Hypothesis Testing

Week 03 - Day 01

What

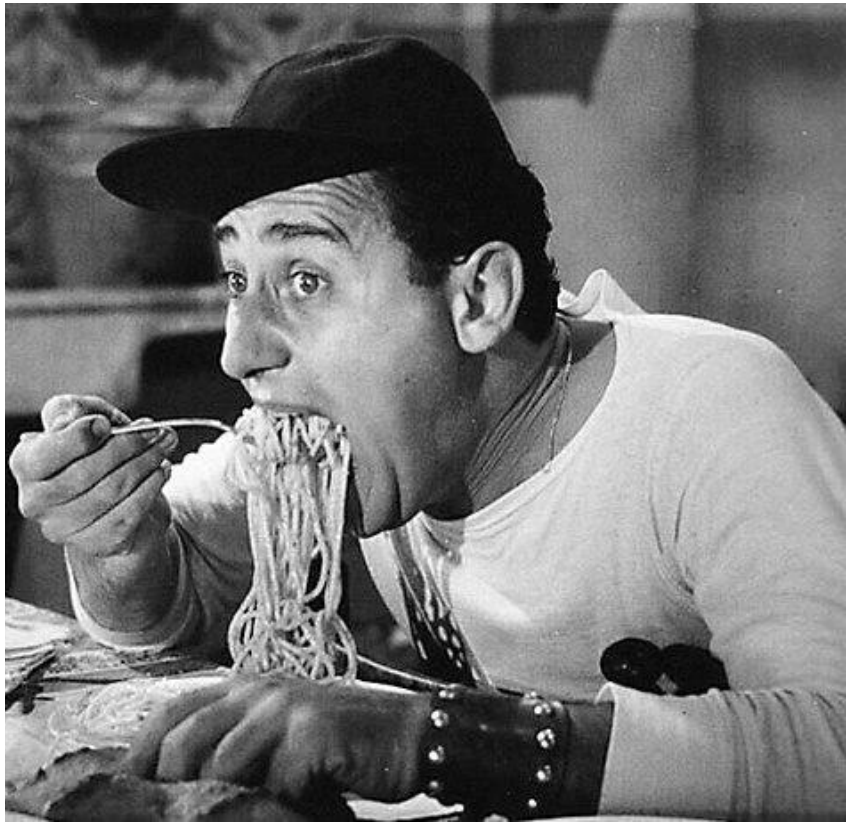
Are two groups **really** different?

