

A/B тестирование

Занятие 2

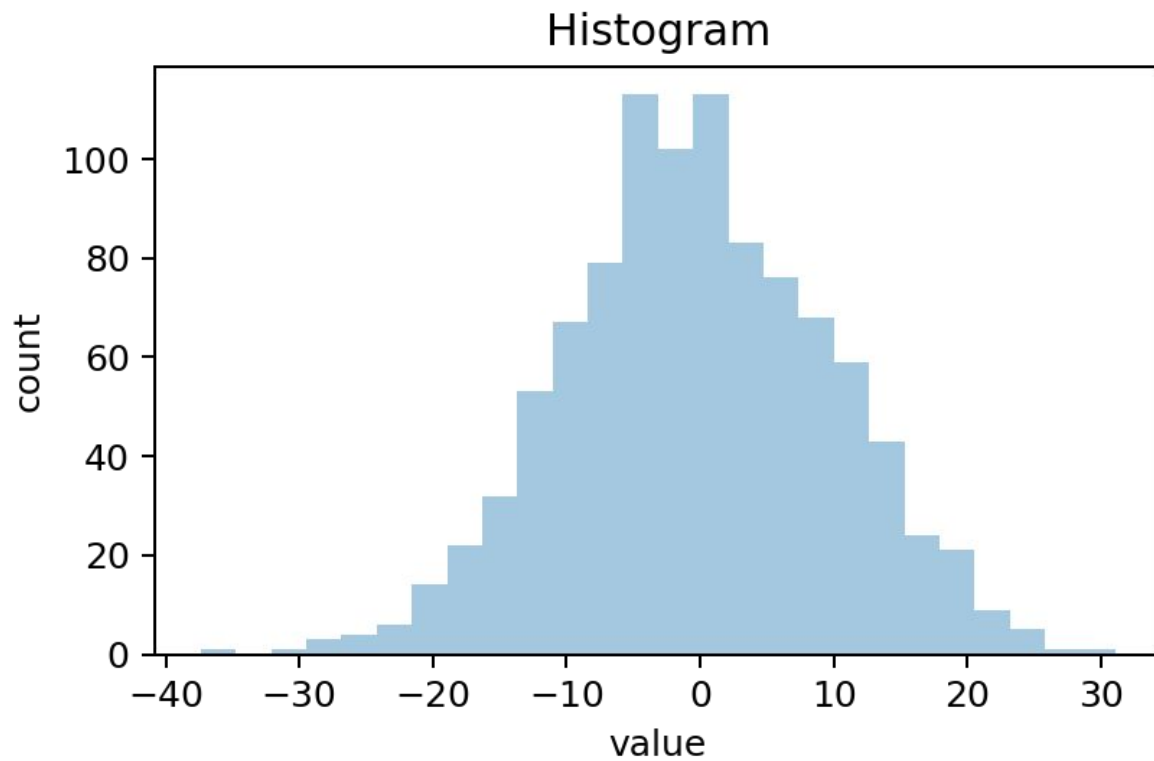
Мурашкин Вячеслав
mvjacheslav@gmail.com

Содержание

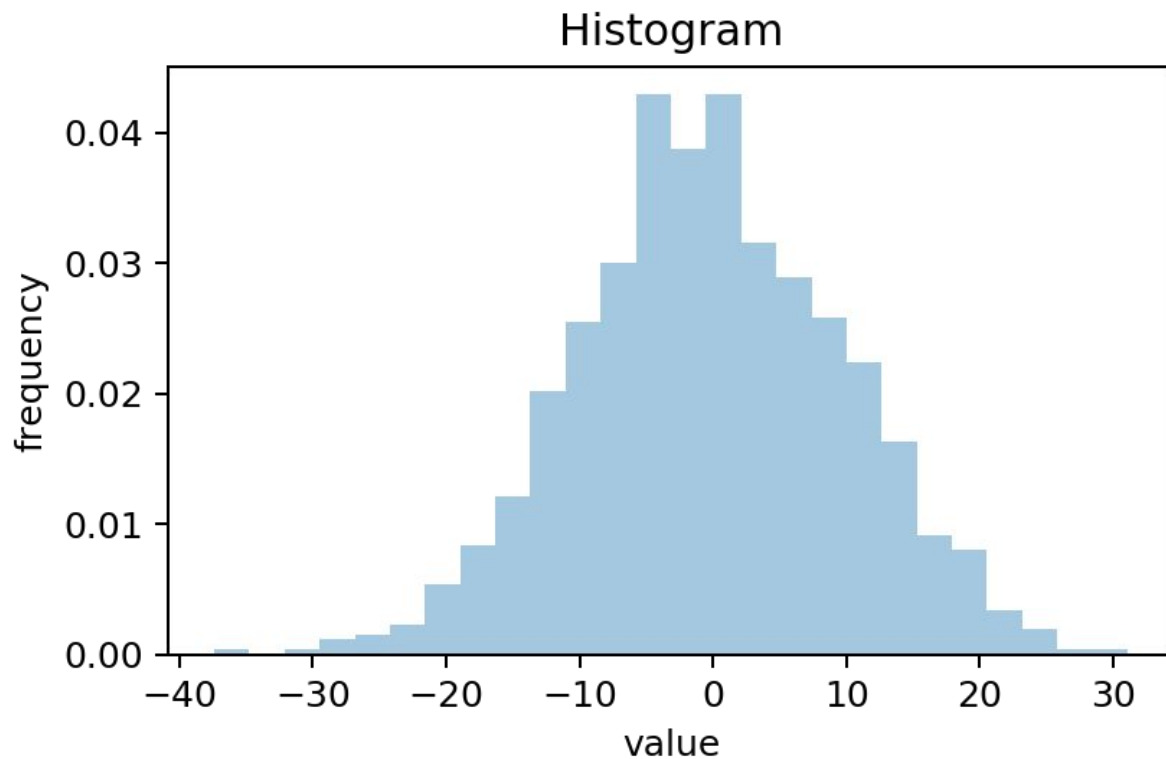
- Гистограммы и плотность распределения
- Выборочное среднее
- Проверка статистических гипотез
- Обзор инструментов A/B тестирования

