Institute of Psychiatry, Psychology and Neuroscience

KING'S College LONDON

Dr Silia Vitoratou

Department: Biostatistics and Health Informatics

Module Title: Introduction to Statistics

Session Title: Research hypotheses

Topic title: Confidence and significance (II)

Topic materials:
Silia Vitoratou

Contributions:
Zahra Abdulla

Improvements:
Nick Beckley-Hoelscher
Kim Goldsmith
Sabine Landau

Learning Outcomes

- To understand the idea of stating hypotheses in research
- To understand the null and the alternative hypotheses

Stating Hypotheses

Observations → Hypothesis → Experiment → Outcome

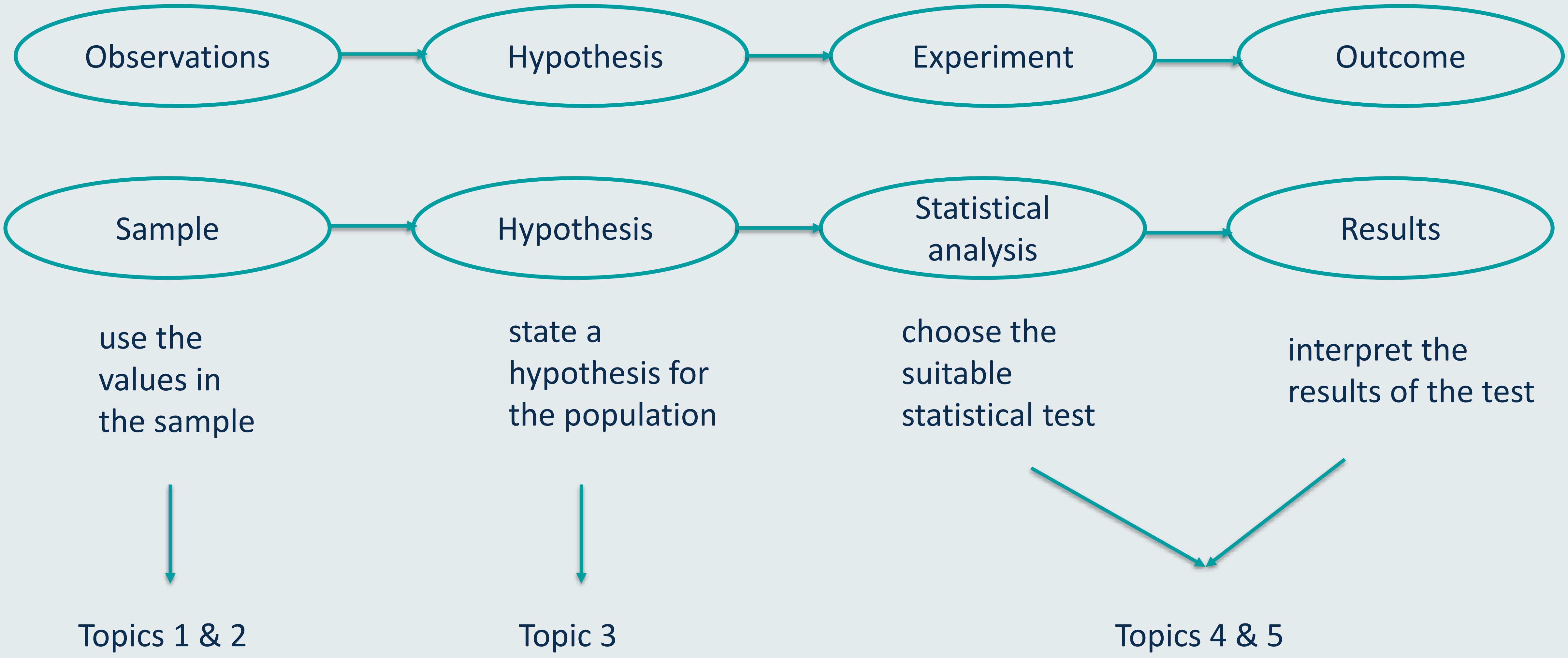
In research, it all begins by making **observations** on our topic of interest.

