

A/B тестирование

Занятие 2

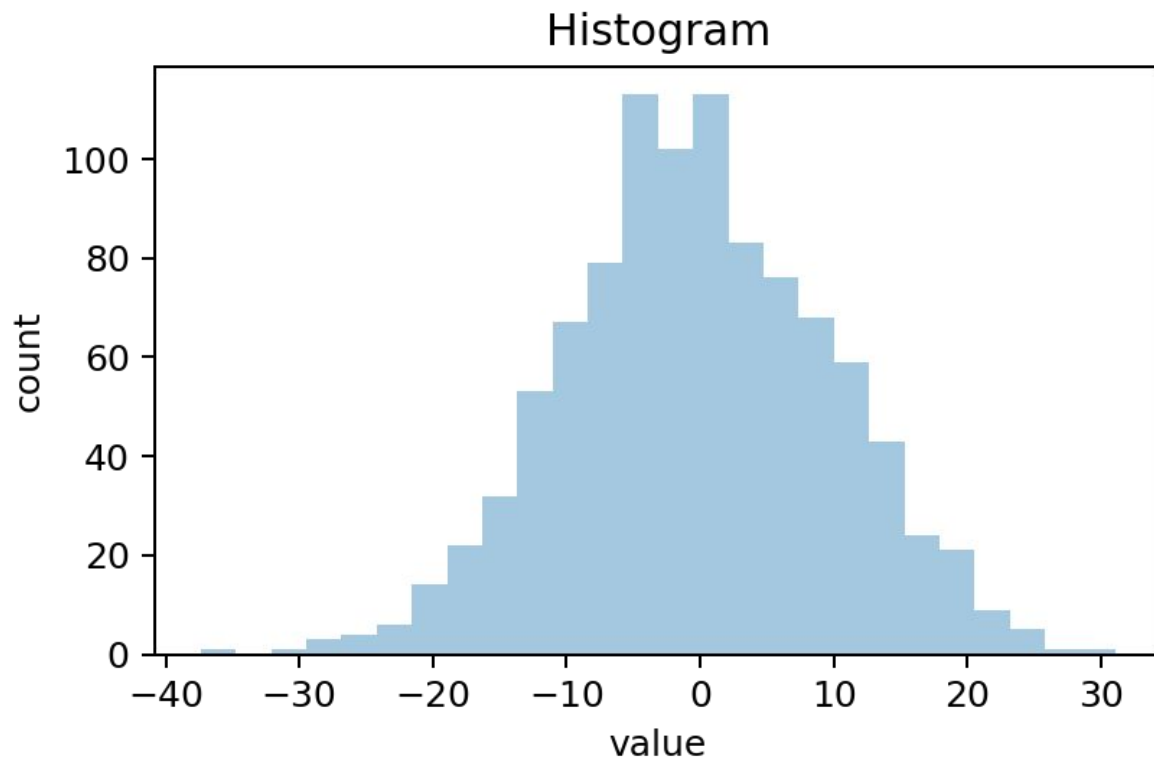
Мурашкин Вячеслав
mvjacheslav@gmail.com

Содержание

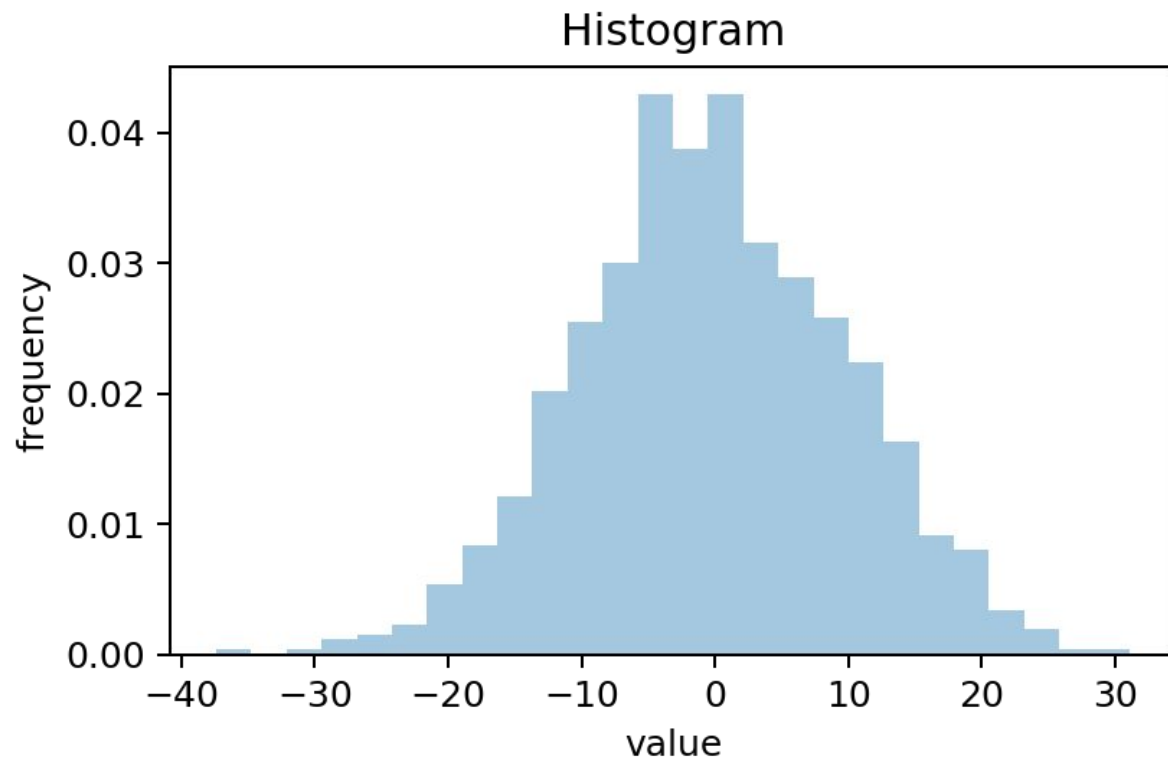
- Гистограммы и плотность распределения
- Выборочное среднее
- Проверка статистических гипотез
- Обзор инструментов A/B тестирования