Гистограммы и плотность распределения

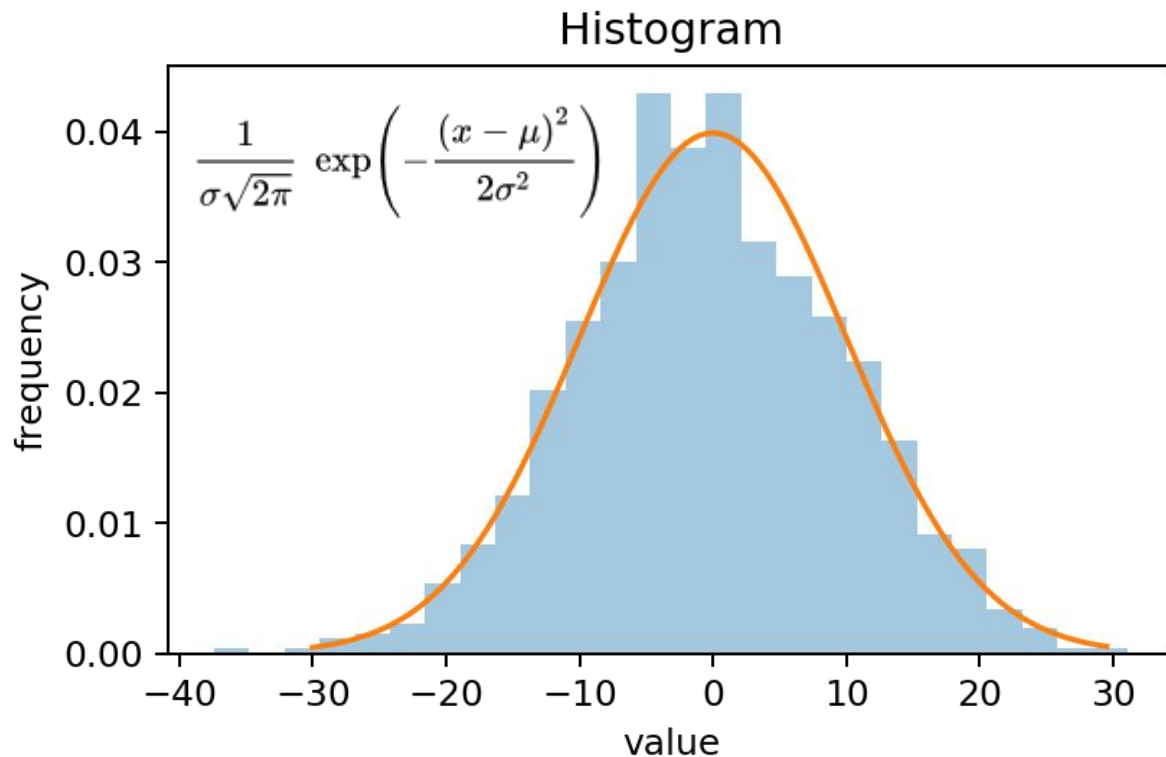
Гистограмма



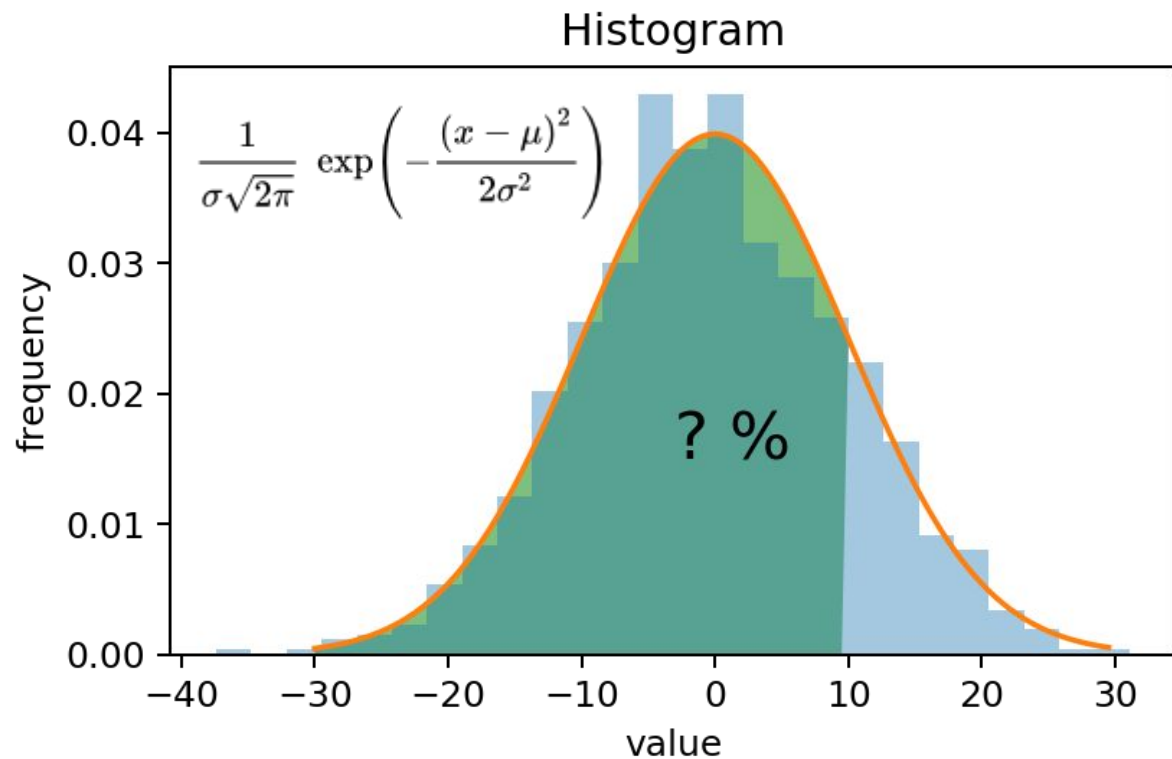
