

ОБЗОР МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ. Часть 3

КУХАЛЬСКИЙ НИКОЛАЙ ГЕННАДЬЕВИЧ

Вопросы занятия

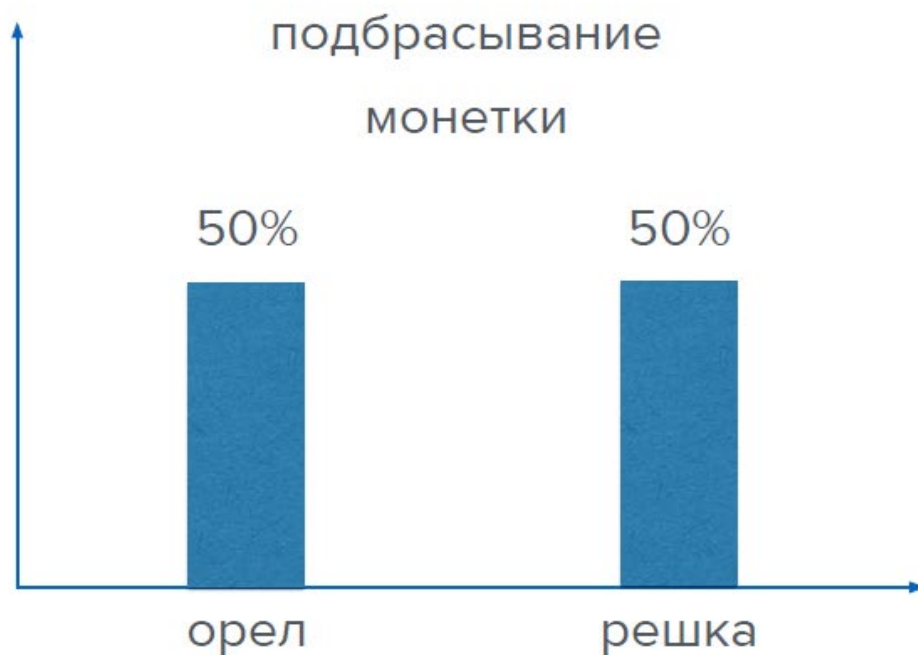
1. Понятия теории вероятностей: основы;
2. Нормальное распределение и правило трех сигм: углубление в теорию;
3. Корреляция и ее применение в различных задачах;
4. Формулировка и проверка гипотез: теория и практика

В КОНЦЕ ЗАНЯТИЯ ВЫ:

- будете знать основные распределения случайных величин;
- познакомитесь с понятием корреляции и ее смыслом в некоторых задачах;
- сможете проверять гипотезы.