Гистограммы, плотность распределения,
квантили

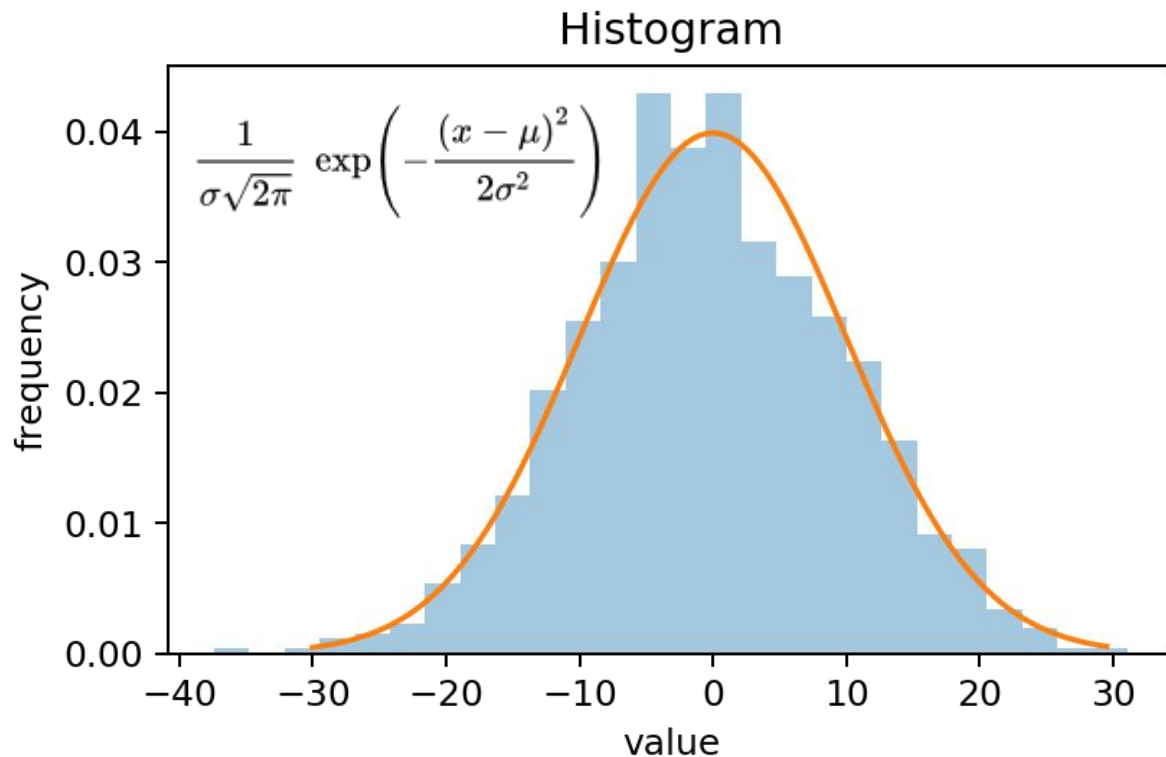
Гистограмма



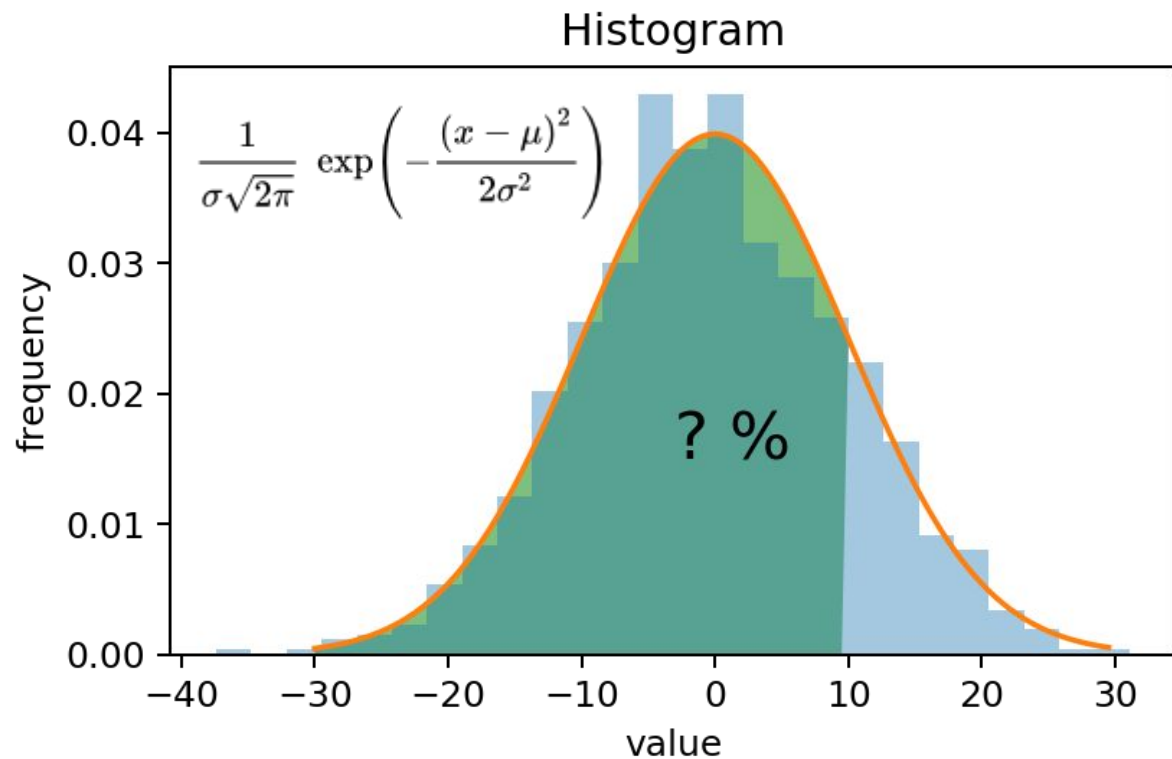
Нормировка гистограммы



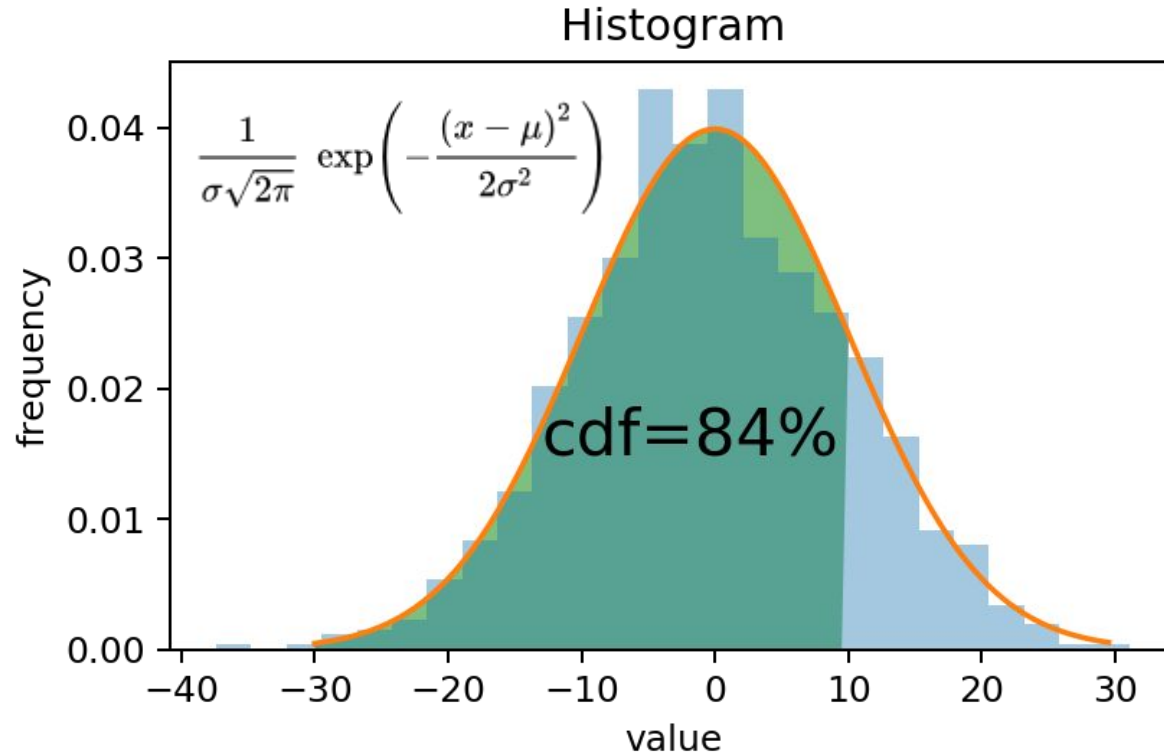