Нормировка гистограммы



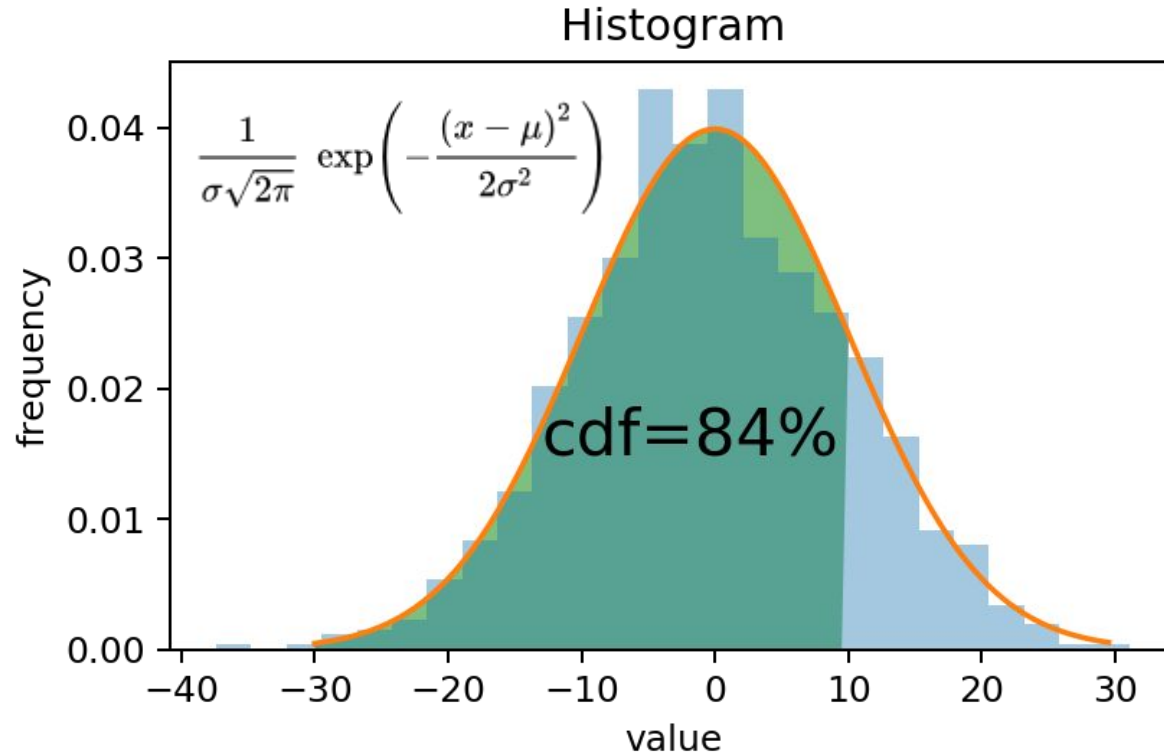
Аппроксимация нормальным распределением



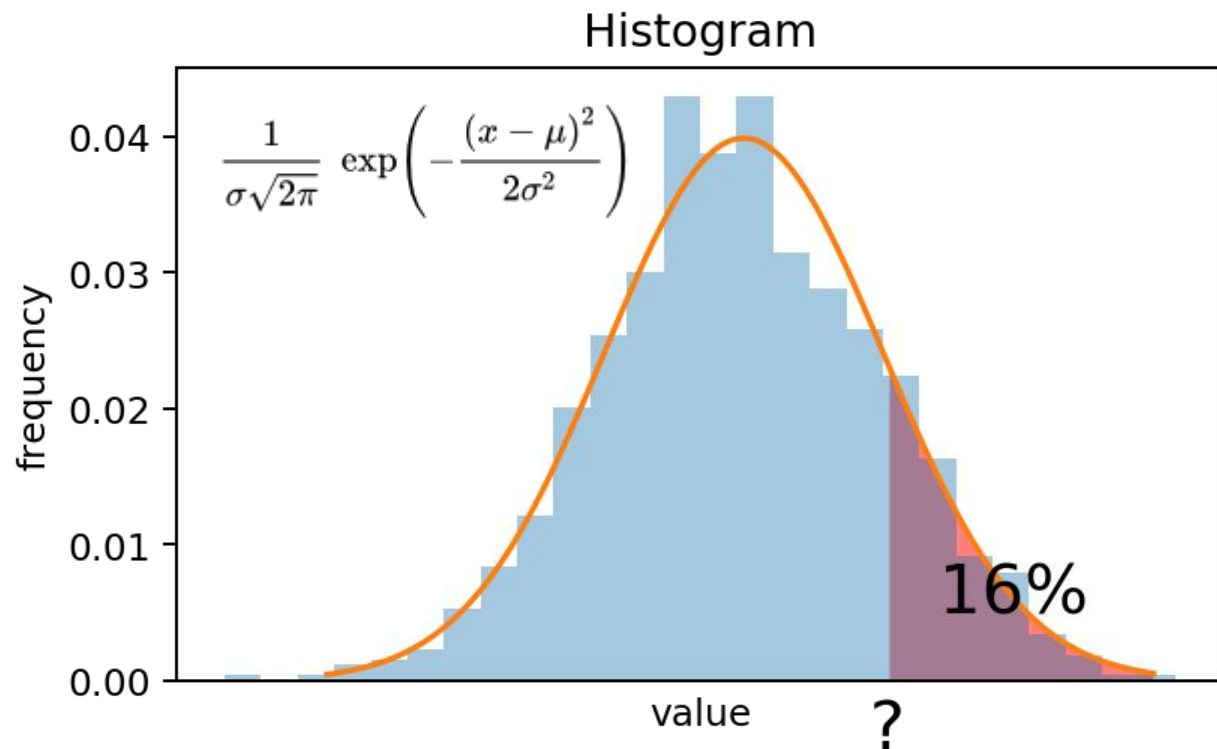
q-Квантиль