Based on these observations, we formulate a **hypothesis**, that is, **a testable statement which entails our beliefs about our observations**.

Then, we may **design an experiment** to verify or falsify our hypothesis.

The result of our experiment will be the **outcome** of our research

Stating Hypotheses



Stating Hypotheses

In statistics, we have two hypotheses: the ‘**null**’ and the ‘***alternative***’.
We always start by stating the **null hypothesis H_0.** It is called **null** because essentially, it states that something ‘**equals zero**’, or in other words there is no finding

H_0: The population mean hours of exercise are **zero**

H_a: The population mean hours of exercise are not zero

H_0: The population mean for **males** is **not different than** that of **females**

H_a: The population mean for males is different than that of females

H_0: The **correlation** of height and weight **is zero**

H_a: The correlation of height and weight is not zero

H_0: There is **no association** between gender and height

H_a: There is association between gender and height.

H_0: The **effect** of smoking in lung cancer is **zero**

H_a: The effect of smoking in lung cancer is non zero

H_0: There is **nothing** there.

H_a: There is something there.

Stating Hypotheses

To summarise, you may remember the two hypotheses as:

H_0: The characteristic I am studying **equals** zero.

H_a: The characteristic I am studying is different than zero.

Two sided test: ≠0, anything different that zero

One sided test: either larger (>0) or smaller (<0) than zero.

Stating Hypotheses

Imagine that you are archaeologists. Say you were given a map, where a treasure might be hidden.