Аппроксимация нормальным распределением



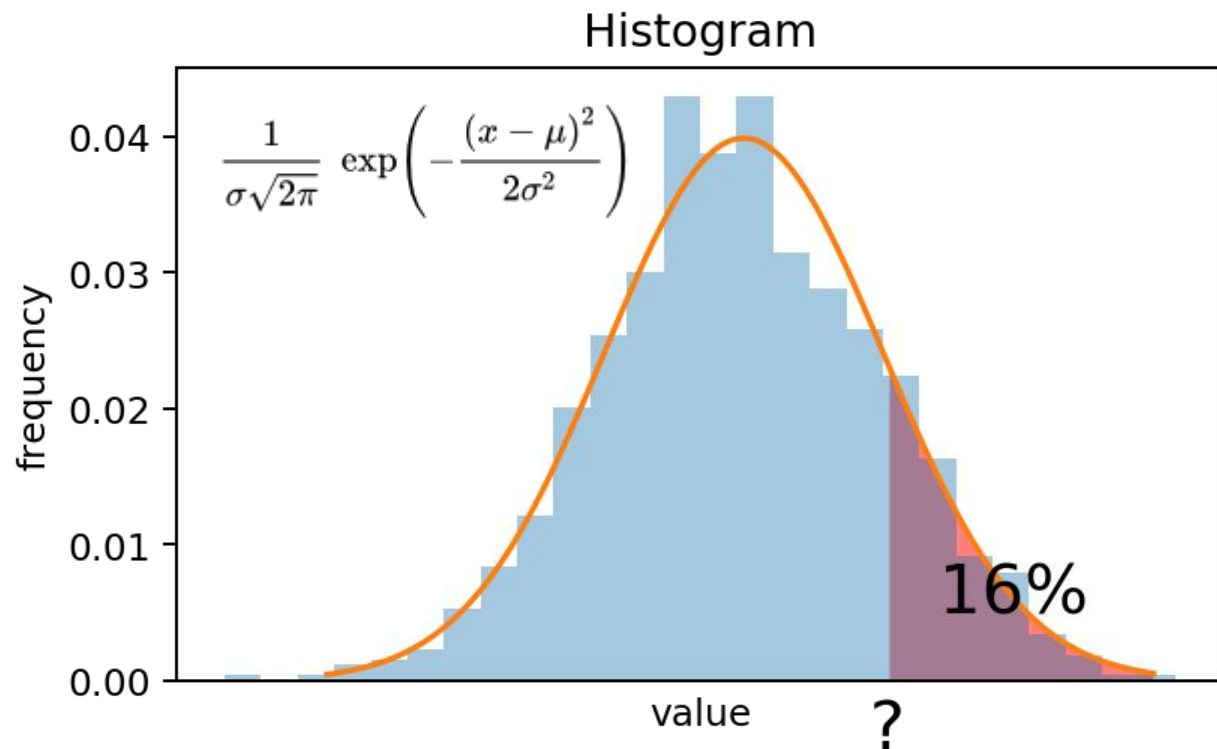
q-Квантиль



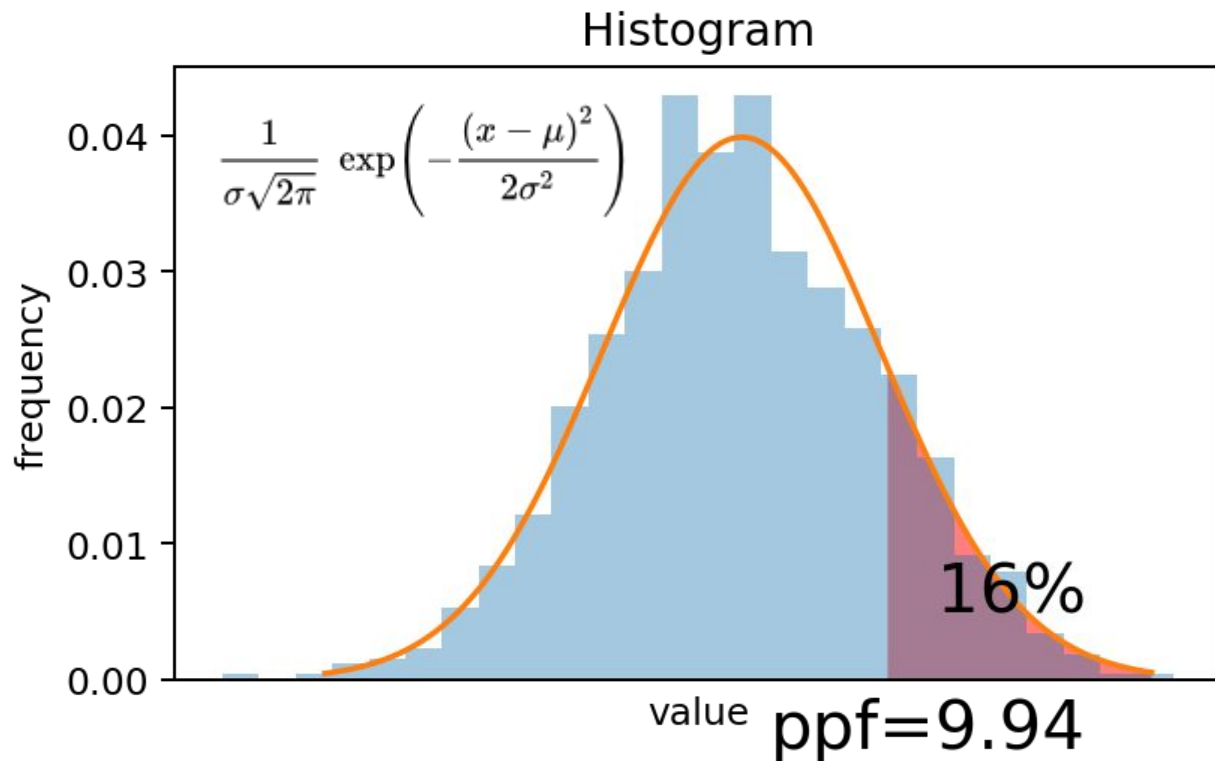
q-Квантиль (cumulative distribution function)