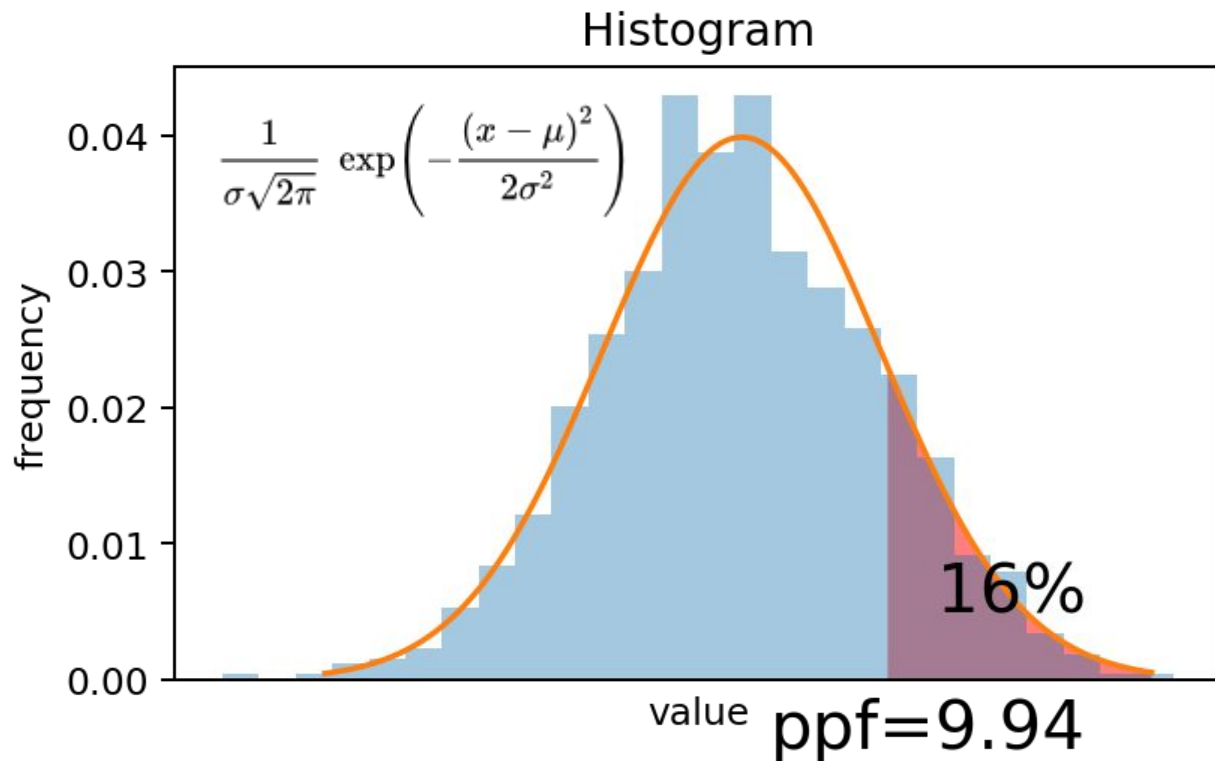
q-Квантиль (cumulative distribution function)



Процентиль



Процентиль (percent point function)

