MTAT.03.319

Business Data Analytics

Lecture 7

Which ad is better?

First : ?
Second: ?
Third: ?

Which ad is better?

https://www.nytimes.com/2017/11/02/magazine/how-facebooks-oracular-algorithm-determines-the-fates-of-start-ups.html

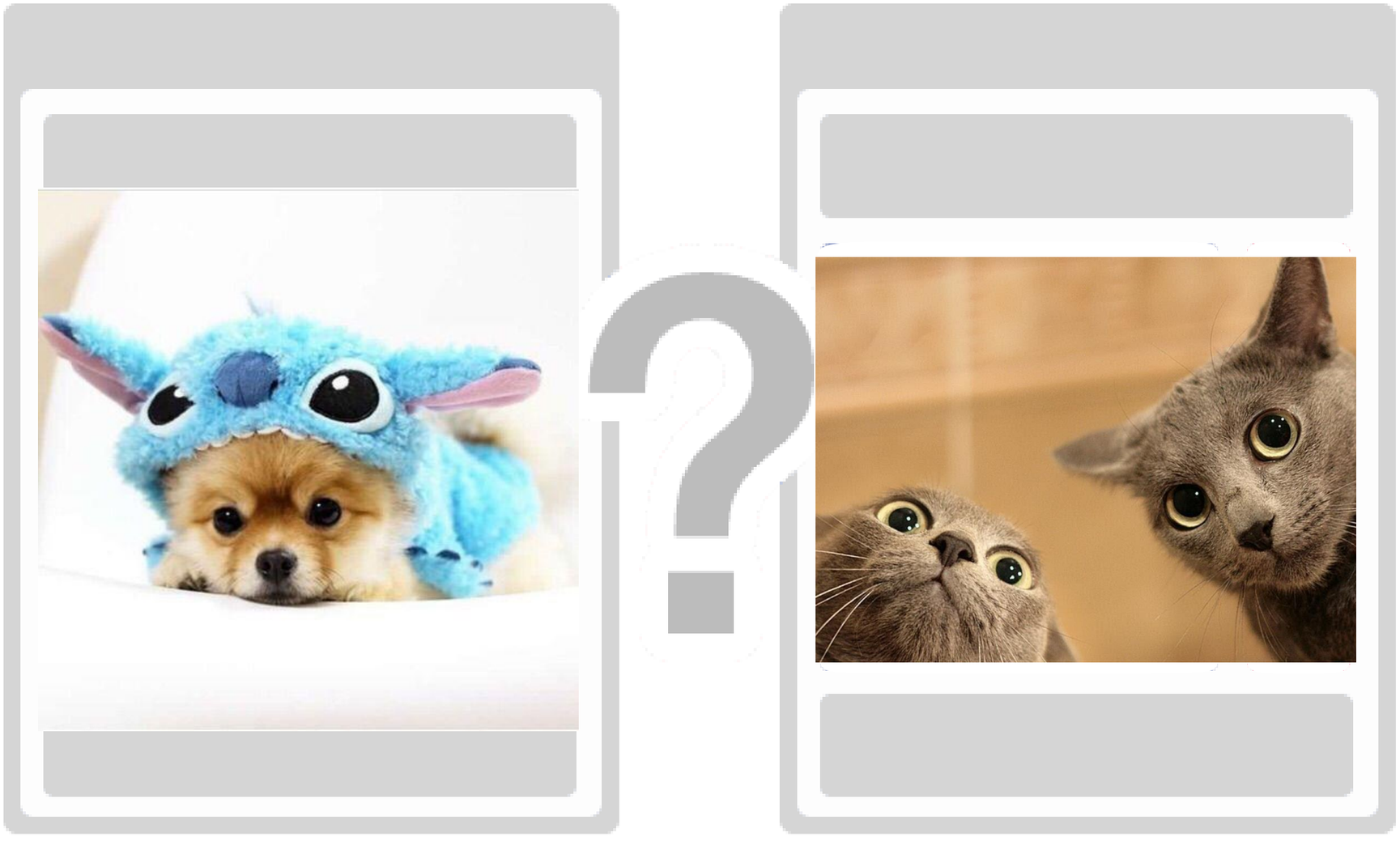
Marketing and Sales

A/B Testing

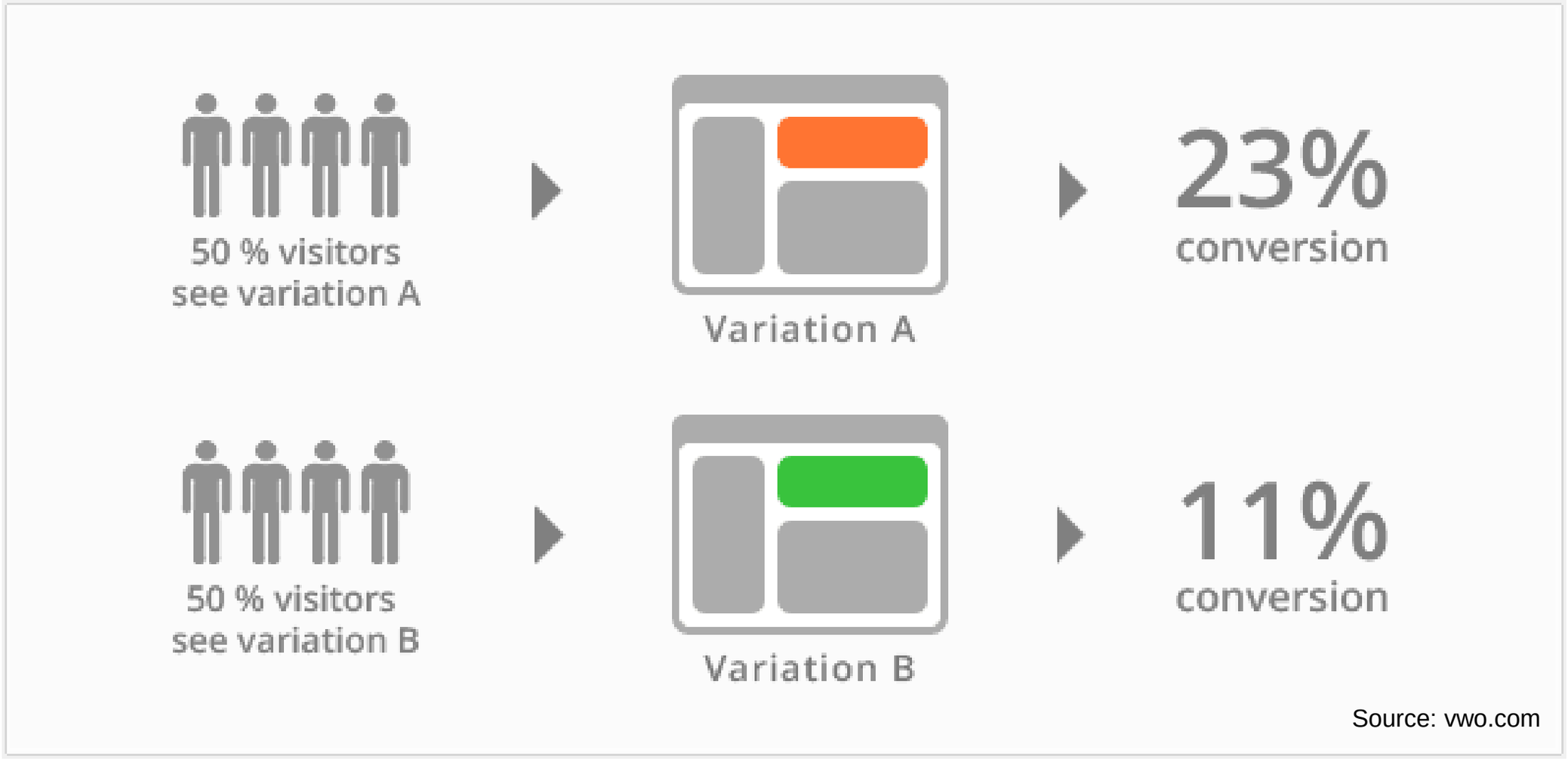
A/B Testing in a nutshell

website variation I

website variation II



A/B Testing in a nutshell



The conversion rate is the percentage of users who take a desired action (eg: bought an item etc.)

A/B Testing in a nutshell

AIM: High Conversion Rate

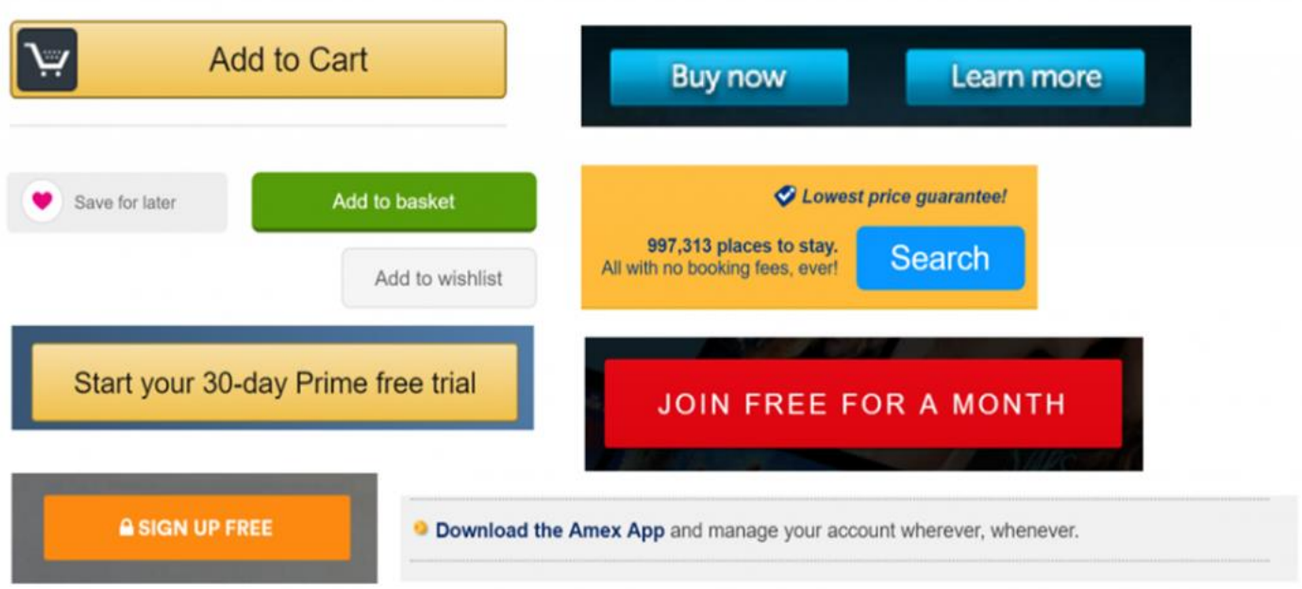
Source: vwo.com

The **conversion rate** is the percentage of users who take a **desired action** (eg: bought an item etc.) through **CTA** (Call to Action)

Conversion

- Call to Action
  - Link to encourage visitors to take action
- Grab the attention of your visitors and really entice them to click.