Процентиль



Процентиль (percent point function)

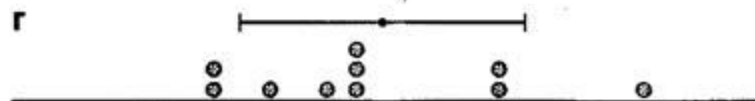
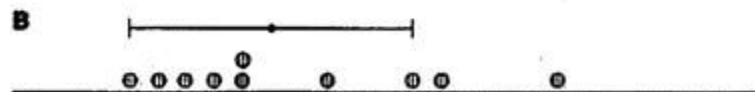
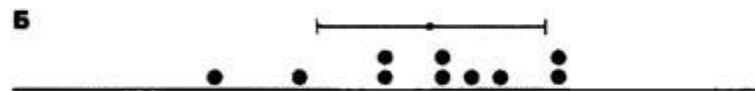
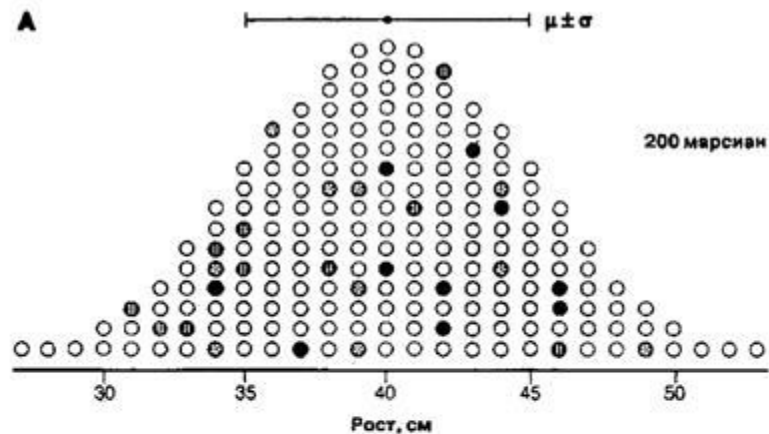


Выборочное среднее и центральная предельная
теорема

Выборочное среднее

- приближение теоретического среднего распределения, основанное на выборке
- дисперсия оценки зависит от числа наблюдений

Выборочное среднее



Дисперсия выборочного среднего и размер семпла

- в процессе эксперимента получаем N значений
- оцениваем среднее полученных значений

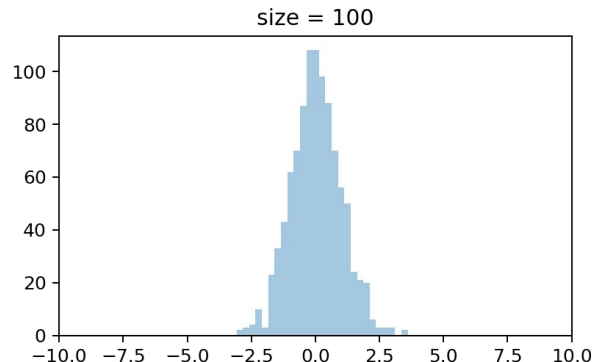
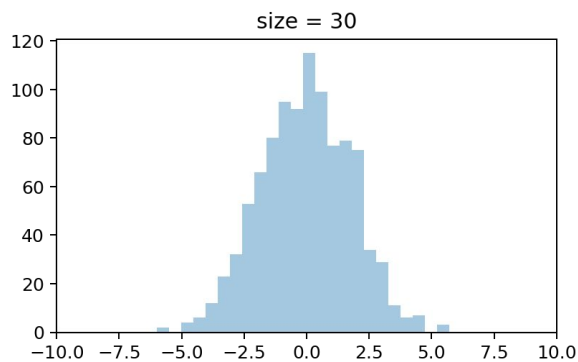
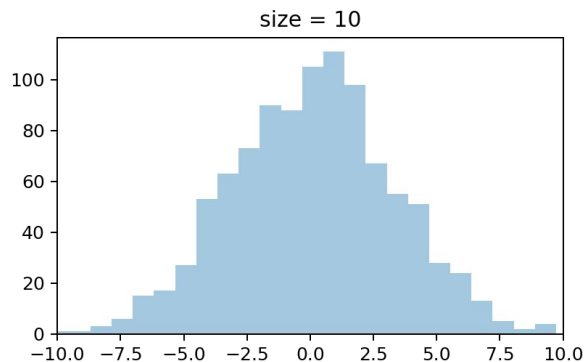
Дисперсия выборочного среднего и размер семпла

- в процессе эксперимента получаем N значений
- оцениваем среднее полученных значений
- повторяем эксперимент M -раз

Дисперсия выборочного среднего и размер семпла

- в процессе эксперимента получаем N значений
- оцениваем среднее полученных значений
- повторяем эксперимент M несколько раз
- получаем M оценок средних
- дисперсия этих оценок зависит от числа наблюдений в эксперименте

Дисперсия выборочного среднего и размер семпла



ЦПТ для выборочного среднего

$\{X_1, X_2, \dots\}$ - независимые наблюдения одинаково распределенных случайных величин

$E[X_i] = \mu$ - мат. ожидание случайной величины

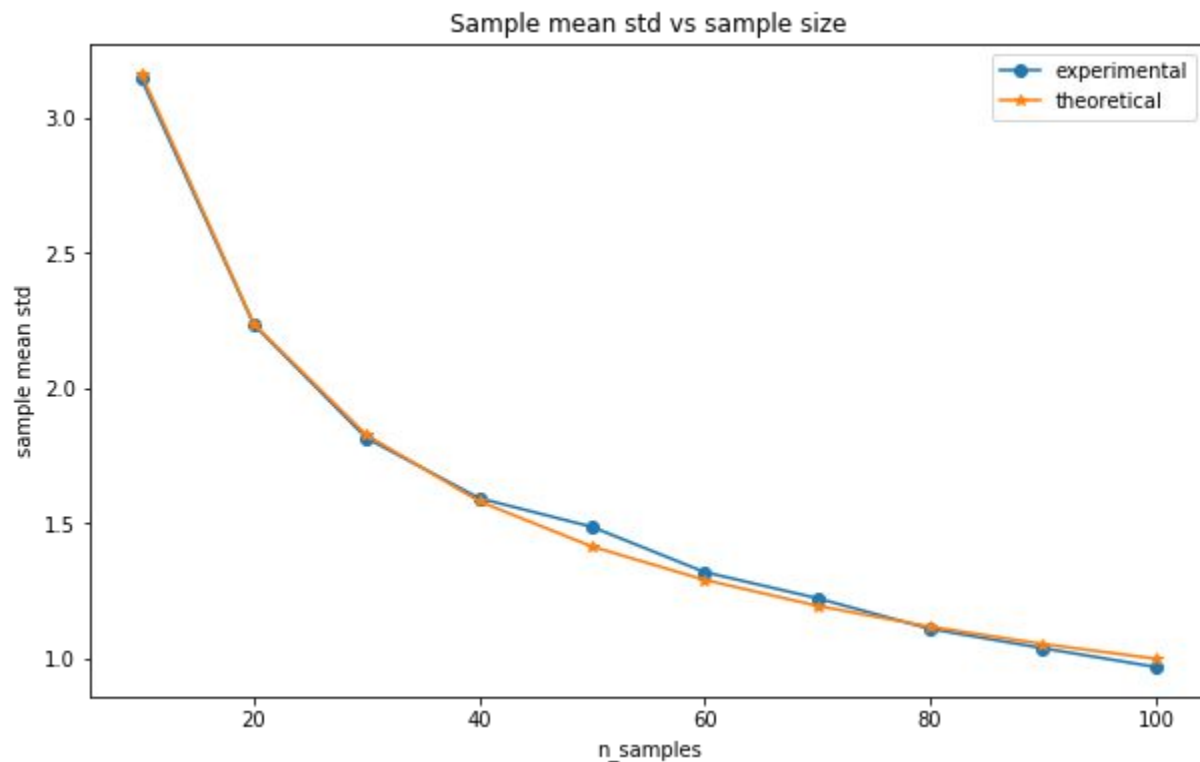
$\text{Var}[X_i] = \sigma^2$ - дисперсия генеральной совокупности

