

# A/B тестирование

Занятие 2

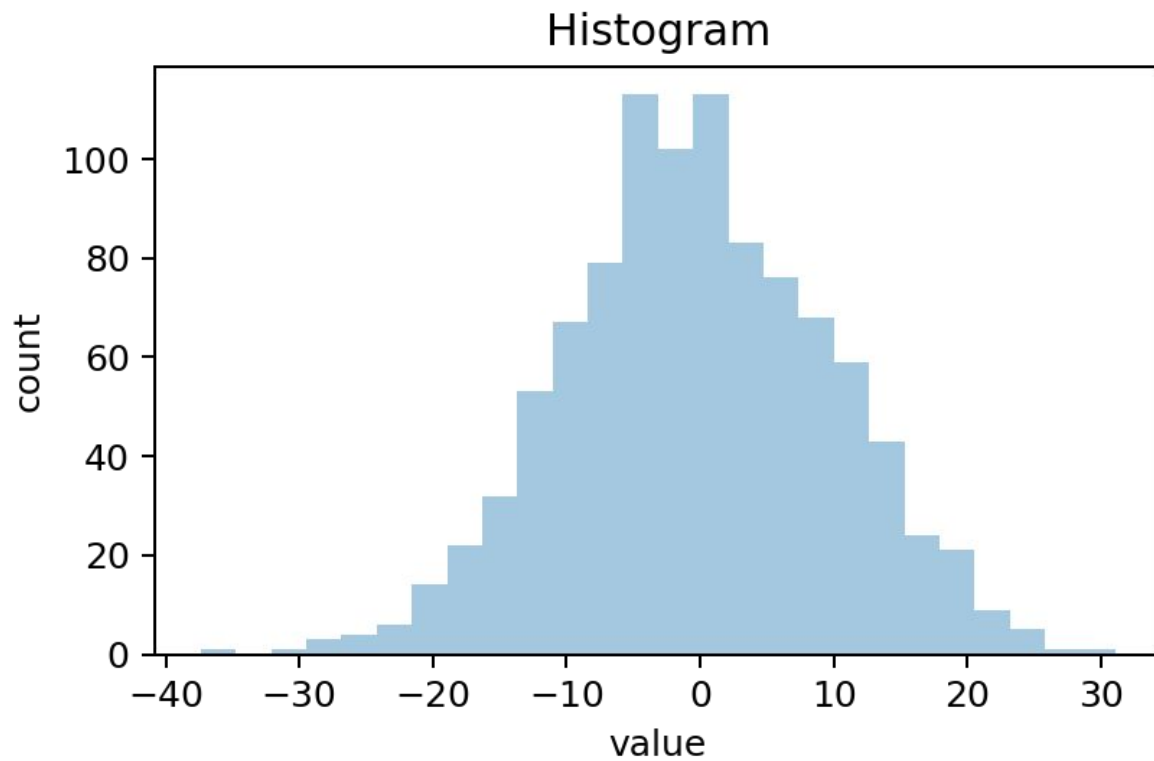
По материалам Вячеслава Мурашкина

# Содержание

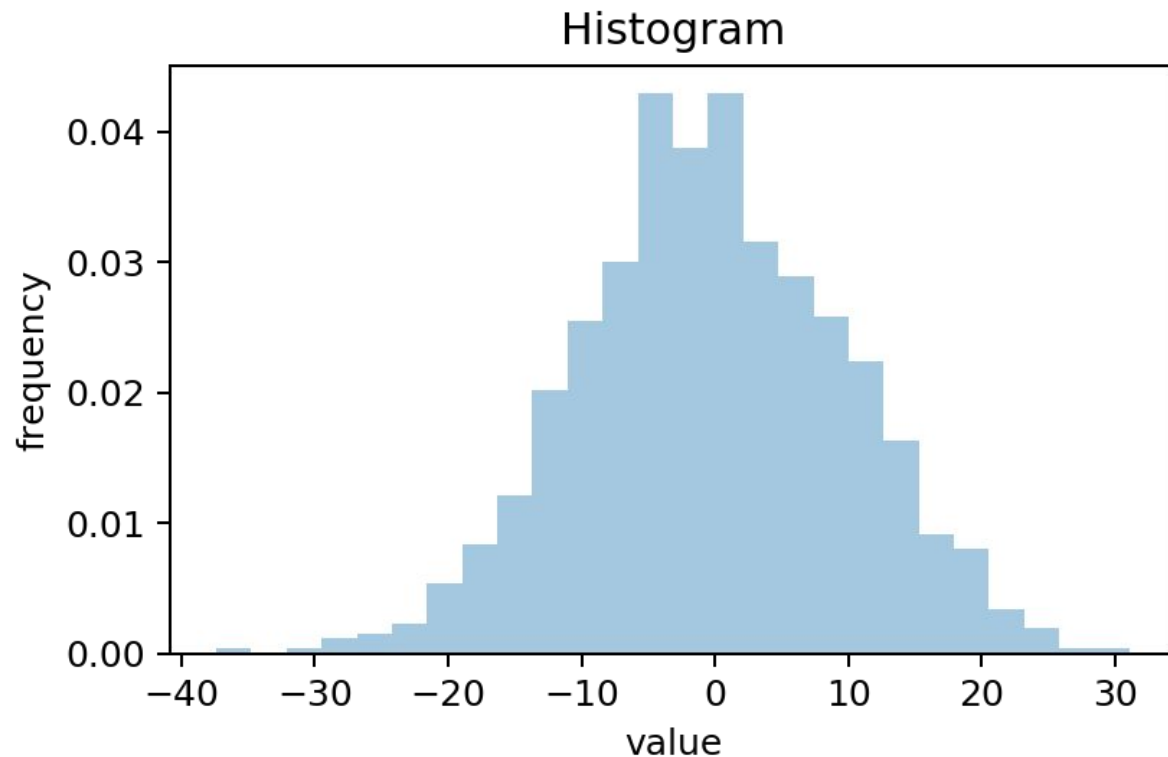
- Гистограммы и плотность распределения
- Выборочное среднее
- Проверка статистических гипотез
- Обзор инструментов A/B тестирования

