

# Вариационная статистика

## Основные показатели вариационной статистики:

<b>N</b>	– объем выборки
<b>M</b>	– среднее значение
<b>D<sub>x</sub></b>	– дисперсия
<b>σ (S<sub>x</sub>)</b>	– среднеквадратическое отклонение
<b>m</b>	– стандартная ошибка
<b>min (max)</b>	– минимальное (максимальное) значение выборки
<b>R</b>	– размах вариации
<b>CV</b>	– коэффициент вариации
<b>95%ДИ</b>	– 95% доверительный интервал
<b>Me</b>	– медиана
<b>Mo</b>	– мода
<b>As</b>	– асимметрия
<b>Ex</b>	– эксцесс
<b>K<sub>0</sub>, K<sub>2</sub>, ..., K<sub>4</sub></b>	– квартили (квантили)
<b>P<sub>1</sub>, P<sub>2</sub>, ..., P<sub>99</sub></b>	– перцентили (процентили)

# Вариационная статистика

*А вот, незадача ?*

*Средний возраст детей равняется ( $M \pm \sigma$ )*

*3,6 $\pm$ 4,2 года*

*Средний уровень креатинина в крови ( $M$ ; 95%ДИ)*

*158,0 (-23,0; 339,0)*

*Значит, что-то неправильно в нашей выборке !*

*Причины:*

- ✓ Расчет неправильный – пересчитать заново!*
- ✓ Недостаточное кол-во данных в выборке –  
увеличить кол-во наблюдений!*
- ✓ В выборке имеются «выскакивающие» величины –  
очистить выборку от них !*

# Вариационная статистика

*Как избавиться от «выскакивающих» величин ?*

1. Перепроверить работу оператора
2. Перепроверить первоисточник
3. Уточнить методику измерения
4. Применить статистические критерии:
  - Критерий  $\sigma$
  - Критерий Романовского
  - Критерий Шарля
  - Критерий Диксона
  - Процентильный способ
  - И .т.д

$$r_{max} = \frac{x_{max} - M}{\sigma_x}; r_{min} = \frac{M - x_{min}}{\sigma_x}$$

## Следующий этап включает «разведывательный анализ»

- Проверить выборки на рандомизированность
- Проверить выборки на повторность
- Проверить выборки на «опустошенность»
- Сравнить средние данные
- Сравнить медианы
- Просмотреть процентильные распределения
- На основании ДИ сделать предварительные заключения
- Если нужно, пересмотреть:
  - ✓ перераспределение данных,
  - ✓ перераспределение групп,
  - ✓ классификацию
  - ✓ критерии исключения и т.д.
- Формировать «нулевую гипотезу»
- Планировать примерные направления статистического анализа

***Один из важных моментов в планировании  
исследования !***

***Вычисление минимального кол-ва исследования***

$$n = \frac{t^2 \sigma^2}{\Delta^2} = \frac{t^2}{\Delta^2 / \sigma^2}$$

$$\Delta = \frac{t m (100 - k\%)}{100}$$

## Разведывательный анализ:

### Как проверить выборки на рандомизированность ?

Если не являются критериями разделения то, показатели сравниваемых групп статистически **не должны** различаться:

- ✓ по полу;
- ✓ по возрасту;
- ✓ по месту жительства;
- ✓ по социальному статусу и .т.д.
- ✓ .
- ✓ .
- ✓ .
- ✓ и т.д.
- ✓ в продольных исследованиях: исходные клинические и лабораторные показатели



# Разведывательный анализ

## Как проверить выборки на опустошенность?

До	7 дней	30 дней
11,2	10,6	9,3
8,3	8,5	7,3
9,6	8,8	9,2
8,1	9,6	11,3
-	9,6	10,3
4,2	-	5,5
5,3	-	6,9
4,5	-	-
5,5	-	-
6,9	8,6	7,5
5,5	12,2	-
4,2	9,6	-
-	-	9,8
-	10,3	10,3
-	-	11,5



# Разведывательный анализ: Как сравнить средние данные?

$$M \pm m$$

$$65,0 \pm 5,0$$

«Волшебная формула» наших любимых профессоров !