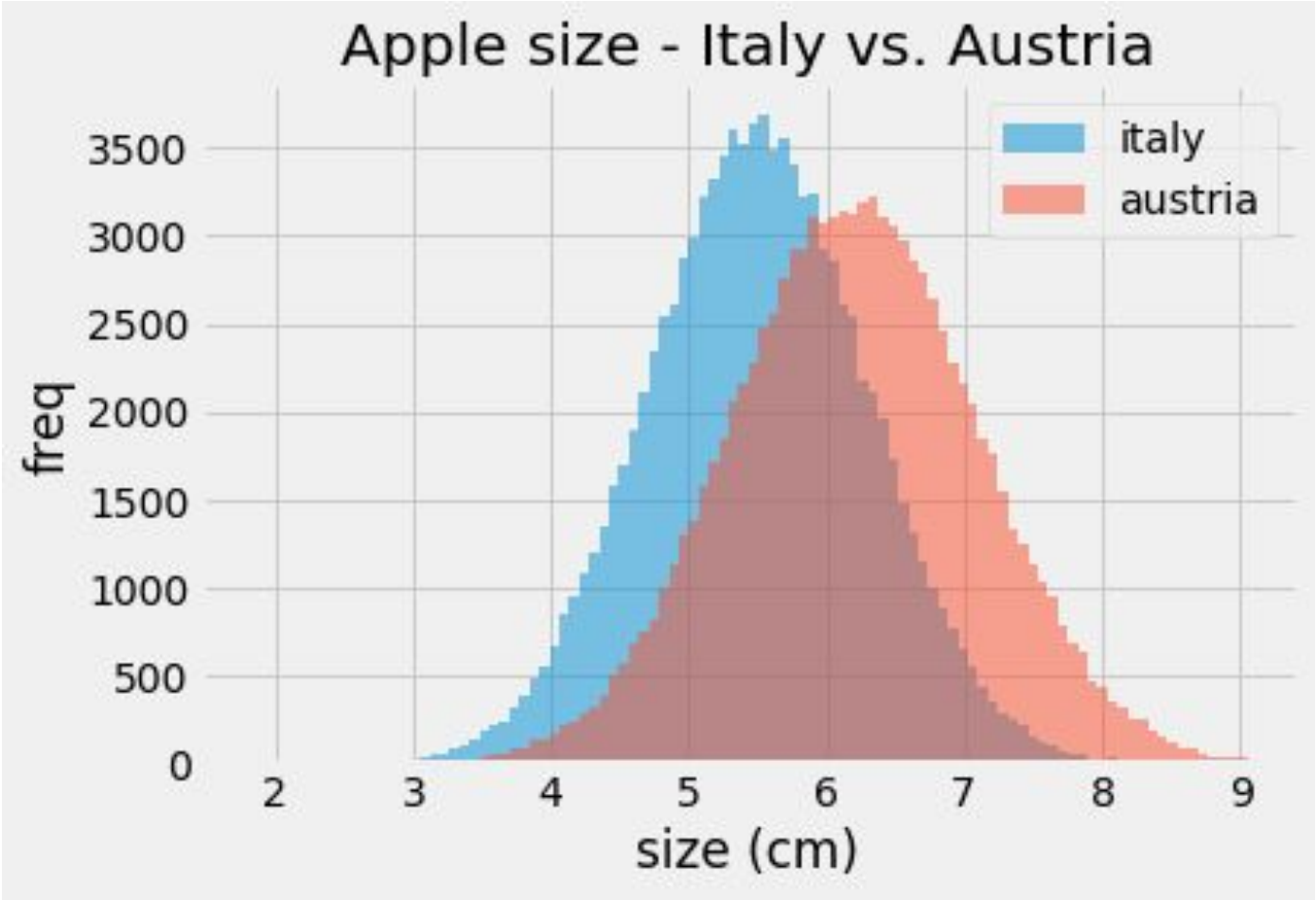
Example #1

Let’s talk about apples

Austria

Italy

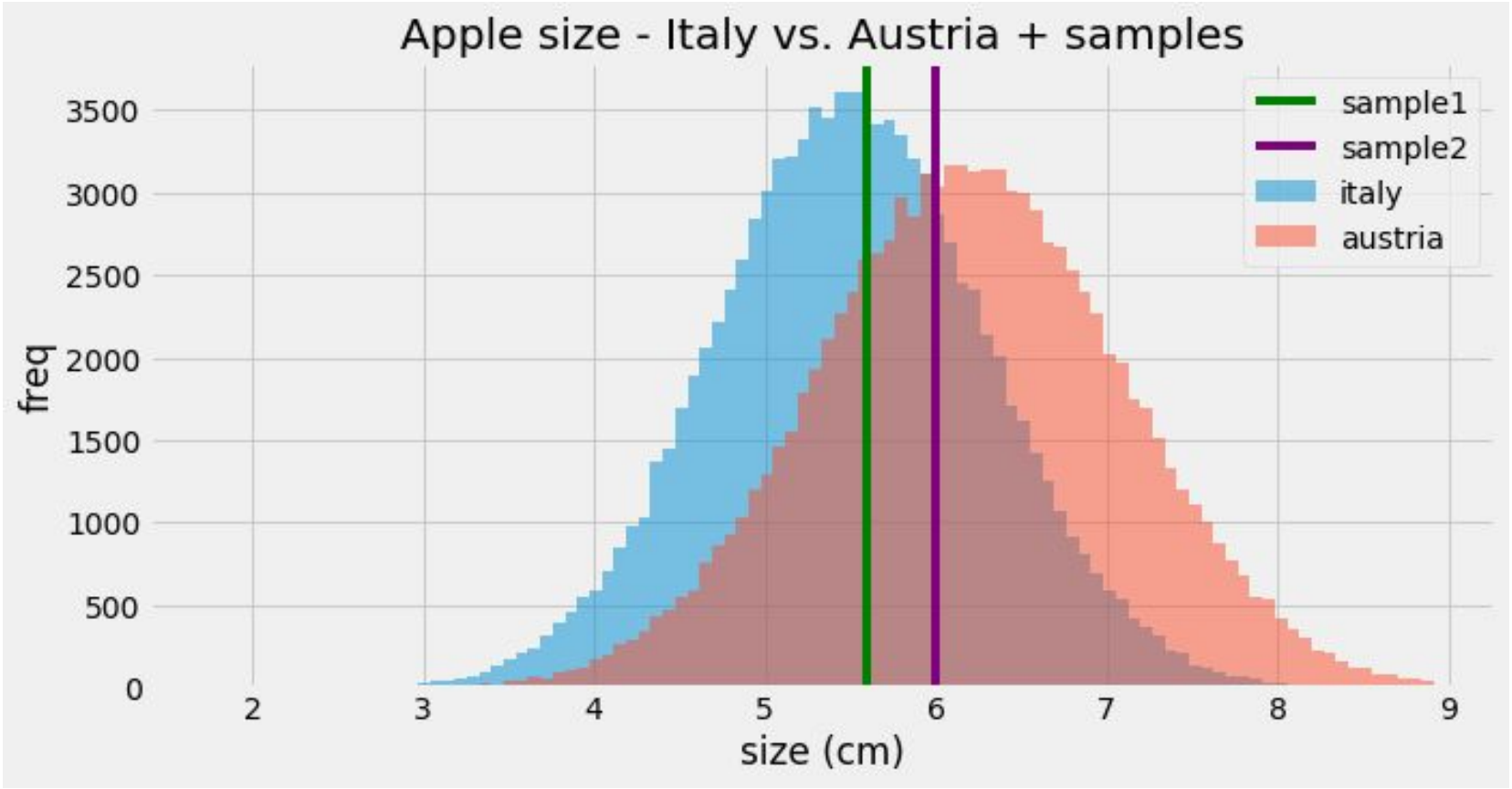


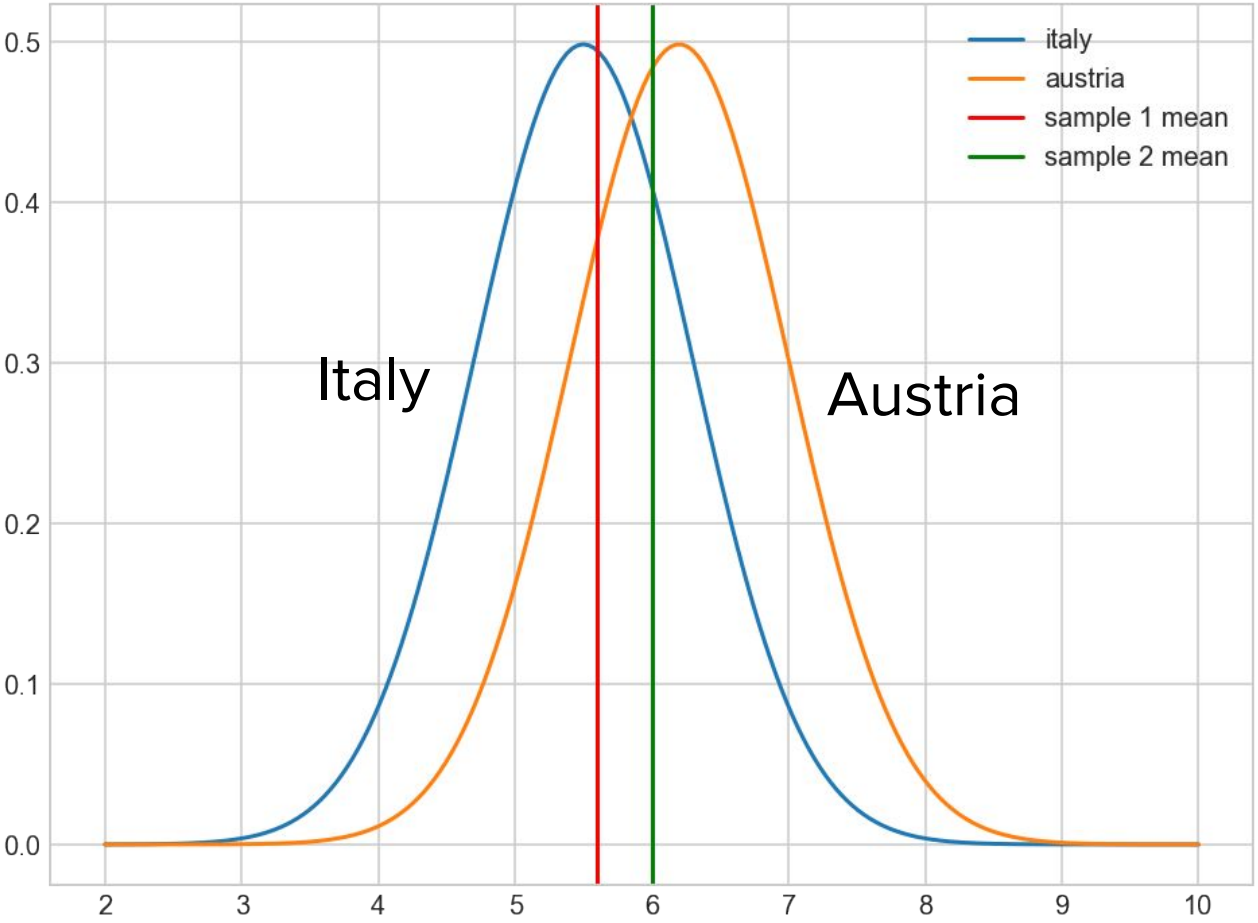

Apple size - Italy vs. Austria
italy
austria
freq
3500
3000
2500
2000
1500
1000
500
0
2
3
4
5
6
7
8
9
size (cm)

Origin	Size Mean	Size STD
Italy	5.5 cm	0.8 cm
Austria	6.2 cm	0.9 cm

2 samples with unknown origin

Sample	Size Mean	Size STD
#1	5.6 cm	0.8 cm
#2	6.0 cm	0.8 cm


Apple size - Italy vs. Austria + samples
sample1
sample2
italy
austria
freq
3500
3000
2500
2000
1500
1000
500
0
2
3
4
5
6
7
8
9
size (cm)


italy
austria
sample 1 mean
sample 2 mean
Italy
Austria
0.5
0.4
0.3
0.2
0.1
0.0
2
3
4
5
6
7
8
9
10

Do they come from Italy or Austria??

10 apples in each sample

Vs.

10000 apples in each sample

