Parametric Statistics

Week 4 - Significance Tests

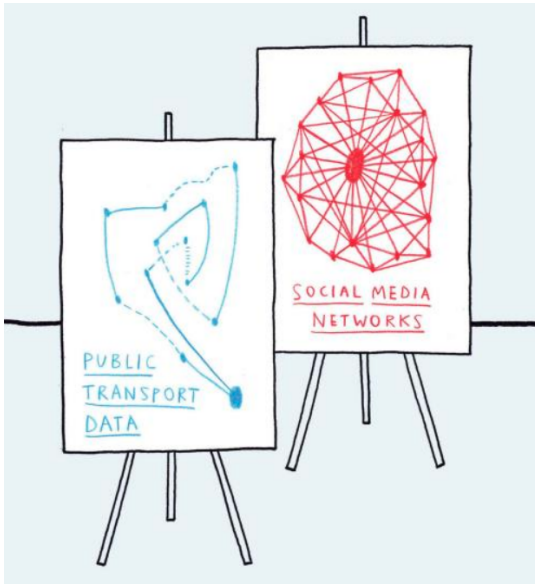
Juan Carlos Medina Serrano

Technische Universität München
Hochschule für Politik
Political Data Science

Munich, 13. November 2019

political data science

https://politicaldatascience.blogspot.de



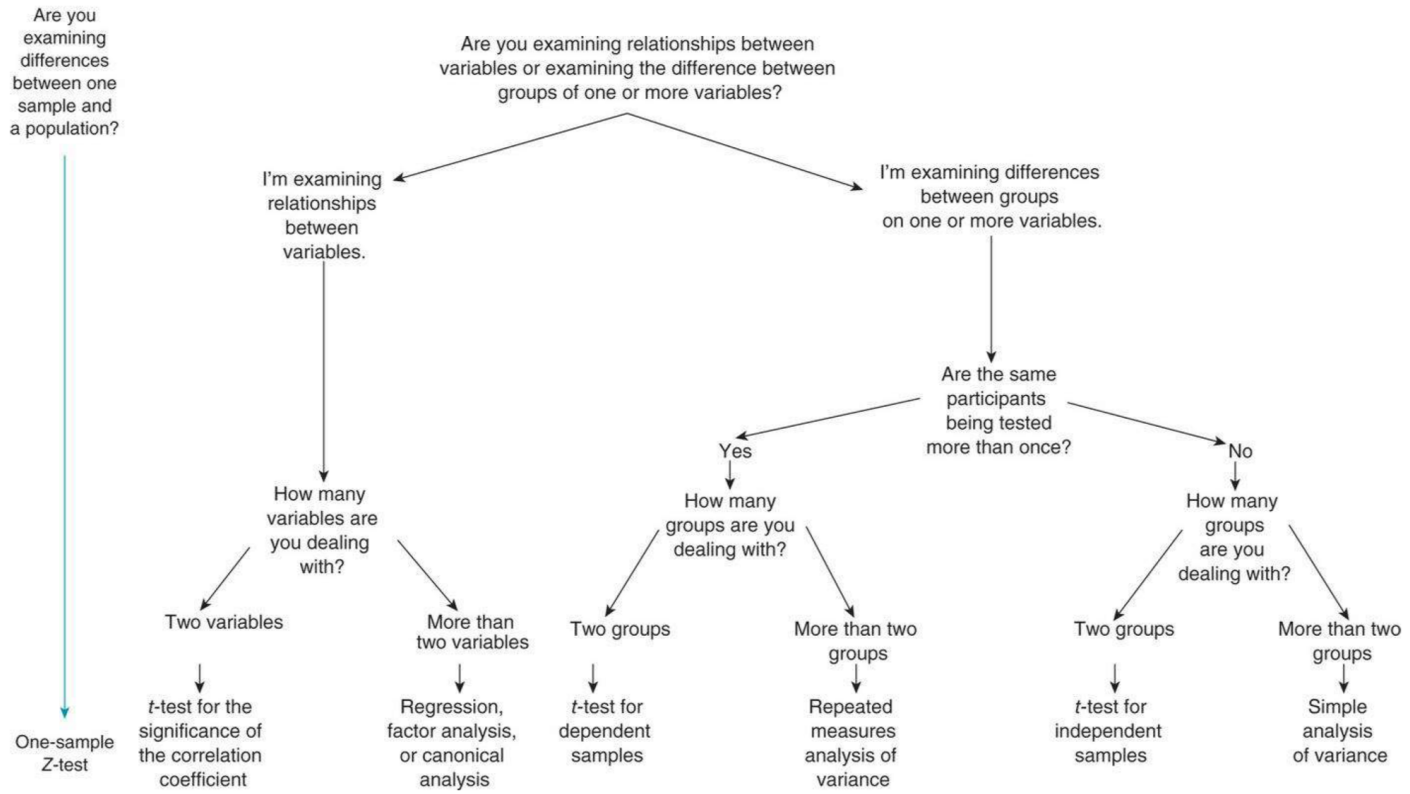
Significance Tests

Significance tests are the core of hypothesis testing. They model the hypothesis and allow us to find significance in the data.

Types of significance tests:

- **Comparing means:** To compare means between two groups.
  Answer the question - Is there a difference?
- **Asses correlation:** Find if there is a relationship between variables
  Answer the question - Is there a relationship?
- **Check assumptions:** Every significance test has assumptions that have to be met in order to use the test. These assumptions can be tested with significance tests (which also have assumptions...inception again)
  Answer the question - Has the collected data a specific property?