Гистограммы, плотность распределения,  
квантили

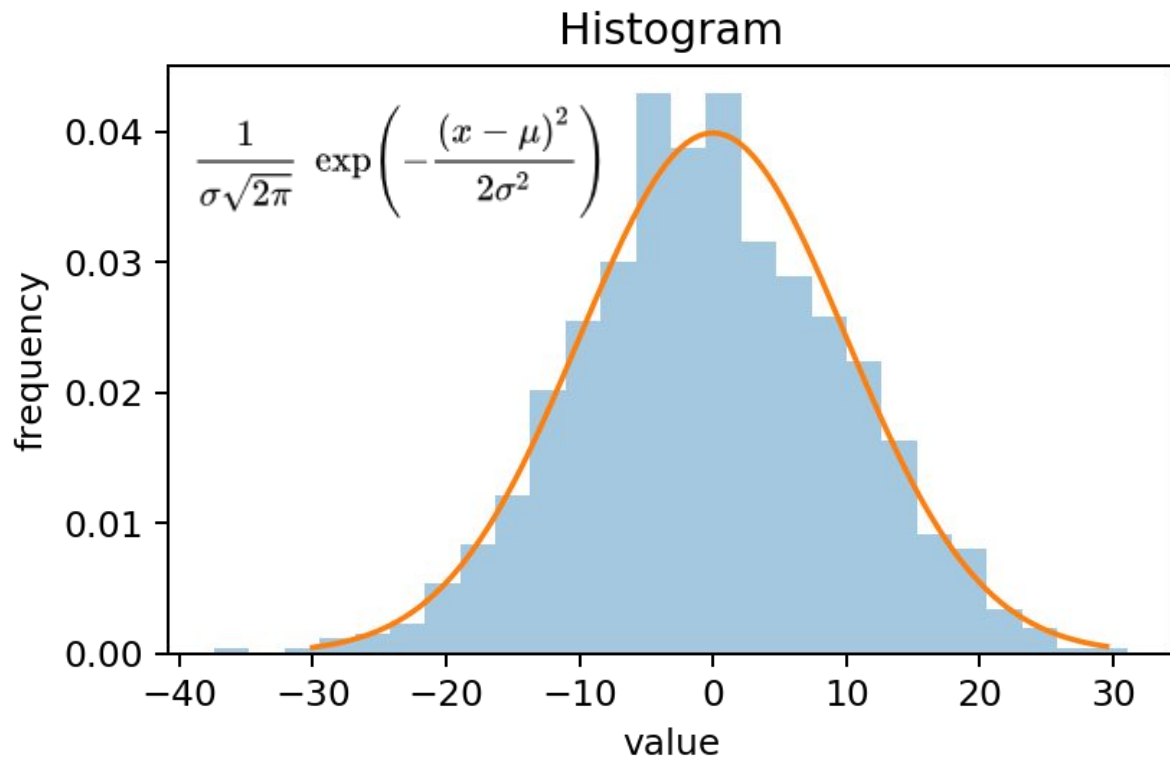
# Гистограмма



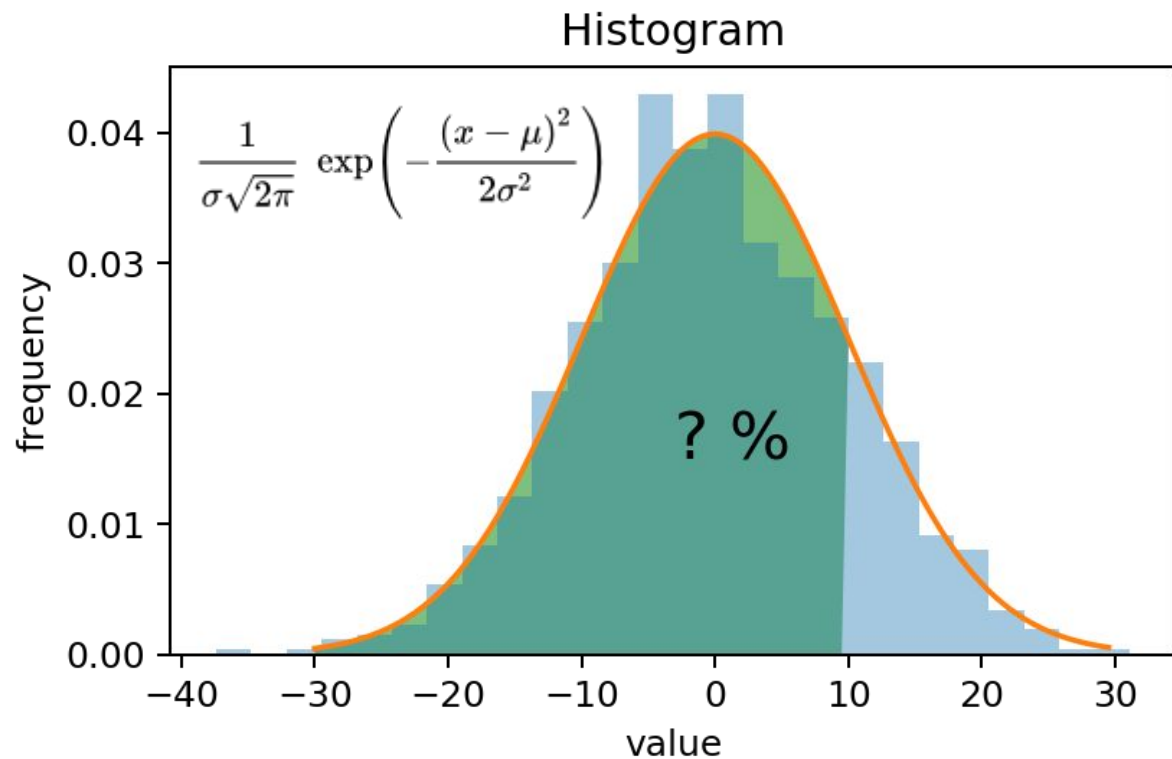
# Нормировка гистограммы



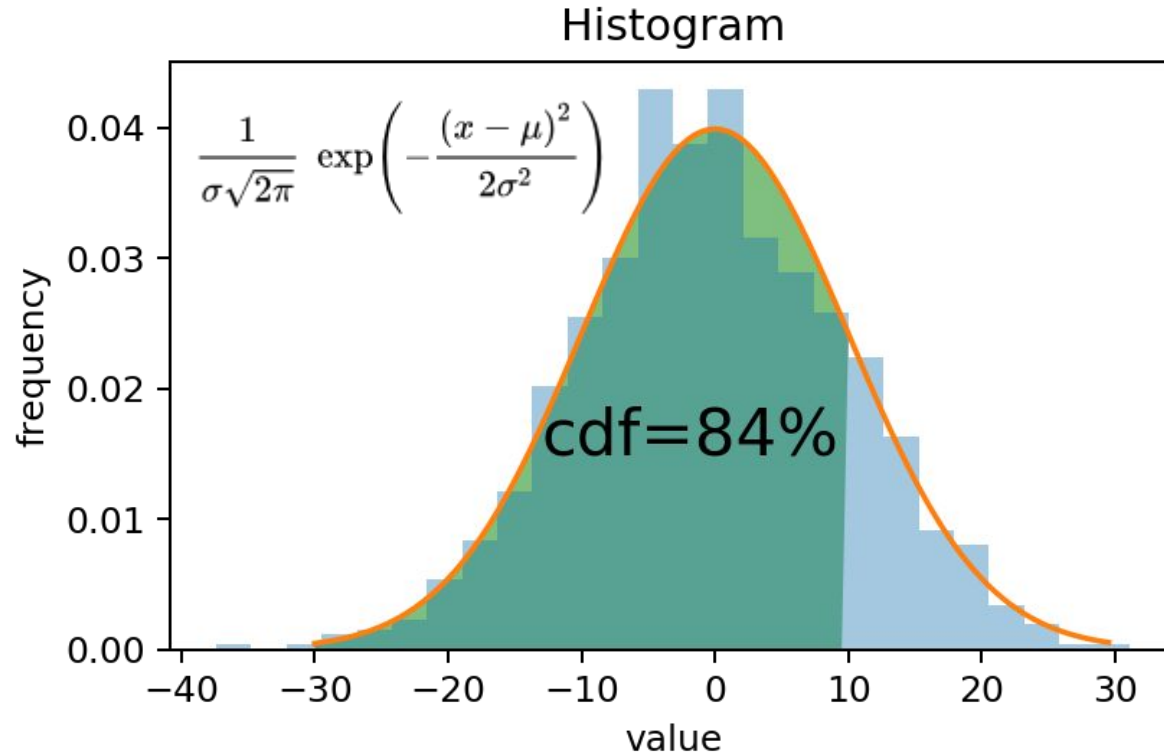
# Аппроксимация нормальным распределением