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

$$M = (\min + \max)/2 ?$$

$M_1$

$M_2$

$M_1 - m_1$

$M_1 + m_1$

$M_2 - m_2$

$M_2 + m_2$

## Все-таки t-критерий Стьюдента универсален !

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}}$$

При обязательных условиях:

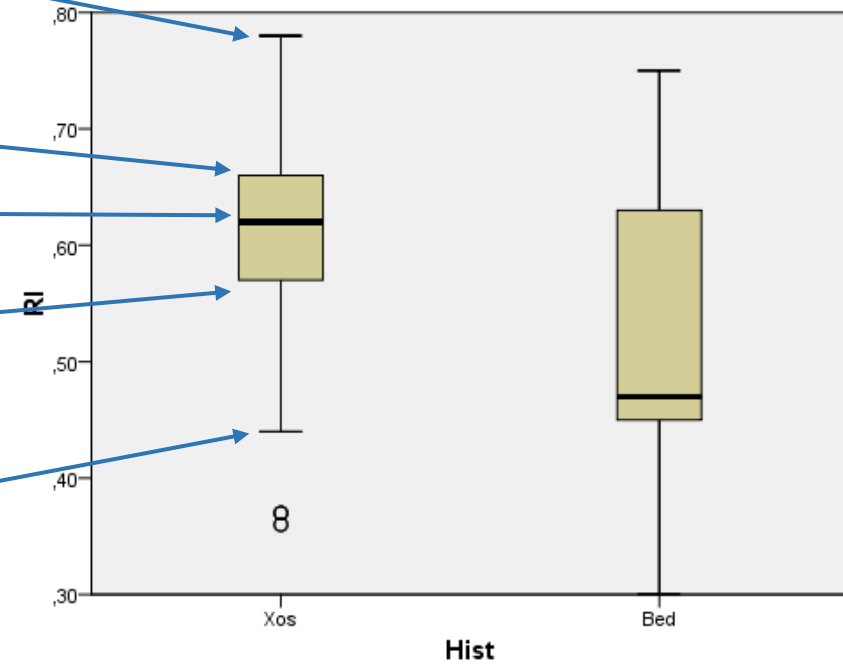
1. Обе выборки должны принадлежать одной и той же генеральной совокупности
2. Распределение выборок должны удовлетворять нормальному закону
3. Среднеквадратические отклонения должны быть примерно одинаковыми ( $\sigma_1 \approx \sigma_2$ )
4. Количество выборок должны быть одинаковыми ( $n_1 = n_2$ )

При невыполнении только 4-го условия ( $n_1 \neq n_2$ ) можно применить следующую формулу:

$$t = \frac{|M_1 - M_2| \sqrt{\frac{n_1 n_2}{n_1 + n_2}}}{\sqrt{\frac{m_1^2 n_1 (n_1 - 1) + m_2^2 n_2 (n_2 - 1)}{n_1 + n_2 - 2}}}$$

# Разведывательный анализ: Процентильные распределения

4,20	p1		
4,20	p5		
5,57	p10		
8,17	p25	Q1	
12,92	p50	Q2	Me
20,52	p75	Q3	
27,52	p90		
32,50	p95		
53,79	p99		



## Оставшиеся этапы разведывательного анализа

- Проверить выборки на рандомизированность
- Проверить выборки на повторность
- Проверить выборки на «опустошенность»
- Сравнить средние данные
- Сравнить медианы
- Просмотреть процентильные распределения
- На основании ДИ сделать предварительные заключения
- Если нужно, пересмотреть:
  - ✓ перераспределение данных,
  - ✓ перераспределение групп,
  - ✓ классификацию
  - ✓ критерии исключения и т.д.
- Формировать «нулевую гипотезу»
- Планировать примерные направления статистического анализа

