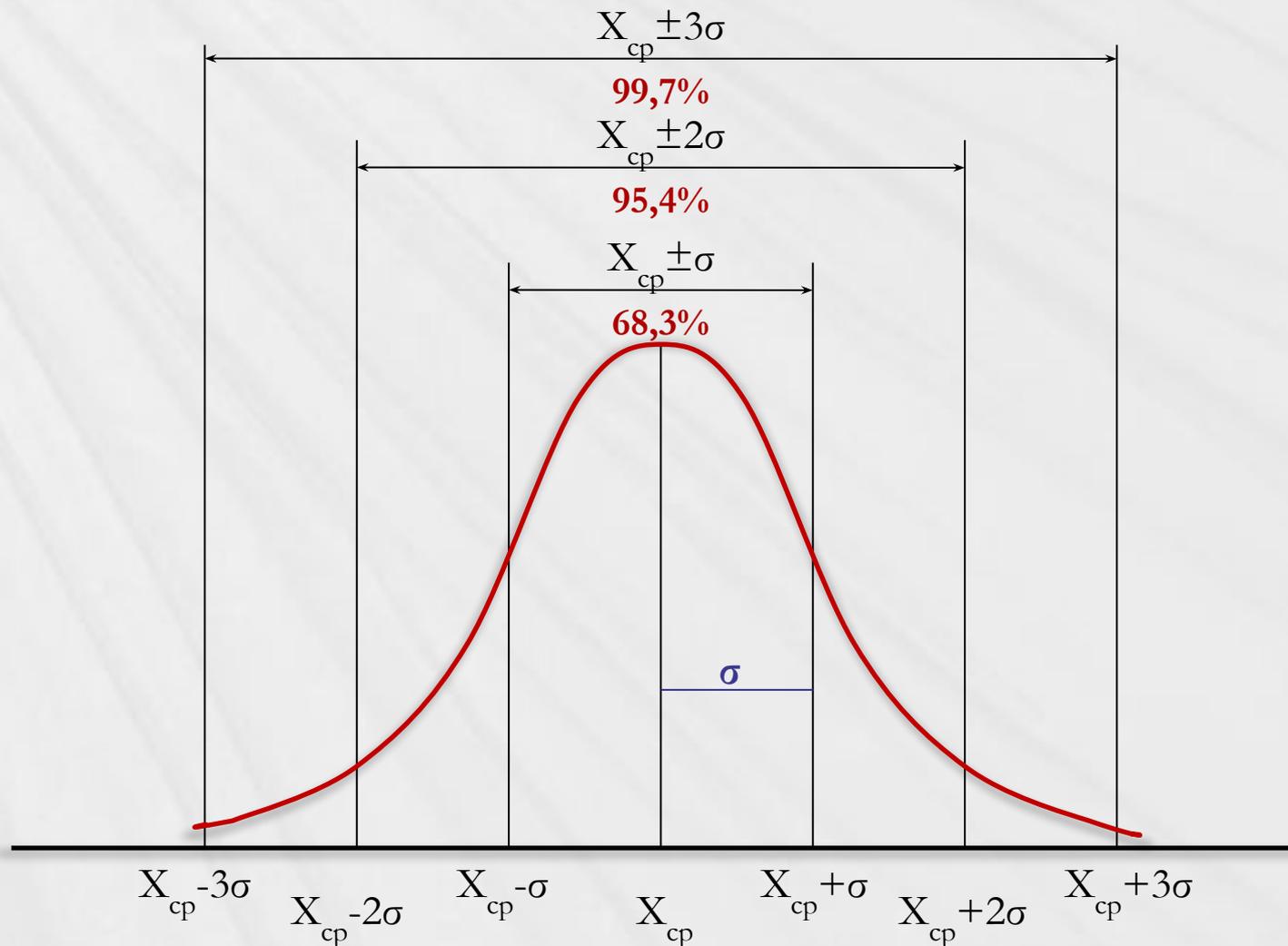
- ❖ **Минимум** – минимальное значение в ряду данных
- ❖ **Максимум** – максимальное значение в ряду данных
- ❖ **Разброс (размах)** – разность между максимальной и минимальной величинами данного ряда значений