# q-Квантиль

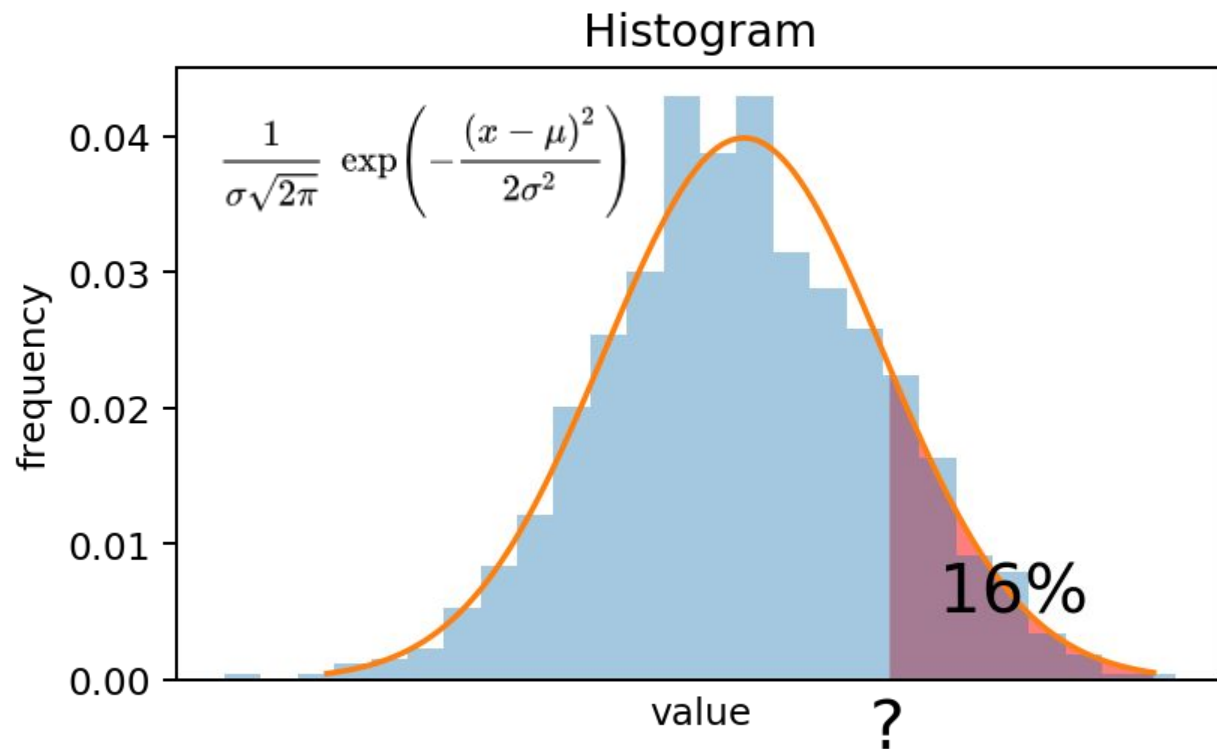


# q-Квантиль (cumulative distribution function)

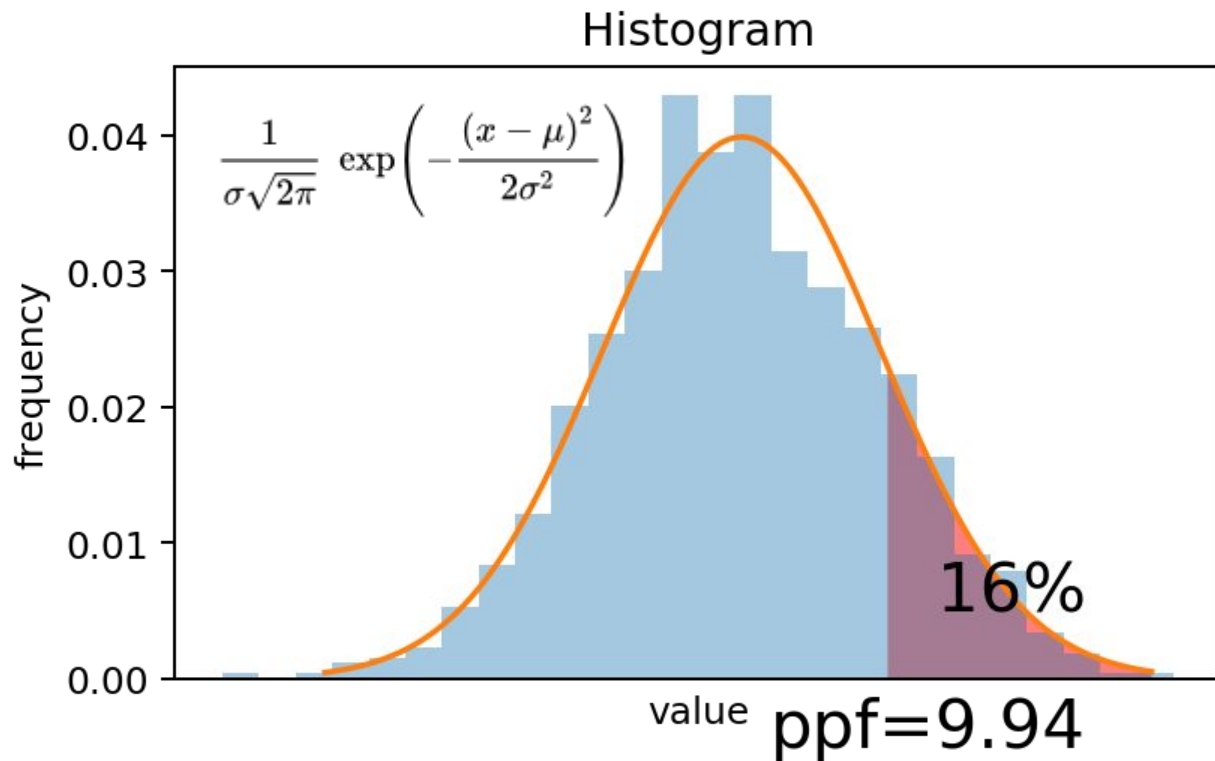




# Процентиль



# Процентиль (percent point function)

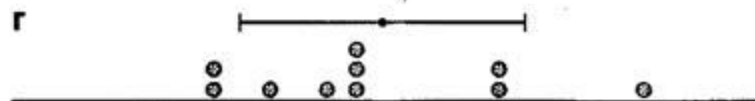
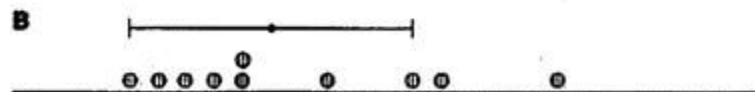
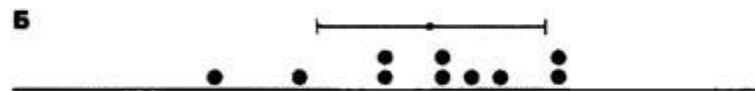
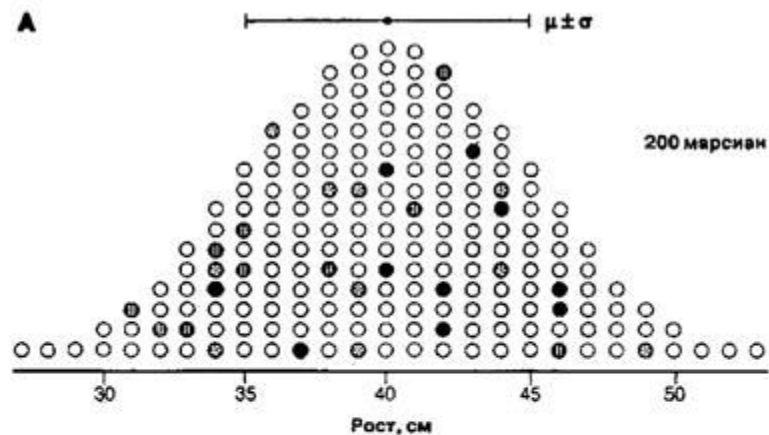