- Desired Action (Conversion)
  - Sign up for your newsletter,
  - Get a demo of your product
  - Buy a product



Best practices for creating strong calls-to-action

1. Make them Action-Oriented

2. Use Persuasive Text

3. Create a Sense of Urgency

4. Make them Easy to Find

5. Include Strong Visuals

Why we need A/B Testing

even small changes can result in significant increases (or decreases) in leads generated, sales and revenue

THE TEST

We used Google's new split testing tool, Optimize 360, to test the impact of changing our regular buttons on our homepage to ghost buttons. The screenshots below show what was tested.

Original design with a pink button.

Variation with a ghost button.

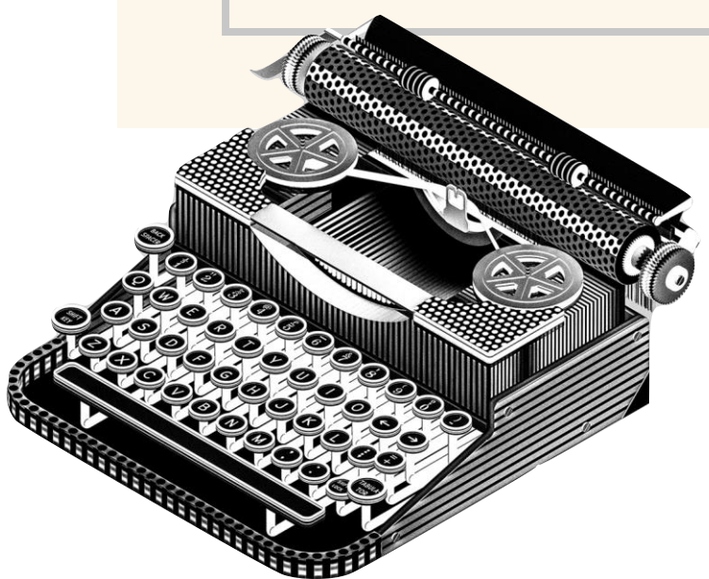
We ran this test over the course of 10,000+ visits to the homepage and measured clicks on any of the CTAs on the homepage.

THE RESULTS

We reached statistical significance for the test, which saw a 20% decrease in clicks on the buttons in the ghost variation.

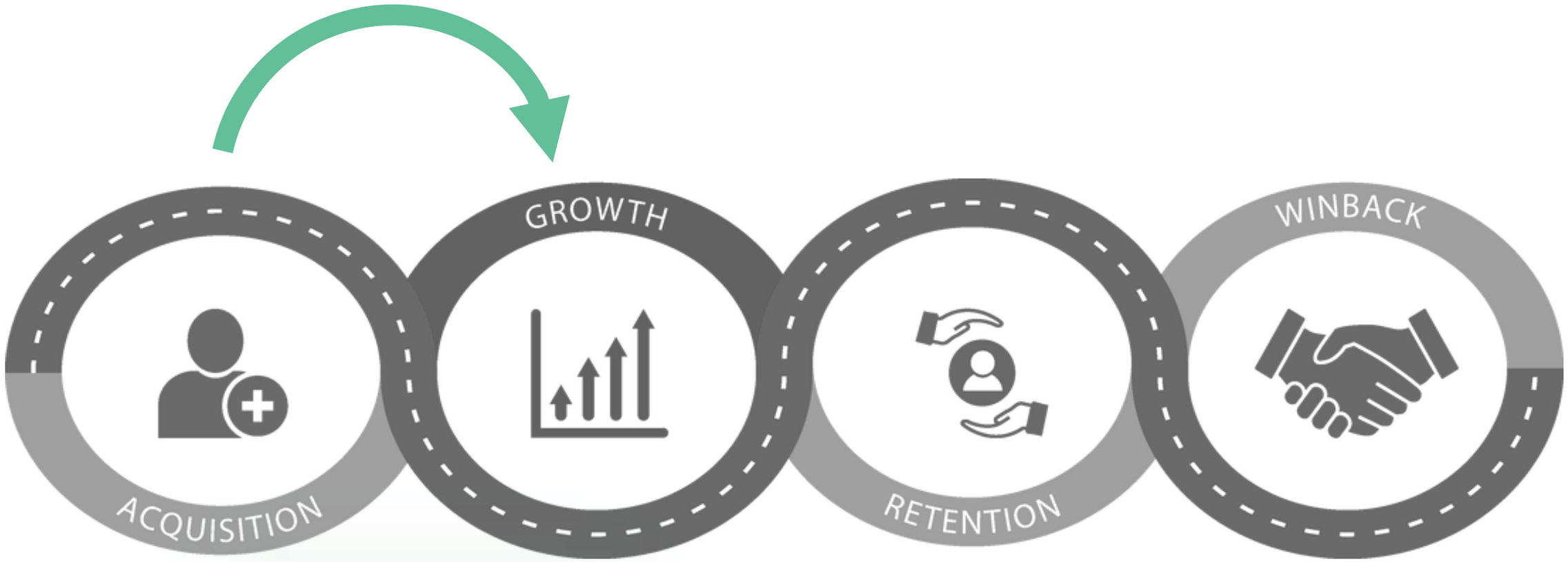
Source: https://conversionxl.com/blog/ghost-buttons/

“A/B testing helps us gain confidence in the change we’re making. It helps us validate new ideas and guides decision making. Without A/B testing, we’re leaving much of what we do up to chance.”



stackoverflow product manager

Customer lifecycle



conversion rate

A/B Testing process

1.Study your data. Identify problem areas or areas with the potential improvement

The homepage has a high bounce rate

2.Use various descriptive tools to form hypotheses

The homepage does not have an action button on the lending page

3.Construct the hypothesis

Placing the button on the landing page will increase conversion rates

A/B Testing process

1. Study your data. Identify problem areas or areas with the potential improvement

2.Use various descriptive tools/note down possible reasons

3. Construct the hypothesis

4. Construct A/B test: use control and focus group. Calculate sample sizes, current conversion rate and the power of test.

5. Perform the test and draw conclusions. If the results inconclusive, go back to 3. and rework your hypothesis.

A/B Testing process

1.Study your data. Identify problem areas or areas with the potential improvement

Example: The homepage has a high bounce rate

2.Use various descriptive tools/note down possible reasons

H1: The homepage does not have an action button on the lending page

H2: The homepage colors are mild

H3: …….:

H4: …….:

A/B Testing process

1.Study your data. Identify problem areas or areas with the potential improvement

The homepage has a high bounce rate

2.Use various descriptive tools/note down possible reasons

The homepage does not have an action button on the lending page

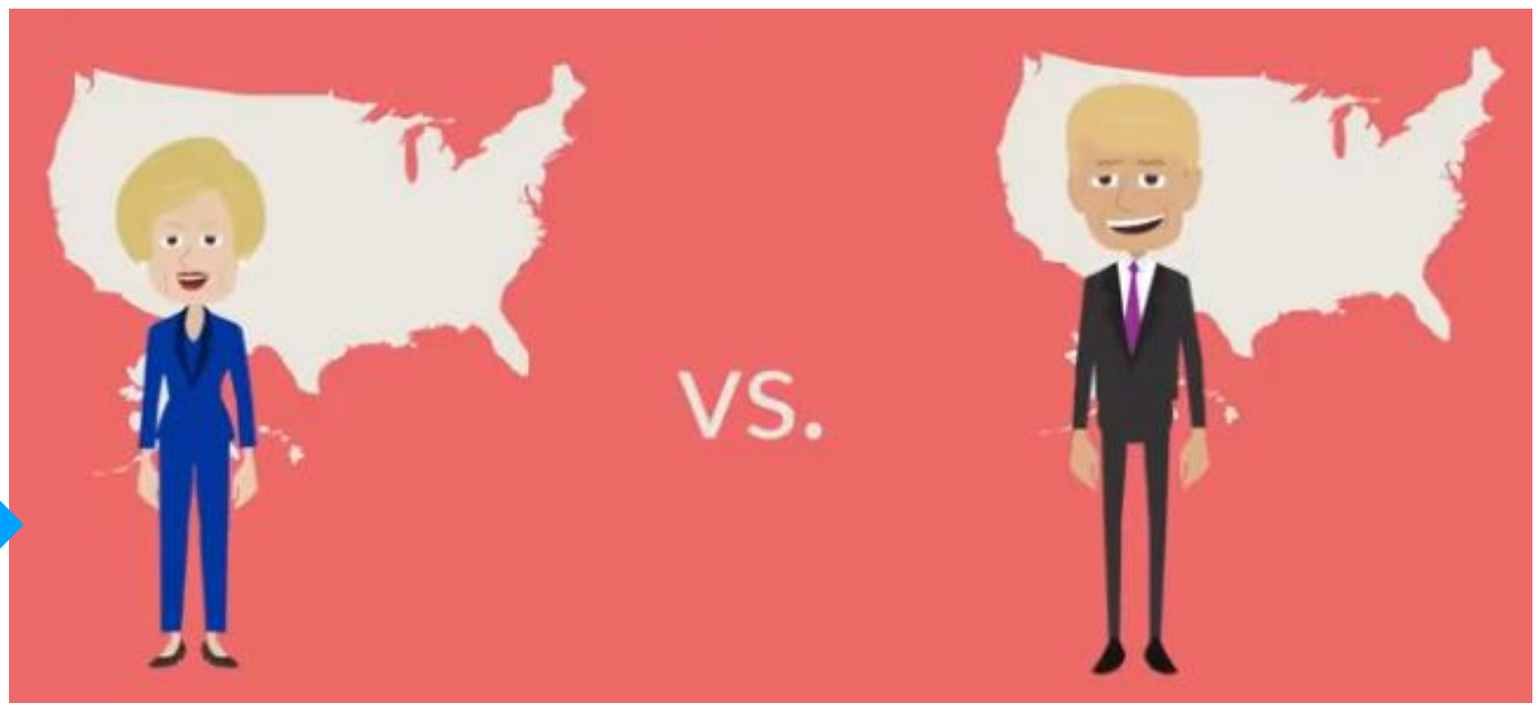
NOTE: Can you construct Hypothesis for anything ?

