

## Урок 5. Применение математической статистики для проверки гипотез в реальной жизни для популярных метрик

### Задание 1.

Вы провели эксперимент с упрощением формы заказа в магазине Утконос и получили результаты по метрике конверсий в покупку. Выберите метод оценки и оцените есть ли стат.значимые различия между конверсиями в двух группах при  $\alpha = 5\%$ . Дайте краткие рекомендации команде.

Результаты:

- 1) Число юзеров в группах, которые заходили на сайт в период эксперимента:  $n_1 = 15550$  и  $n_2 = 15550$ .
- 2) Число юзеров в группах, которые совершили хотя бы одну покупку за период эксперимента:  $n_1 = 164$  и  $n_2 = 228$
- 3) Конверсии:  $conv_1 = 1.05\%$   $conv_2 = 1.47\%$ .

Гипотезы можно проверять с помощью доверительного интервала или p-value. Мы будем проверять с помощью p-value.

Нулевая гипотеза: между выборками (конверсиями  $conv_1$  и  $conv_2$ ) нет статистически значимой разницы.

Альтернативная гипотеза: между выборками (конверсиями  $conv_1$  и  $conv_2$ ) есть статистически значимая разница.

Ниже приведено решение с помощью двух различных онлайн сервисов, из которых видно, что между конверсиями / выборками есть статистически значимая разница и выборка 2 более удачная.

$p\text{-value} (0.0011) < \alpha (0.05)$

Выбираем альтернативную гипотезу и работаем по результатам второй выборки.



(/)

Is your test result significant? Does it have enough power?

Play with the controls and get a better feel for how a lower confidence level will boost the power or how an increase in test size can make a small CR-difference significant!

Pre-test calculation or post-test evaluation?

- ☐ Pre-test analysis
- ☒ Test evaluation

Test data

Visitors A

15550

Conversions A

164

Visitors B

15550

Conversions B

228

Apply changes

Settings

Hypothesis (?)

- ☐ One-sided
- ☒ Two-sided

Confidence (?)

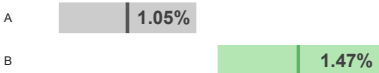
- ☐ 90%
- ☒ 95%
- ☐ 99%

Test result

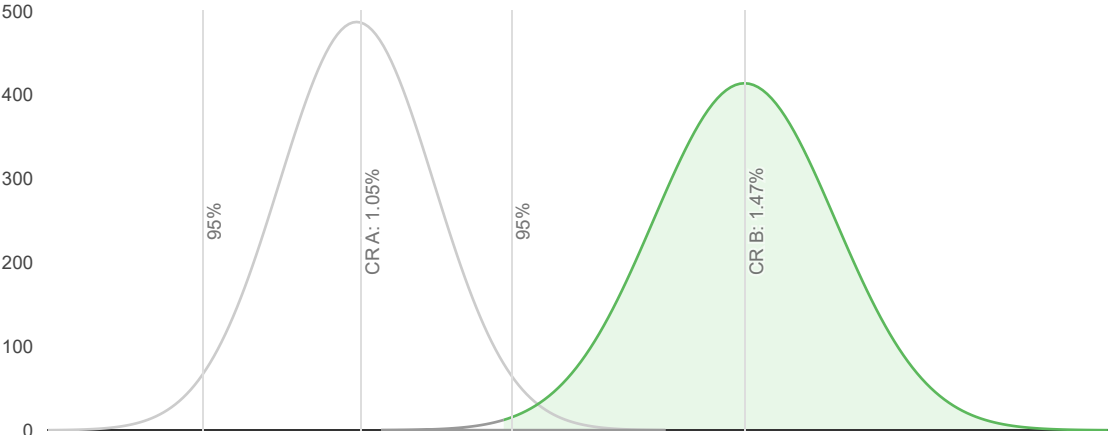
save & share url

Significant test result!

Variation B's observed conversion rate (1.47%) was 39.02% higher than variation A's conversion rate (1.05%). You can be 95% confident that this result is a consequence of the changes you made and not a result of random chance.



The expected distributions of variation A and B.



1.05%

Conversion Rate B

Conversions B / Visitors B

1.47%

Relative uplift in Conversion Rate

$CR_B - CR_A / CR_A$

39.02%

Observed Power

99.54%

p value

0.0011

Z-score

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

3.2536

Standard error A

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

0.000819

Standard error B

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

0.000964

Std. Error of difference

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

0.001265

More tools

A/B Test Size Calculator - How many visitors do you need? (<https://abtestguide.com/abtestsize/>)  
Bayesian A/B-test Calculator - What is the probability that your test variation beats the original? (<https://abtestguide.com/bayesian/>)  
False Discovery Rate Calculator (<https://abtestguide.com/fdr/>)

Recommended reading

AB-testing tech note determining sample-size (<http://37signals.com/svn/posts/3004-ab-testing-tech-note-determining-sample-size>)  
A clear picture of power and significance in AB-tests/ (<http://www.win-vector.com/blog/2014/05/a-clear-picture-of-power-and-significance-in-ab-tests/>)

Evan's Awesome A/B Tools ([home](#)):

[Sample Size Calculator](#) | **Chi-Squared Test** | [Sequential Sampling](#) | [2 Sample T-Test](#) | [Survival Times](#) | [Count Data](#)

*Question:* Does the rate of success differ across two groups?