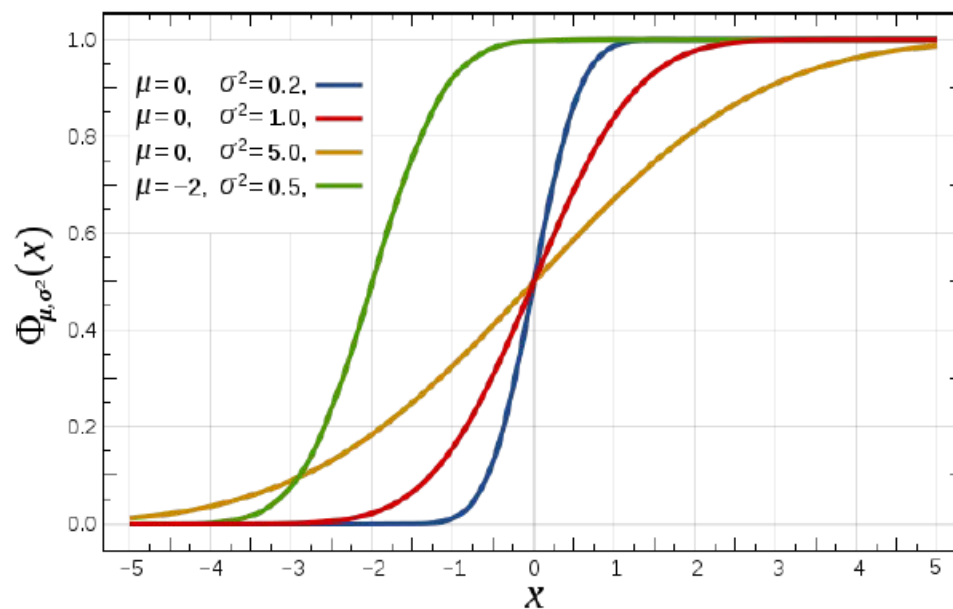
РАСПРЕДЕЛЕНИЕ

Закон, описывающий область значений случайной величины и вероятности их исхода



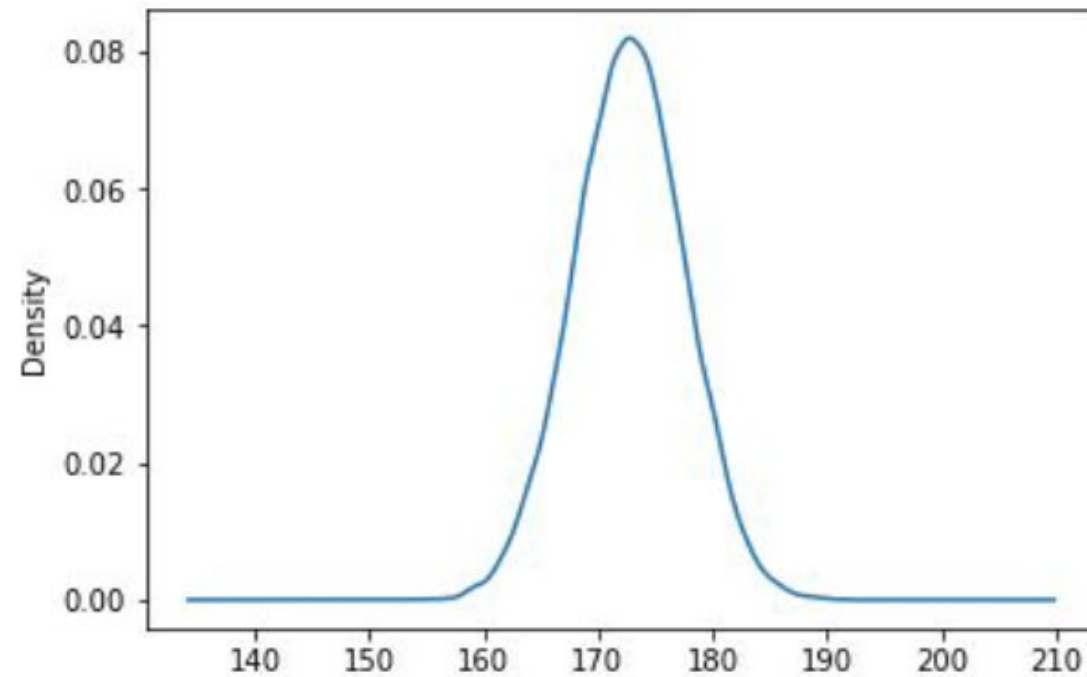
ФУНКЦИЯ РАСПРЕДЕЛЕНИЯ

Вероятность $F(x)$, что случайная величина примет значение, меньшее или равное x



ПЛОТНОСТЬ ВЕРОЯТНОСТИ

Производная функции распределения



РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

НОРМАЛЬНОЕ РАСПРЕДЕЛЕНИЕ

Плотность вероятности:

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

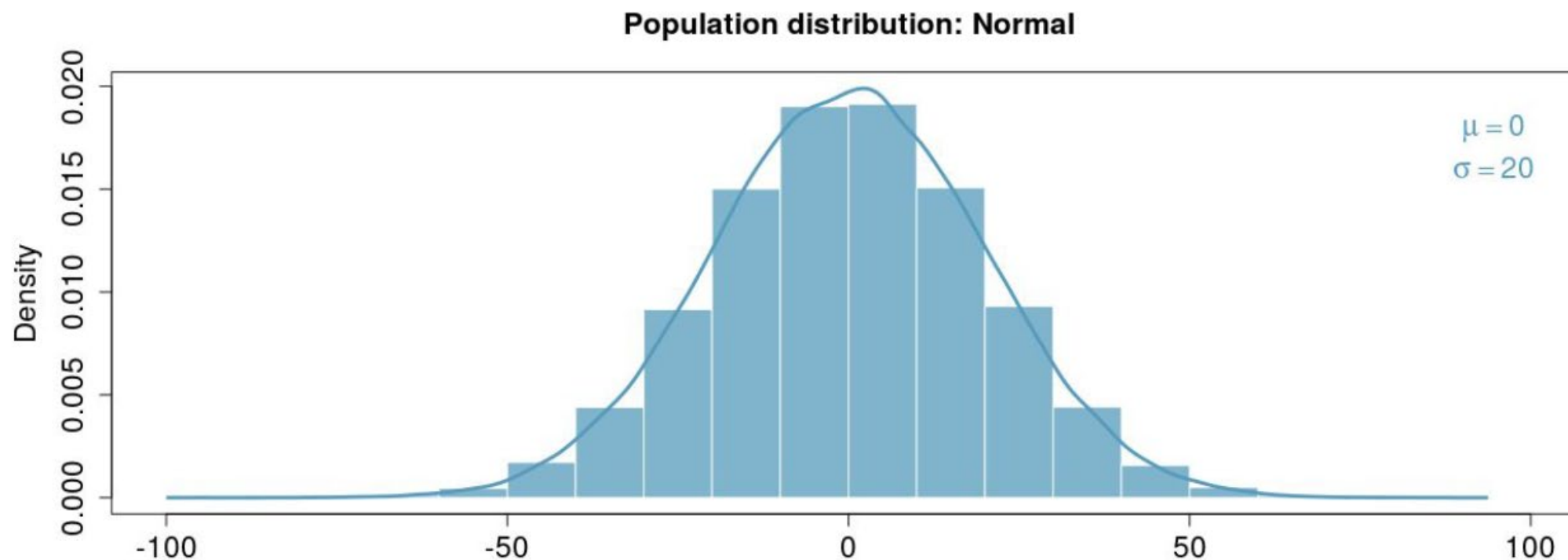
μ — среднее значение, σ — среднеквадратическое отклонение

ЦПТ

Сумма n независимых одинаково распределённых случайных величин имеет распределение, близкое к нормальному

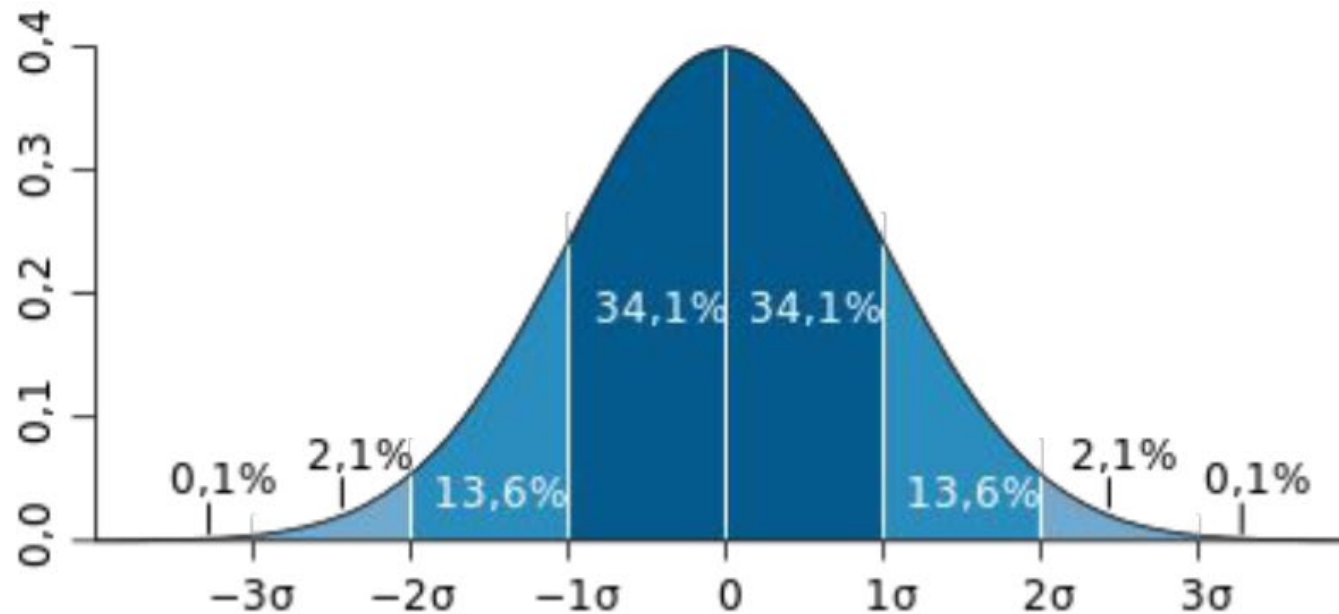
РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