Comparing Means and Correlation Tests



Assumptions

Why bother?

If even one of the assumptions for a significance test is not met...the complete test can be wrong.

Assumptions of Parametric Data:

- **Normally distributed data:** Something in the data has to be normally distributed. Which property depends on the test (e.g. the data, the sample distribution, the errors in the model)
- **Homogeneity of variance:** Variances should be the same throughout the data. Again the exact property depends on the test.
- **Independence:** Some property in the data has to be independent. For example, the observations have to be independent from each other.

Significance Tests

Types of significance tests:

- **Comparing means**
- Asses correlation
- Check assumptions

Comparing means

Parametric tests to compare means between groups:

- Z-test
- independent t-test
- Welch's t-test
- dependent t-test
- Analysis of variance (ANOVA)

Z-tests

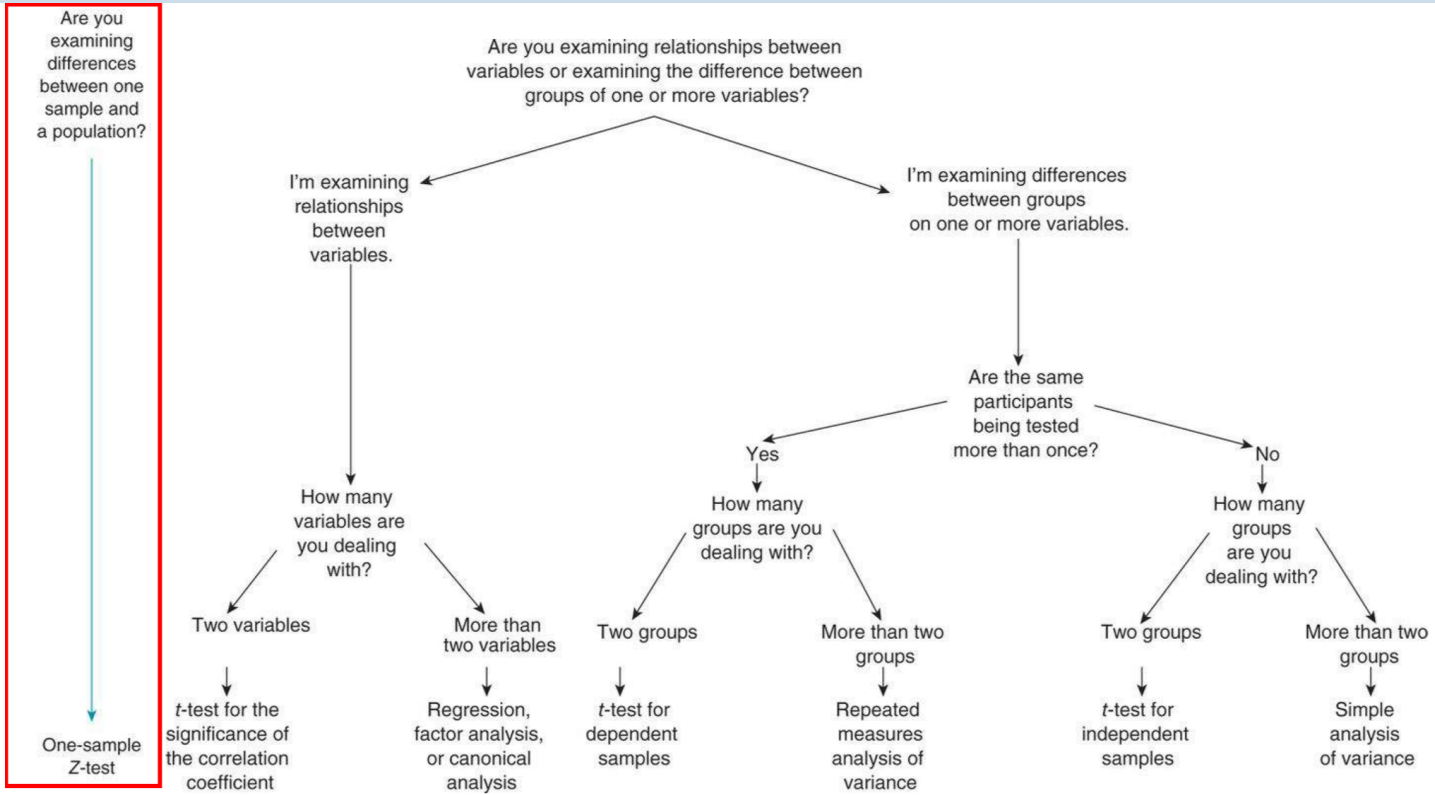
Usage

Comparing the mean of a sample and the population. Is the sample representative of the population? Is the sample mean significantly smaller/bigger than the population?