Example #2

Clinical trial for the new Viagra

Group	Satisfaction
experimental	50%
placebo	30%

Does it work?

Case 1

Group	Satisfaction	Individuals
experimental	50%	10
placebo	30%	10

Case 2

Group	Satisfaction	Individuals
experimental	50%	500
placebo	30%	500

Case 3

Group	Satisfaction	Individuals
experimental	35%	100,000
placebo	30%	100,000

Case 4

Group	Satisfaction	Individuals
experimental	90%	20
placebo	10%	20

Effect size

+

Sample size

Can we be 100% sure about our conclusions?

Solution: Uncertainty!

Uncertainty = Probability

6 ways of saying the same thing

1.

We’re 80% sure the new Viagra really has a positive effect

2.

We’re 80% sure the new Viagra has a positive effect not by chance

3.

We’re 80% sure control group

and effect group are really/statistically different

4.

We’re 80% sure control group and effect group come from different populations

5.

We accept a 20% **risk** that we're selling a new viagra that is not better than placebo

6.

There is a 20% chance that we can observe that difference (or a larger one) in two samples coming from the same exact population

There is a **20% chance** that we can observe that difference (or a larger one) in two samples coming from the same exact population

There is a 20% chance that we can observe **that difference** (or a larger one) in two samples coming from the same exact population

There is a 20% chance that we can observe that difference (or a larger one) in two samples coming from **the same exact population**

What is a good p to say viagra works?