$S_n := \frac{X_1 + \dots + X_n}{n}$ - среднее значение семпла n случайных величин

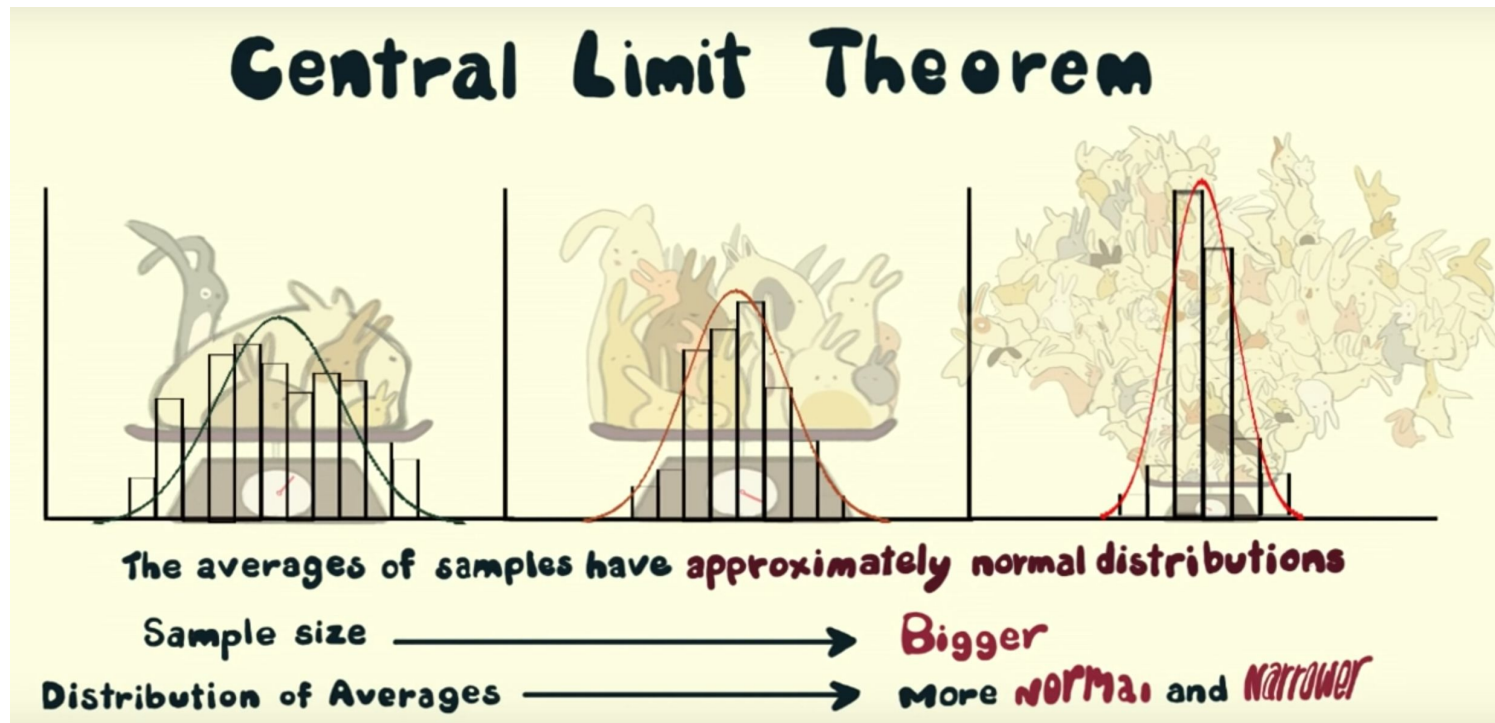
$\sqrt{n} (S_n - \mu) \xrightarrow{d} N(0, \sigma^2)$ - связь распределения средних с нормальным распределением

https://en.wikipedia.org/wiki/Central_limit_theorem

ЦПТ для выборочного среднего



ЦПТ для выборочного среднего



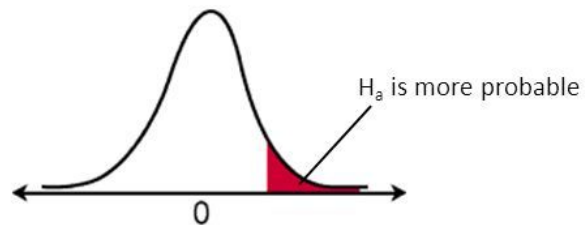
<https://www.youtube.com/watch?v=jvoxEYmQHNM>

Проверка статистических гипотез

Проверка статистических гипотез

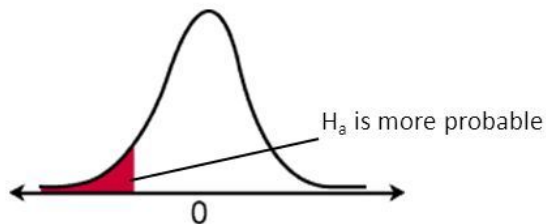
- вычисляется разность средних в сплитах
- эта разность сравнивается с нулем
- нулевая гипотеза: “Разницы в средних нет”
- альтернативная гипотеза: “Разность средних отлична от нуля”