- If you have no possibility of acquiring data, then it is just an idea.
- If you have the data or if it can be collected then it is a hypothesis.
- It is important that your hypothesis can be tested using some data.

Hypothesis Terminologies

Hypothesis

Just an idea



A/B Testing process

1.Study your data. Identify problem areas or areas with the potential improvement

The homepage has a high bounce rate

2.Use various descriptive tools/note down possible reasons

The homepage does not have an action button on the lending page

3.Construct the hypothesis

Placing the button on the landing page will increase conversion rates

Terminologies

Hypothesis: Example 1

Hypotheses	Notation
Null Hypothesis	H_0
Alternative Hypothesis	H_A or H_1

Example

Data Scientist Salary in US

Hypotheses	Notation & Example
Null Hypotheses (Also the status quo)	$H_0 : \mu = 113{,}000\$$
Alternative Hypotheses	$H_1 : \mu \neq 113{,}000\$$

True until proven wrong

NOTE: Depending upon the problems, you can construct hypothesis in different ways. Some more examples in coming slides

2 Types of Tests

- 2 Tail Tests

$H_0 : \mu = 113{,}000\$$

$H_1 : \mu \neq 113{,}000\$$

Salary can be more than or less than μ. It can go in either of the directions

- 1 Tail Tests

$H_0 : \mu > 113{,}000\$$

$H_1 : \mu <= 113{,}000\$$

If Salary is not more than μ, then it will go only in one direction.

A/B Testing process

1.Study your data. Identify problem areas or areas with the potential improvement

The homepage has a high bounce rate

2.Use various descriptive tools to form hypotheses

The homepage does not have an action button on the lending page

3.Construct the hypothesis

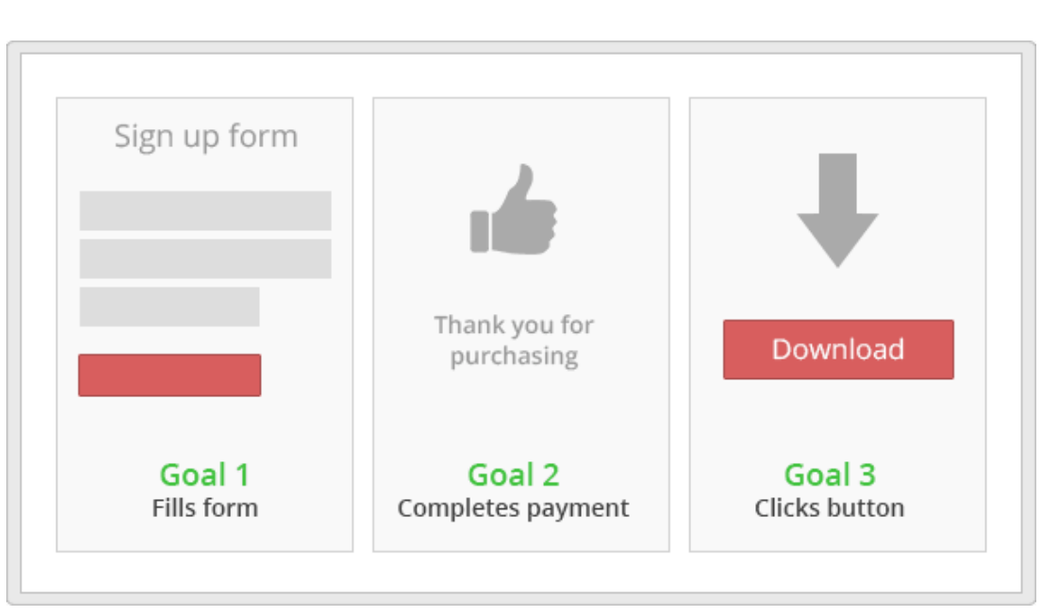
Placing the button on the landing page will increase conversion rates

A/B Testing process

4.Construct A/B test: use control and focus groups. Calculate sample sizes, current conversion rate and the power of the test

Page A: a page with the button, B: original page

define the goal of conversion



Source: vwo.com

Terminologies

A **focus/experimental/treatment group** is a small, but demographically diverse group of people and whose reactions are studied especially in market research or political analysis in guided or open discussions about a new product or something else to determine the reactions that can be expected from a larger population (receive the experimental treatment)

Source: https://en.wikipedia.org/wiki/Focus_group

The **control group** is composed of participants who do not receive the experimental treatment. When conducting an experiment, these people are randomly selected to be in this group. They also closely resemble the participants who are in the experimental group or the individuals who receive the treatment.

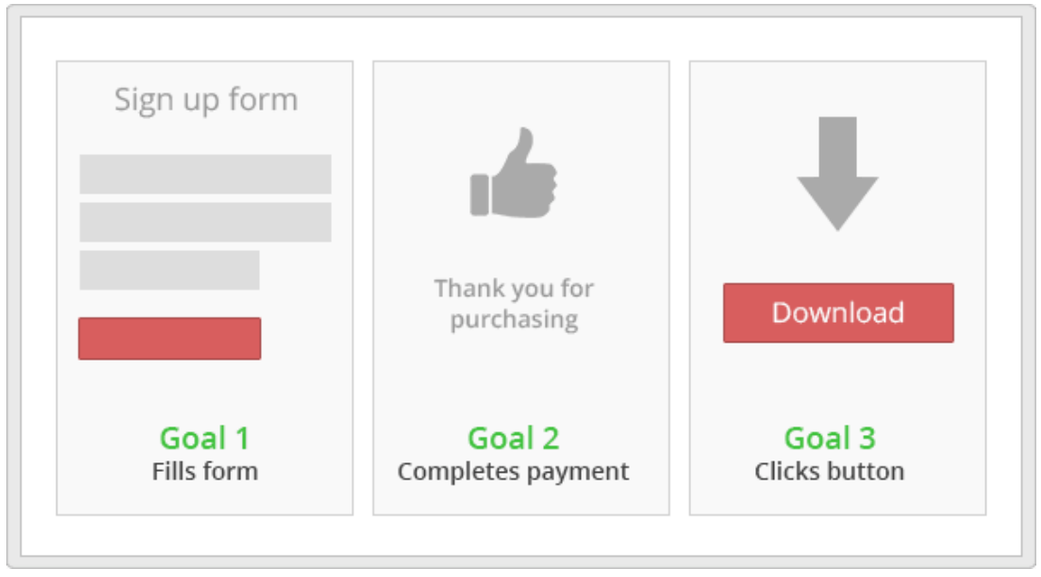
Source: https://www.verywellmind.com/what-is-the-control-group-2794977

A/B Testing process

4.Construct A/B test: use control and focus groups. Calculate sample sizes, current conversion rate and the power of the test

A: a page with the button, B: original page

define the goal of conversion



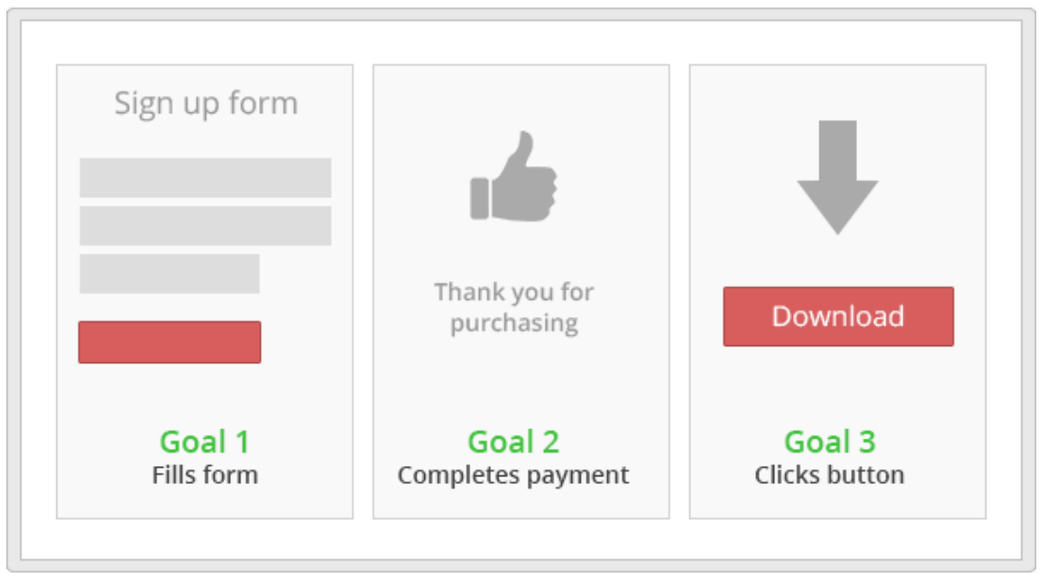
Source: vwo.com

A/B Testing process

4.Construct A/B test: use control and focus groups. Calculate sample sizes, current conversion rate and the power of the test

Page A: a page with the button, B: original page

define the goal of conversion



Source: vwo.com

Terminologies

Definition 1: The **power** of a **hypothesis test** is the probability of rejecting the null **hypothesis** (H_0) when the alternative **hypothesis** (H_A) is true.

Definition 2: The **power** of a **hypothesis test** is the probability of making the correct decision if the alternative **hypothesis** is true.

