Parametric Statistics

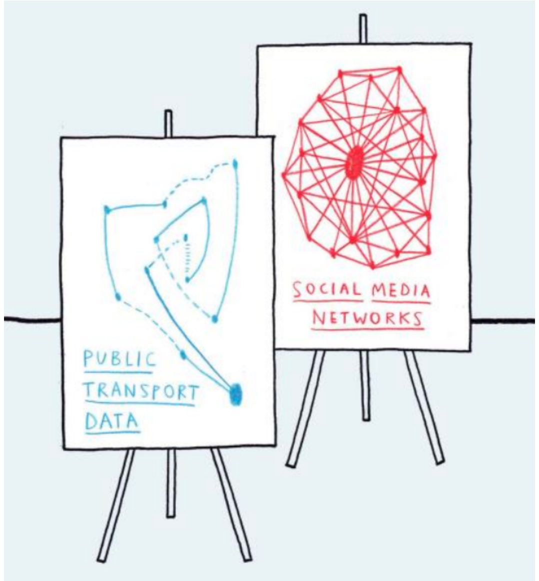
Week 3 - Hypothesis Testing

Juan Carlos Medina Serrano

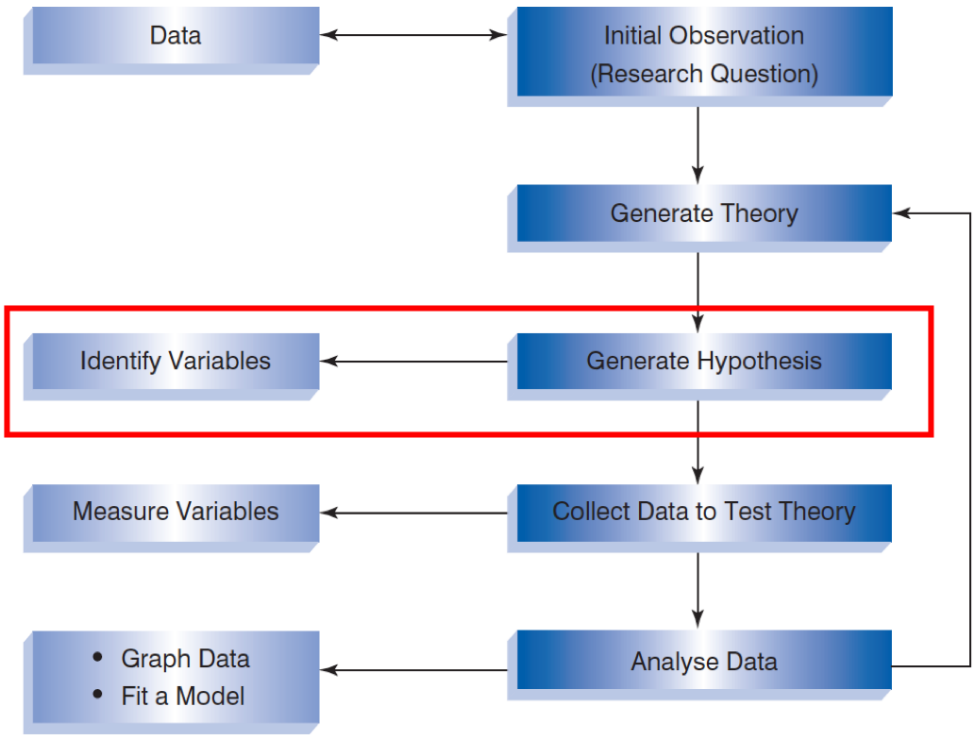
Technische Universität München
Hochschule für Politik
Political Data Science

Munich, 05. November 2019

political
data
science
https://politicaldatascience.blogspot.de



The Research Process



Hypothesis Testing

Steps in a hypothesis test:

1. **State the null hypothesis and the alternative hypothesis.**

The Null Hypothesis

$$H_0$$

- Serves as a benchmark for the research. Starting point which is accepted in the absence of any other information.
- Stated as a lack of relationship or difference between varibles ("There is no relation between X and Y", "There is no difference between X and Y)
- If there are differences, the null hypothesis assumes that they are because of chance.
- The only way to disprove the null hypothesis is to find that it is almost impossible that the difference/relationship was just by chance

Objective Find enough evidence to reject the null hypothesis in favor of an alternative hypothesis.