(0.001%, 0.1%, 1%, 5%, 10%, 20%, 30%)

Less than 5%? Good enough!

5% = 1 out of 20

QUIZ

How to calculate p?

Sample 1, Sample 2

Are they statistically different?

The hacker way

Let’s assume two samples are coming from the same population

What is the distribution of mean(sample1)-mean(sample2)?

1 - generate one population (of apples)

Norm(mean,std)

1 - generate one population (of apples)

2 - extract two samples and calculate mean(sample1) - mean(sample2)

Sample 1 --> mean=5.4

Sample 2 --> mean=4.9

mean1-mean2 = 0.5

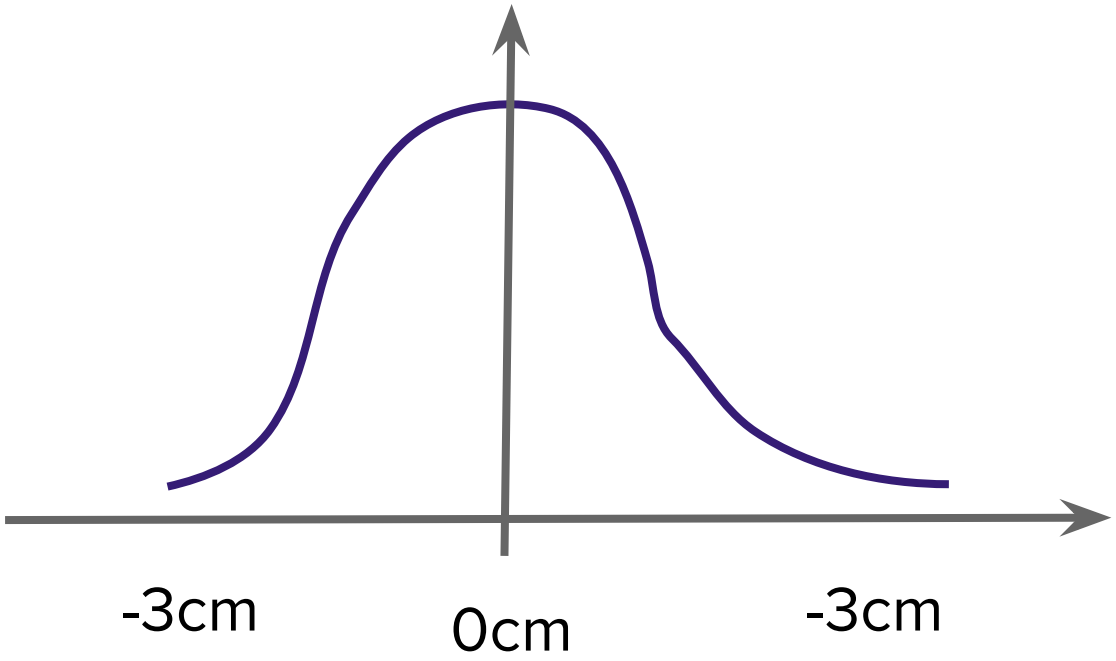
1 - generate one population (of apples)

2 - extract two samples and calculate mean(sample1) - mean(sample2)

3 - repeat the step 2) 100,000 times

[0.5, -0.7, 0.8, -0.2, ..., 0.9]

mean(sample1) - mean(sample2) distribution



1 - generate one population (of apples)