Example: t-test is one such test which is used for rejecting or accepting a particular type of hypothesis tests (we will see t-tests in coming slides)

using the standard α = 0.05 cutoff, the null hypothesis is rejected when p < .05 and not rejected when p > .05.

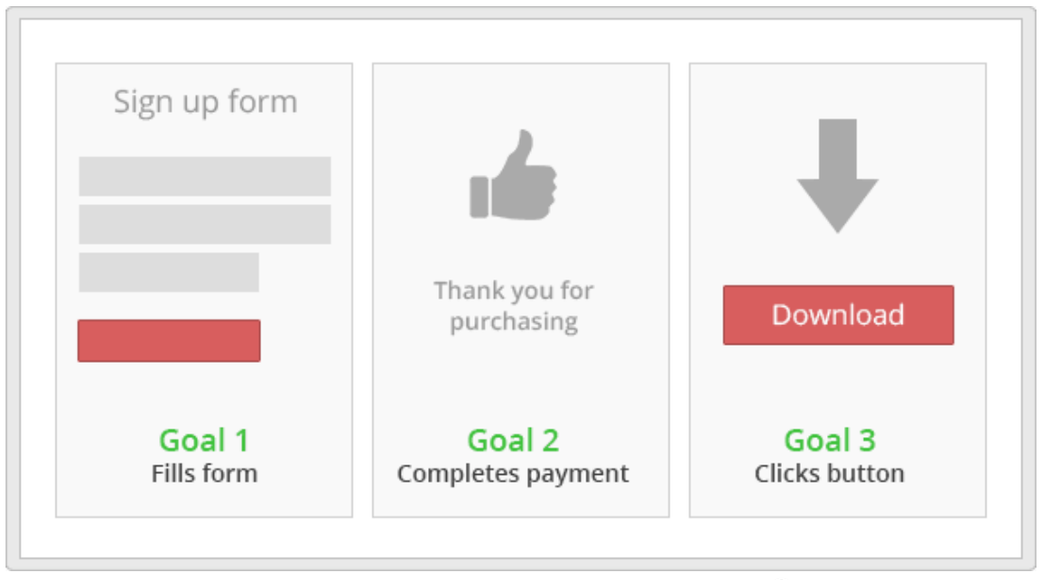
Source: https://www.verywellmind.com/what-is-the-control-group-2794977

A/B Testing process

4.Construct A/B test: use control and focus groups. Calculate sample sizes, current conversion rate and the power of the test

A: a page with the button, B: original page

define the goal of conversion



Source: vwo.com

5. Perform the test and draw conclusions. If the results inconclusive, go back to 3. and rework your hypothesis.

Terminologies

The **power** of a **hypothesis test** is the probability of making the correct decision if the alternative **hypothesis** is true. That is, the **power** of a **hypothesis test** is the probability of rejecting the null **hypothesis** H_0 when the alternative**hypothesis** H_A is the **hypothesis** that is true.

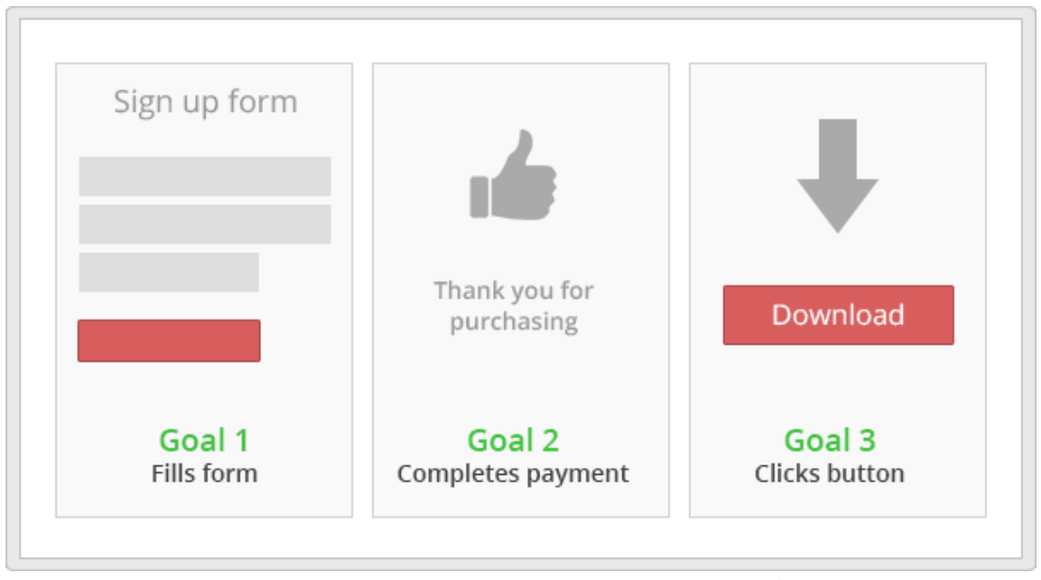
Source: https://www.verywellmind.com/what-is-the-control-group-2794977

A/B Testing process

4.Construct A/B test: use control and focus groups. Calculate sample sizes, current conversion rate and the power of the test

A: a page with the button, B: original page

define the goal of conversion



Source: vwo.com

5. Perform the test and draw conclusions. If the results inconclusive, go back to 3. and rework your hypothesis.

A/B Testing process

1.Study your data. Identify problem areas or areas with the potential improvement

The homepage has a high bounce rate

2.Use various descriptive tools to form hypotheses

The homepage does not have an action button on the lending page

3.Construct the hypothesis

Placing the button on the landing page will increase conversion rates

What is null hypothesis and alternative hypothesis?

Sleeping habits

Hypothesis example 2

4 hours of sleep

8 hours of sleep

exam performance

Hypothesis testing

H0: People who sleep 4 hours at night perform the same at exam than those who sleep 8

H1: 4 hours sleepers perform different at exam than those who sleep 8

Hypothesis testing

- 24 participants
- randomly divided into one of each group
- after, the same exam is taken and results recorded

	group1	group2
1	6.173134	8.082474
2	7.562668	7.519545
3	5.286290	7.645783
4	5.989726	7.525387
5	6.752221	7.788239
6	7.455438	8.572076
7	7.132399	8.028798
8	6.275678	8.225362
9	6.328979	7.308410
10	6.606508	8.116942
11	7.040576	8.204206
12	7.433520	7.307770

Hypothesis testing

group1

H0: People who sleep 4 hours at night perform the same at exam than those who sleep 8

group2

H1: 4 hours sleepers perform different at exam than those who sleep 8

H0: $\text{mean}(\text{score}_{\text{group1}}) = \text{mean}(\text{score}_{\text{group2}})$

H1: $\text{mean}(\text{score}_{\text{group1}}) \neq \text{mean}(\text{score}_{\text{group2}})$

Hypothesis testing

Imagine you want to see the difference between two versions of the website - the current version and the one with the different color scheme. Construct the hypothesis to test this.

What is null hypothesis and alternative hypothesis ?

Hypothesis testing

group1

H0: People who sleep 4 hours at night perform the same at exam than those who sleep 8

group2

H1: 4 hours sleepers perform different at exam than those who sleep 8

H0: $\text{mean}(\text{score}_{\text{group1}}) = \text{mean}(\text{score}_{\text{group2}})$

H1: $\text{mean}(\text{score}_{\text{group1}}) \neq \text{mean}(\text{score}_{\text{group2}})$

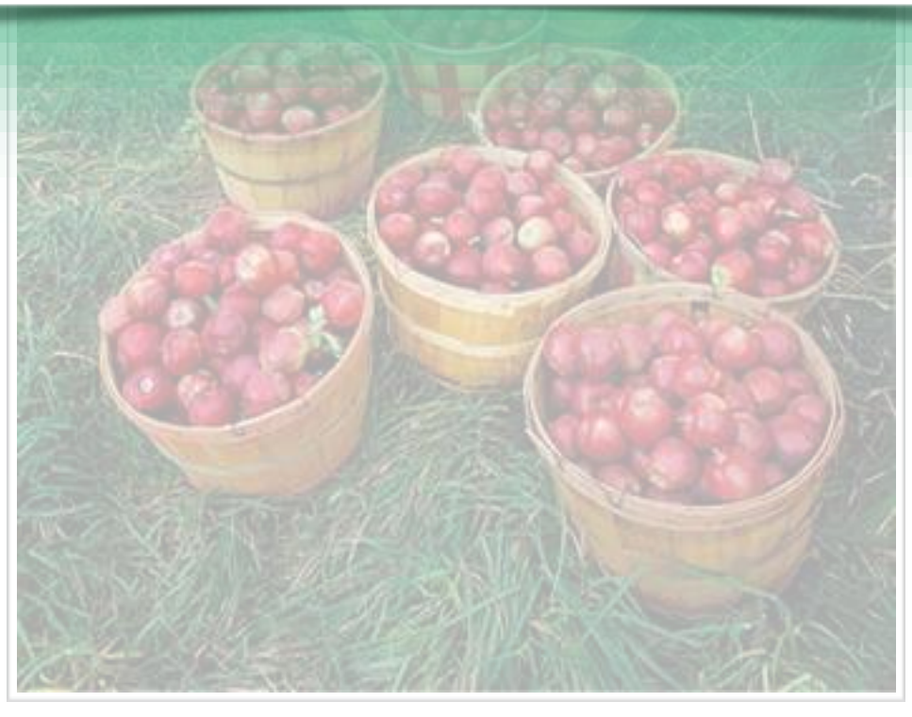
Hypothesis testing

Collect the data

Hypothesis testing

Collect the data

When enough is enough?



A/B Testing process

4. Construct **A/B test**: use control and focus groups. Calculate sample sizes, current conversion rate and the power of the test

A: a page with the button, B: original page

- how sure we need to be that we are measuring a real change
- how big the change we expect to see because of the new version,

compared to the baseline?

Source: vwo.com

4. a) how sure we need to be that we are measuring a real change?

- What percentage of the time are we willing to miss a real effect?

  (power of the test)

- What percentage of the time are we willing to see an effect by random chance?

  (significance level, or the probability of rejecting the null hypothesis)

how sure we need to be that we are measuring a real change?

- What percentage of the time are we willing to miss a real effect?