The Alternative Hypothesis

$$H_1$$

Think of this as your research hypothesis. A definite statement that there is a relationship or difference between variables.

A good hypothesis should:

- be stated in declarative form
- posit a relationship between variables
- reflect a theory on which they are based
- be brief
- be testable

Two types of Alternative Hypothesis

Nondirectional Hypothesis: Difference between groups, but the direction is not specified.

$$H_0 : \theta_1 = \theta_2$$

$$H_1 : \theta_1 \neq \theta_2$$

Needs a two-way hypothesis test.

Directional Hypothesis: Difference between groups and direction is specified.

$$H_0 : \theta_1 = \theta_2$$

$$H_1 : \theta_1 > \theta_2$$

Needs a one-way hypothesis test.

Hypothesis Testing

Steps in a hypothesis test:

1. State the null hypothesis and the alternative hypothesis.
2. **Set the level of significance.**

Significance

Effect: What we are trying to measure in our hypothesis. Is there an effect? (Difference, relationship or other phenomena)

Significant: The effect we observe is due to a systematic influence and not due to chance.

Statistical Significance: Degree of risk you are willing to take that you will reject a null hypothesis when it is actually true. This type of error is called Type I error, referred to as α.

Significance Level: The risk associated with not being 100% confident.
Significance at the 0.05 α-level means that we are 95% confident of not commiting a Type I error
Common significance levels: 0.1, 0.05, 0.01

Type II error: We fail to reject the null hypothesis even though it is false. Referred to as β.

Type of errors

		Action You Take	
		Accept the Null Hypothesis	**Reject the Null Hypothesis**
True nature of the null hypothesis	The null hypothesis is really true.	1 ☺ Bingo, you accepted a null when it is true and there is really no difference between groups.	2 ☹ Oops—you made a Type I error and rejected a null hypothesis even when there really is no difference between groups. Type I errors are also represented by the Greek letter alpha, or α.
	The null hypothesis is really false.	3 ☹ Uh-oh—you made a Type II error and accepted a false null hypothesis. Type II errors are also represented by the Greek letter beta, or β.	4 ☺ Good job, you rejected the null hypothesis when there really are differences between the two groups. This is also called power, or $1 - \beta$.

Ideally, we want to minimize both Type I and Type II errors. But... we can only control Type I, by setting a significance level. Type II depend on sample size since is caused by lack of evidence. As the number of observations increases this error decreases.

Hypothesis Testing

Steps in a hypothesis test:

1. State the null hypothesis and the alternative hypothesis.
2. Set the level of significance.
3. **Select an appropriate test statistic.**

Test Statistics

The next weeks we will learn many test statistics and how to choose between them.
But before...what are these mysterious entities?
Remember: Systematic variation (can be explained by the model) vs. Unsystematic variation (can not be explained by the model)

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

Test Statistics

The next weeks we will learn many test statistics and how to choose between them.
But before...what are these mysterious entities?
Remember: Systematic variation (can be explained by the model) vs. Unsystematic variation (can not be explained by the model)

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

Main idea: We select a model, for which we know its probability distribution. We can calculate the probability of obtaining one value. Is this value an extreme value? If its not extreme then its possible that the observations is only caused by pure chance (null hypothesis).