Assumptions

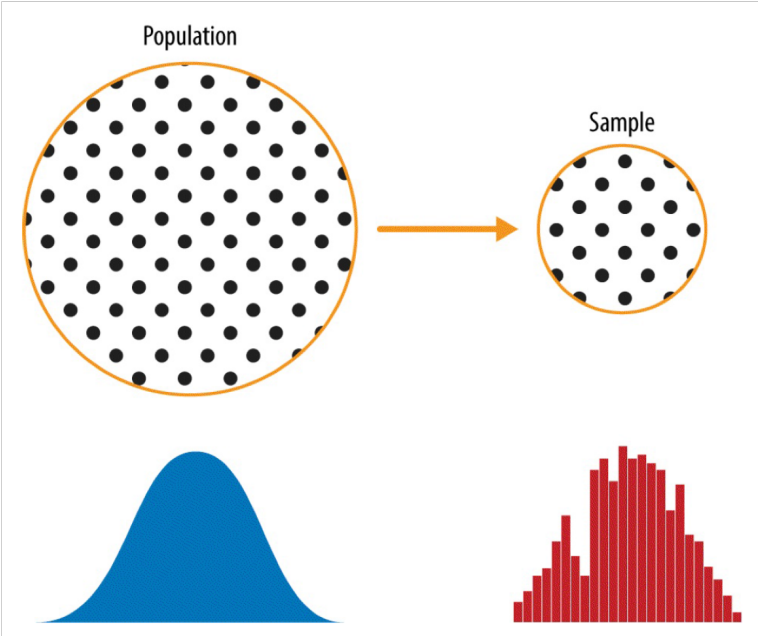
- Distribution from which the sample is drawn and population distribution are normally distributed.
- Population mean and variance are known (Most of the time we don't know them)

Z-tests


Are you examining differences between one sample and a population?
One-sample Z-test
Are you examining relationships between variables or examining the difference between groups of one or more variables?
I'm examining relationships between variables.
I'm examining differences between groups on one or more variables.
How many variables are you dealing with?
Two variables
t-test for the significance of the correlation coefficient
More than two variables
Regression, factor analysis, or canonical analysis
Are the same participants being tested more than once?
Yes
How many groups are you dealing with?
Two groups
t-test for dependent samples
More than two groups
Repeated measures analysis of variance
No
How many groups are you dealing with?
Two groups
t-test for independent samples
More than two groups
Simple analysis of variance

Z-tests

Sample Data: mean $\overline{X}$, standard deviation s and N observations.
Population Data: mean μ and standard deviation σ.



Z-tests

Remember: What is a test statistic?

$$test\ statistic = \frac{variance\ explained\ by\ the\ model}{variance\ not\ explained} = \frac{effect}{error}$$

The Z-test statistic:

$$Z = \frac{\mu - \overline{X}}{SE}$$

We are modeling the difference of means, this is the effect.
We know from two weeks ago, what standard error for a sample mean is:

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

Last time, we used s to estimate the SE, but in this case we have the real SE from the population.

$$Z = \frac{\mu - \overline{X}}{\frac{\sigma}{\sqrt{N}}}$$

Should we learn the formulas by heart? **NO!** R and Google are always there for us, important is only to understand them

Z-tests

Hypothesis test for Z-tests:

$$H_0 : \overline{X} = \mu$$

$$H_1 : \overline{X} \neq \mu$$

OR

$$H1 : \overline{X} < \mu$$

OR

$$H_1 : \overline{X} > \mu$$

Z-tests

Effect size for Z-tests:
For this test, we use Cohen's d to calculate the standardized effect size:

$$d = \frac{\overline{X} - \mu}{\sigma}$$

Hmmm... Similar to Z-score...Yes! But no N, the effect size is totally independent of the sample size!

Cohen's d guidelines:

- d= 0 to .2 (small effect size)
- d = .2 to .5 (middle effect size)
- d = .5 or more (large effect size)

Remember: Effect size can helps us find the minimum of participants for an experiment after defining the power we want.

Z-tests

Example Exam results for a school group vs. the exam results of the whole state

	Size	Mean	Standard Deviation
Sample	36	100	5.0
Population	1,000	99	2.5

Z-tests

Example Exam results for a school group vs. the exam results of the whole state

	Size	Mean	Standard Deviation
Sample	36	100	5.0
Population	1,000	99	2.5

$$Z = \frac{100 - 99}{\frac{2.5}{\sqrt{36}}} = 2.38$$

Remember the standardized mean distribution? 1.96 was the critical value for an alpha level of .5 (95%confidence). Reject the null hypothesis!

$$d = \frac{100 - 99}{2.5} = .4$$

We find significance and the effect has a medium size.