(power of the test)

typical value: 80%

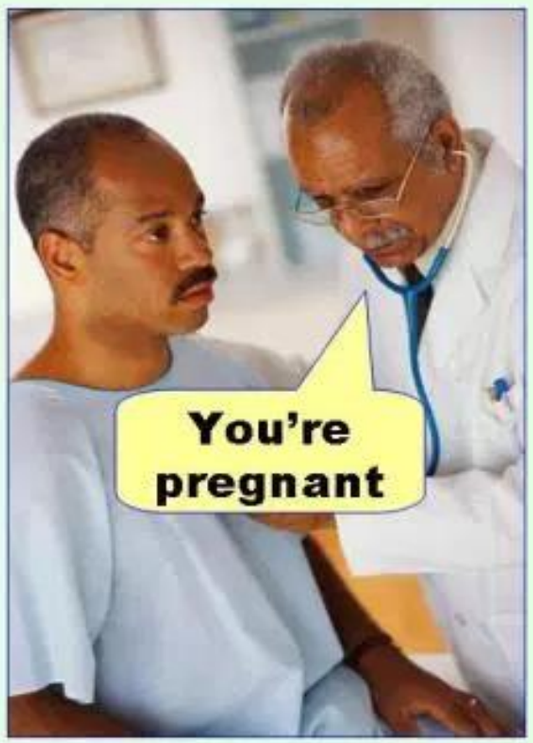
- What percentage of the time are we willing to see an effect by random chance?

(significance level, or the probability of rejecting the null hypothesis)

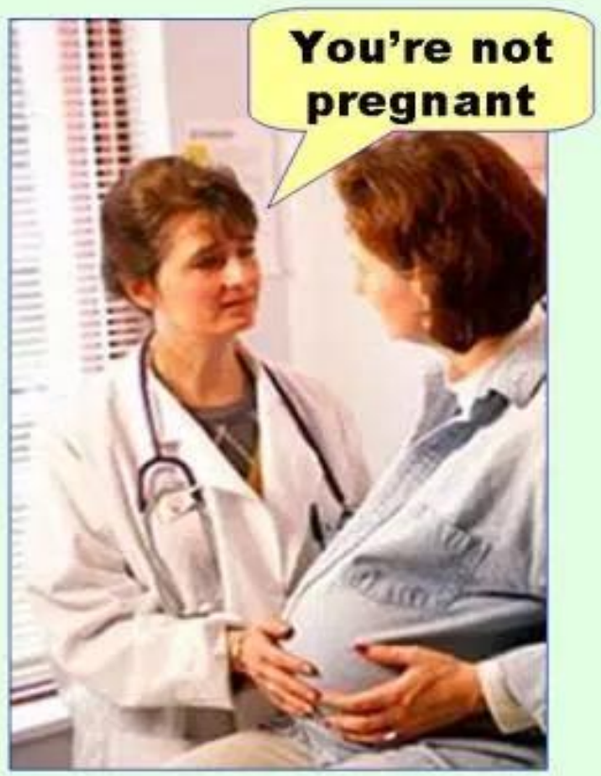
typical value: 5%

What can go wrong ?

Type I error
(false positive)



Type II error
(false negative)

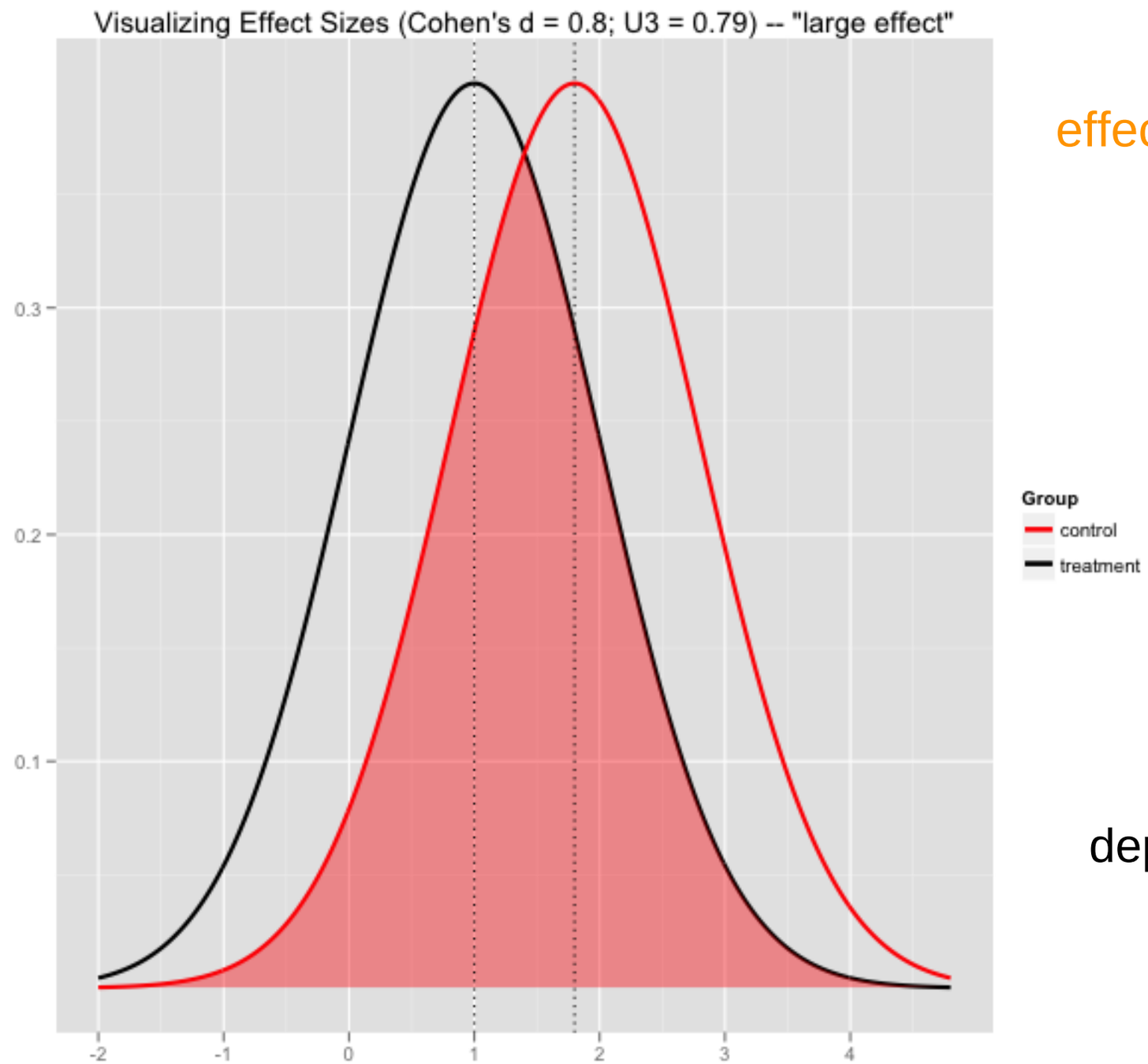


Rejecting the null hypothesis (H_0) even though it is true

Failure to reject a false null hypothesis

	H_0 is actually	
	True	False
Reject H_0	Type I Error	Correct
Accept H_0	Correct	Type II Error

4.b) how big is the change we expect to see?



effect size - expected difference between groups

depends on the business context

A/B Testing process

4. Construct **A/B test**: use control and focus groups. Calculate sample sizes, current conversion rate and the power of the test

A: a page with the button, B: original page

- how sure we need to be that we are measuring a real change
- how big the change we expect to see because of the new version,

compared to the baseline?

Source: vwo.com

5. Perform the test and draw conclusions. If the results inconclusive, go back to 3. and rework your hypothesis.

Statistical hypothesis testing is a general framework of testing claims.

1. state the relevant null and alternative hypotheses ☑

2. decide which test is appropriate for this type of hypothesis

1) Independent samples
(unpaired-samples t-test or Between samples)

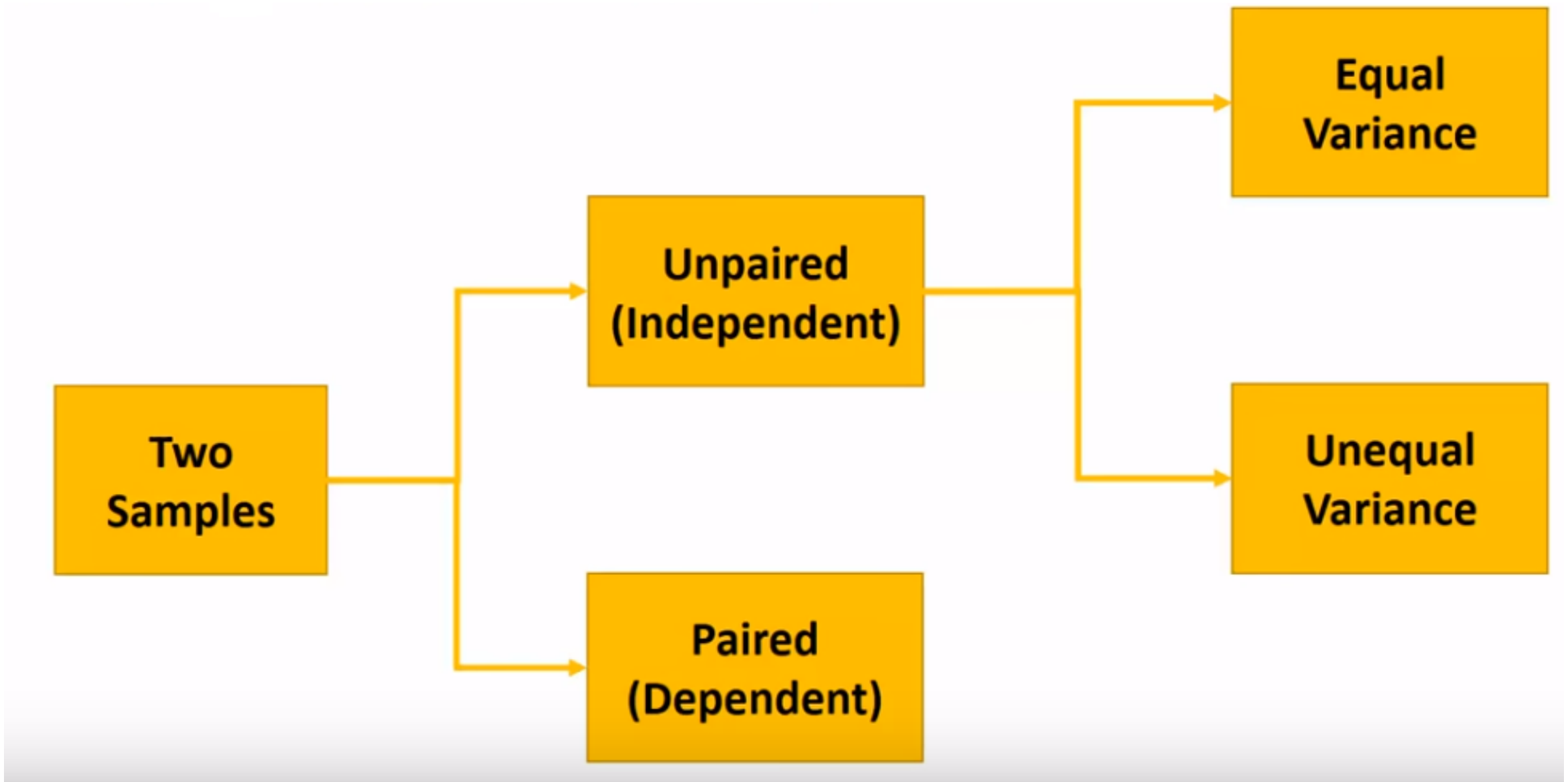
2) Paired samples
(within-subjects , repeated measures, dependent-samples)