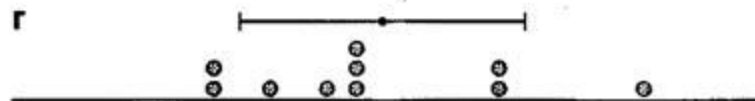
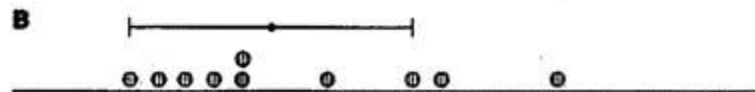
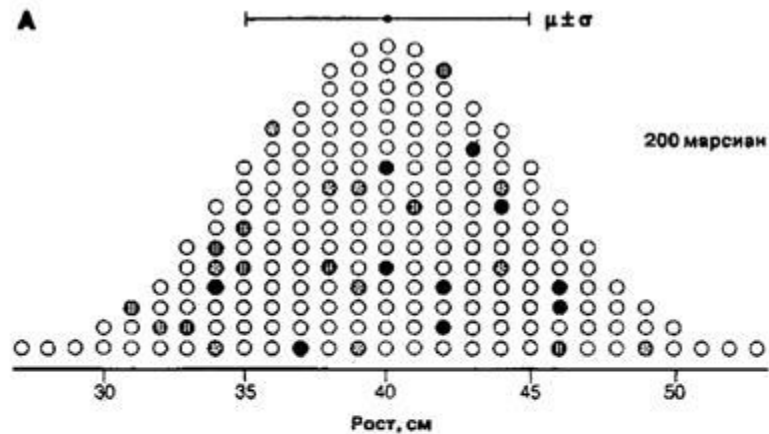


Выборочное среднее

Выборочное среднее

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

Выборочное среднее



Дисперсия оценки среднего от размера семпла

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

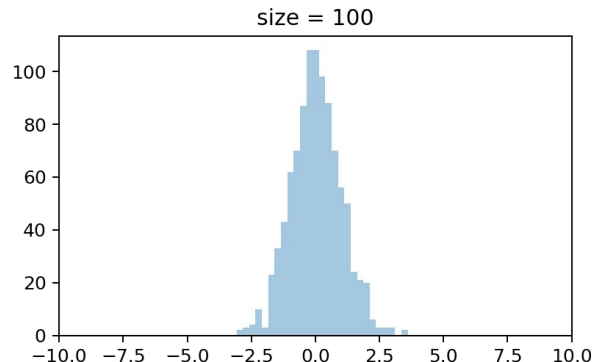
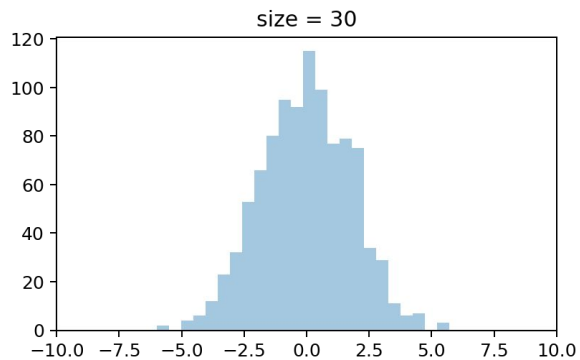
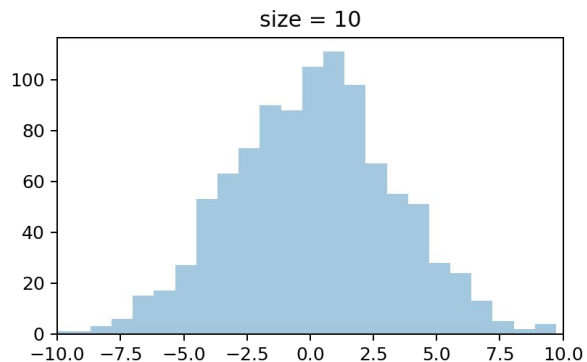
Дисперсия оценки среднего от размера семпла

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

Дисперсия оценки среднего от размера 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 оценки среднего от размера семпла



Распределение Стьюдента

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

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

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

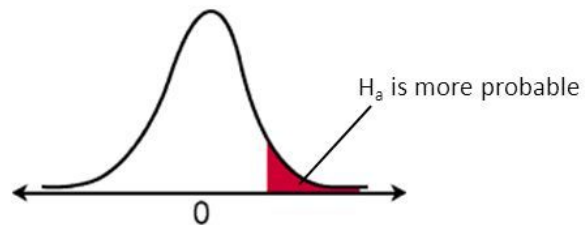
Упражнение: зависимость дисперсии оценки среднего от числа наблюдений
в эксперименте

Проверка статистических гипотез

Проверка статистических гипотез

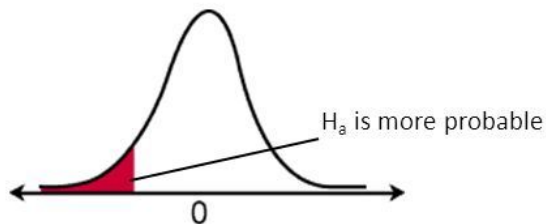
- вычисляется разность средних в сплитах
- эта разность сравнивается с нулем
- нулевая гипотеза: “Разницы в средних нет”
- альтернативная гипотеза: “Разность средних отлична от нуля”