ЦПТ



РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

ПРАВИЛО ТРЕХ СИГМ



Доля значений

σ – 68.2%

2σ – 95.4%

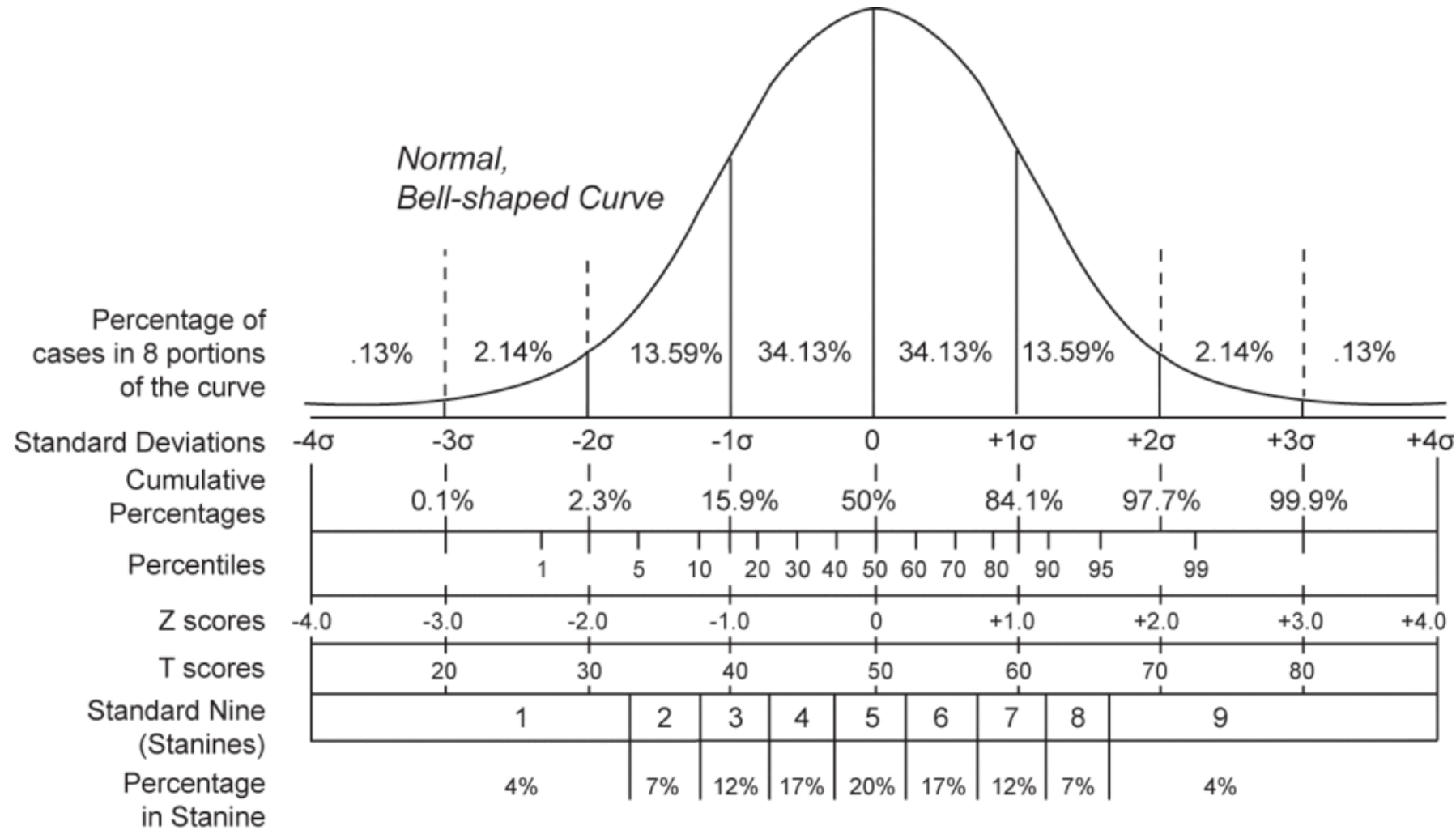