Independent t-tests

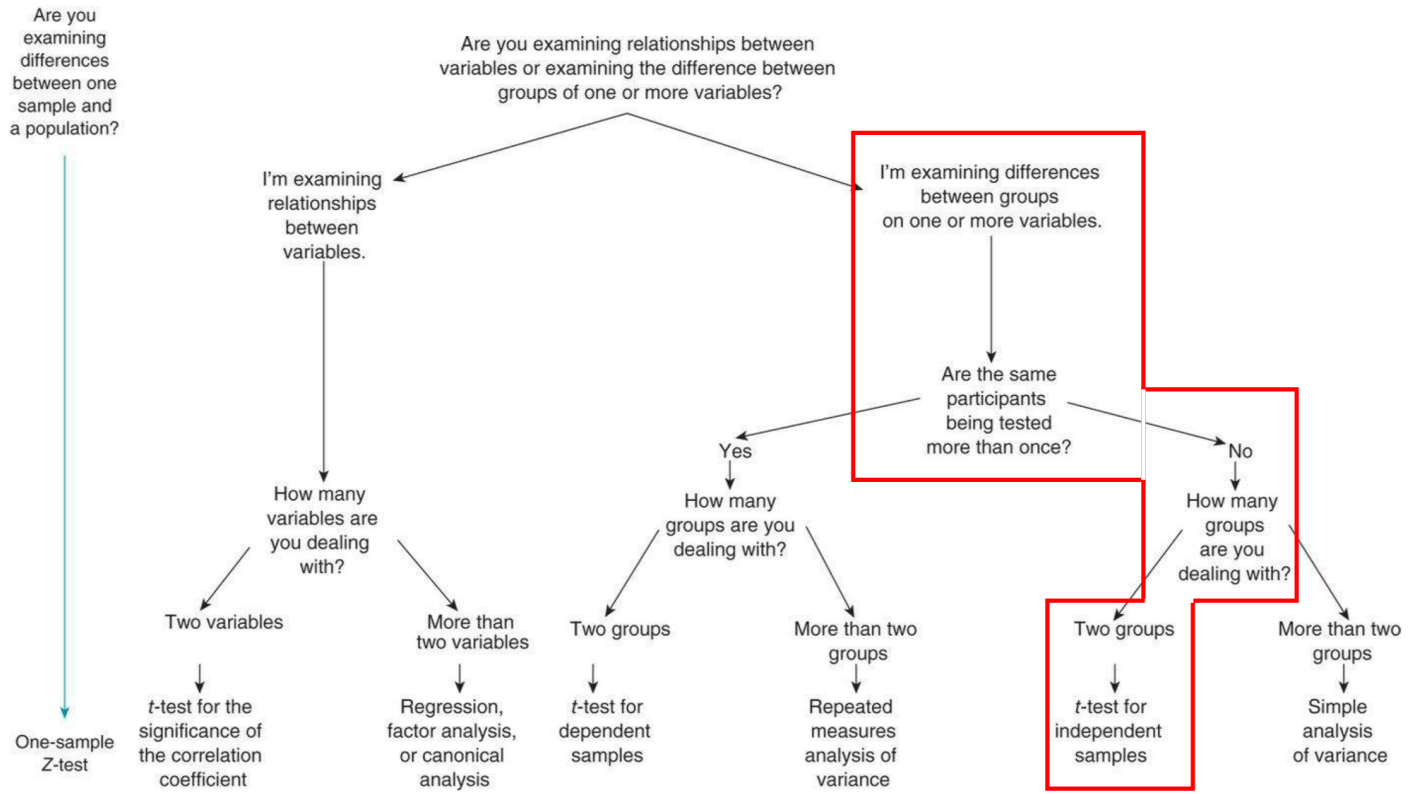
Usage

Comparing the mean of two independent groups (Different people/observations in each group). For example, the control and the treatment group.

Assumptions

- Distributions from which the samples are drawn are normally distributed.
- The samples in the groups are independent.
- Homogeneity of variance in case the sample size of the groups are unequal.

Independent t-tests



Independent t-tests

Group 1 Data: mean $\overline{X_1}$, standard deviation s_1 and N_1 observations. Group 2 Data: mean $\overline{X_2}$, standard deviation s_2 and N_2 observations.
t-statistic:

$$t = \frac{\overline{X_1} - \overline{X_2}}{f(N_1, s_1, N_2, s_2)}$$

$f(a, b)$ means a complicated function that depends on a and b. Not worth the effort writing it down or learning it...

The function depends on the standard errors of each group:

$$SE_1 = \frac{s_1}{\sqrt{N_1}}$$

$$SE_2 = \frac{s_2}{\sqrt{N_2}}$$

Independent t-tests

Hypothesis test for independent t-tests:

$$H_0 : \overline{X_1} = \overline{X_2}$$

$$H_1 : \overline{X_1} \neq \overline{X_2}$$

OR

$$H1 : \overline{X_1} < \overline{X_2}$$

OR

$$H_1 : \overline{X_1} > \overline{X_2}$$

Independent t-tests

To select a critical value, we need to define the Degrees of Freedom of the t-distribution. In this test:

$$df = N_1 - 1 + N_2 - 1$$

Effect size: We use the r Pearson's coefficient

$$r = \sqrt{\frac{t^2}{t^2 + df}}$$

Where we have seen that:

- small effect: r = .1
- medium effect: r = .3
- large effect: r= .5

(Again...you don't need to learn this r formula)

Independent t-tests

Reporting an independent test

Example Are you afraid of spiders? To one group of 12 people, you show a picture of a spider and to a second group of the same size, you show a real spider.

On average, participants experienced greater anxiety from real spiders ($M=47$, $SE=3.18$), than from pictures of spiders ($M=40$, $SE=2.68$). This difference is significant ($p<.05$). Moreover, it represents a medium-size effect $r=.34$.

Welch's t-tests

Usage

Comparing the mean of two independent groups (Different people/observations in each group). For example, the control and the treatment group. The variance in the two groups is different from each other.

Assumptions

- Distributions from which the samples are drawn are normally distributed.
- The samples in the groups are independent.

Welch's t-tests

Same as independent t-test! Only that it relaxes one assumption.

There is a correction in the degrees of freedom, to make the test more conservative (cautious).