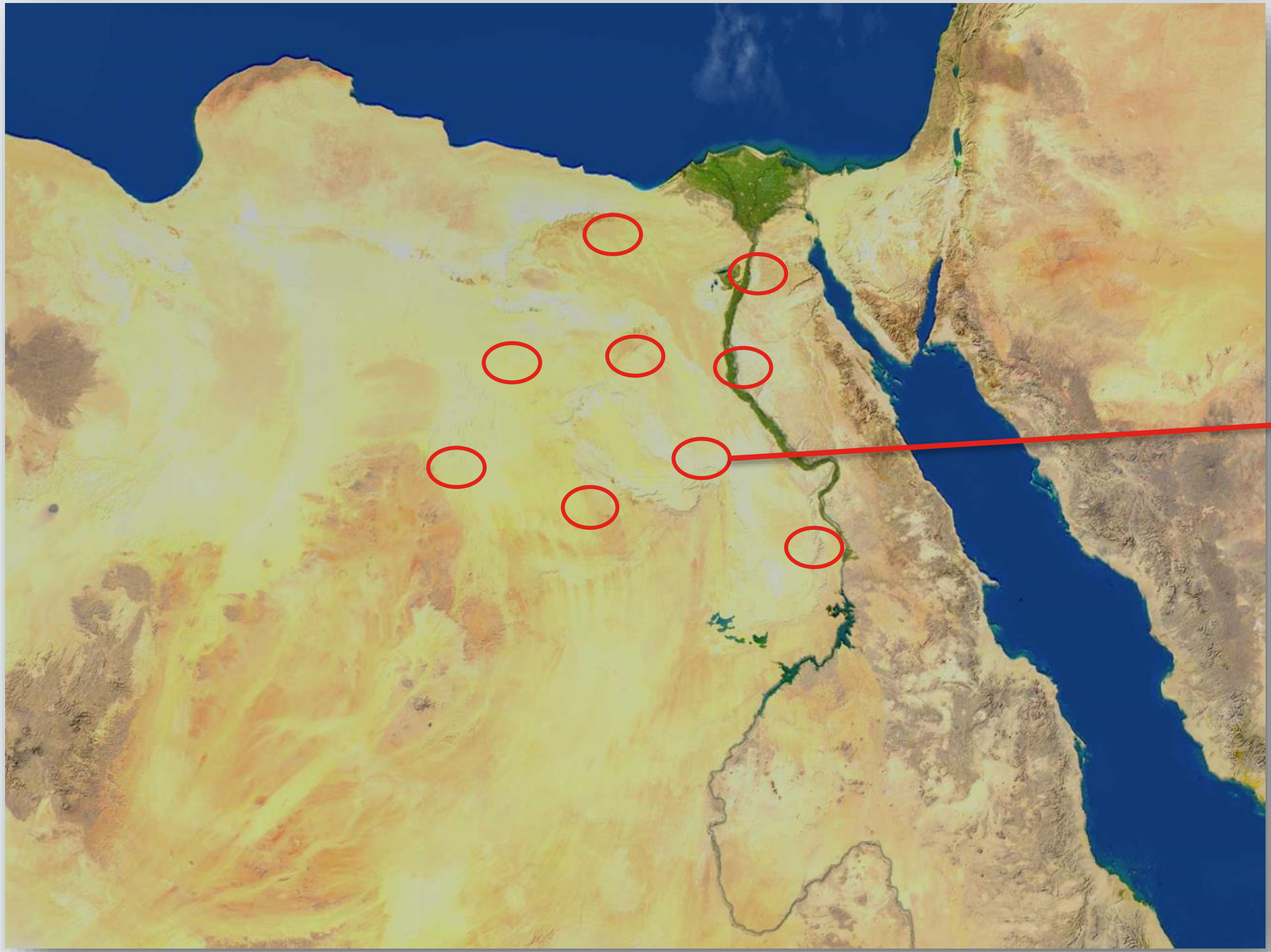
H_0: There is no treasure

H_a: There is a treasure

Stating Hypotheses

You need to start your research, but there are limited resources.

So you can only search (sample) parts of the area.