## Разведывательный анализ

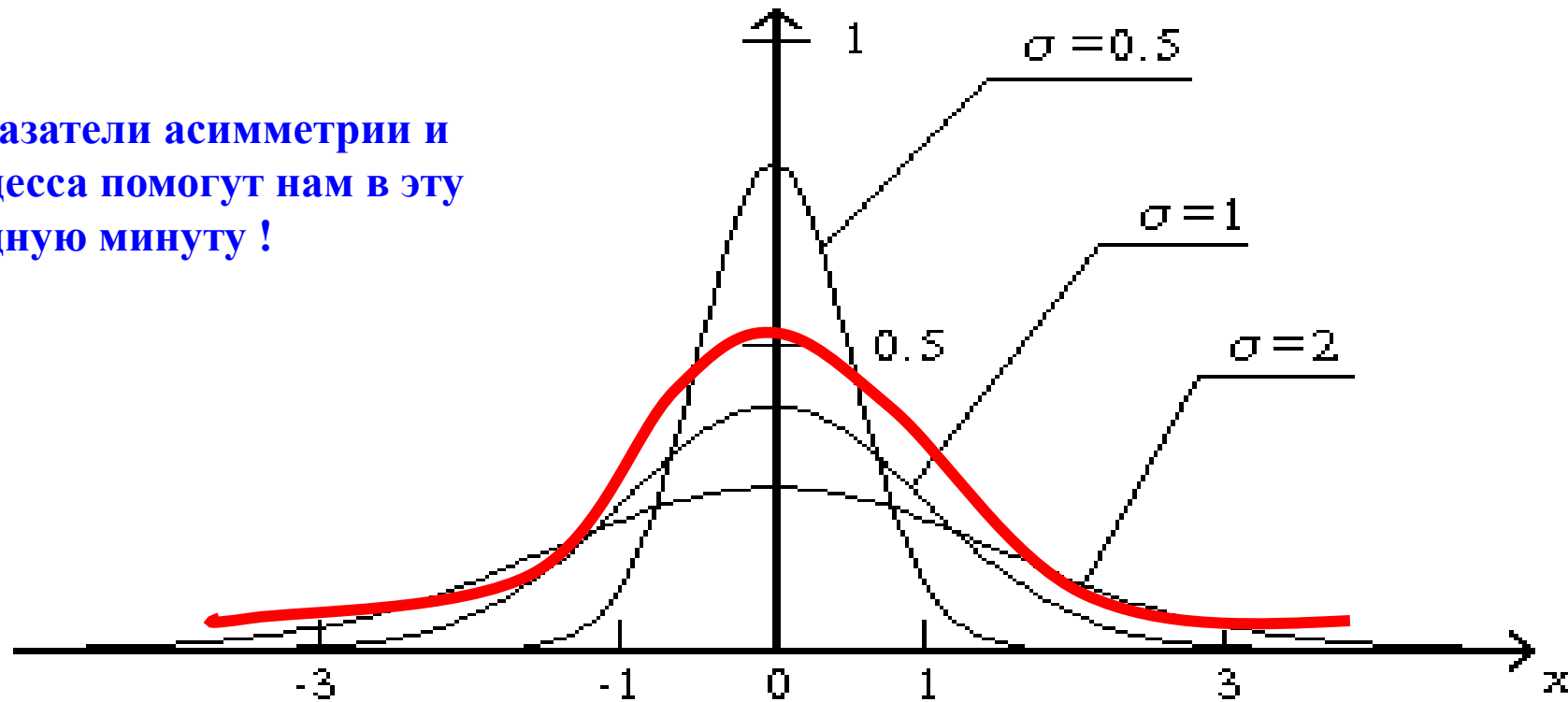
Как дальше быть?