	# successes	# trials	Confidence interval	[ <a href="#">clear</a> ] [ <a href="#">link</a> ]
Sample 1:	164	15550	0.9% – 1.2%	
Sample 2:	228	15550	1.3% – 1.7%	

*Verdict:*

**Sample 2 is more successful**

(p = 0.00114)

Confidence level:  95%

*If the experiment is repeated many times, the confidence level is the percent of the time each sample's success rate will fall within the reported confidence interval.*

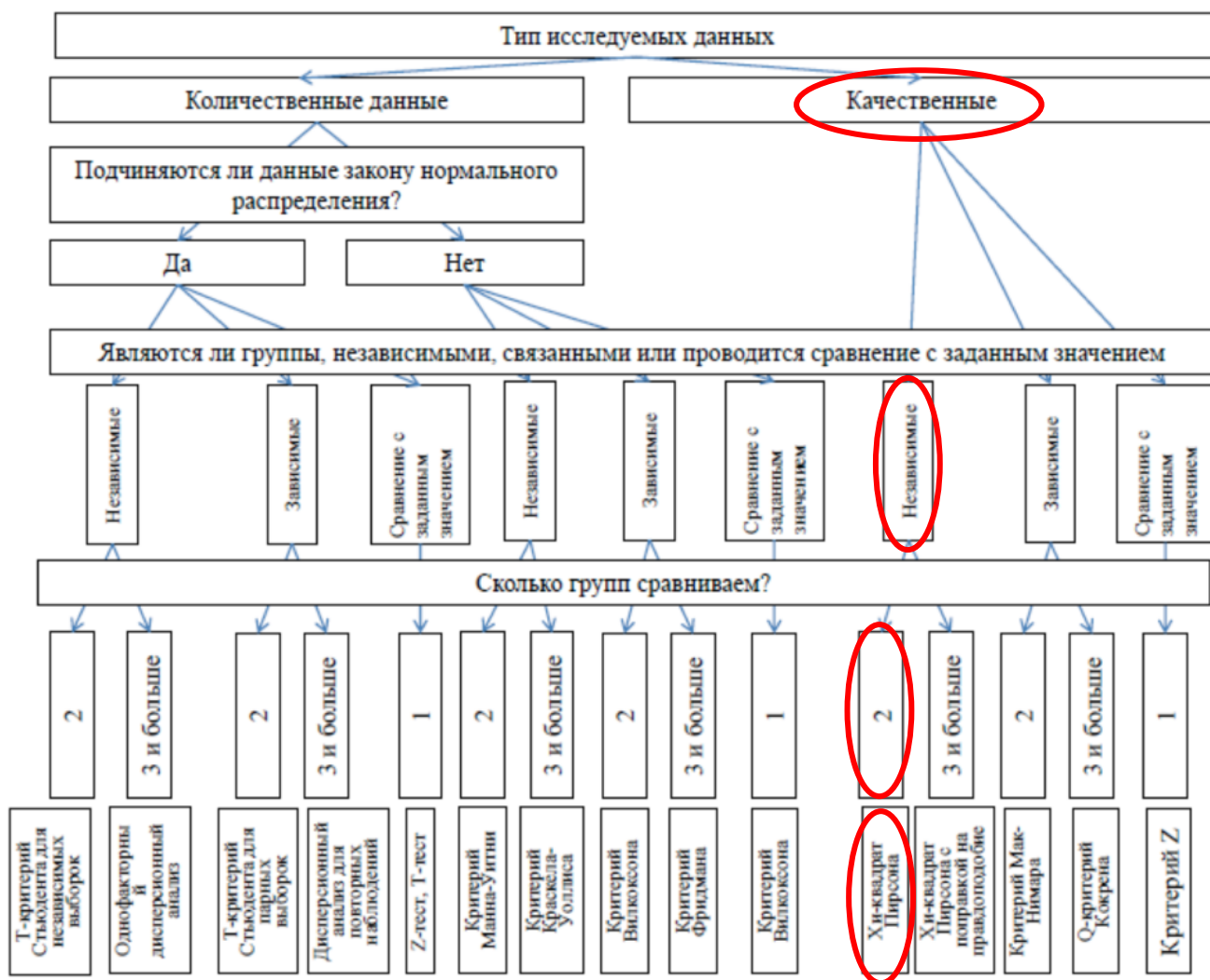
*It is also the percent of the time no difference will be detected between the two groups, assuming no difference exists.*

*If you like this, check out [Wizard](#) — the easy Mac statistics app.*

## Задание 2.

Сравниваем метрику конверсия в покупку. Размер выборки - 10000 элементов в каждой группе. Какой статистический критерий тут лучше всего подойдёт и почему?

Для определения какой статистический критерий больше подойдёт воспользуемся таблицей из лекции.



Таким образом получается, что наилучшим будет Хи-квадрат.

Теперь проверим достаточно ли выборка в 10000 для этого метода. Для этого посчитаем минимальную выборку (расчёт приведён ниже). И она будет составлять 1093, что говорит о достаточном размере выборки.

```
from statsmodels.stats.proportion import proportion_effectsize
from statsmodels.stats.power import TTestIndPower

baseline_cr = 0.2 # базовый уровень конверсии
min_effect = 0.05 # минимальный значимый результат

effect_size = proportion_effectsize(baseline_cr, baseline_cr + min_effect)

alpha = 0.05 # уровень значимости
power = 0.8 # уровень мощности
power_analysis = TTestIndPower()
sample_size = power_analysis.solve_power(effect_size, power=power, alpha=alpha, alternative='two-sided')

print(f"Необходимый размер выборки: {sample_size:.0f}")
```

Необходимый размер выборки: 1093