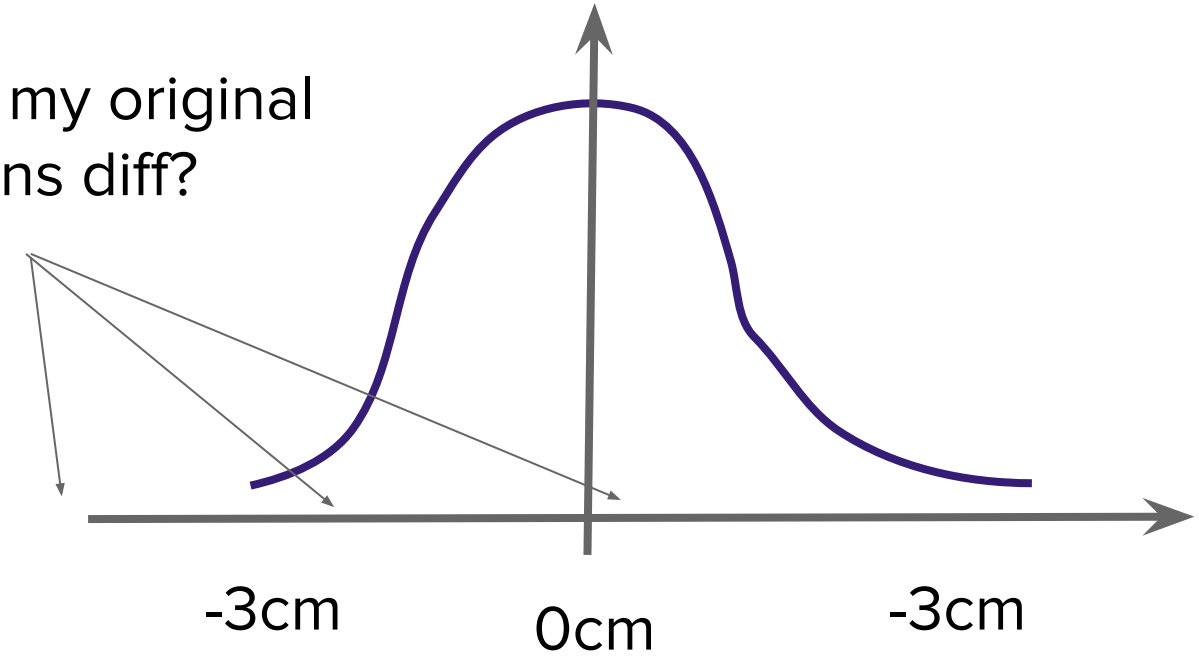
2 - extract two samples and calculate mean(sample1) - mean(sample2)

3 - repeat the step 2) 100,000 times

4 - check the original means diff against this distribution

mean(sample1) - mean(sample2) distribution

Where is my original means diff?



The boring way

Use t-test

(t-statistic + t-distribution)

$$t = \frac{\bar{x}_E - \bar{x}_C}{\sqrt{s^2(\frac{1}{n_E} + \frac{1}{n_C})}}$$

$$s^2 = \frac{\sum_{i=1}^{n_E}(x_i - \bar{x}_E)^2 + \sum_{j=1}^{n_C}(x_j - \bar{x}_C)^2}{n_E + n_C - 2}$$

Check your t-statistics

against the t-distribution

Nice video about t-statistics

The right way

stats.ttes_ind(sample1, sample2)

ttest_indResult(statistic=-1.23, pvalue=0.218)

