

Politics. Economics. Philosophy, 2018-2019
Data Analysis in the Social Sciences
Lecture 8. Hypotheses testing. (24 January)

Alla Tambovtseva

Hypothesis testing: basic terms

A **null hypothesis** (H_0): a statement about a population (its distribution or parameters).

An **alternative hypothesis** (H_1): a statement about a population that is opposite to the null hypothesis.

A null hypothesis is usually about equality. An alternative hypothesis is usually about inequality. Let us consider some examples.

Example 1. Null hypothesis: a proportion of people who are going to participate in elections is 0.55. Alternative hypothesis: a proportion of people who are going to participate in elections is not 0.55.

$$H_0 : p = 0.55 \text{ and } H_1 : p \neq 0.55$$

Example 2. Null hypothesis: a mean number of children in a family in Russia is 2. Alternative hypothesis: a mean number of children on a family in Russia is greater than 2.

$$H_0 : \mu = 2 \text{ and } H_1 : \mu > 2$$

Example 3. Null hypothesis: a number of children in a family in Russia is not associated with family's income¹. Alternative hypothesis: a number of children in a family in Russia is associated with family's income.

$$H_0 : \text{corr}(x, y) = 0 \text{ and } H_1 : \text{corr}(x, y) \neq 0$$

All the examples above illustrate different types of alternatives. Usually alternative hypotheses are divided into two groups: *two-sided* and *one-sided*.

Two-sided alternatives contain inequality with \neq : we are not interested in the direction of deviation from the value in the null hypothesis (either greater or less). One-sided alternatives contain inequality with a definite direction, with $<$ or with $>$. This time we will work with two-sided alternatives, but then we will discuss one-sided as well.

¹We will talk about correlation later, but as I guess, you discussed it briefly during other courses.

Hypothesis testing: logic

When a null hypothesis is stated, we can start to test it. First, the same thing as with confidence intervals, we should choose an error rate we allow while making conclusions based on our data. This error is called a significance level and is denoted as α .

A **significance level** (α) shows the following. Suppose we chose $\alpha = 0.05$. It means that if we independently repeat the research many times, in 5% of cases we will reject the null hypothesis based on our data provided it is true. In the social sciences we usually take 5% significance level or, sometimes 1% or 10%.

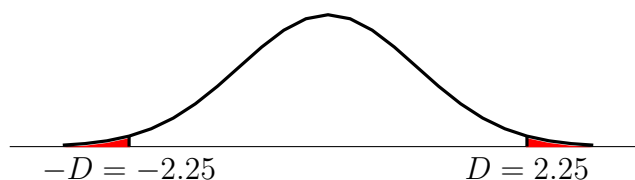
The idea of choosing a significance level is close to the idea of choosing a confidence level. When we take 95% level of confidence, we assume that in 5% of cases confidence intervals will not include a true value of a population parameter. In hypotheses testing we set a significance level 5% and assume that in 5% of cases we will reject the true null hypothesis about a population parameter based on our data.

Now we are ready to test a null hypothesis. So, we need a measure of likelihood of our null hypothesis based on the data given. How to get this measure?

Suppose we want to test the hypothesis that the proportion of people who are going to participate in elections is equal to 0.6. On our sample we got the proportion equal to 0.5. Is it far enough to reject our null hypothesis? And what if a sample proportion equals to 0.65? To understand whether a divergence from the value in the null hypothesis is large, we need a value that evaluates this difference compared to the variability of an estimate. This normalized difference D is calculated differently for different tests, we will not discuss them in detail, but we will need a probability connected with such values.

Suppose we got the normalized difference equal to 2.25. As we do not specify which value, one from H_0 or one calculated based on a sample, is greater, we consider two values, 2.25 and -2.25 ($\hat{p} > p_0$ or $\hat{p} < p_0$). If the null hypothesis holds, the distribution of this difference is standard normal. What values of the difference are less typical than we got on our data? Those that lie in two tails:

the distribution of normalized difference provided H_0 holds



What is the probability of getting such values? It is the area of these two tails. Thus, we have come to the concept called p-value. Formally, **p-value** is the probability that the normalized difference will take values greater or equal to D we got on our data. Less formally, p-value is the likelihood of the null hypothesis that we estimate on the data given. Hence, the higher is the p-value, the more likely our null hypothesis is. If we go back to the graph, p-value is namely the shaded area, area of tails. And the more space

this area takes, the higher is the "viability" of our null hypothesis.

How small should be p-value so we had grounds to reject the null hypothesis? It is logical to compare p-value with the error α we chose. If the likelihood of the null hypothesis is higher than the probability of mistakenly rejecting the null hypothesis, we have no grounds to reject it.

- $p - value > \alpha \Rightarrow H_0$ is not rejected;
- $p - value < \alpha \Rightarrow H_0$ is rejected.

Hypothesis testing: general algorithm and interpretation

1. Formulate a null hypothesis.
2. Formulate an alternative hypothesis (it is important since later we will see how p-value changes when we choose a one-sided alternative instead of a two-sided one).
3. Choose a significance level α (usually $\alpha = 5\%$).
4. Calculate p-value on the data given. Compare it with α .
5. Formulate a statistical conclusion.

Example of such a conclusion: At the 5% significance level we have grounds to reject the null hypothesis on the data given.

6. Formulate a substantial conclusion.

Example of such a conclusion: The percent of people who are going to participate in elections is not (or greater/less than, depends on the alternative) 60%.

NB: We neither accept null hypothesis based on the results obtained nor claim that our null hypothesis is true. We cannot prove it using standard statistical testing. All we can do is to decide whether we have grounds to reject the null hypothesis or not.

Types of errors

| | | |
|------------------|---------------|-------------------|
| | H_0 is true | H_0 is not true |
| reject H_0 | Type I error | OK |
| not reject H_0 | OK | Type II error |

Type I error: the probability to reject H_0 provided it is actually true (the significance level α we choose).

Type II error: the probability not to reject H_0 provided it is not true (sometimes denoted as β).