H_0: There is no treasure

H_a: There is a treasure

Stating Hypotheses

If you find the treasure, you can

reject the null hypothesis & **accept the alternative** hypothesis

~~H_0: There is no treasure~~

H_a: There is a treasure

Stating Hypotheses

If you DO NOT find the treasure, can you

~~**accept the null** hypothesis & reject the alternative hypothesis~~

H_0: **There is no treasure**

H_a: **There is a treasure**

Stating Hypotheses

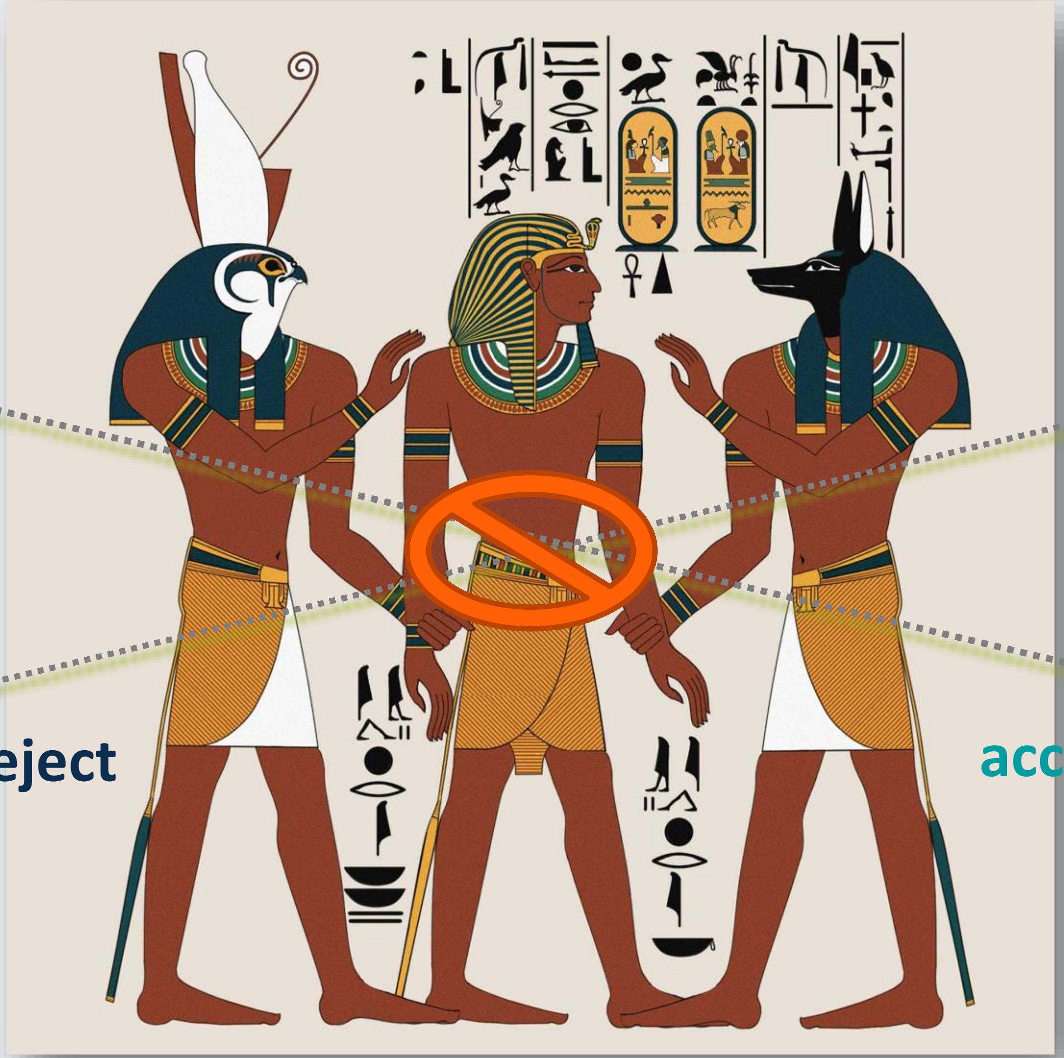
If you do not find the treasure you cannot say there is not treasure. So **you cannot accept the null hypothesis**, you simply: Do not **reject** the null hypothesis

H_0: There is no treasure

H_a: There is a treasure

Stating Hypotheses

H_0: There is no treasure
H_a: There is a treasure



Null

Alternative

reject or not reject

accept or not accept

Hypothesis Testing

What can go wrong with your quest?

H_0: There is no treasure

H_a: There is a treasure

To be under the impression that you found a treasure, but to turn out that the treasure was thin air.

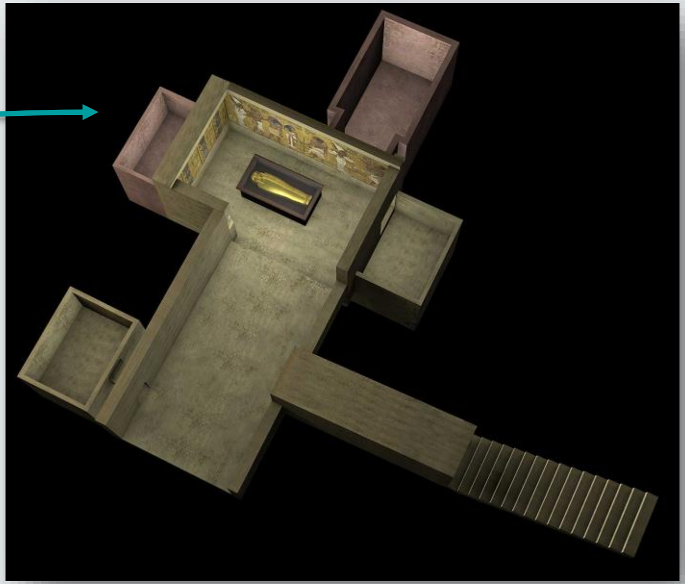
Hypothesis Testing

What can go wrong with your quest?