As a test statistic gets bigger... the probability of occurring is smaller

Hypothesis Testing

Steps in a hypothesis test:

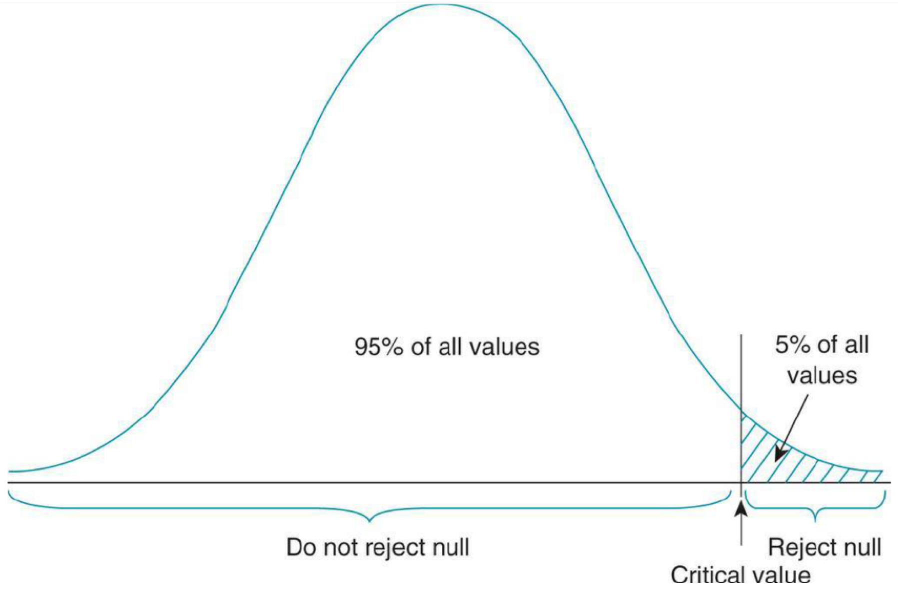
1. State the null hypothesis and the alternative hypothesis.
2. Set the level of significance.
3. Select an appropriate test statistic.
4. **Calculate a critical value and the test statistic value.**

Critical and Test Statistic Values

Calculate them using the selected model / probability distribution.

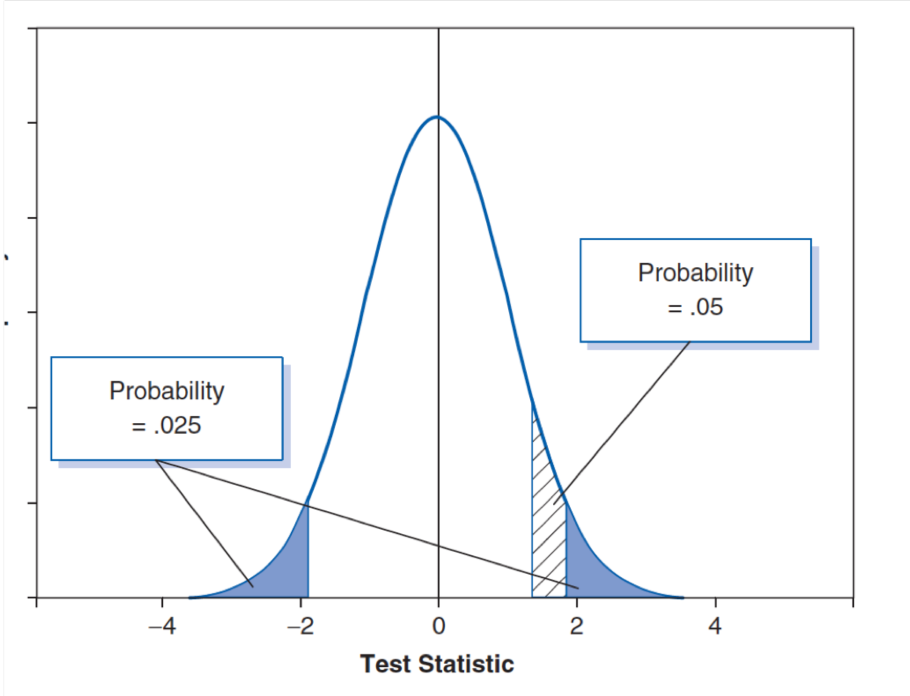
- Critical value depends on the significance level .
- Test statistic value depends on the data.