3σ – 99.6%

Z-СТАНДАРТИЗАЦИЯ

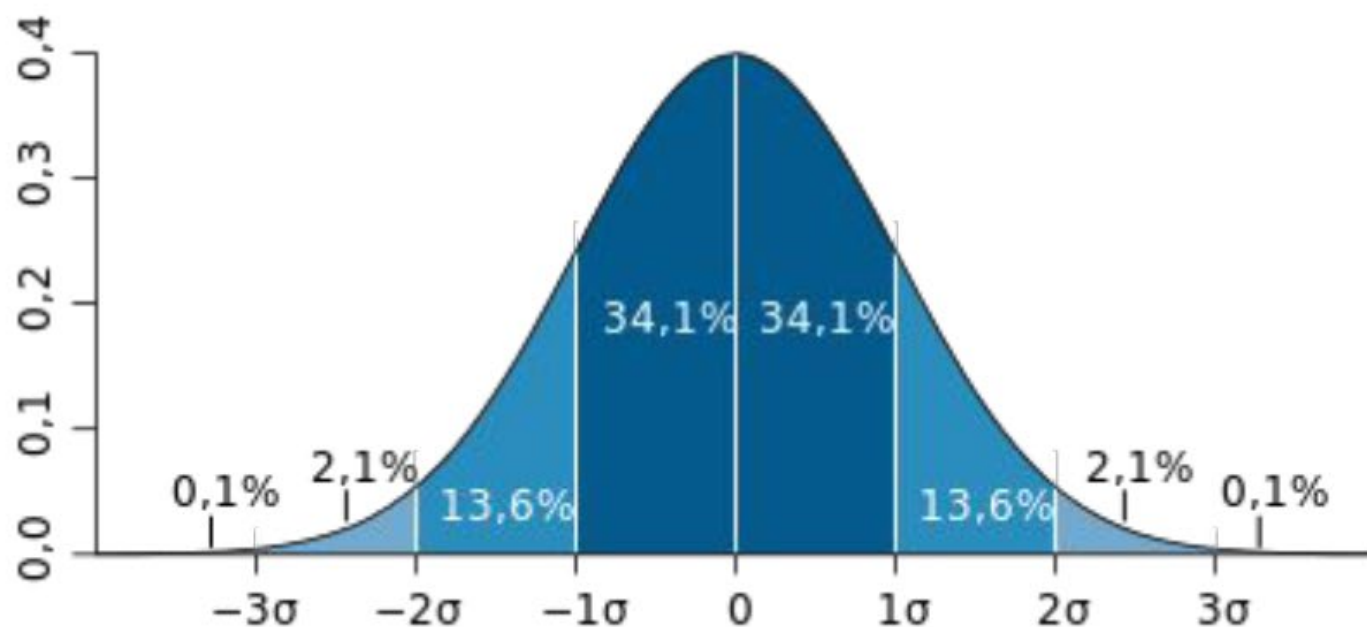
Приведение набора данных к нулевому среднему и $STD = 1$

