Utrecht University

CENTRE FOR DIGITAL HUMANITIES

19 September 2023

Basics of Statistics

Session three

training for researchers and teachers in the Humanities

Kirsten Schutter
k.schutter@uu.nl
Centre for Digital Humanities, Utrecht University

1

Hypothesis testing

A hypothesis is a statement about a relationship between variables that can further be tested by a statistical model

2

Hypothesis testing

Hypothesis testing in three steps

1. Formulate hypotheses
2. Compute a test statistic (some number that says something about group differences, for example)
3. Compute the probability of that test statistic occurring under the null hypothesis (p-value)

3

Hypothesis testing

- Null hypothesis (H0)

Typically, a "no effect" hypothesis
There is no relationship between variables, no difference between groups, etc.

- Alternative hypothesis (HA)

Statement that contains your prediction
There is a relationship between variables, difference between groups, etc.

4

Hypothesis testing

In statistical analyses, it is always H0 that is put to the test

The goal of hypothesis testing is rejecting H0
Of course, you never expect that there is nothing going on in the population, that is, that the null hypotheses is true.

So why do we test it?

5

Hypothesis testing

"The reason that we need the null hypothesis is because we cannot prove the alternative hypothesis using statistics, but we can reject the null hypothesis. If our data give us confidence to reject the null hypothesis, then this provides support for our experimental hypothesis."

A. Field (2009) p.27

6

Hypothesis testing

Two types of wrong decisions

- Type I error

You reject H0 when you should not
There is no effect in the population, but your sample data suggests so
The probability of making a Type I error is denoted by alpha (α)

- Type II error

You do not reject H0 when you should
There is an effect in the population, but your sample data do not reveal it
The probability of making a Type II error is denoted by beta (β)

7

Significance

- The p-value is the probability that you reject the null hypothesis while it is true (Type I error)
- The probability level at which we reject the null hypothesis is called the **α-level**
- The conventional value for α = 0.05 (Fisher, 1925)
- For an α-level of 0.05, you allow a 5% chance that you will falsely reject H0
- When we find a p-value < 0.05 we reject the null hypothesis, and we say we have found a statistically significant effect

8

Significance

- In normal English, 'significant' means 'important'
- In statistics 'significant' means 'probably true' or 'not due to chance'
- A research finding may be true without being important / relevant

9

Significance

- Finding a significant effect is not sufficient
- Significance (actually the p-value) is dependent on sample size
- A large enough sample will almost always show a significant effect
- When you find a statistically significant effect, it does not necessarily mean that it's meaningful
- How do you know your significant effect is also relevant?

10

Effect size

- Effect size is a numerical way of expressing the strength or magnitude of an effect
- Amount of difference standardized to amount of dispersion
- The reporting of effect sizes facilitates the interpretation of the substantial, as opposed to the statistical, significance of a research result

11

Effect size

- Effect sizes are calculated on a common scale (standardized), which allows you to compare the effect of:
  - Different variables within a study
  - Same variables across studies
- The reporting of effect sizes facilitates the interpretation of the substantial, as opposed to the statistical, significance of a research result
- Effect sizes complement statistical hypothesis testing
- Effect sizes expresses the relevance of an effect

12

Statistical power

The probability that the statistical test correctly rejects the null hypothesis when the alternative hypothesis is true

13

Statistical power

		Decision based on sample	
		Reject H0	Fail to reject H0
Truth about population	H0 is true	Type I error False positive	Correct decision 1 - α
	HA is true	Correct decision 1 - β	Type II error False negative

Statistical power is inversely related to the probability of making a Type II error (β)
In short, power = 1 – β

14

Relationship between significance, effect size and power

- Significance and power are dependent on sample size
- Increasing sample size increases the chances of finding significant results, but it will not increase effect size
- However, increasing sample size will lead to more accurate estimates (=power), i.e. chances to find subtle effects (small effect sizes) will increase

15

Relationship between significance, effect size and power

- Larger effects are easier to detect than smaller effects
- The smaller the effect you're looking for, in terms of significance, the more data you need to show it
- If your study lacks power, chances are high that you will not be able to detect an effect that is present in the population (Type II error)

16