A few terms

H0 vs. H1

H0 - Null Hypothesis

There is **not** statistical difference between the two samples

H1 - Alternative Hypothesis

The two samples are

statistically different

Rejecting the null hypothesis

p-value<=0.05

H0 is rejected - H0 is false

H1 is true - The samples are different

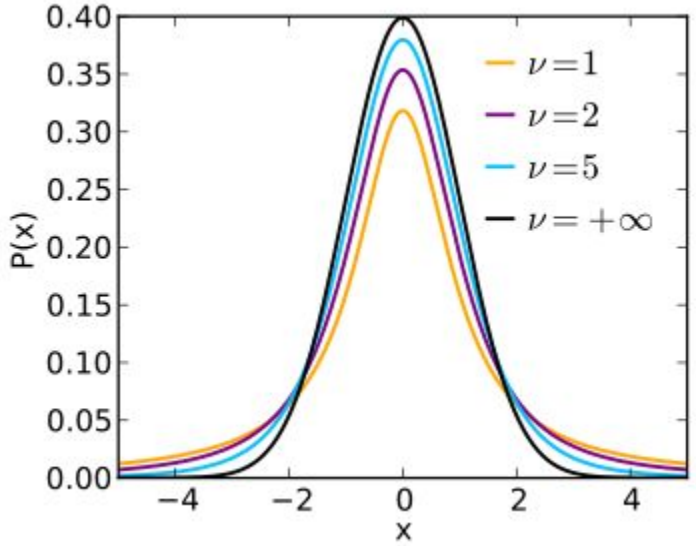
p-value>0.05

H0 is NOT rejected - H0 is true

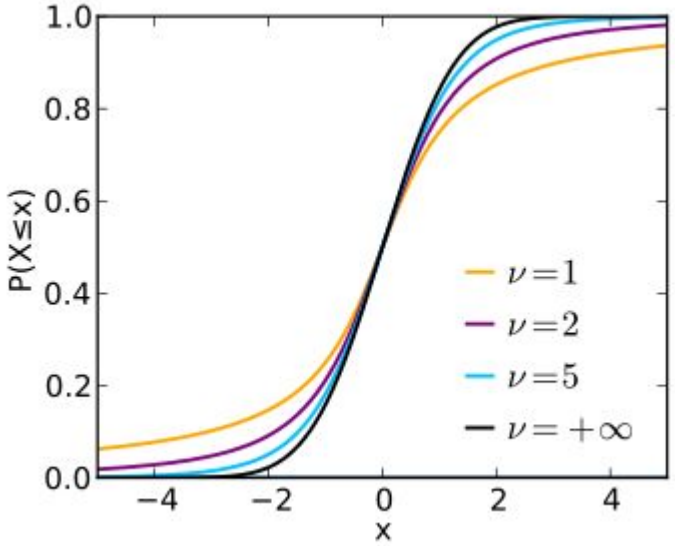
H1 is false - The samples are NOT different

PDF vs. CDF

PDF - Probbility density function



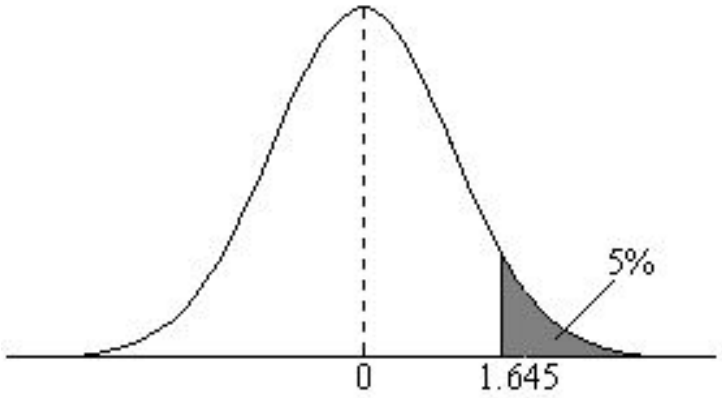
CDF - Cumulative distribution function



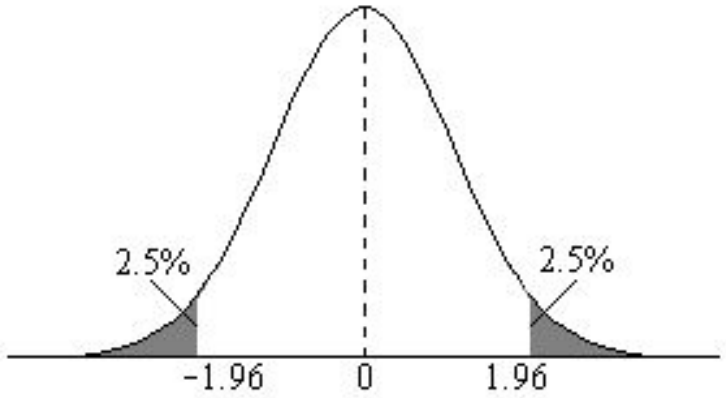
One-tailed tests

vs.

Two-tailed tests



(a) One-tailed test



(b) Two-tailed test

Viagra

Apples

Significance level = alpha = 5%

Parametric tests

Vs.

Non parametric tests

P-Hacking

https://projects.fivethirtyeight.com/p-hacking/

“Psychology journal bans *P* values”

“Test for reliability of results ‘too easy to pass’, say editors.”

https://www.nature.com/news/psychology-journal-bans-p-values-1.17001

Summary

Statistical difference

P-value (<=5%)

T-test

H0 vs H1

The hacker way

5% = 1/20