3) One-sample

Statistical hypothesis testing is a general framework of testing claims.

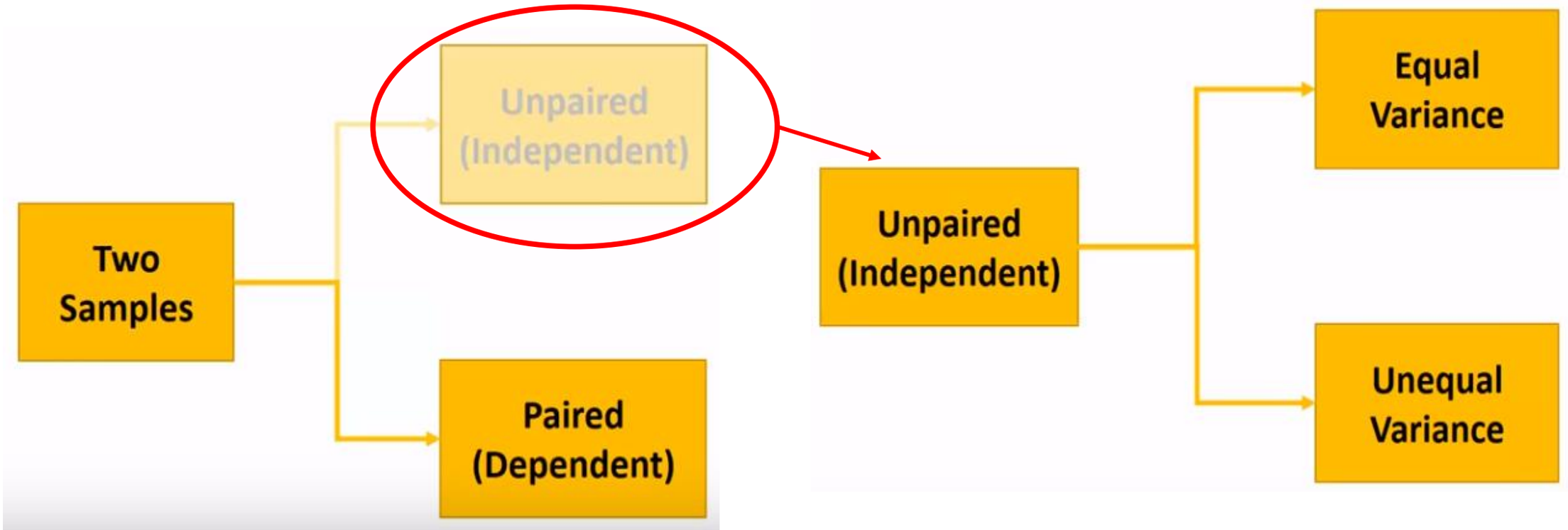
1. state the relevant null and alternative hypotheses ☑

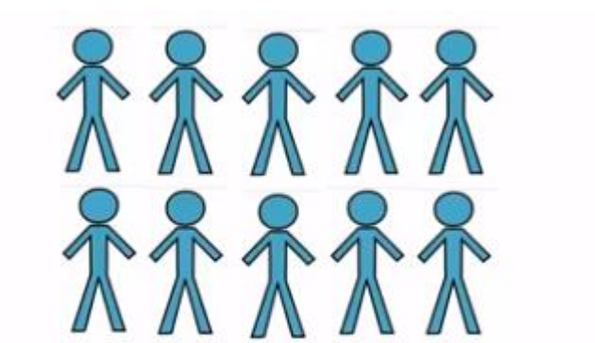
2. decide which test is appropriate for this type of hypothesis



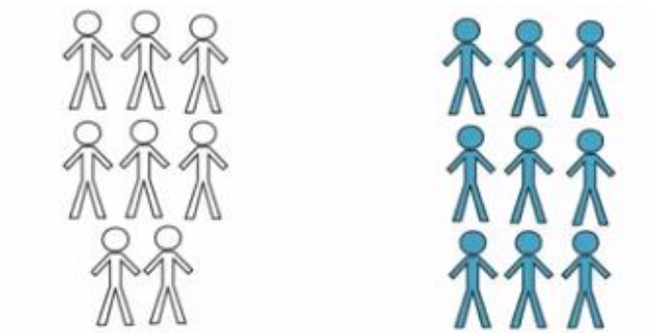
http://web.mit.edu/~csvoss/Public/usabo/stats_handout.pdf

Deciding which t-test ?





One group, tested twice.



Two groups – tested once.

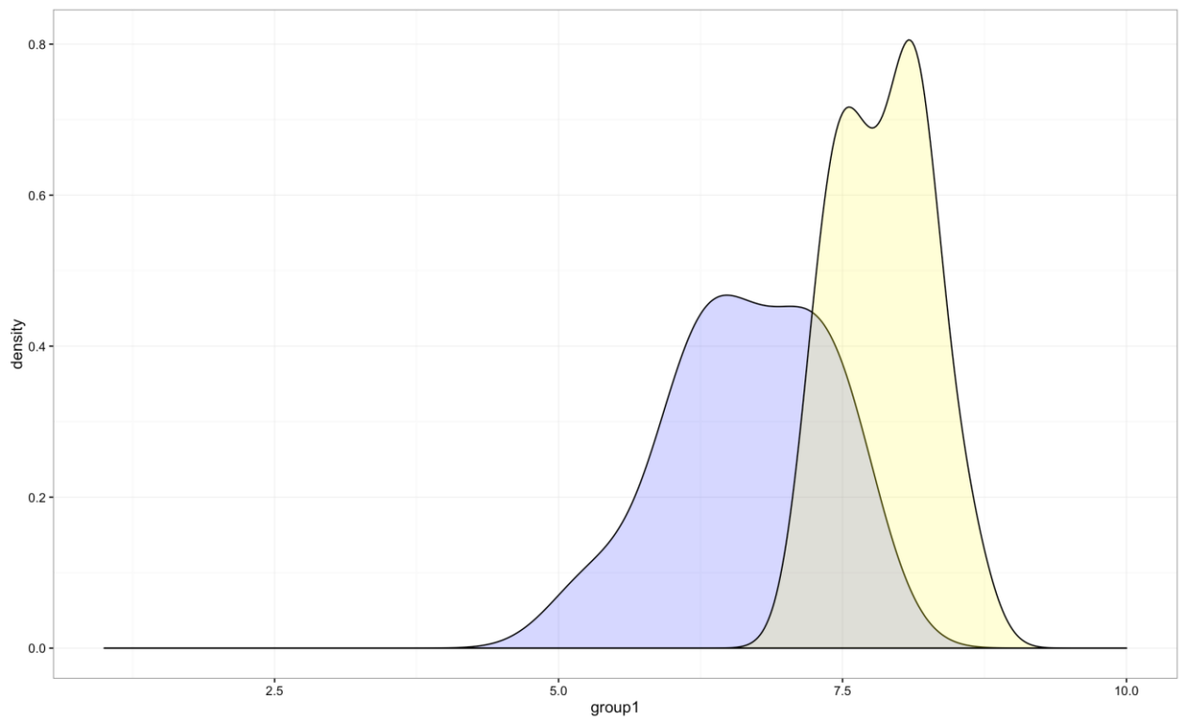
Non exclusive list of tests

- Estimate Population Proportion
- Estimate Population Mean
- One Sample Proportion
- Two Sample Proportions
- One Sample t (Mean)
- Two Sample t (Mean)
- Paired t
- Correlation/Regression Analysis
- One-Way ANOVA
- Two-Way ANOVA
- Chi Square Test
- One Sample Variance
- Two Sample Variances
- Wilcoxon rank-sum test
- Kruskal-Wallis test

Statistical hypothesis testing is a general framework of testing claims.

1. state the relevant null and alternative hypotheses
2. decide which test is appropriate for this type of hypothesis
3. check the statistical assumptions of the test

Normal distribution of both groups ?



Statistical hypothesis testing is a general framework of testing claims.

1. state the relevant null and alternative hypotheses

2. decide which test is appropriate for this type of hypothesis

3. check the statistical assumptions of the test

4. state the relevant test statistic T

two-sample, unpaired, with unequal variances:

$$\textit{Test statistic} = \frac{\text{Observed data} - \text{What we expect if null is true}}{\text{Average Variation}}$$

$$t = \frac{\text{signal}}{\text{noise}}$$

Statistical hypothesis testing is a general framework of testing claims.