Выборочное среднее и центральная предельная  
теорема

# 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

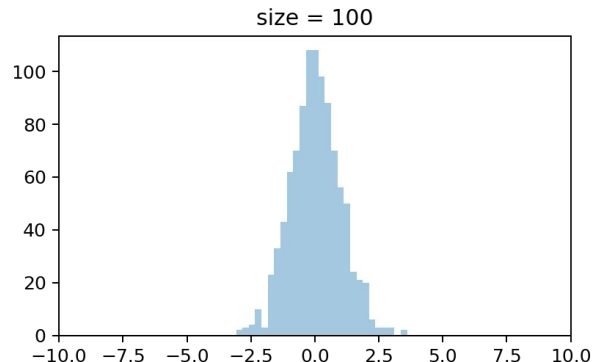
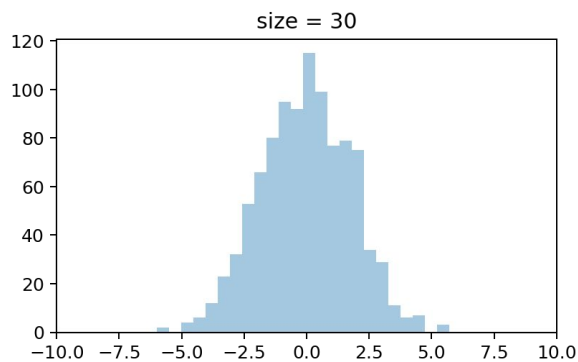
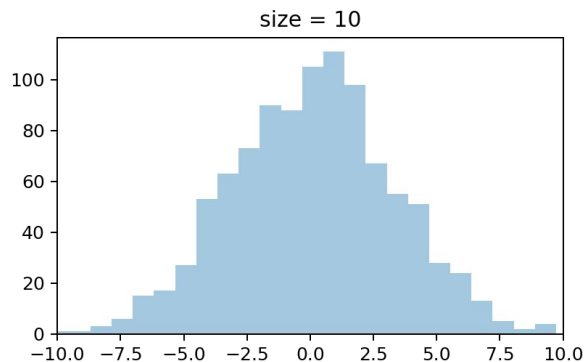
- в процессе эксперимента получаем  $N$  значений
- оцениваем среднее полученных значений
- повторяем эксперимент  $M$ -раз

# 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

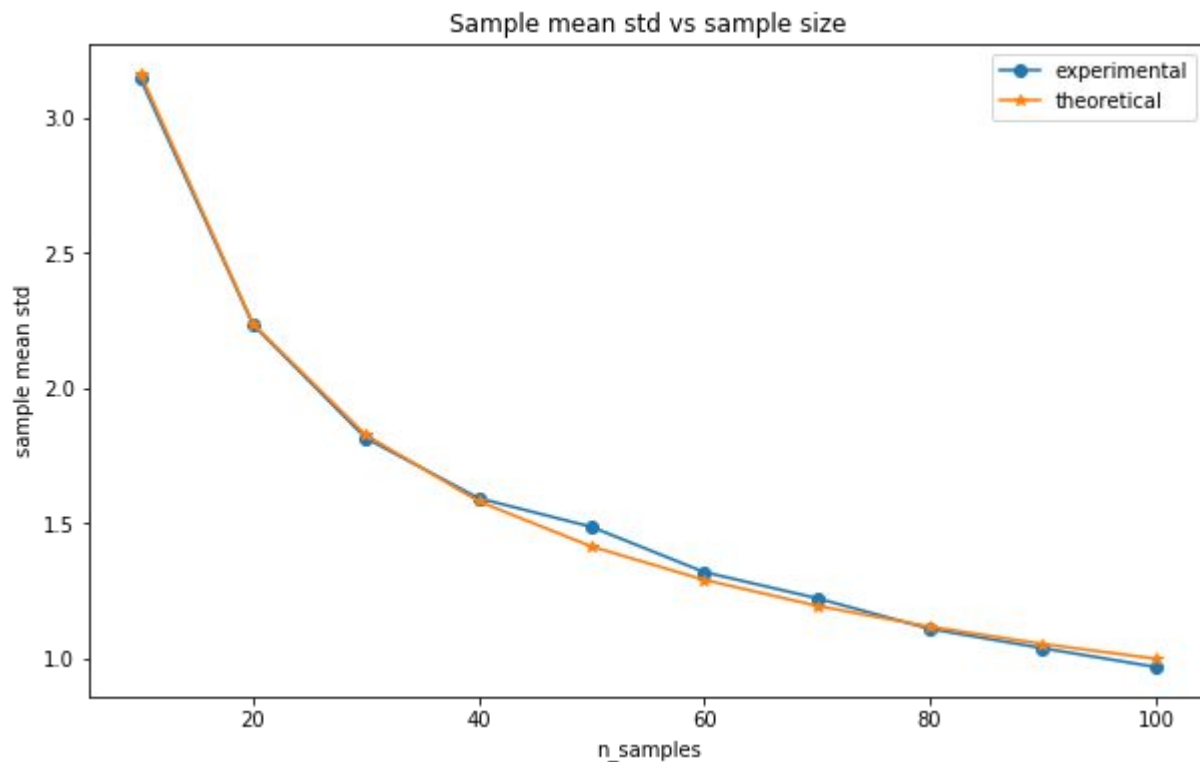
$\text{Var}[X_i] = \sigma^2$  - дисперсия генеральной совокупности

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

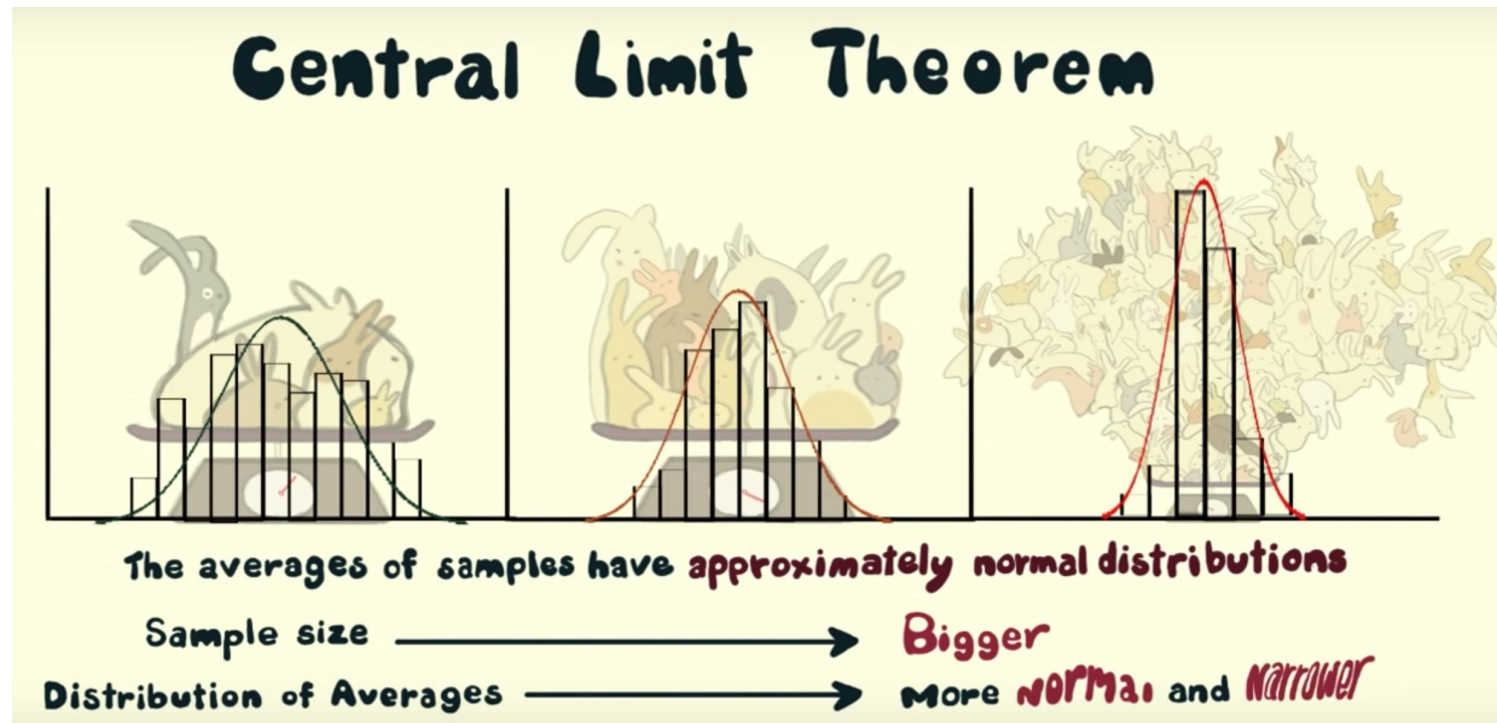
$\sqrt{n} (S_n - \mu) \xrightarrow{d} N(0, \sigma^2)$  - связь распределения средних с нормальным распределением