H_0: There is no treasure

H_a: There is a treasure

To be unable to find the treasure, when in fact there was one (left hidden, undiscovered)



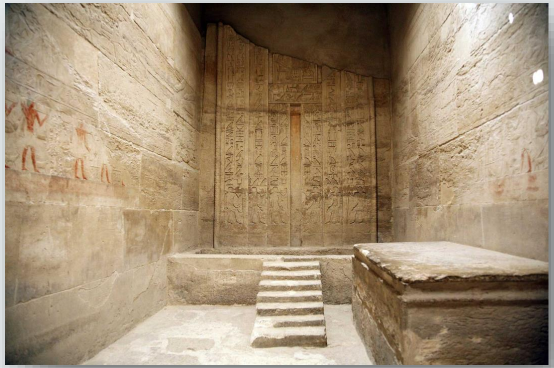
Hypothesis Testing

These are the two types of errors in statistics:

H_0: There is no treasure
H_a: There is a treasure

<u>Type I error: treasure was worthless</u>

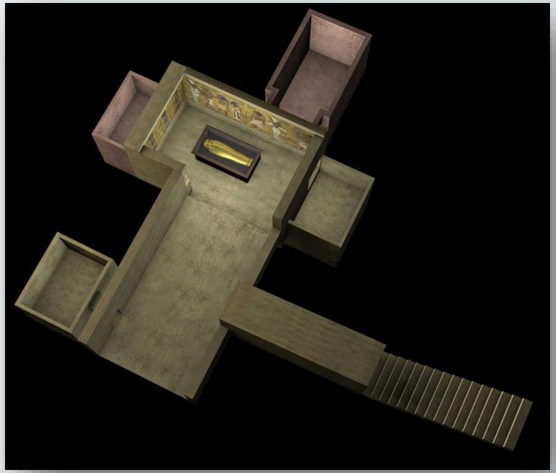
To think you found a treasure, but to turn out to be wrong.



H_0 rejected but was true

<u>Type II error: treasure left hidden</u>

To think that there is no treasure, but in fact there was one.



H_0 not rejected but was false

Hypothesis Testing

H_0: There is no difference **H_a: There is a difference**

		Sample	
		H_0 rejected	H_0 not rejected
Population	H_0 was true	Type I error **(with probability α)**	Correct inference **(1-α)**
	H_0 was false	Correct inference **(1-β)**	Type II error **(with probability β)**

Type I error and significance level α:
By convention the probability of Type I error (α) is set to 0.05, that is: if I say that there is a difference in the population I am at least 95% confident that indeed there is a difference or I allow myself a 5% change to be wrong if I reject the null.

1-Type II error or Power (1-β):
By convention set to 0.80, that is: if there was a difference in the population I am at least 80% confident that I would be able to reveal it.

Hypothesis Testing

What affects our chances to make those errors?