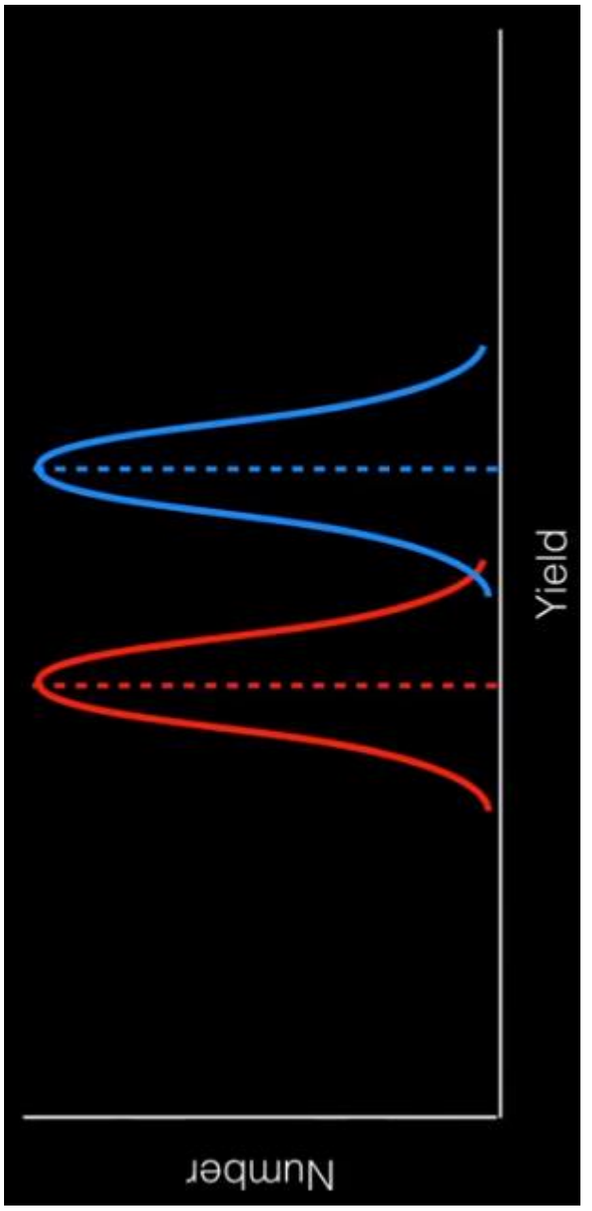
1. state the relevant null and alternative hypotheses
2. decide which test is appropriate for this type of hypothesis
3. check the statistical assumptions of the test
4. state the relevant test statistic T

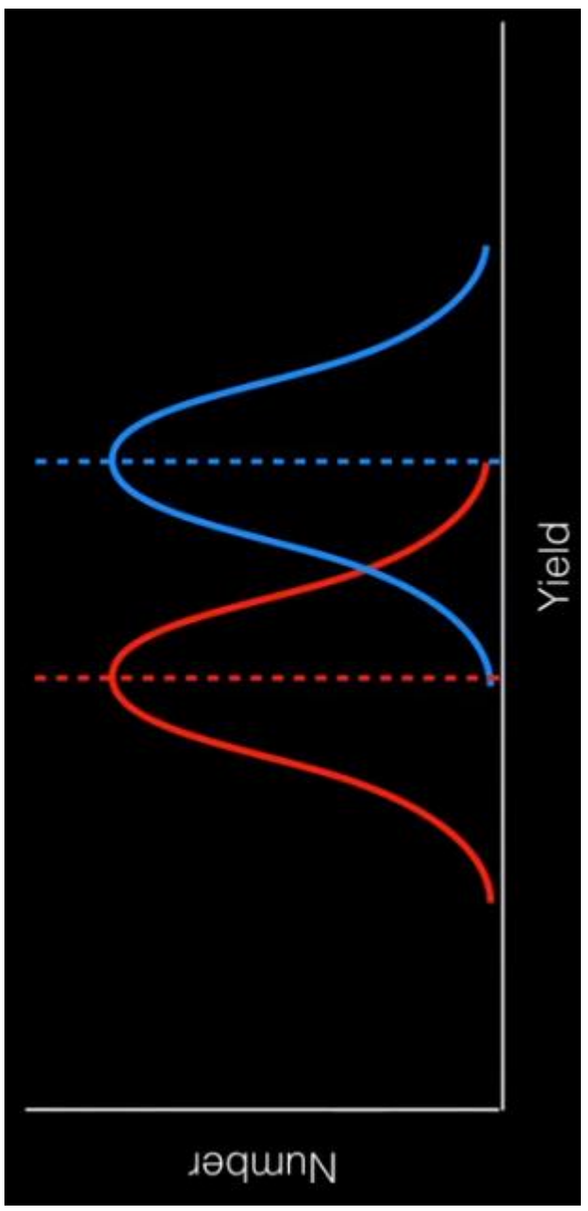
two-sample or unpaired or independent with unequal variances:

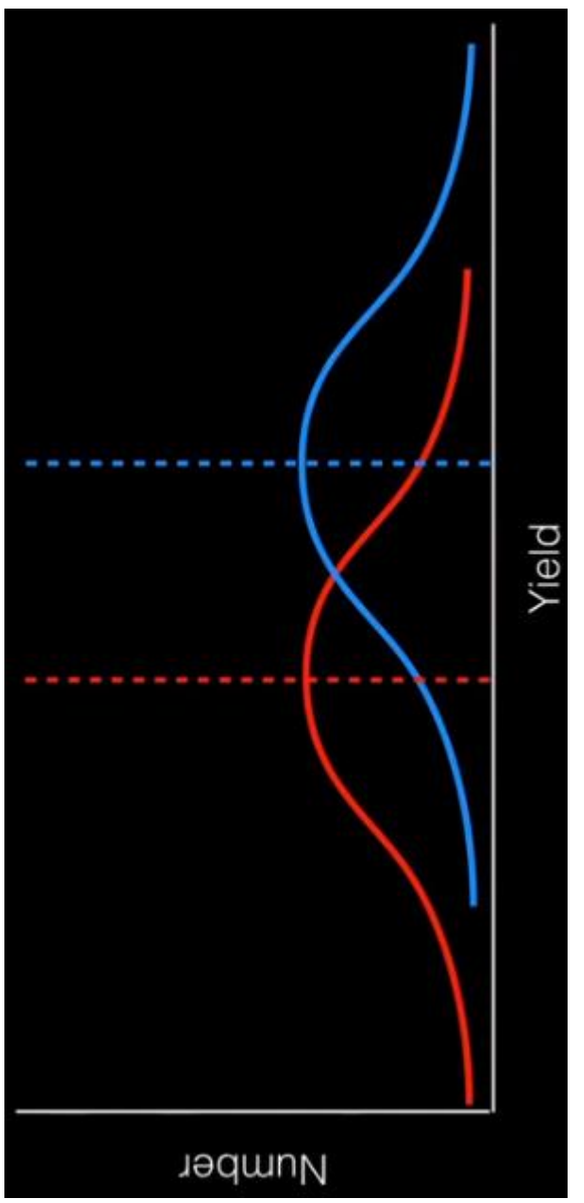
$$t = \frac{\text{signal}}{\text{noise}}$$

$$t = \frac{\text{Difference between group means}}{\text{Variability of groups}}$$

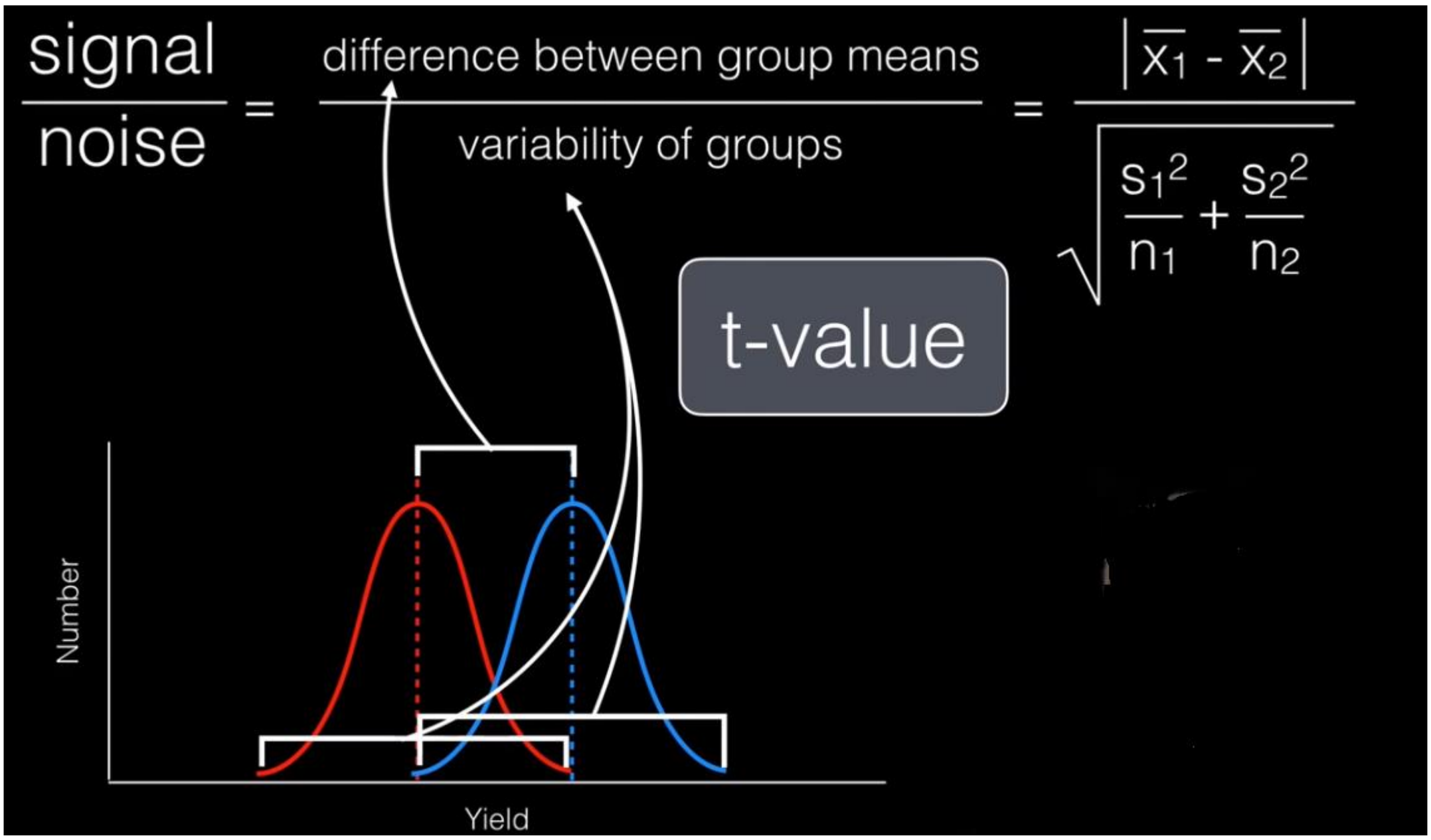
$$t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$


Yield
Number


Yield
Number


Yield
Number

t-test


signal
noise
=
difference between group means
variability of groups
=
$\left|\overline{x_1} - \overline{x_2}\right|$
$\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$
t-value
Number
Yield

Limitations

1. Results of inferential statistics can only be applied to populations that resemble the sample that was tested.

2. Sample should be roughly be normal distribution (resembling bell curve)

3. Each group should have about the same number of datapoints.

4. All data points should be independent of each other (and scores should not be influenced by each other)

Statistical hypothesis testing is a general framework of testing claims.