[https://en.wikipedia.org/wiki/Central\\_limit\\_theorem](https://en.wikipedia.org/wiki/Central_limit_theorem)

# ЦПТ для выборочного среднего



# ЦПТ для выборочного среднего



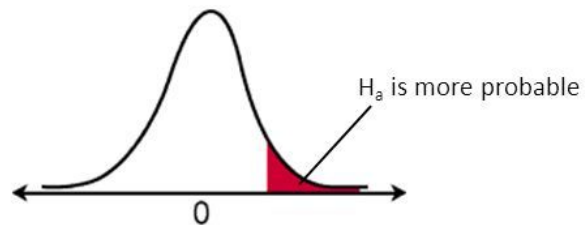
<https://www.youtube.com/watch?v=jvoxEYmQHNM>

# Проверка статистических гипотез

# Проверка статистических гипотез

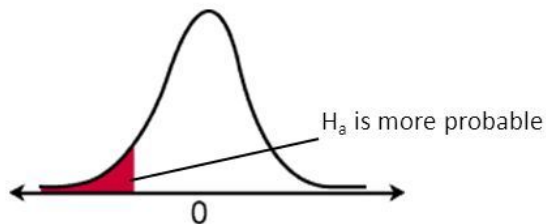
- вычисляется разность средних в сплитах
- эта разность сравнивается с нулем
- нулевая гипотеза: “Разницы в средних нет”
- альтернативная гипотеза: “Разность средних отлична от нуля”