2. Распределение переменной

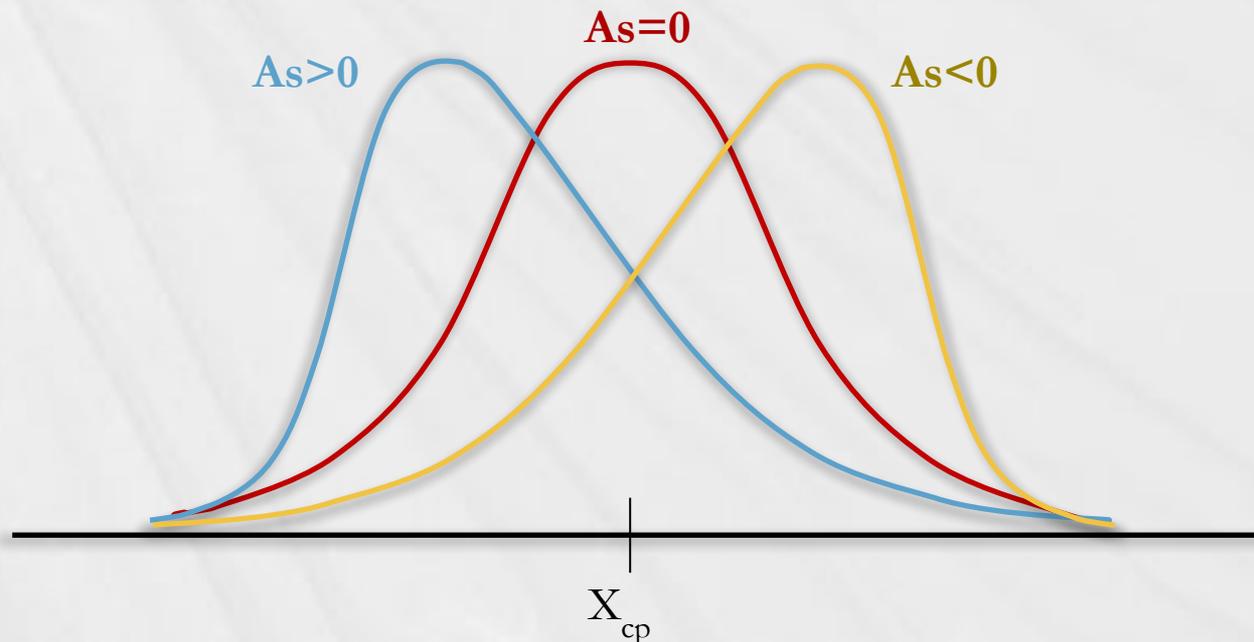
- ❖ Важным способом "описания" переменной является **форма ее распределения**, которая показывает, с какой частотой значения переменной попадают в определенные интервалы. Эти интервалы, называемые интервалами группировки, выбираются исследователем.
- ❖ **Нормальное распределение** – распределение, зависящее от двух параметров: **среднего арифметического** как точки отсчета и **стандартного отклонения** как масштаба (шага интервалов).

График нормального распределения



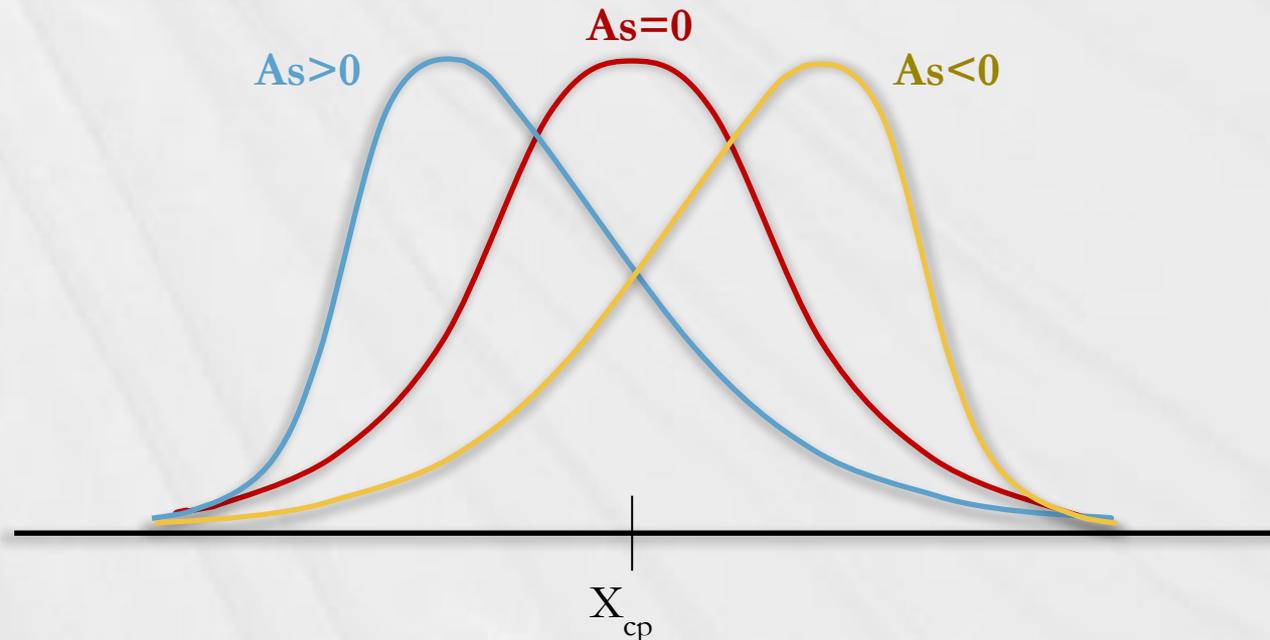
Статистические показатели распределения переменной