How to decide if there is statistically significant ?

- We can calculate the critical t-value and if our t-statistic is greater than the critical value we reject the null hypothesis
- We can calculate the p-value from our t-statistic and we can reject the null hypothesis if the p-value is smaller than our chosen alpha value

Hypothesis for Landing page example

group1

H0: Landing page button version has same conversion rate as of version without landing page

group2

H1: Landing page button web page had different Conversion rate compared to the page without landing page button

H0: $P_{group1} = P_{group2}$

H1: $P_{group1} \neq P_{group2}$

Test of proportions

Defects:

Assembly Line 1: 28 out of 350
Assembly Line 2: 32 out of 500

Are two results comparable ?

H0: Two proportions are equal
H1: Two proportions are not equal

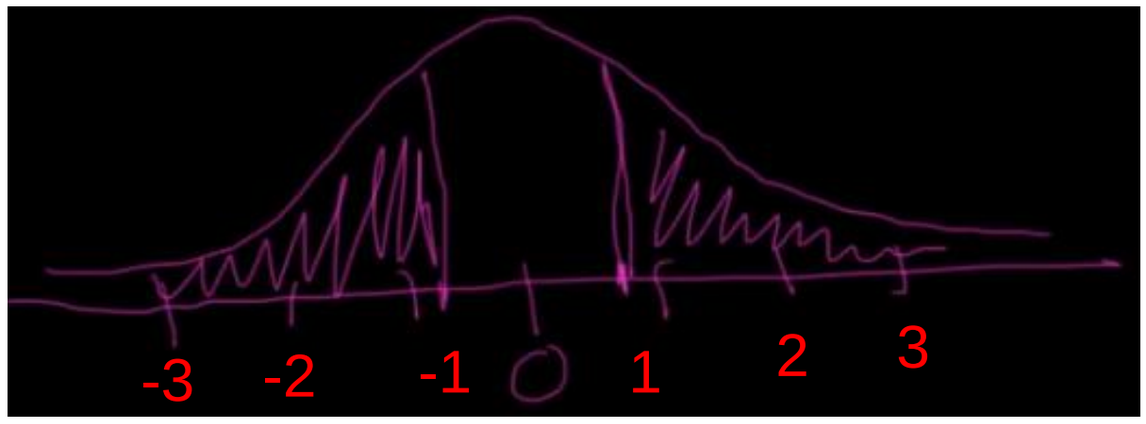
Line 1: 28/350 = .08
Line 2: 32/500 = .064

Overall proportions (p): (28+32)/(350+500) = 0.0706

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad = \quad \frac{\mathbf{(0.08 - .064) - 0}}{\sqrt{\mathbf{0.07(1-0.07)}\left(\frac{\mathbf{1}}{\mathbf{350}} + \frac{\mathbf{1}}{\mathbf{500}}\right)}} \quad \mathbf{= .8939}$$

Test of Proportions

$$Z = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$ **= .8939**



P-value= .3714

reject H0 if the p-value is less than α

.3714 < 0.05 ? No

Z Table

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239	.5279	.5319	.5359
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636	.5675	.5714	.5753
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026	.6064	.6103	.6141
0.3	.6179	.6217	.6255	.6293	.6331	.6368	.6406	.6443	.6480	.6517
0.4	.6554	.6591	.6628	.6664	.6700	.6736	.6772	.6808	.6844	.6879
0.5	.6915	.6950	.6985	.7019	.7054	.7088	.7123	.7157	.7190	.7224
0.6	.7257	.7291	.7324	.7357	.7389	.7422	.7454	.7486	.7517	.7549
0.7	.7580	.7611	.7642	.7673	.7704	.7734	.7764	.7794	.7823	.7852
0.8	.7881	.7910	.7939	.7967	.7995	.8023	.8051	.8078	.8106	.8133
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315	.8340	.8365	.8389
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554	.8577	.8599	.8621
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770	.8790	.8810	.8830

Why do we need testing for significance

Statistical significance means highly likely that the differences are real, repeatable, and not due to random chance

Why do we need testing for significance

Statistical significance means highly likely that the differences are real, repeatable, and not due to random chance

Infinite monkey theorem

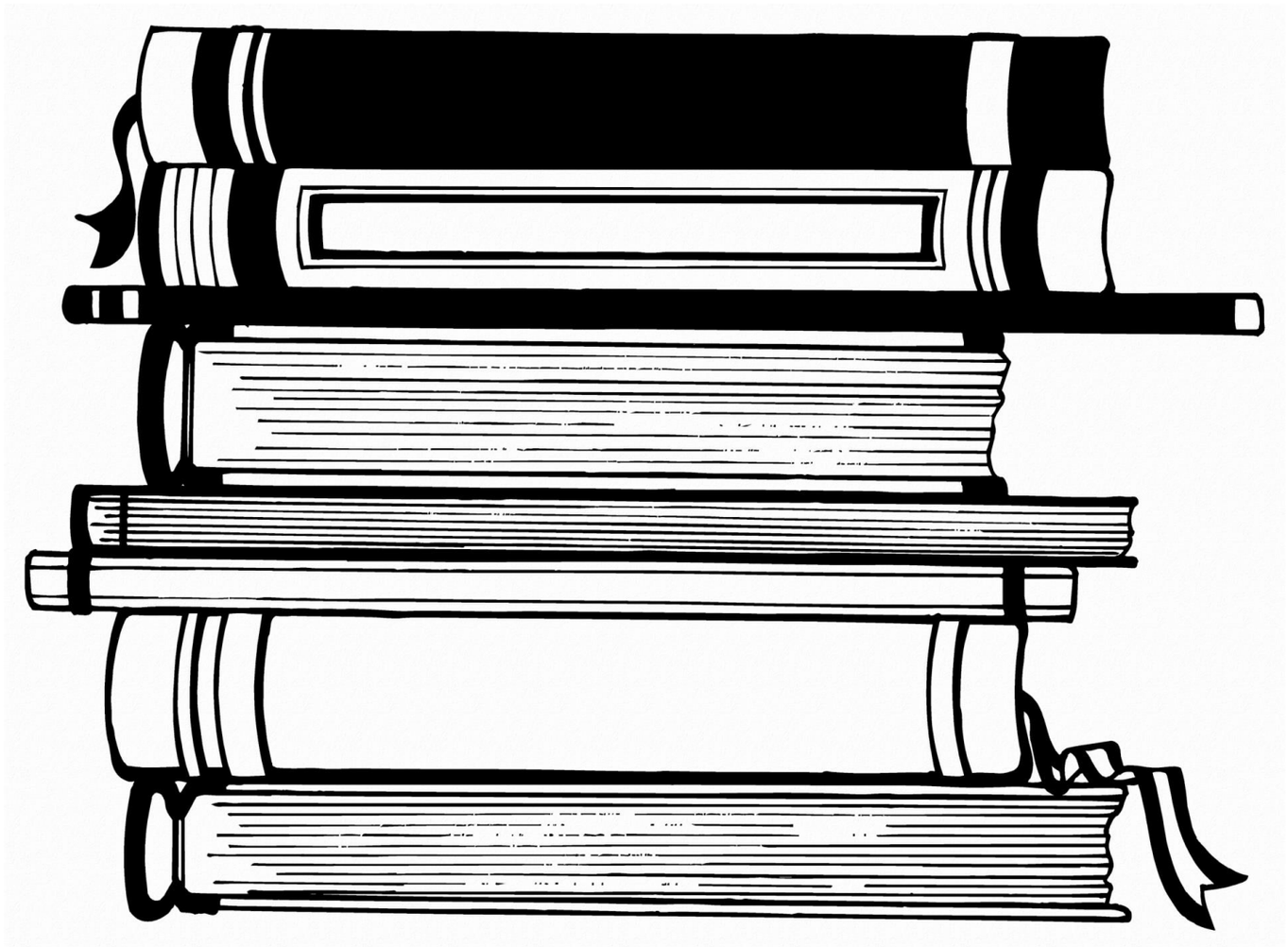
From Wikipedia, the free encyclopedia

The **infinite monkey theorem** states that a monkey hitting keys at random on a typewriter keyboard for an infinite amount of time will almost surely type a given text, such as the complete works of William Shakespeare.

Common tests

Assumed Distribution	Example Case	Standard Test	Alternative Test
Gaussian	Average Revenue Per Paying User	Welch's t-test (Unpaired t-test)	Student's t-test
Binomial	Click Through Rate	Fisher's exact test	Barnard's test
Poisson	Transactions Per Paying User	E-test[7]	C-test
Multinomial	Number of each product Purchased	Chi-squared test	
Unknown	--	Mann–Whitney U test	Gibbs sampling

Source: wikipedia



https://conversionxl.com/blog/ab-testing-guide/