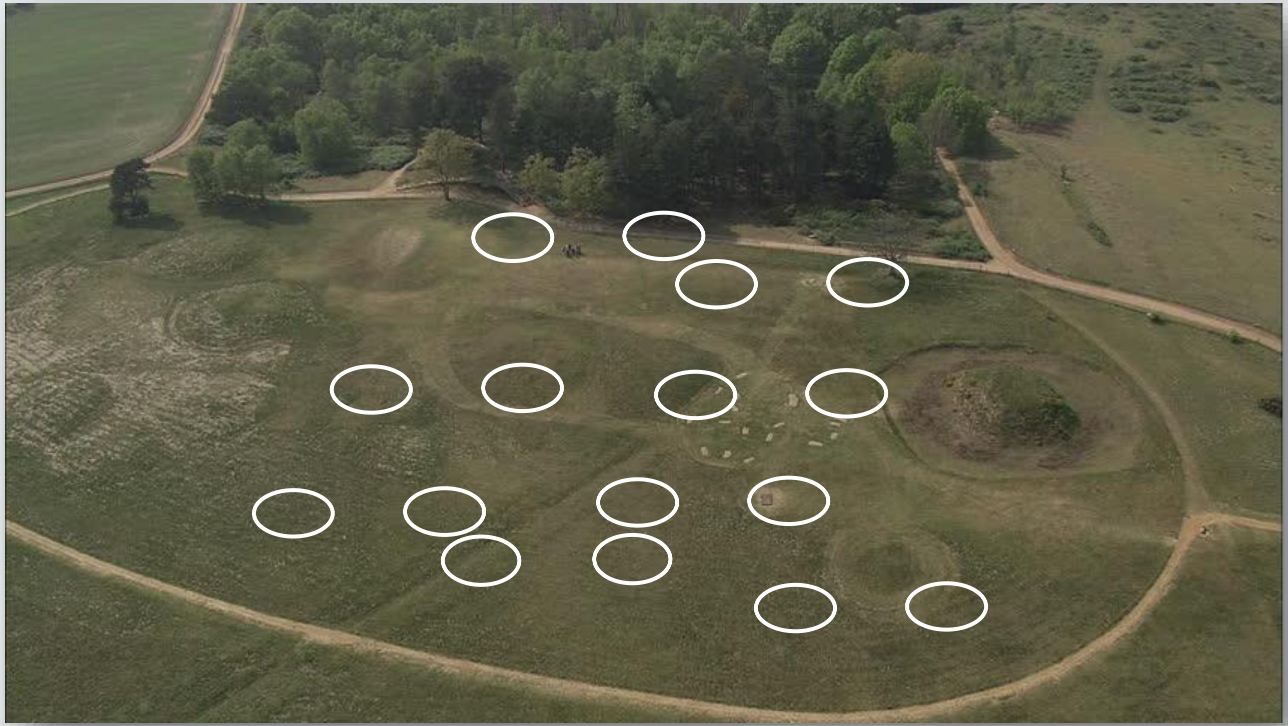
All other things being equal, is your power to find the treasure the same if:

H_0: There is no treasure

H_a: There is a treasure

a) You search 8 areas

b) You search 13 areas



Anglo-Saxon ship burial at Sutton Hoo, Suffolk, one of the most important discoveries in British archaeology

Hypothesis Testing

What affects our chances to make those errors?

All other things being equal, is your power to find the treasure the same if:

H_0: There is no treasure

H_a: There is a treasure

a) the treasure is a buried ship

The imprint of the ship found by Basil Brown and Edith Pretty. Photograph by Charles W Phillips (British museum blog).

b) the treasure is a tiny belt buckle?

Peggy Piggott excavates the gold buckle. Photograph by John Brailsford.

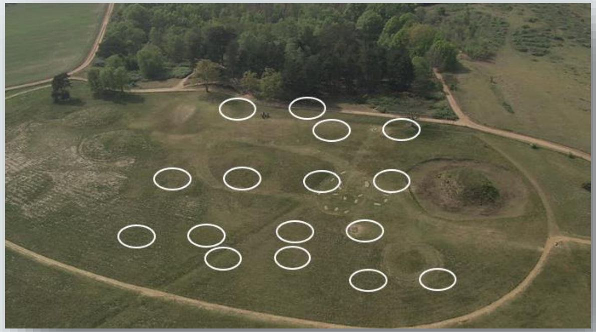
Hypothesis Testing

The larger the sample size, the more able I am to trace the difference.

H_0: **There is no treasure**

H_a: **There is a treasure**

vs



The larger the difference, the more able I am to trace the difference.

vs

If the null hypothesis is false (there is a treasure), my power to correctly reject it (1-β), is larger if the treasure is big or if I sample more areas.

Hypothesis Testing: Power Analysis

How do I know the sufficient sample size for a research? We do so using power analysis.