Проверка статистических гипотез



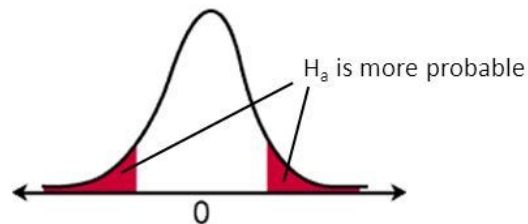
Right-tail test

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



Left-tail test

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



Two-tail test

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

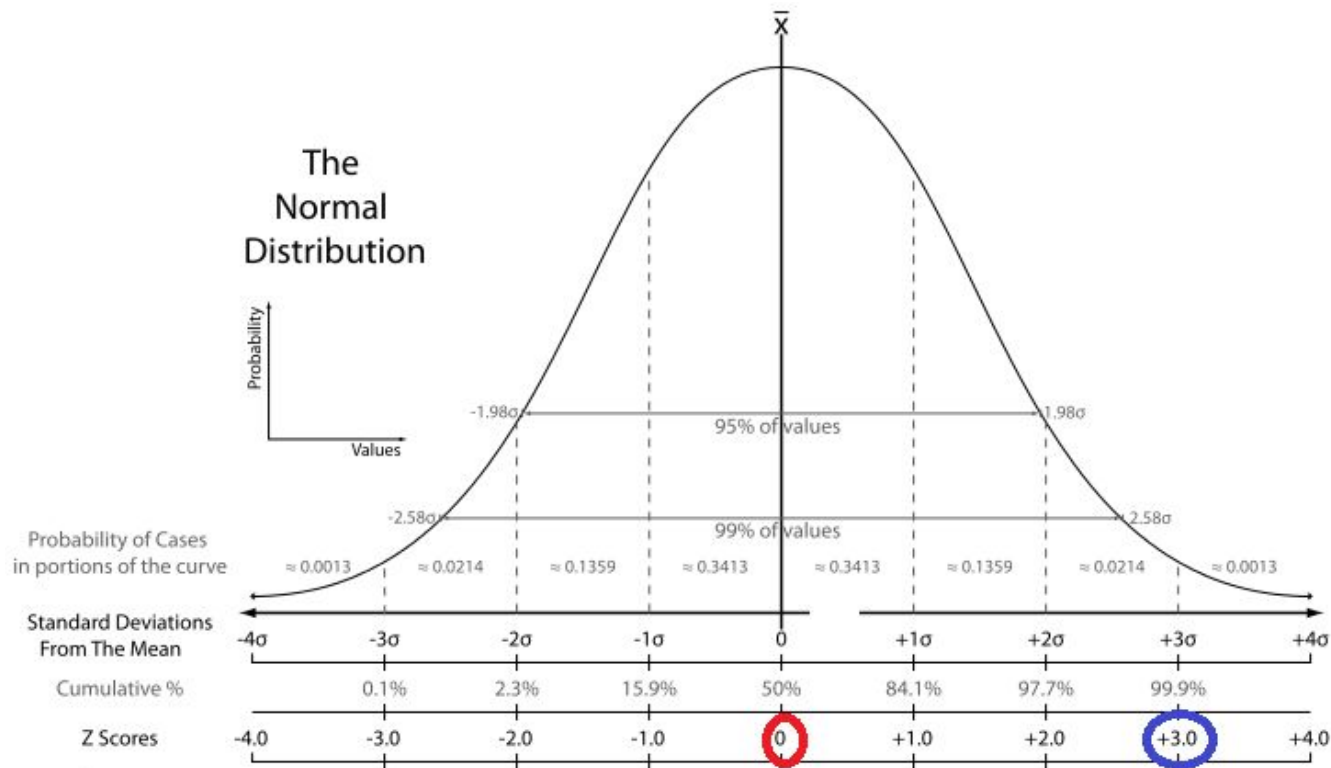
Проверка статистических гипотез

- 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

Проверка статистических гипотез



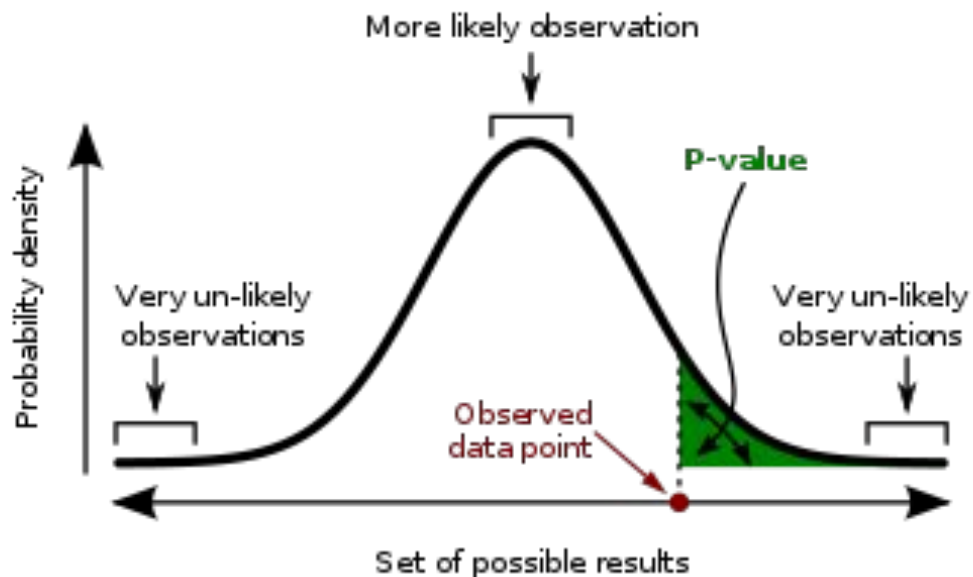
Дисперсия суммы (разности) независимых с. в.