Dependent t-tests

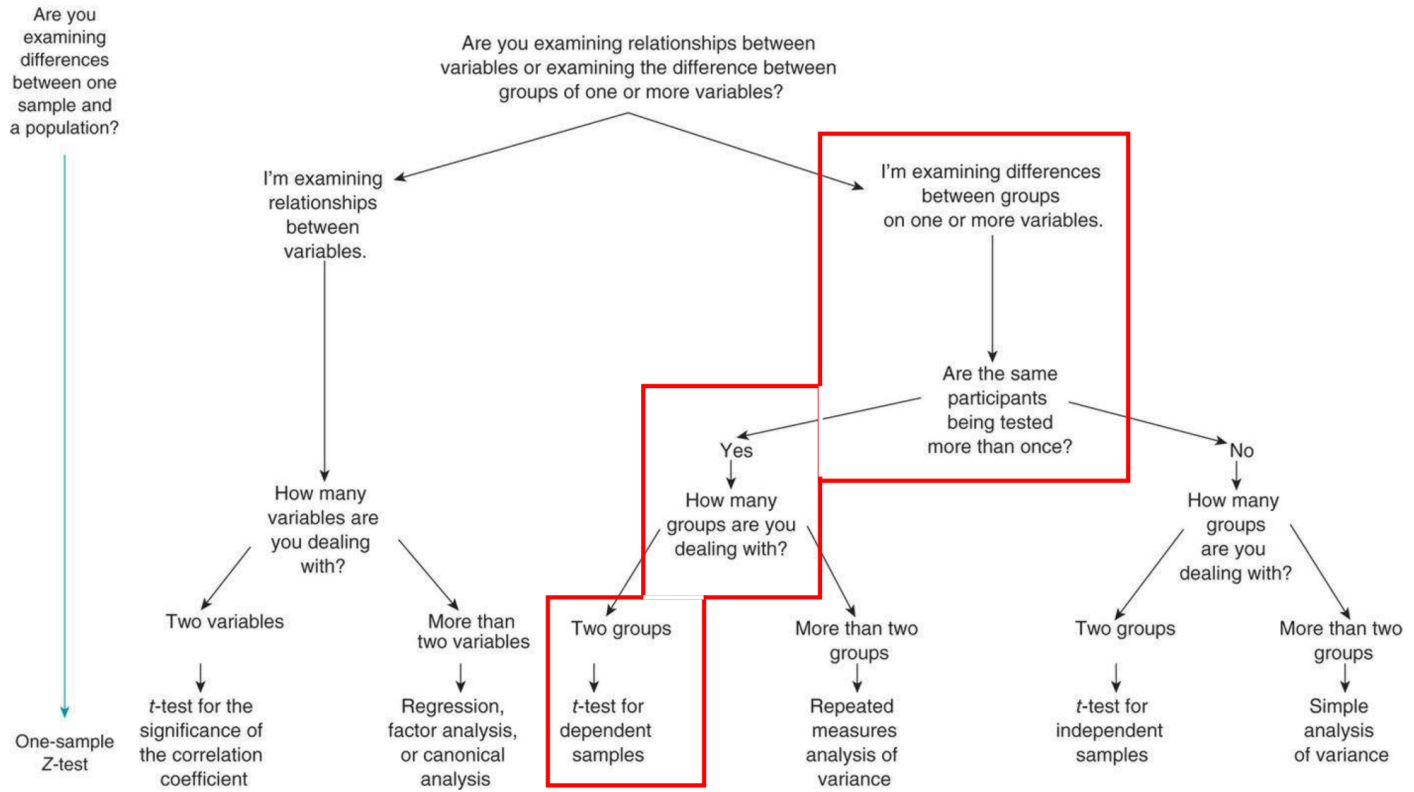
Usage

Comparing the mean of two dependent groups. The same people/observations are measured before and after some time or some treatment.

Assumptions

- The differences between values before and after the treatment are normally distributed.
- Sample sizes are equal in both groups (Same participants).

Dependent t-tests



Dependent t-tests

Group 1 Data (before treatment): mean $\overline{X_1}$, standard deviation s_1 and N observations. Group 2 Data (after treatment): mean $\overline{X_2}$, standard deviation s_2 and N observations.
t-statistic:

$$t = \frac{\sum_i X_{2_i} - X_{1_i}}{f(N,D)} = \frac{\sum D}{f(N,D)}$$

where D is the difference between each value before and after treatment. f(N,D) is the standard error of differences.

The degrees of freedom in this case are $N-1$

Dependent t-tests

Hypothesis test for dependent t-tests:

$$H_0 : \overline{X_1} = \overline{X_2}$$

$$H_1 : \overline{X_1} \neq \overline{X_2}$$

OR

$$H1 : \overline{X_1} < \overline{X_2}$$

OR

$$H_1 : \overline{X_1} > \overline{X_2}$$

Same as for the independent t-tests. Effect size also uses the same formula.