		Sample	
		H_0 rejected	H_0 not rejected
Population	H_0 was true	Type I error **α**	Correct inference **1-α**
	H_0 was false	Correct inference **1-β**	Type II error **β**

Power

a=0.05

(probability of error allowed)

Sample size n

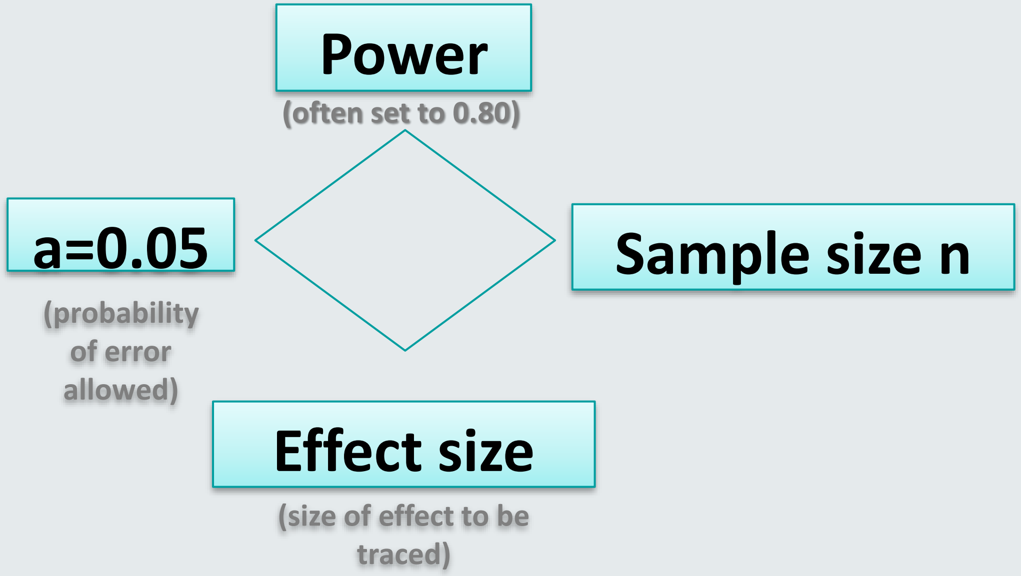
Effect size

(size of effect to be traced)

Closed system of four: when you know 3 of these, you can compute the 4th .

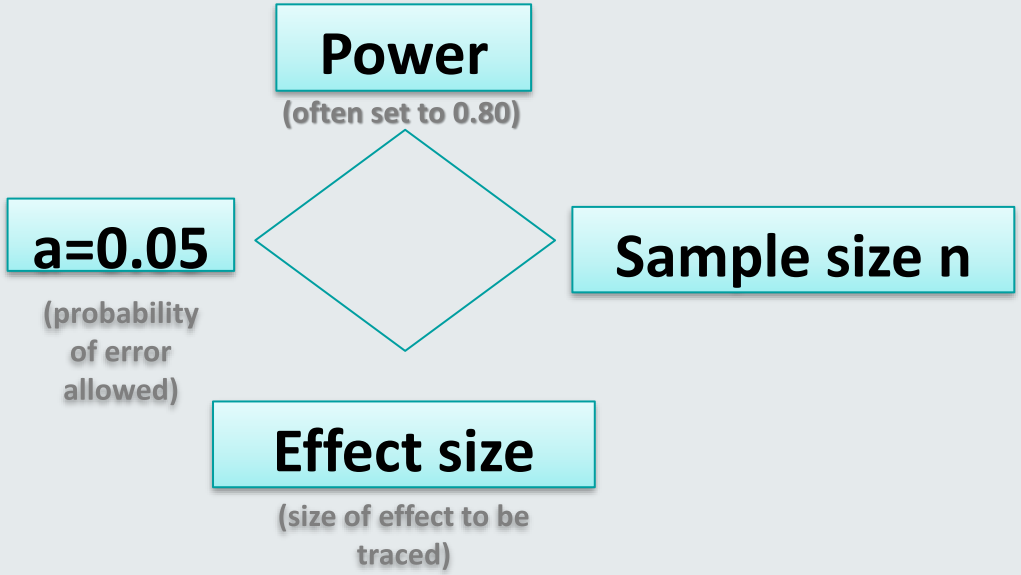
Hypothesis Testing: Power Analysis

A-Priori Power Analysis: before conducting our experiment we compute the sample size needed to have at least 80% power to detect a difference, with error margin of 0.05. We estimate the effect size based on previous knowledge.