# Проверка статистических гипотез



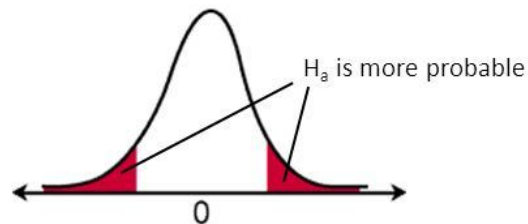
Right-tail test

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



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%



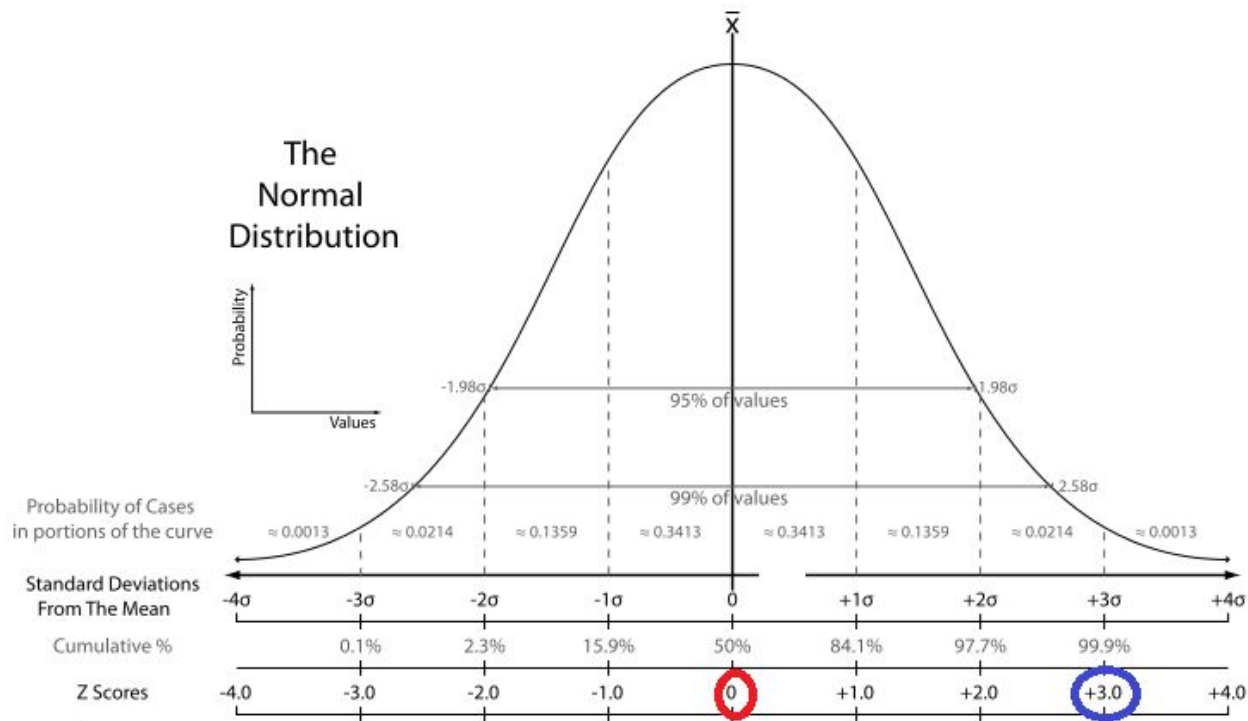
# Проверка статистических гипотез

- 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

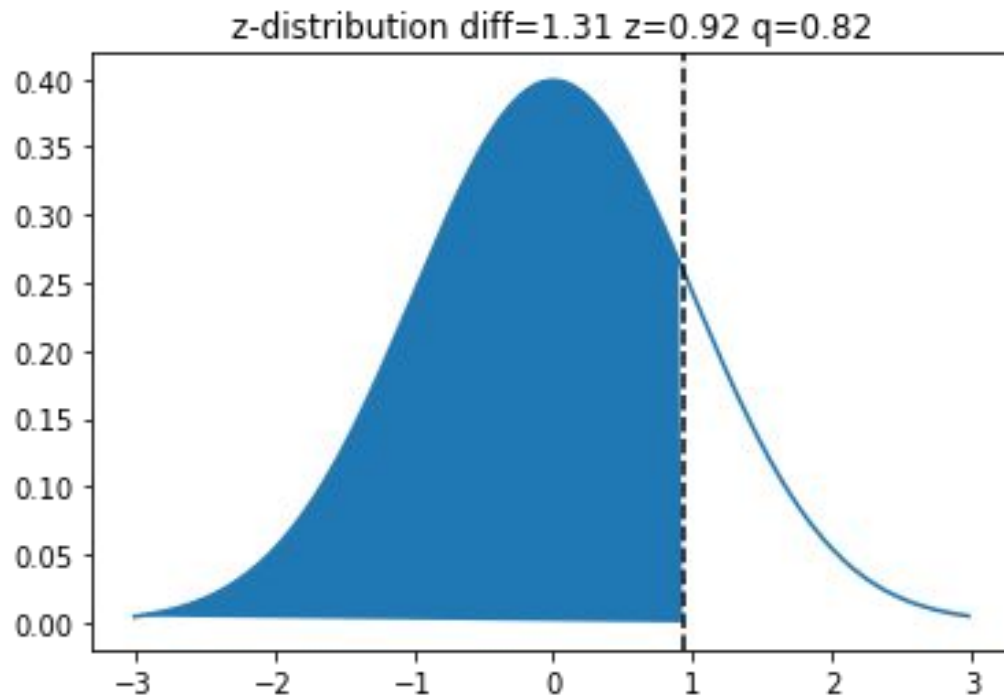
$$SE_{AB} = \sqrt{SE_A^2 + SE_B^2}$$

- формула для стандартной ошибки разности

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

- связь стандартной ошибки и дисперсии

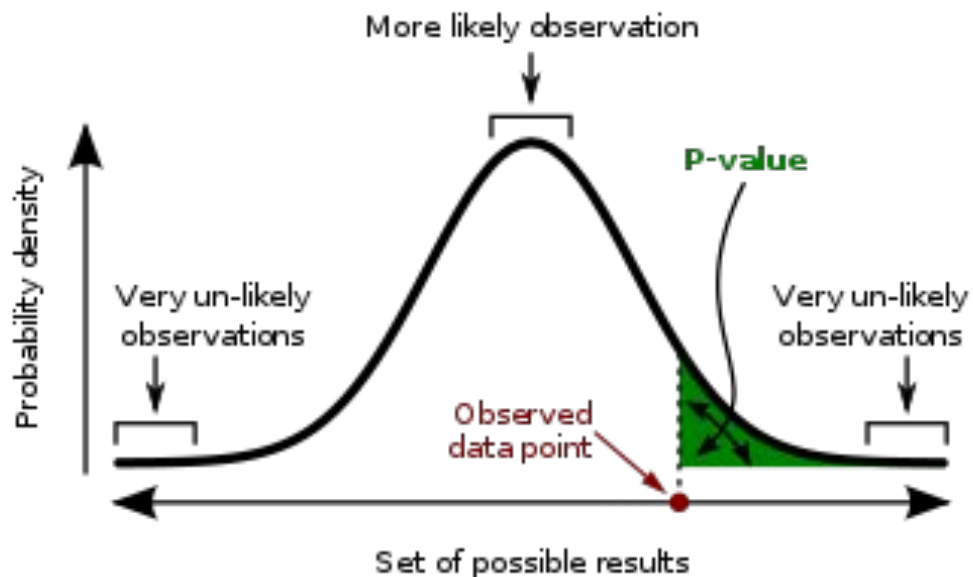
# Z-test



# P-value

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

# P-value



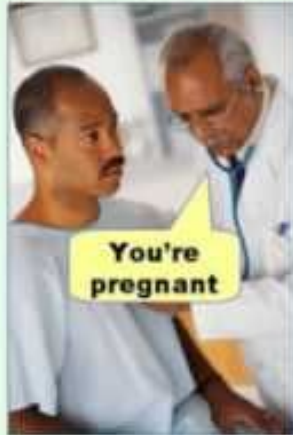
A **p-value** (shaded green area) is the probability of an observed (or more extreme) result assuming that the null hypothesis is true.

# Ошибки первого и второго рода

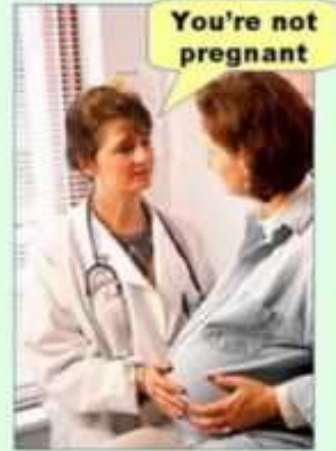
$H_0$  : You are not pregnant

$H_A$ : You are pregnant

**Type I error**  
(false positive)



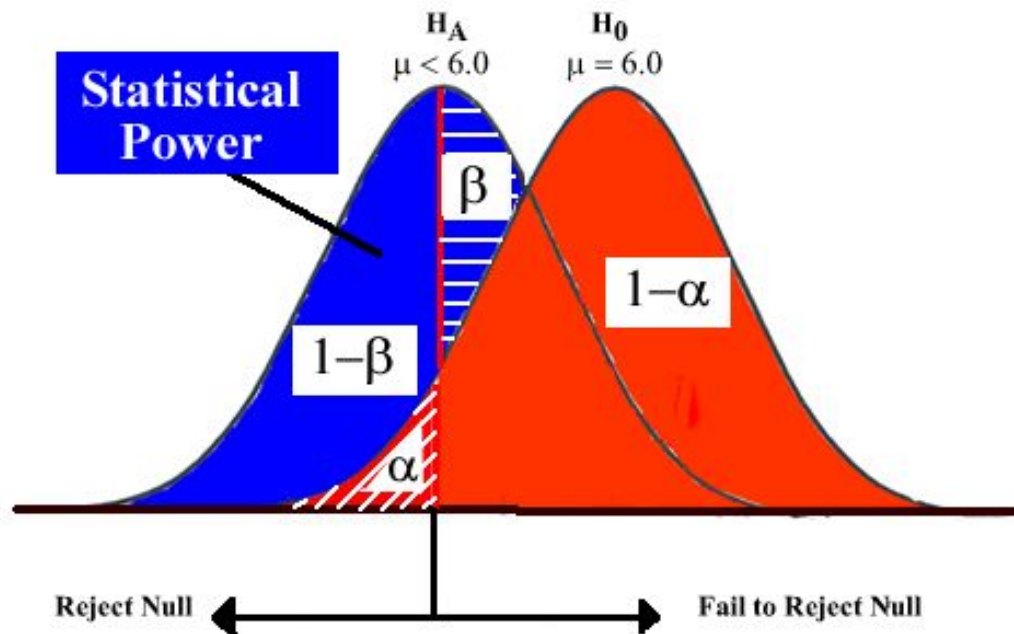
**Type II error**  
(false negative)