Проверка статистических гипотез



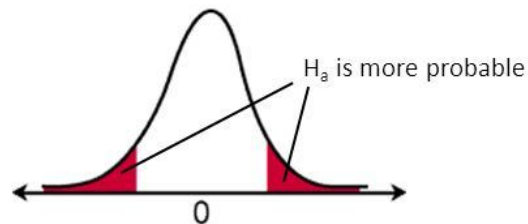
Right-tail test

$$H_a: \mu > \text{value}$$



Left-tail test

$$H_a: \mu < \text{value}$$



Two-tail test

$$H_a: \mu \neq \text{value}$$

Проверка статистических гипотез

- оцениваем дисперсию разности
- задаемся порогом “уверенности”, например 95%

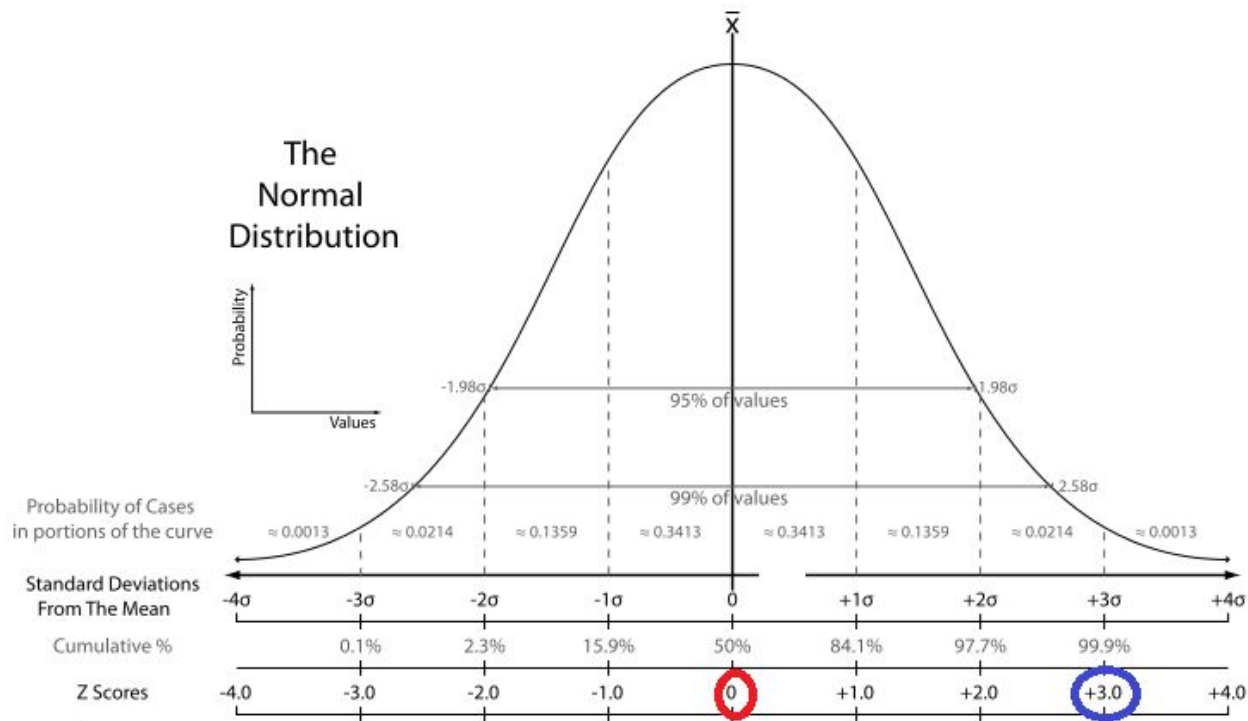
Проверка статистических гипотез

- оцениваем дисперсию разности
- задаемся порогом “уверенности”, например 95%
- определяем порог для выбранного значения “уверенности”
- если разность больше порога - считаем верной альтернативную гипотезу