Example The model is a normal distribution. Significance level 0.05. One-way hypothesis:



Critical and Test Statistic Values

Example The model is a t-distribution. Significance level 0.05. Two-way hypothesis (blue background), one-way hypothesis (stripes):



Hypothesis Testing

Steps in a hypothesis test:

1. State the null hypothesis and the alternative hypothesis.
2. Set the level of significance.
3. Select an appropriate test statistic.
4. Calculate a critical value and the test statistic value.
5. **Compare the values to reject or not the null hypothesis.**

Rejecting the Null

Decision: If test statistic (absolute) value bigger than critical (absolute) value: reject null hypothesis.

Important Message 1: Rejection does not prove the alternative hypothesis, it only supports it.

Important Message 2: In reality: reject the null is better expressed as

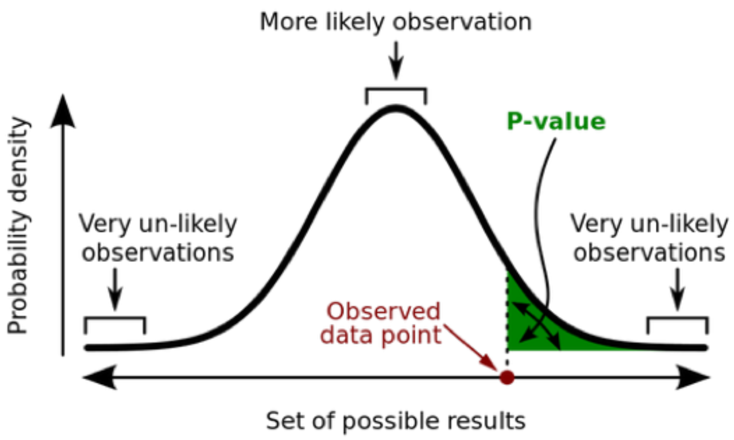
The chances of obtaining the data we collected assuming the null hypothesis are very low!

Important message 3: Not rejecting the null does not imply that the null is true! We only had no evidence to contradict it.

Rejecting the Null: p-values

Alternate Decision: If p-value smaller than the significance level (α-level) reject the null.

p-value: Probability that, given a model, results at least as extreme as those obtained in a sample can occur. (Area under the curve)



Beware of the p-value! *"p-values do not measure the probability that the studied hypothesis is true or the probability that the data were produced by random chance alone"* - American Statistical Assosiation

Effect Size

Hypothesis tests do not give insights on how meaningful or important an effect is. It just tells that there is an effect. Can we do better?

Effect Size: The size of the effect we find between variables.

Effect size serves to tell us if the effect we find is **meaningful** or not.

Note: The effect size is estimated! We only have data on the sample, not on the population.

They can be standardized to be able to compare effect sizes across different studies. There are different standardization procedures for each significance test.

Power

In experiments, we often select at the beginning the minimum size of the effect that we hope to detect in the hypothesis test.

Power: The probability for the hypothesis test to detect a specified effect size with a given sample size. (Over 80% is good)

Remember: Type II error (β) is the probability of failing to detect an effect when it does exist (failing to reject the null hypothesis even though it is false).
Power is the probability of rejecting the null when it is indeed false. This is equal to $1 - \beta$
Power help us determine how many observations/participants are needed to achieve a desired level of power.

We will explore during the course how to calculate effect size and how to use power to select the right number of observations in an experiment.

Effect Size and Power

Example We want to do a survey, with $\alpha = 0.05$ and power (1 - β) = 0.8. How many participants do we need?

We use standardized effect sizes with Pearson's correlation coefficient (WHAAAAAT???), where:

- small effect: r = 0.1
- medium effect: r = 0.3
- large effect: r= 0.5

According to statistical tables (or our friend R), to detect a

- small effect, we need 783 participants.
- medium effect, we need 85 participants.
- large effect, we need 28 participants.

Question Why is a good α less than 0.05 (5%) and a β of 0.2 (20%) good enough?
(This slide is hard stuff...we will study it again in the near future)

Literature

- Discovering Statistics using R : Chapter 2 (Starting in 2.6)
- Statistics for people who hate statistics : Chapter 7, 9