One-way ANOVA

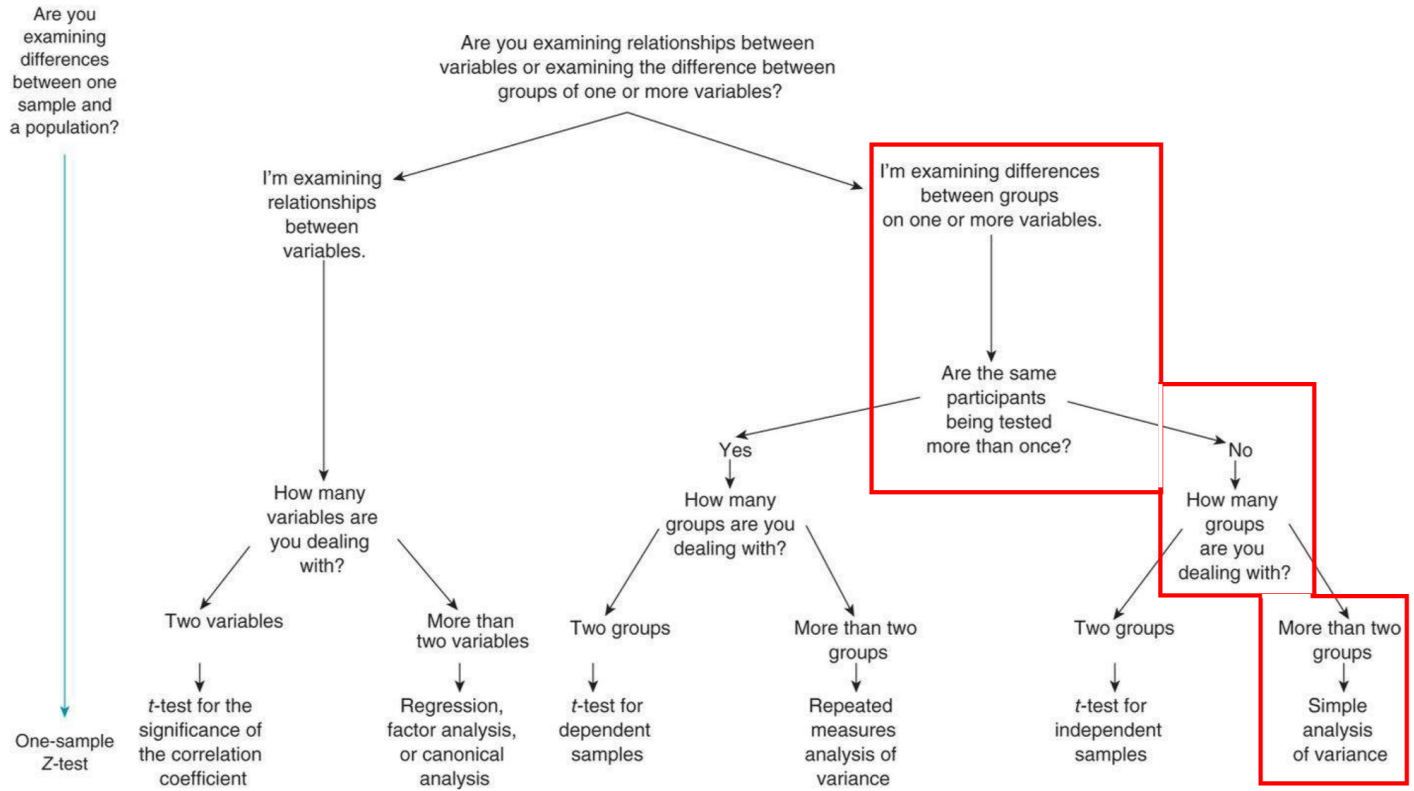
Usage

Comparing the mean of more than two independent groups (Different people/observations in each group). For example, the control and more than one treatment groups.

Assumptions

- Distributions from which the samples are drawn are normally distributed.
- The samples in the groups are independent.
- Homogeneity of variance in case the sample size of the groups are unequal.

One-way ANOVA



One-way ANOVA

If we want to compare many groups, why don't we use t-tests many times?
We loose confidence...a lot!

Example Three groups would need three t-tests, between 1 and 2, 2 and 3, and 1 and 3. If we set a .05 significance level, the overall confidence is:

$$.95^3 = .857$$

Now the real significance level is .14. 14% is not small!

We need a better way...
One-way ANOVA tells if the groups are different, without losing confidence.
However, it doesn't tell us which of the groups are exactly different, only that a difference exists.

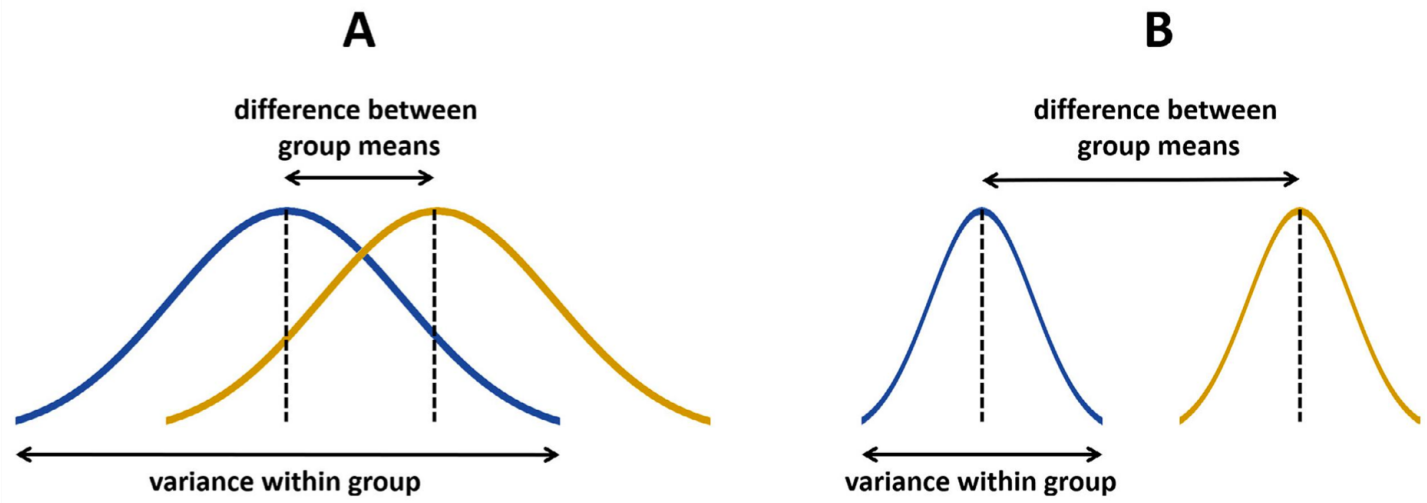
One-way ANOVA

ANOVA uses the **F-statistic**:

$$F = \frac{variance\ between\ groups}{variance\ within\ groups} = \frac{MS_{between}}{MS_{within}} = \frac{model}{error}$$

What do we mean by between and within?