$$Z_i = \frac{X_i - \bar{X}}{\sigma_x}$$

РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН



ДОВЕРИТЕЛЬНЫЙ ИНТЕРВАЛ



95% всех наблюдений
лежат в интервале ± 1.96
 σ

РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Т-КРИТЕРИЙ СТЬЮДЕНТА

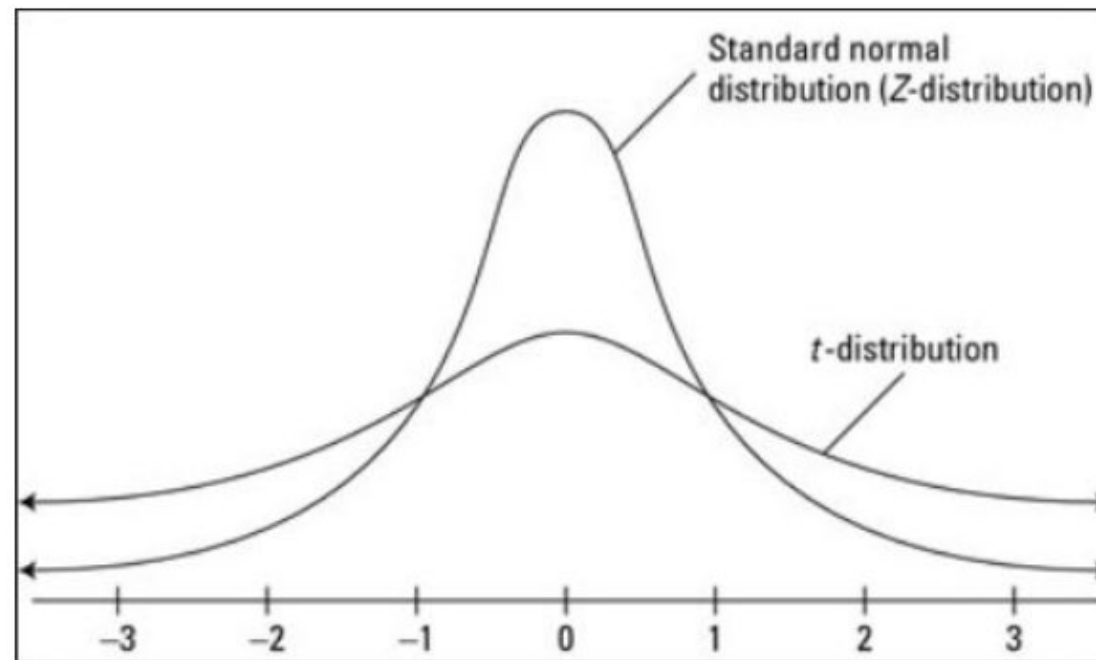
Исходные данные должны иметь нормальное распределение

$N < 30$

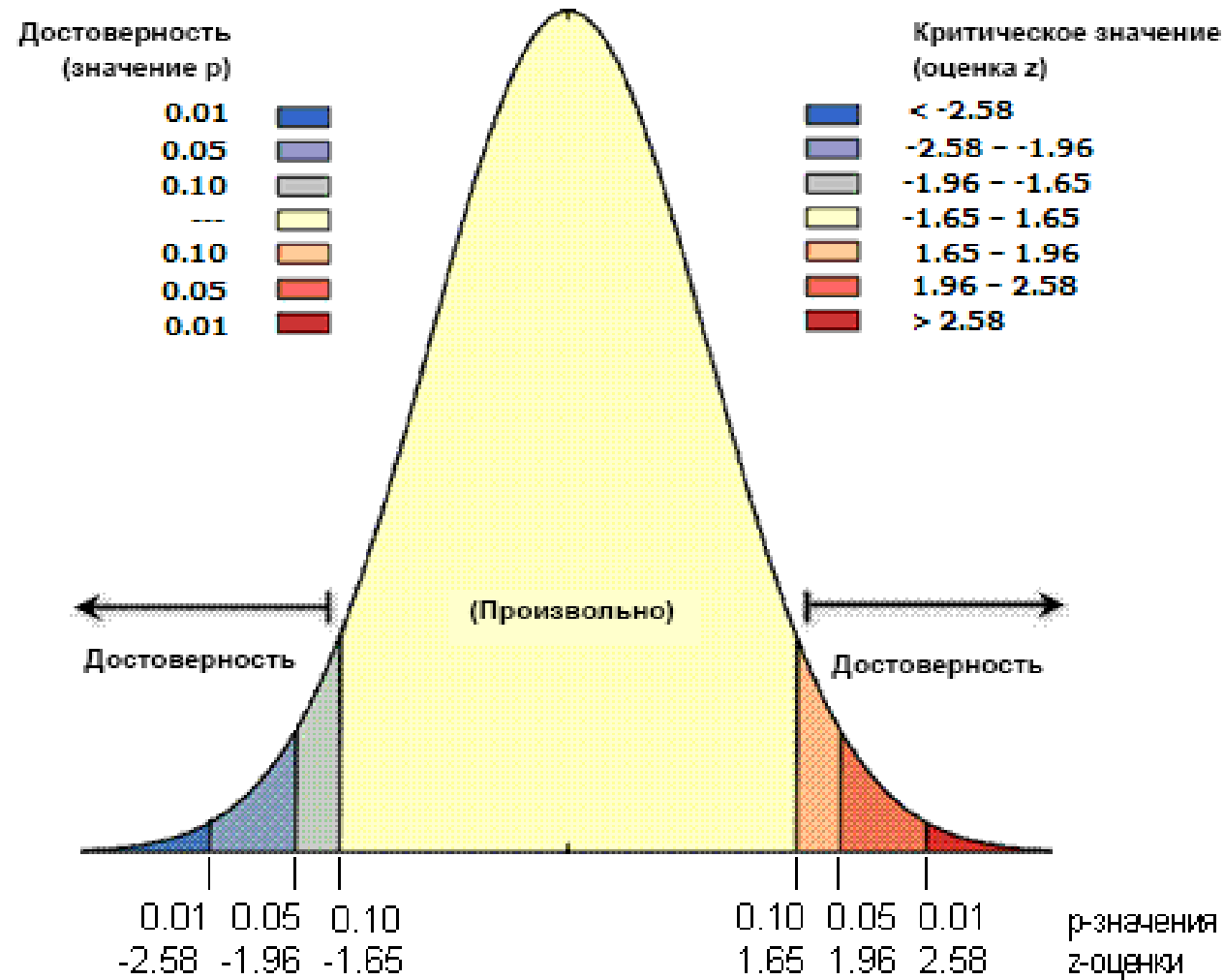
Нарушается предположение о том, что выборочные средние будут вести себя в соответствии с нормальным законом

РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН

Т-РАСПРЕДЕЛЕНИЕ

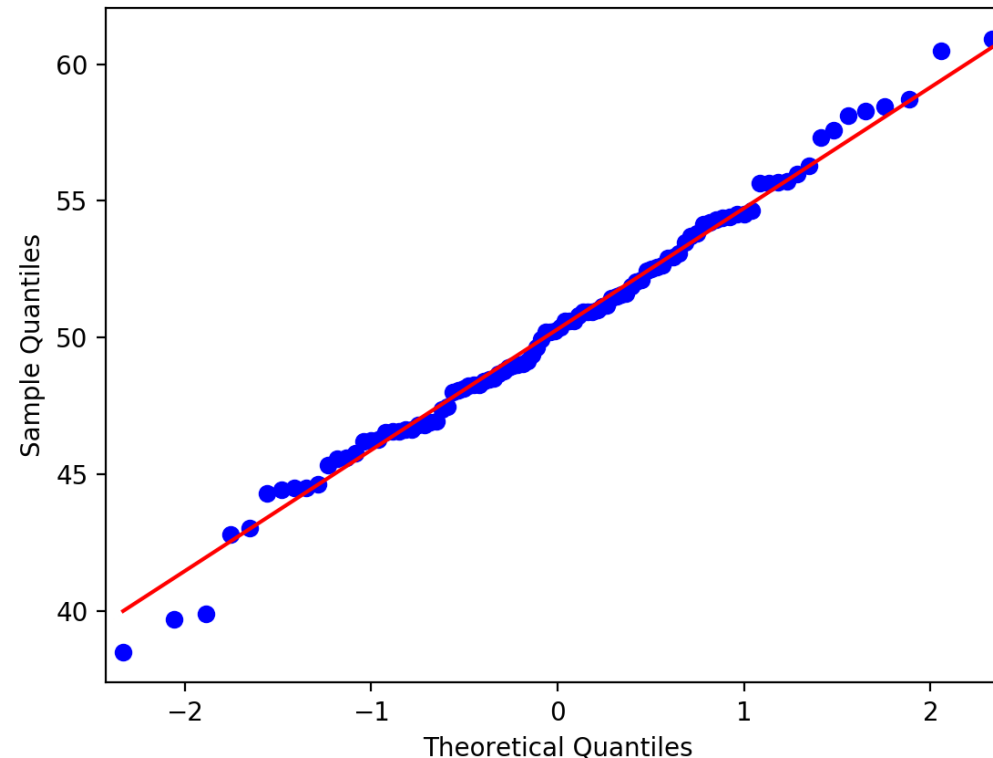


РАСПРЕДЕЛЕНИЯ СЛУЧАЙНЫХ ВЕЛИЧИН



ГРАФИКИ КВАНТИЛЬ-КВАНТИЛЬ

Общий график КК позволяет оценить сходство распределений двух наборов данных. Такие графики создаются так же, как и нормальные графики КК. Единственным отличием является возможность использования любого набора данных в качестве второго набора данных. Если наборы данных имеют одинаковые распределения, точки на общем графике КК выстроятся в прямую (под углом 45 градусов) линию.



ОБЗОР МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ. Часть 3

КУХАЛЬСКИЙ НИКОЛАЙ ГЕННАДЬЕВИЧ