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

P-value

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

P-value

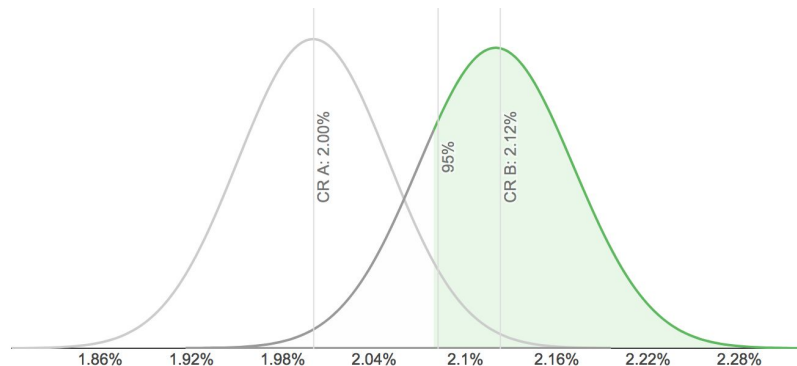


A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

Упражнение: калькулятор необходимого размера выборки

Обзор инструментов

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

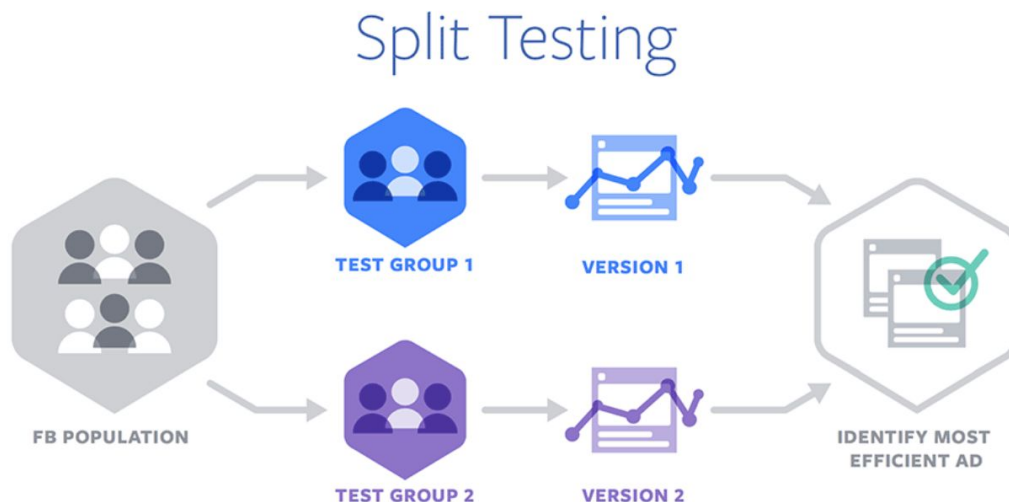
$SE_{\text{difference}} = (SE_A^2 + SE_B^2)^{1/2}$

0.00071

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

Facebook Business

Facebook Business

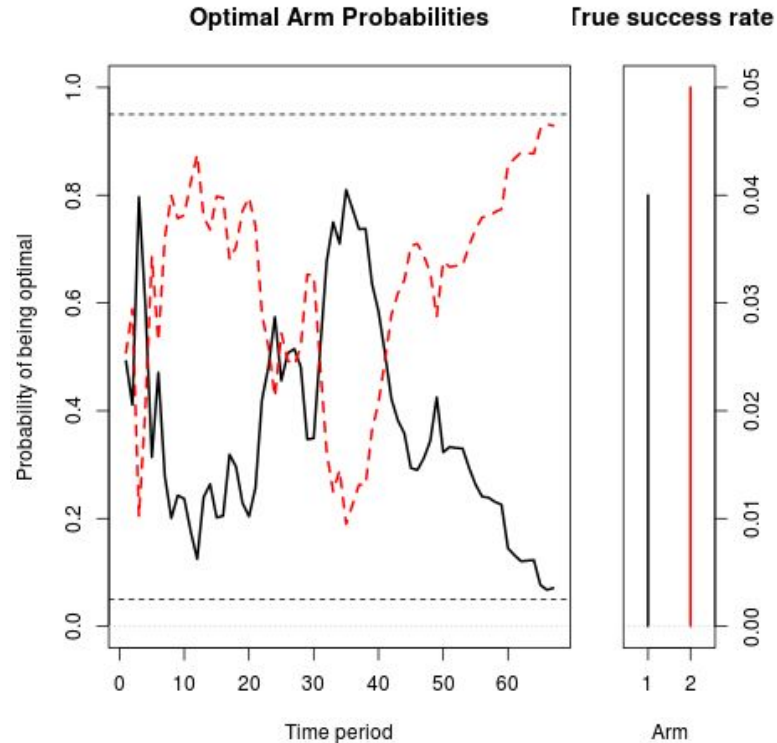


If you opt in, we'll let you know which version of your ad is performing best as soon as we have enough data. If you prefer to wait until all data has been compiled, you'll get this information once your campaign is complete. You can apply these learnings by restarting the winning campaign or when creating future campaigns.