Дисперсия суммы (разности) независимых с. в.
равна сумме дисперсий

$$\sigma_{-} = \sqrt{\sigma_x^2 + \sigma_y^2}$$

Стандартное нормальное распределение



Z-test

$$z = \frac{\bar{X}_A - \bar{X}_B}{SE_{AB}}$$

- z-статистика, используется для оценки уровня значимости

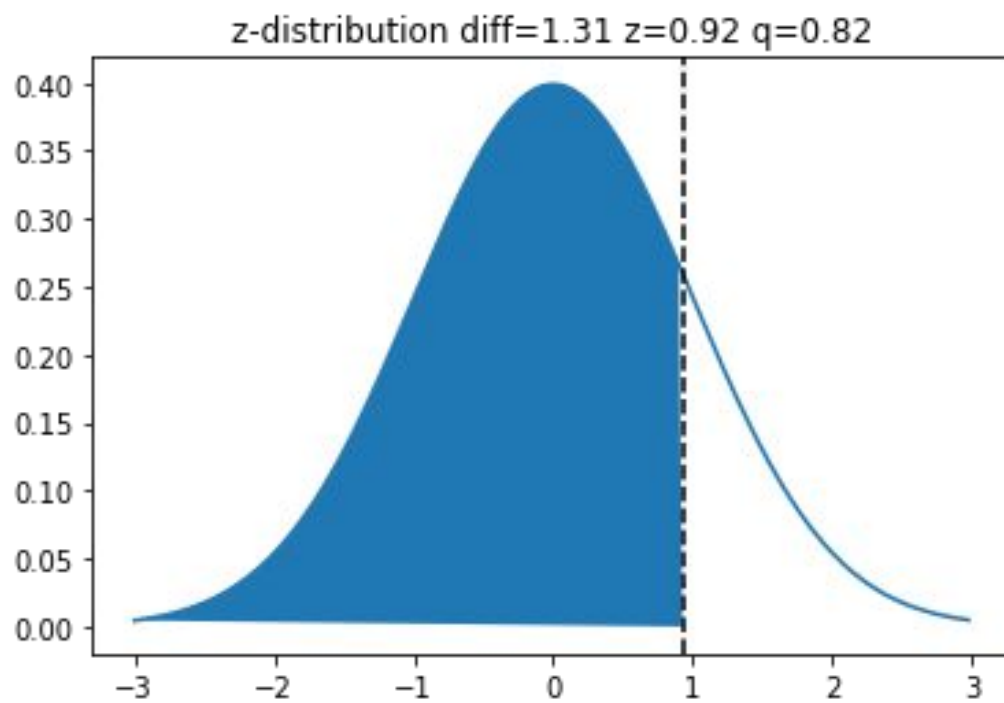
$$SE_{AB} = \sqrt{SE_A^2 + SE_B^2}$$

- формула для стандартной ошибки разности

$$SE = \frac{\sigma}{\sqrt{n}}$$

- связь стандартной ошибки и дисперсии

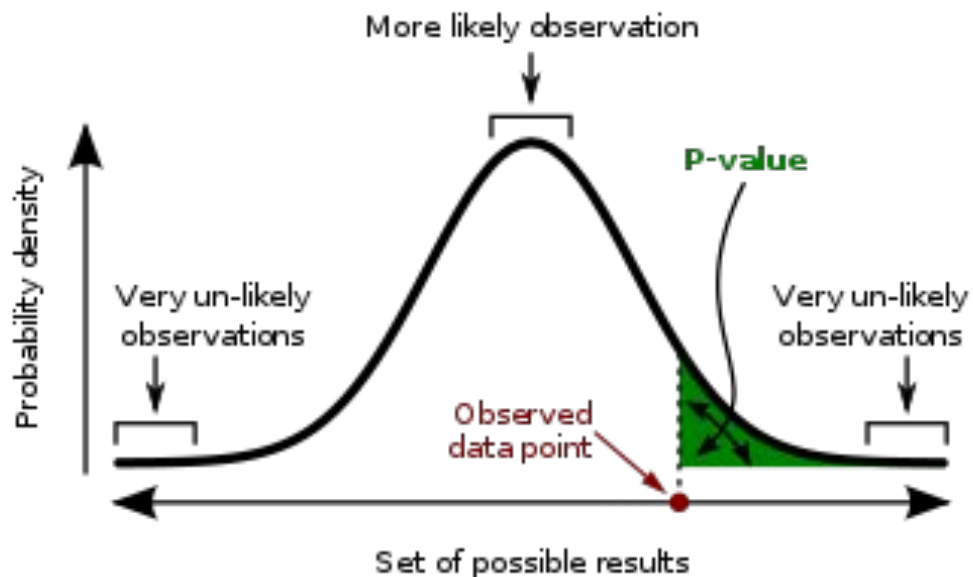
Z-test



P-value

- вероятность получить тоже значение разницы средних или больше, при условии верности нулевой гипотезы
- вероятность ошибки отклонения нулевой гипотезы

P-value



A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

Распределение Стьюдента

- мало примеров в выборке
- дисперсию сложно оценить
- используем функцию `ppf` для получения квантилей/процентилей (t-статистика)
- степень свободы

Распределение Стьюдента

$$\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$$

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{2}{n}}}$$

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\frac{\bar{X} - \mu}{S/\sqrt{n}}$$

$$s_p = \sqrt{\frac{s_{X_1}^2 + s_{X_2}^2}{2}}$$

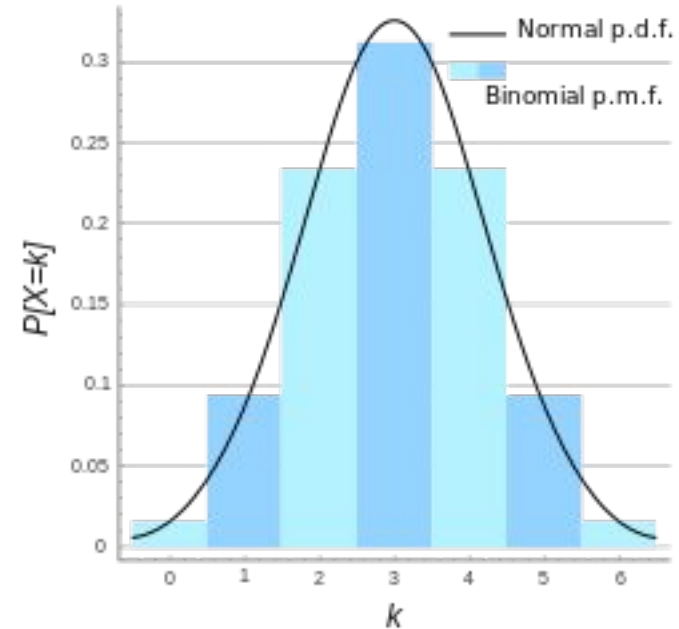
ЦПТ для пропорций

$$p = \frac{x}{n}$$

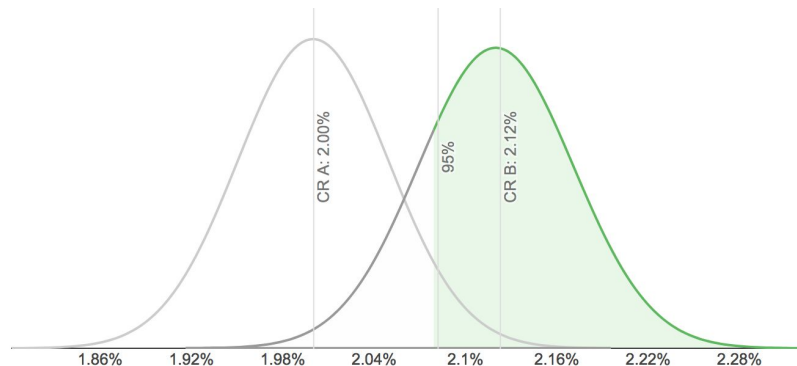
- среднее значение числа положительных событий в семпле