- ❖ **Асимметрия** – степень отклонения графика распределения частот от симметричного вида относительно среднего значения. Для симметричного распределения асимметрия равна 0.



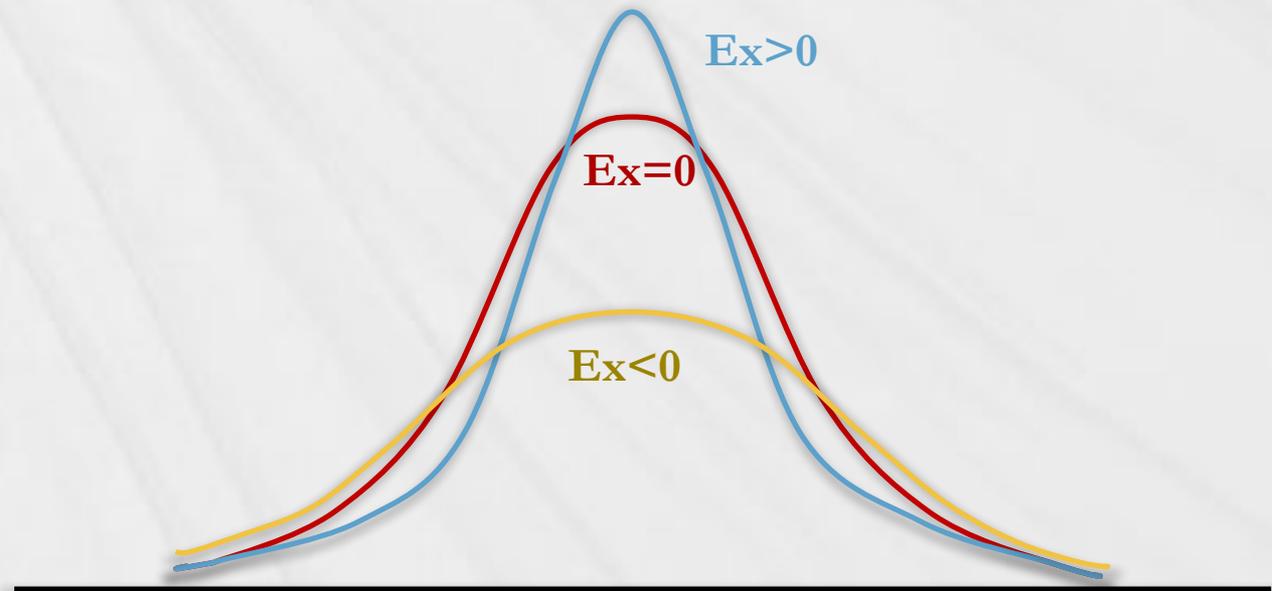
Статистические показатели распределения переменной

- ❖ Чем больше отклонение от нуля, тем больше асимметрия.
- ❖ При $A > 0$ левосторонней (положительной) асимметрии чаще встречаются низкие значения признака.
- ❖ При $A < 0$ правосторонней (отрицательной) чаще встречаются высокие значения признака.



Статистические показатели распределения переменной

- ❖ **Эксцесс** – мера плосковершинности или остроконечности графика распределения измеренного признака. Островершинность характеризуется **положительным** эксцессом, плосковершинность – **отрицательным**.





Критерии нормальности распределения

Данные, представленные выборкой, можно считать нормально распределенными, если:

- ❖ Выборка является симметричной
- ❖ Медиана и мода практически совпадают со средним арифметическим
- ❖ Показатели асимметрии и эксцесса близки к нулю



Способы оценки нормальности распределения

1 способ. **По соотношению
основных параметров
распределения** (среднего
арифметического, моды и медианы).
При нормальном распределении
значения среднего арифметического,
моды и медианы совпадают



Способы оценки нормальности распределения

2 способ. **По показателям асимметрии и эксцесса** – значения асимметрии и эксцесса должны стремиться к нулю. Допустимыми считаются их значения в пределах **от – 1 до 1** (в исключительных случаях от – 2 до 2). При расчетах в программе SPSS, значения асимметрии и эксцесса должны быть меньше, чем значения их стандартных ошибок по модулю.