- between - the differences between groups
- within - in each group the differences



Remember: Variance

$$s^2 = \frac{\sum(x - \overline{X})^2}{n - 1}$$

Same formula as the standard deviation but without the square root bracket.

The numerator is called **Sum of squares**:

$$SS = \sum(x - \overline{X})^2$$

The denominator represents the number of **Degrees of Freedom** (Number of observations that are free to vary). The mean of the sample is already used as an estimator of the population's mean. We can only vary n - 1 observations.

One-way ANOVA

Different variances require different sum of squares and different degrees of freedom. With k groups, each with N_k samples:

- Sum of Squares between: Differences between the mean of each group and the combined mean (general G mean).

$$SS_{between} = \sum^{k} N_k * (\overline{X_k} - \overline{X_G})^2$$

  Why do we multiplicate by N_k? Take each point in the group into consideration.

- Sum of Squares within: Difference between each observation and the mean of the group it belongs to:

$$SS_{within} = \sum^{k} \sum^{N_k} (x_{ik} - \overline{X_k})^2$$

- Sum of Squares total: Difference between each observation and the general mean:

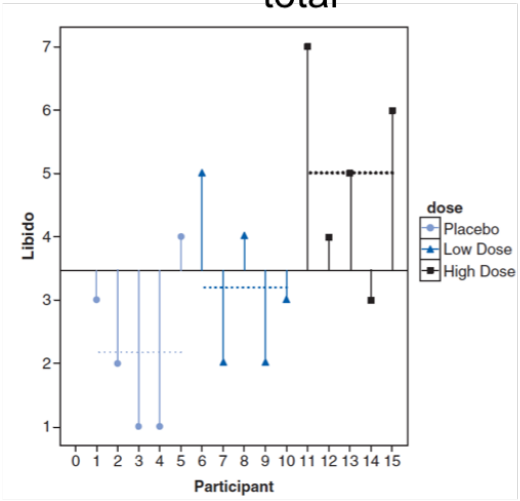
$$SS_{total} = \sum^{N} (x_i - \overline{X_G})^2$$

$$SS_{total} = SS_{between} + SS_{within}$$

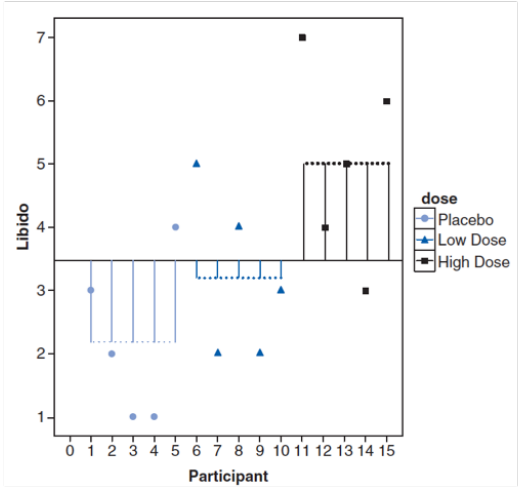
One-way ANOVA

Example Testing if Viagra increases libido with one control group, one with low dosis and one with high dosis.

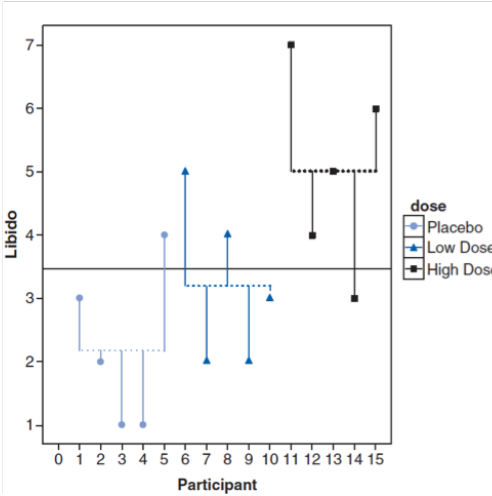
SS_{total}



$SS_{between}$



SS_{within}



One-way ANOVA

Degrees of freedom:

- between - $df_{between} = k - 1$
- within - $df_{within} = N - k$

$$df_{total} = N - 1 = df_{between} + df_{within}$$

Variances:

- between variance- $MS_{between} = \frac{SS_{between}}{df_{between}}$
- within variance- $MS_{within} = \frac{SS_{within}}{df_{within}}$