$$\sigma = \sqrt{\frac{p(1-p)}{n}}$$

- стандартное отклонение среднего



Online Calculator



Conversion Rate Control

Conversions A / Visitors A

2.00%

Conversion Rate B

Conversions B / Visitors B

2.12%

Relative uplift in Conversion Rate

$CR_B - CR_A / CR_A$

6.00%

Observed Power

77.57%

p value

0.0455

Z-score

$(CR_B - CR_A) / SE_{\text{difference}}$

1.6897

Standard error A

$(CR_A * (1 - CR_A) / \text{Visitors}_A)^{1/2}$

0.000495

Standard error B

$(CR_B * (1 - CR_B) / \text{Visitors}_B)^{1/2}$

0.000509

Std. Error of difference

$SE_{\text{difference}} = (SE_A^2 + SE_B^2)^{1/2}$

0.00071

<https://abtestguide.com/calc/>

Непараметрические методы

Тест Манна-Уитни

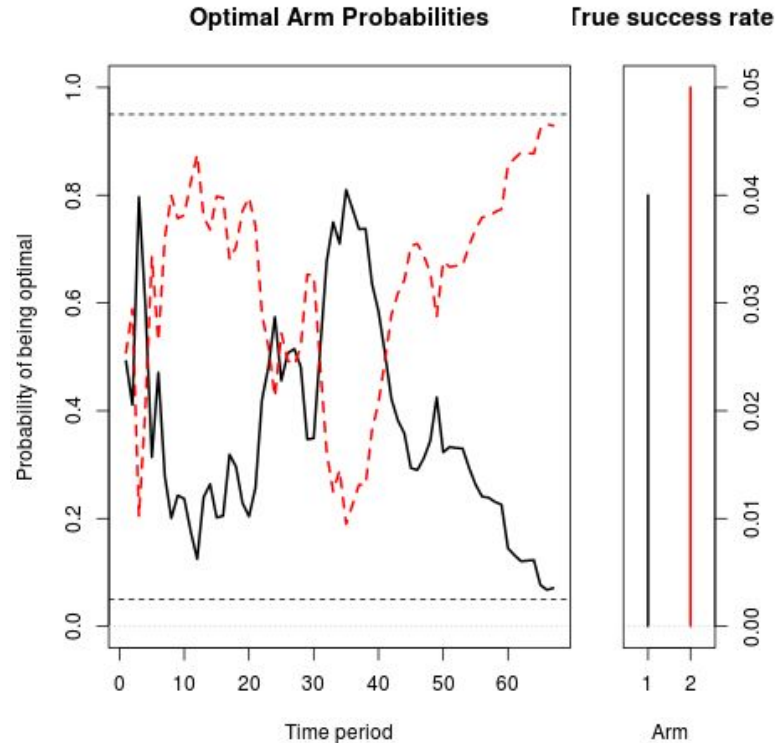
Placebo	7	5	6	4	12
New Drug	3	6	4	2	1

$$U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$$

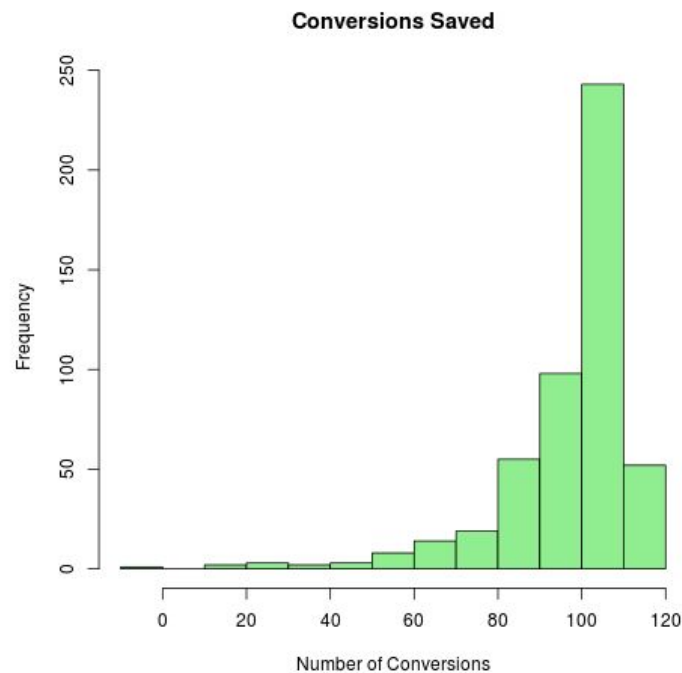
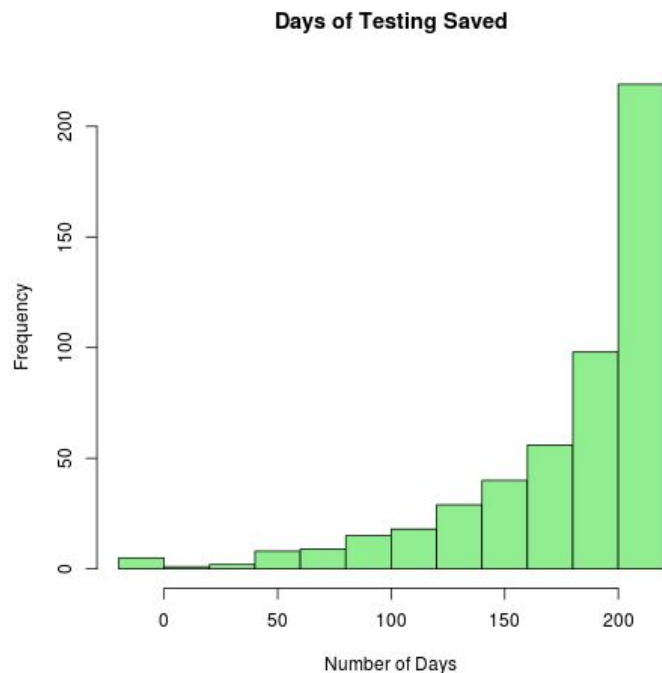
$$U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$$

Multi-armed bandit experiments

Multi-armed bandit experiments



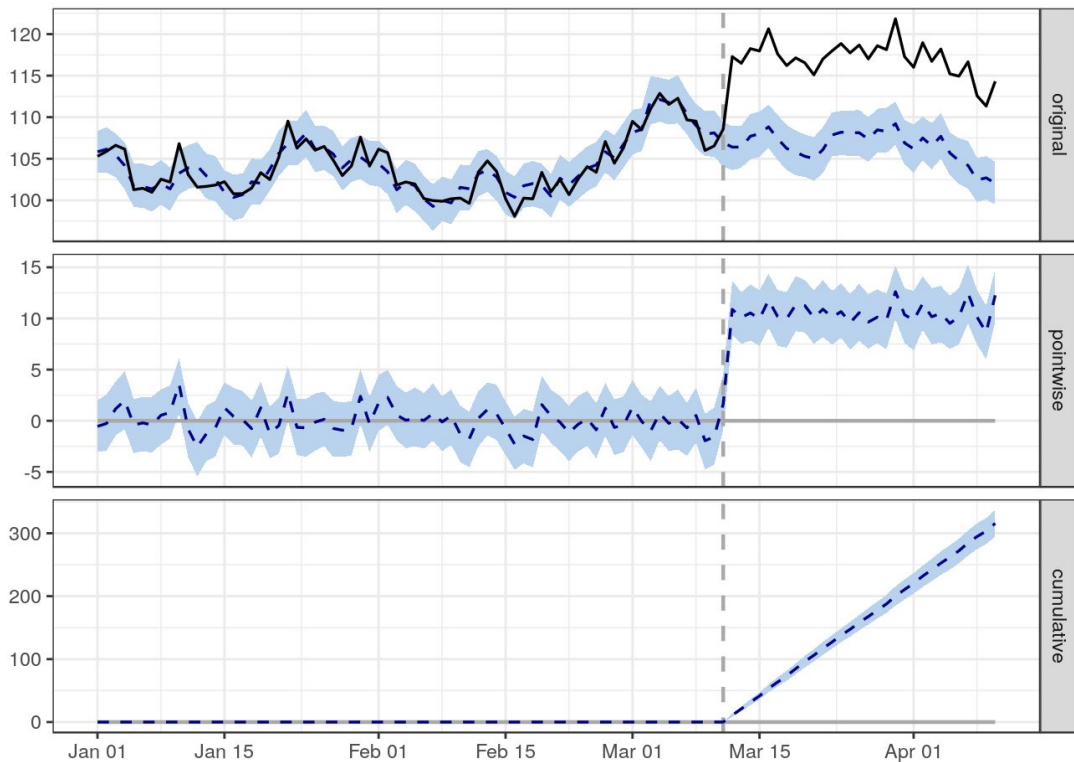
Multi-armed bandit experiments



<https://support.google.com/analytics/answer/2844870?hl=en>

Causal Impact

Causal Impact



<https://google.github.io/CausalImpact/CausalImpact.html>

Полезные материалы

- [Central Limit Theorem](#)
- [OpenIntro Statistics 3rd Edition, 2015](#)
- [Introduction to Probability and Data](#)
- [Mathematical Biostatistics Boot Camp](#)
- [Probability Cheatsheet](#)