Способы оценки нормальности распределения

3 способ. **По расчету критерия Колмогорова-Смирнова** – критерий, сравнивающий эмпирическое распределение переменной с теоретическим (нормальным) распределением. Если присутствуют значимые отличия между ними (уровень значимости меньше 0,05), то эмпирическое распределение **не соответствует нормальному виду**



Условия применения критерия Колмогорова-Смирнова:

- ❖ Измерение может быть произведено в шкале интервалов или отношений (количественных шкалах)
- ❖ Объем выборки должен быть более 50 человек. С увеличением объема выборки точность критерия повышается



3. Статистические гипотезы

- ❖ **СТАТИСТИЧЕСКАЯ ГИПОТЕЗА** – утверждение относительно неизвестного параметра генеральной совокупности на основе выборочного исследования.
- ❖ Это предположение о том, что сходство (или различие) некоторых параметрических или функциональных характеристик случайно, или, наоборот, неслучайно.

Виды статистических гипотез

- ❖ **НУЛЕВАЯ СТАТИСТИЧЕСКАЯ ГИПОТЕЗА H_0** – это гипотеза о сходстве, об отсутствии различий (или связи).
- ❖ **АЛЬТЕРНАТИВНАЯ СТАТИСТИЧЕСКАЯ ГИПОТЕЗА H_1** – это гипотеза о значимости различий (или о наличии связи).
- ❖ Нулевая и альтернативная гипотезы могут быть **НАПРАВЛЕННЫМИ** (если надо доказать, что в одной из групп индивидуальные значения испытуемых по какому-либо признаку выше, а в другой ниже) и **НЕНАПРАВЛЕННЫМИ** (если надо доказать, что различия существуют без указания направления).



❖ *Ошибка первого рода* произойдет, когда будет принято решение отклонить гипотезу H_0 , хотя в действительности она оказывается верной.

❖ *Ошибка второго рода* произойдет, когда будет принято решение не отклонять гипотезу H_0 , хотя в действительности она будет не верна.

❖ Более «критичной» ошибкой считается статистическая ошибка первого рода

❖ «Судебная» аналогия: Вердикт «Не виновен» или «Виновен»

❖ Ошибка первого рода - невинный обвинен

❖ Ошибка второго рода - виновный освобожден

- 
- ❖ **Пример: Психологу необходимо определить наличие тревожности у ребенка:**
 - ❖ При этом ошибки могут быть двоякого рода:
 - ❖ - принимается решение об отсутствии тревожности у данного индивида при ее объективном наличии (**ошибки первого рода**);
 - ❖ - принимается решение о наличии тревожности при ее объективном отсутствии (**ошибки второго рода**).
 - ❖ В большинстве случаев единственный путь минимизации ошибок заключается в увеличении объема выборки

4. Уровни статистической значимости

- ❖ **СТАТИСТИЧЕСКАЯ ЗНАЧИМОСТЬ** – вероятность получения результата выборочного исследования, при котором верна нулевая статистическая гипотеза для генеральной совокупности (**статистически значимых различий между генеральной и выборочной совокупностью нет**).
- ❖ **P-уровень значимости** – это вероятность ошибочного отклонения нулевой гипотезы (**вероятность того, что мы сочли различия существенными, а они на самом деле случайны**).

- 
- ❖ Обычно используют уровни значимости (обозначаемые p), равные 0,05, 0,01 и 0,001
 - ❖ Например, уровень значимости, равный 0,05, означает, что **допускается не более чем 5%-ая вероятность ошибки**. Т.е. нулевую гипотезу можно отвергнуть в пользу альтернативной гипотезы, если по результатам статистического теста вероятность ошибки, т.е. вероятность случайного возникновения обнаруженного различия (p -уровень) не превышает 5 из 100, т.е. имеется лишь 5 шансов из 100 ошибиться.
 - ❖ Если же этот уровень значимости не достигается (вероятность ошибки выше 5%), считают, что разница вполне может быть случайной и поэтому нельзя отклонить нулевую гипотезу.

Расчет статистических показателей

- ❖ **Уровень значимости** (надежности) – отражает вероятность ошибочности выводов по статистическому вычислению