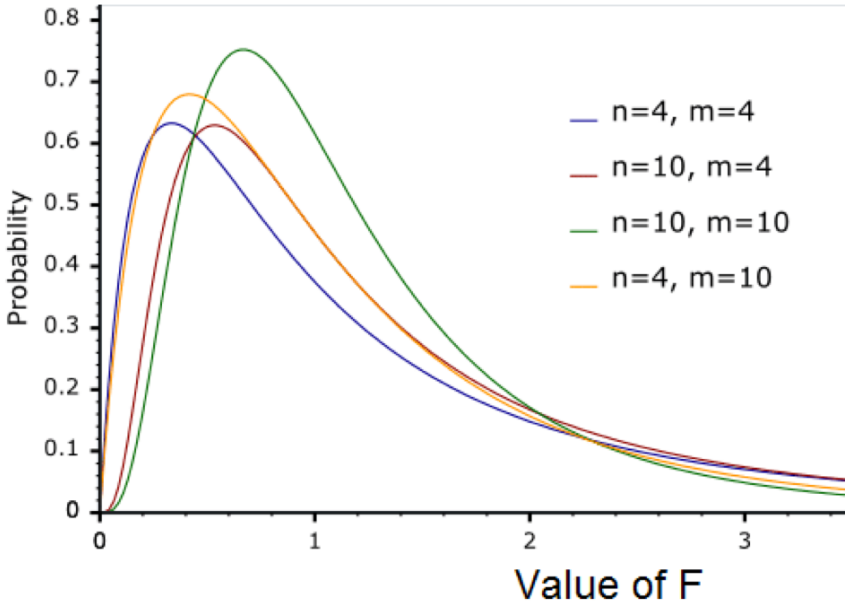
The **F-statistic**:

$$F = \frac{variance\ between\ groups}{variance\ within\ groups} = \frac{MS_{between}}{MS_{within}}$$

The higher the F value, the more statistical significance.

One-way ANOVA

Why F? Say hello to the F distribution, with two parameters, $df_{between}$, df_{within}:

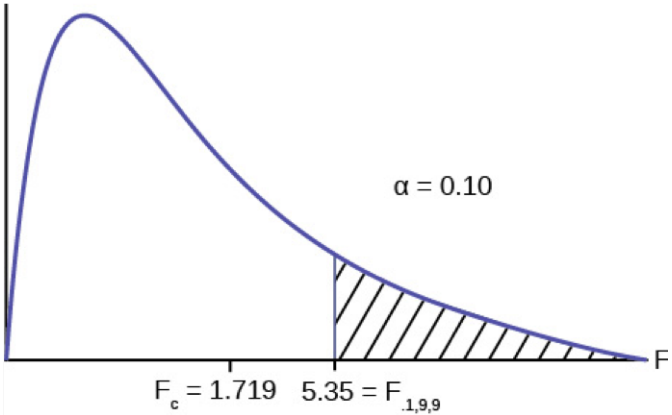


One-way ANOVA

Hypothesis test for one-way ANOVA and three groups:

$$H_0 : \overline{X_1} = \overline{X_2} = \overline{X_3}$$

$$H_1 : \overline{X_1} \neq \overline{X_2} \neq \overline{X_3}$$



One-way ANOVA

Effect size: eta squared

$$\eta^2 = \frac{SS_{between}}{SS_{total}}$$

- small effect .01
- medium effect size .06
- large effect size .14

Post-hoc Tests

ANOVA only tells if there is a difference between the groups, but not which one.

Only if one-way ANOVA is significant, we can figure out exactly which groups were different with **post-hoc tests**.

The most common post-hoc test is the **Bonferroni correction**:
Similar to performing t-tests on each pair of groups, but dividing the significance level α (Type I error) by the number of comparisons C

$$p_{level} = \frac{\alpha}{C}$$

If we conduct 10 tests, we correct the .05 level to .005 as our criterion for significance.

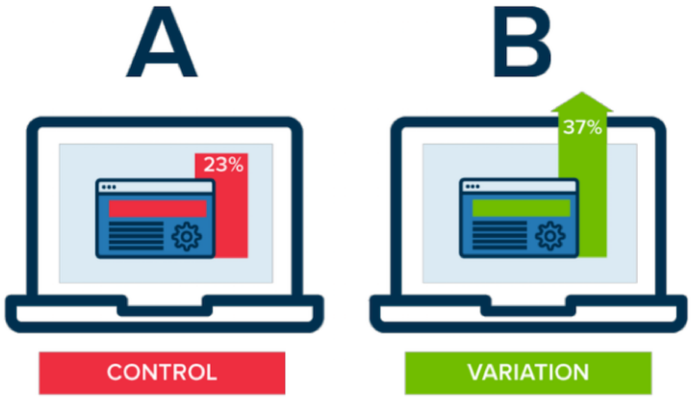
Consequences: the probability of rejecting an effect that does actually exist is increased (Type II error).

An Alternative Approach

Can we do comparison of means tests without knowing anything about distributions?

Alternative way! Only using the data, **NO** assumptions

In data science/ website development, they are part of the famous A/B tests (similar to t-tests)



Permutation Tests

Remember bootstraping? Similar idea, but sampling is done **with** replacement (taking an element out and not putting it back).

A/B Permutation test Algorithm

For two groups A, B:

1. Combine the results from the two groups
2. Shuffle the combined data, then randomly draw (without replacing) a resample of the same size as group A.
3. From the remaining data, randomly draw (without replacing) a resample of the same size as group B. (remaining elements)
4. Calculate the mean now for the resamples, and record the difference of means; this constitutes one permutation iteration.
5. Repeat the previous steps R times (for example 1000 times) to yield a permutation distribution of the mean differences.
6. Decision time: plot the permutation distribution and find where the observed difference is in the plot. **What proportion of time did the resample mean difference exceeded the observed value?** This is the p-value.

Can be extended to more than two groups, like ANOVA: Instead of recording the difference of means, record the variance between group means.

Literature

- Statistics for people who hate statistics: Chapters 10,11,12,13
- Discovering Statistics using R: Chapter 9,10 (more advanced topics)
- Practical Statistics for Data Science: Chapter 2