Выбирать

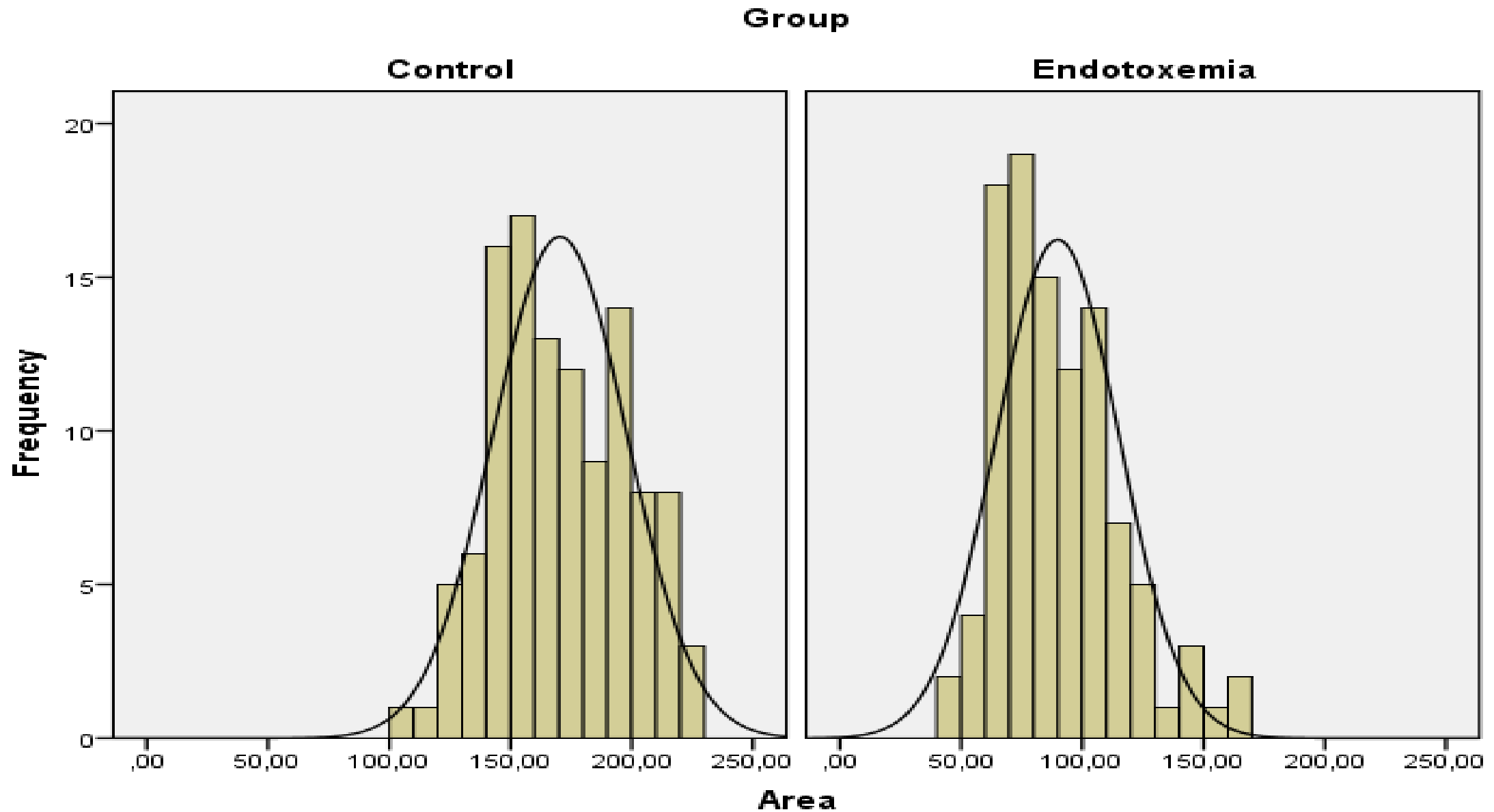
**параметрические или непараметрические**  
методы анализа ?



Показатели асимметрии и эксцесса помогут нам в эту трудную минуту !



# Разведывательный анализ



# Вариационная статистика

## Критерии сравнения количественных значений:

### Для несвязанных выборок

- $t$  – критерий Стьюдента
- $U$  – критерий Уилкоксона (Манна-Уитни)
- $W$  – критерий Уайта
- $X$  – критерий Ван дер Вардена
- $Z$  – критерий Колмогоров-Смирнова
- $VV$  – критерий Вальда-Вольфовича
- $KW$  – критерий Крускала-Уоллиса
- $ST$  – критерий Джонкхир-Терпстр
- $Me$  – медианный критерий
- $M$  – критерий экстремальной реакции Мозеса
- ... и др.

# Вариационная статистика

## Критерии сравнения количественных значений:

### Для связанных выборок

- $d$  – критерий Стьюдента
- $T$  – критерий Уилкоксона
- $N$  – критерий Мак Немара
- $z$  – критерий знаков
- $XL$  – критерий Ходжеса-Лемана
- $W$  – критерий Кендалла
- $F$  – критерий Фридмана
- ... и др.