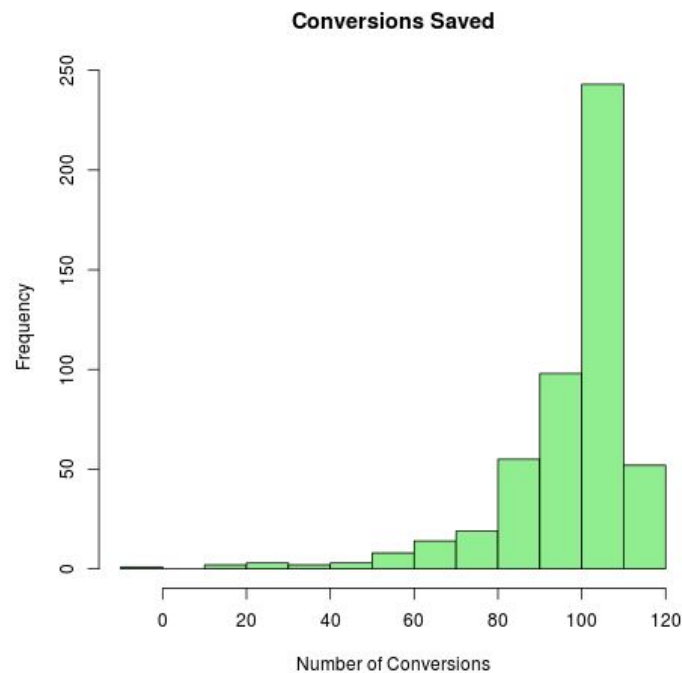
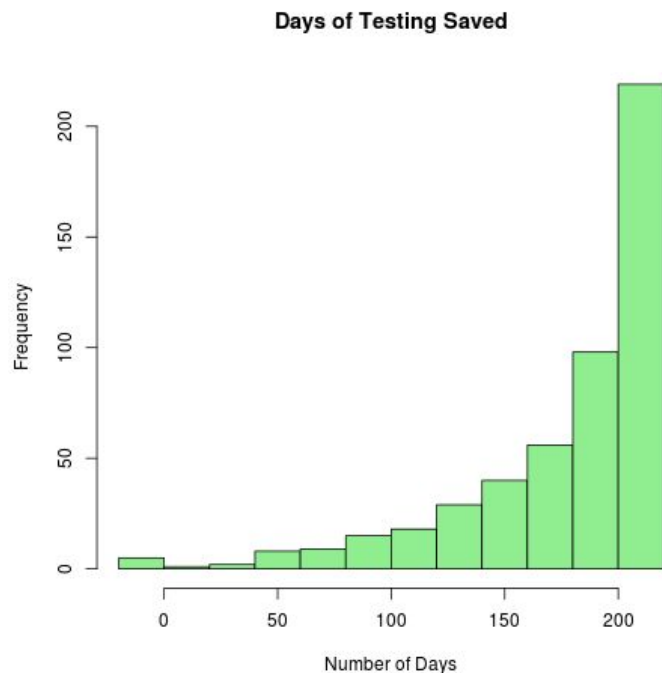
<https://www.facebook.com/business/news/optimize-your-ads-with-split-testing>

Multi-armed bandit experiments

Multi-armed bandit experiments



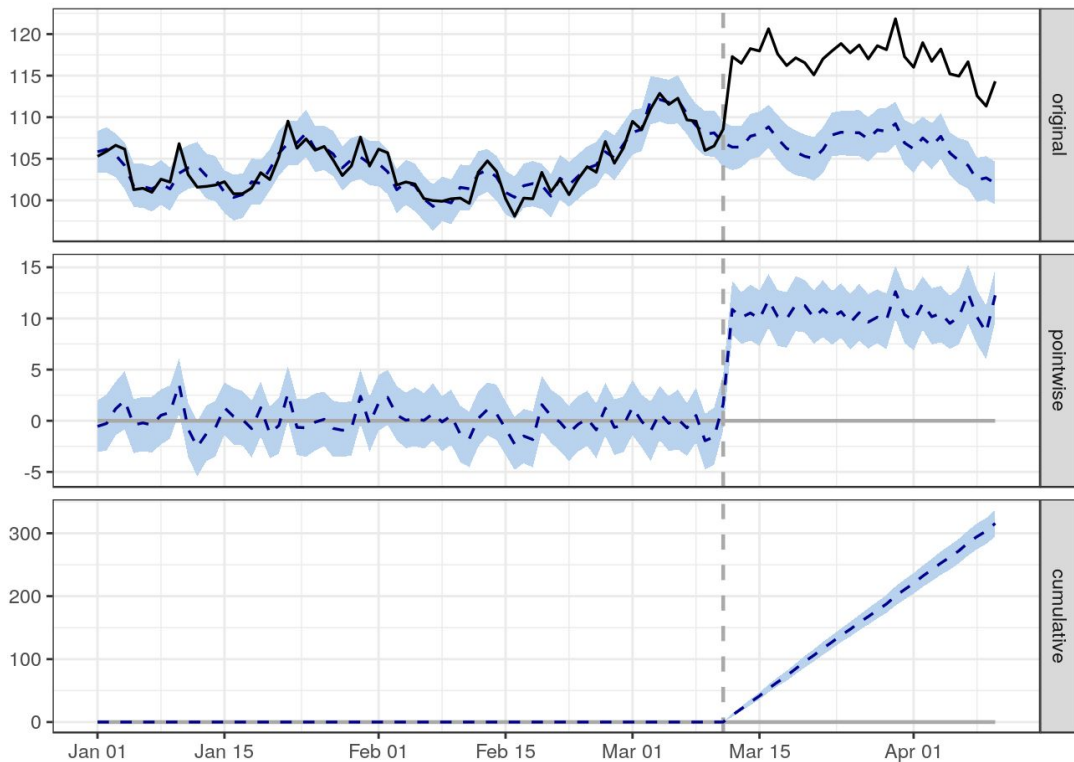
Multi-armed bandit experiments



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

Causal Impact

Causal Impact



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

Полезные материалы

- [Introduction to Probability and Data](#)
- [Mathematical Biostatistics Boot Camp](#)