# Мощность критерия

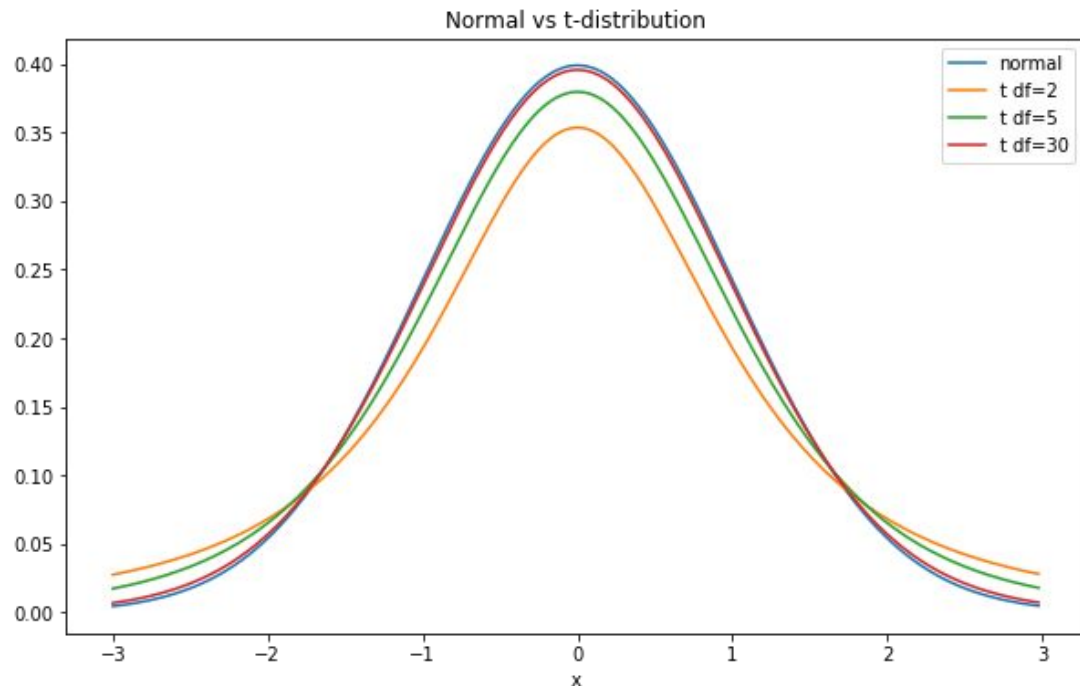
$$\text{power} = \Pr(\text{reject } H_0 \mid H_1 \text{ is true})$$



# Распределение Стьюдента (t-распределение)

- альтернатива нормальному распределению
- используется в случае небольшого числа наблюдений
- дисперсия генеральной совокупности неизвестна, оценивается по семплу

# Распределение Стьюдента vs Нормальное



$$f(t) = \frac{\Gamma(\frac{\nu+1}{2})}{\sqrt{\nu\pi} \Gamma(\frac{\nu}{2})} \left(1 + \frac{t^2}{\nu}\right)^{-\frac{\nu+1}{2}}$$

$$\Gamma(n) = (n-1)!$$

[https://en.wikipedia.org/wiki/Student%27s\\_t-distribution#Probability\\_density\\_function](https://en.wikipedia.org/wiki/Student%27s_t-distribution#Probability_density_function)

# Тест Стьюдента

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_{\bar{\Delta}}}$$

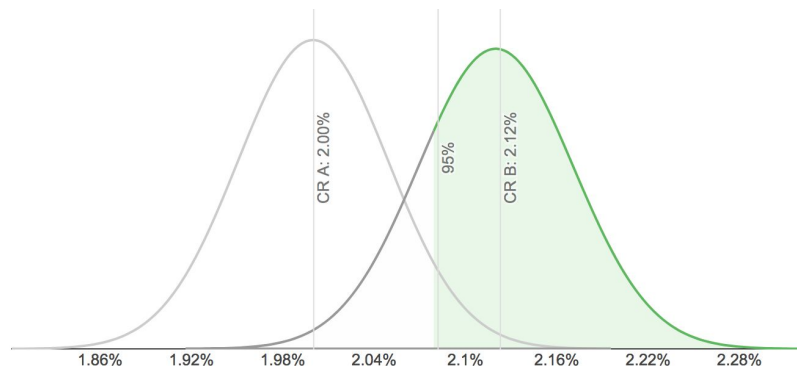
$$s_{\bar{\Delta}} = \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

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

$$\text{d. f.} = \frac{\left( \frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

- t-статистика, используется для оценки уровня значимости
- стандартная ошибка разности
- оценка дисперсии по выборке
- число степеней свободы

# 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

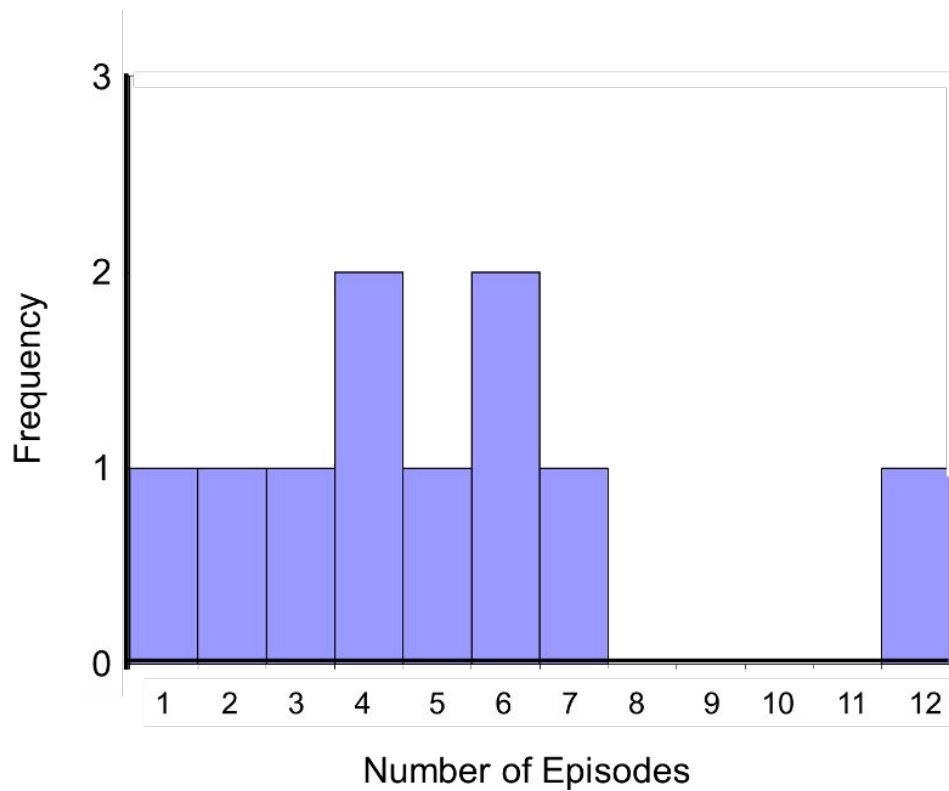
# Тест Манна-Уитни

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

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

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

# Тест Манна-Уитни





# Тест Манна-Уитни

		Total Sample (Ordered Smallest to Largest)		Ranks	
Usual Care	New Program	Usual Care	New Program	Usual Care	New Program
8	9	2		1	
7	8	3		2	
6	7	5		3	
2	8	6	6	4.5	4.5
5	10	7	7	7	7
8	9	7		7	
7	6	8	8	10.5	10.5
3		8	8	10.5	10.5
			9		13.5
			9		13.5
			10		15
				$R_1=45.5$	$R_2=74.5$

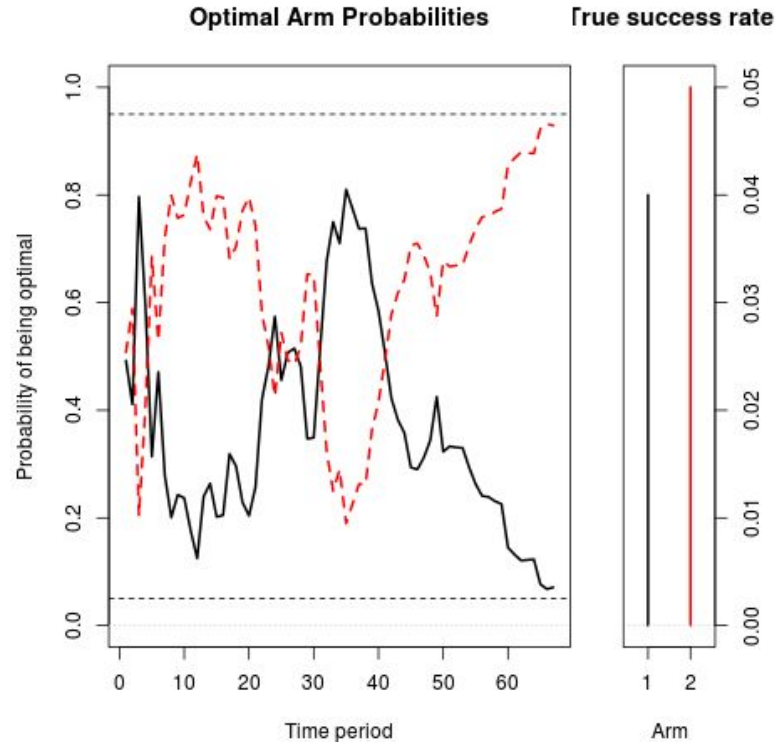
# Уровень значимости оценивается по таблице

**Critical Values of the Mann-Whitney U**  
(One-Tailed Testing)

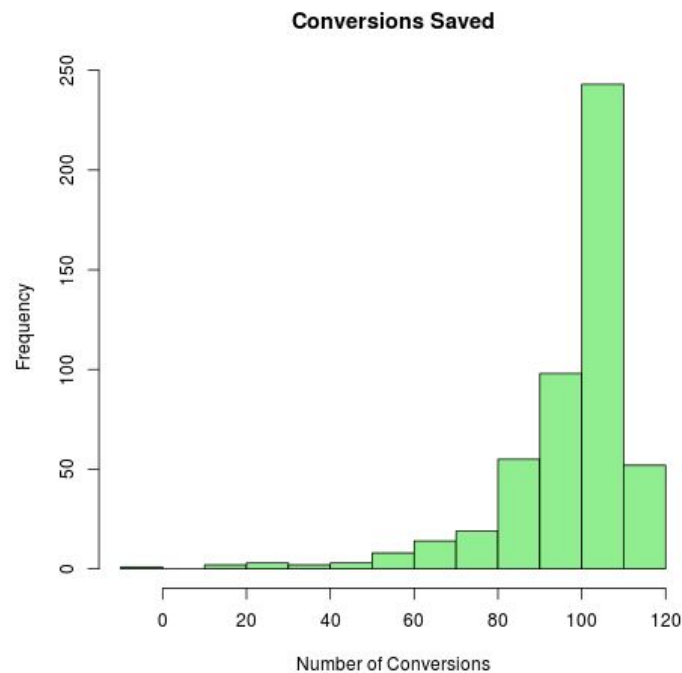
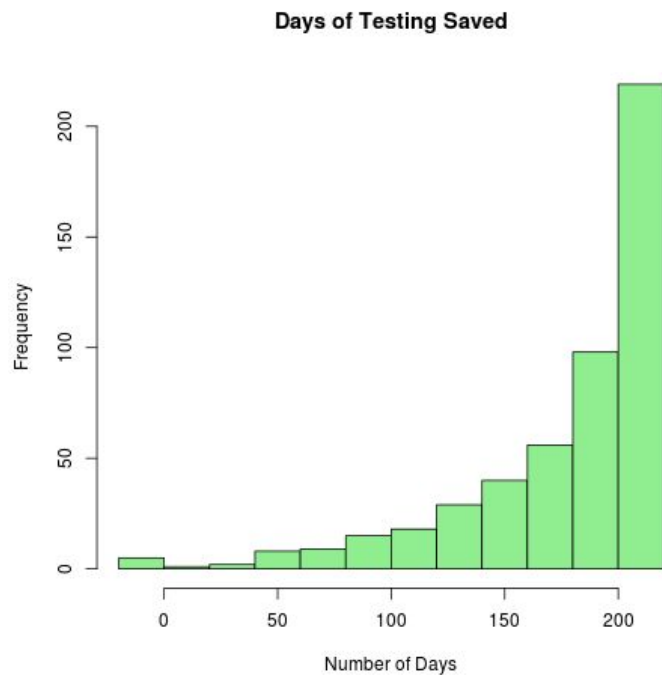
n <sub>2</sub>	$\alpha$	n <sub>1</sub>																	
		3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
3	.05	0	0	1	2	2	3	4	4	5	5	6	7	7	8	9	9	10	11
	.01	--	0	0	0	0	0	1	1	1	2	2	2	3	3	4	4	4	5
4	.05	0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
	.01	--	--	0	1	1	2	3	3	4	5	5	6	7	7	8	9	9	10
5	.05	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
	.01	--	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
6	.05	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
	.01	--	1	2	3	4	6	7	8	9	11	12	13	15	16	18	19	20	22
7	.05	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
	.01	0	1	3	4	6	7	9	11	12	14	16	17	19	21	23	24	26	28
8	.05	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
	.01	0	2	4	6	7	9	11	13	15	17	20	22	24	26	28	30	32	34

# Multi-armed bandit experiments

# Multi-armed bandit experiments



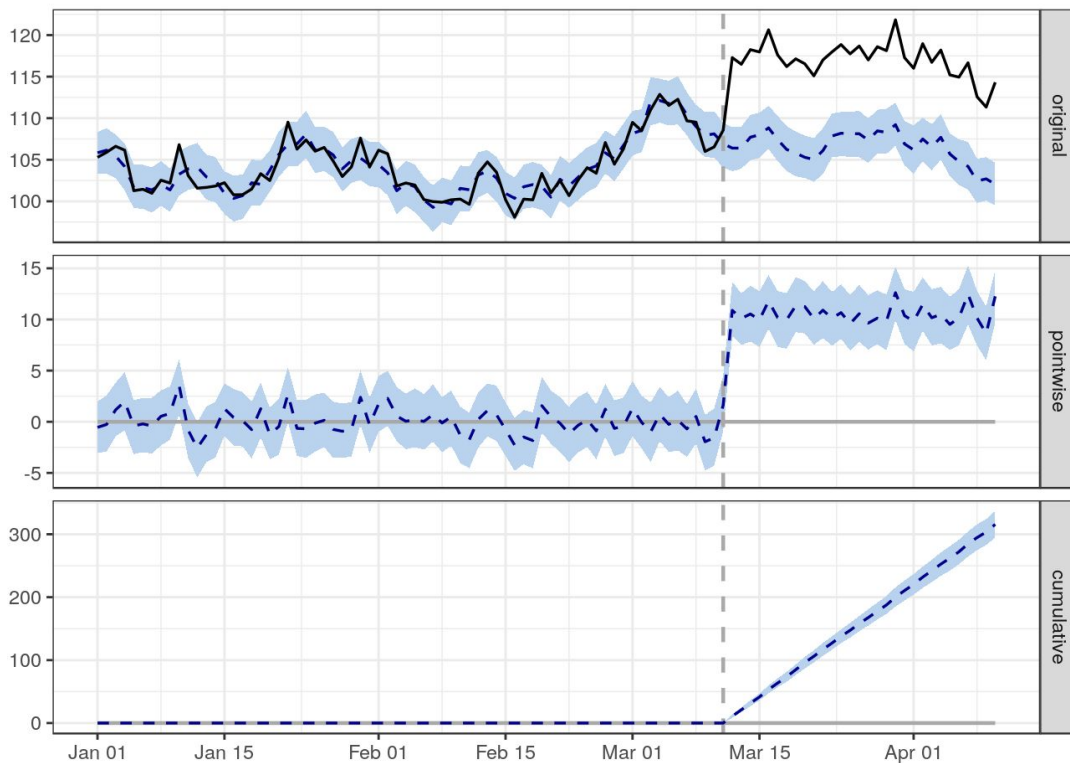
# Multi-armed bandit experiments



<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](#)
- [Nonparametric Tests](#)