Hypothesis Testing: Power Analysis

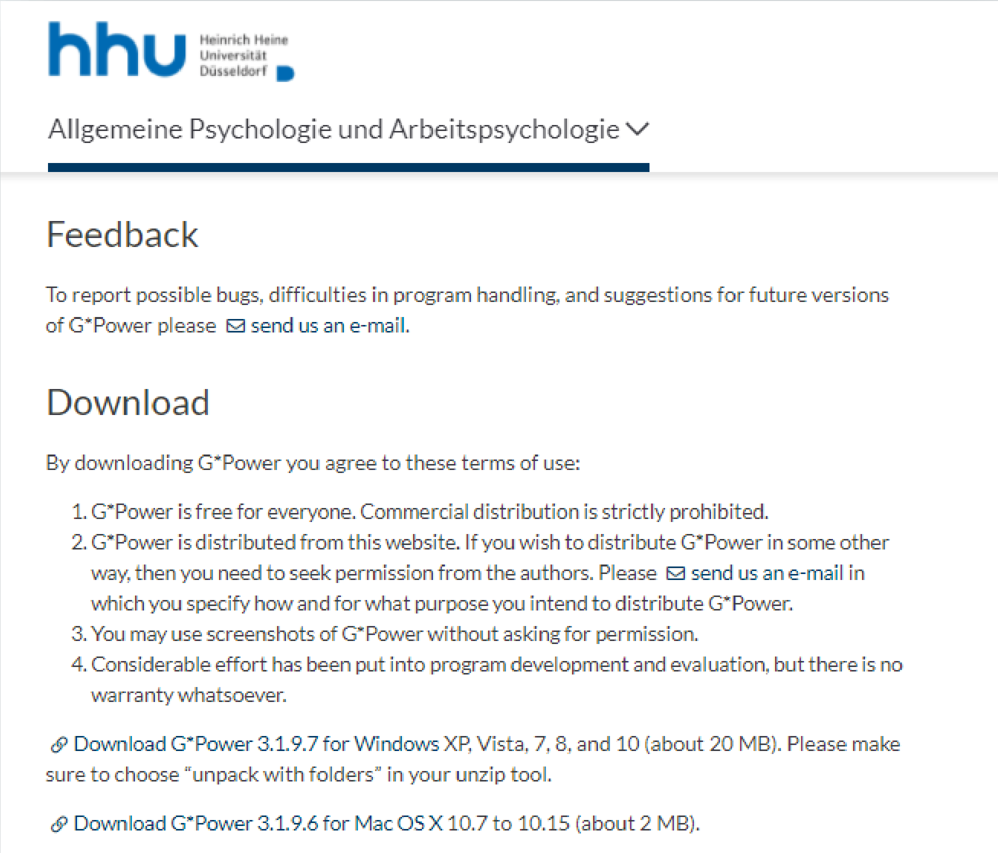
A-Posterior Power Analysis: after conducting our experiment we can compute the power we actually had to detect a difference, with error margin of 0.05, and the sample size and effect size we actually had.



Hypothesis Testing: Power Analysis

Software

Several software exist for the computation of Power. A standalone, freeware software is currently the G*Power, available from the University of Dusseldorf.


hhu Heinrich Heine Universität Düsseldorf

Allgemeine Psychologie und Arbeitspsychologie

Feedback

To report possible bugs, difficulties in program handling, and suggestions for future versions of G*Power please send us an e-mail.

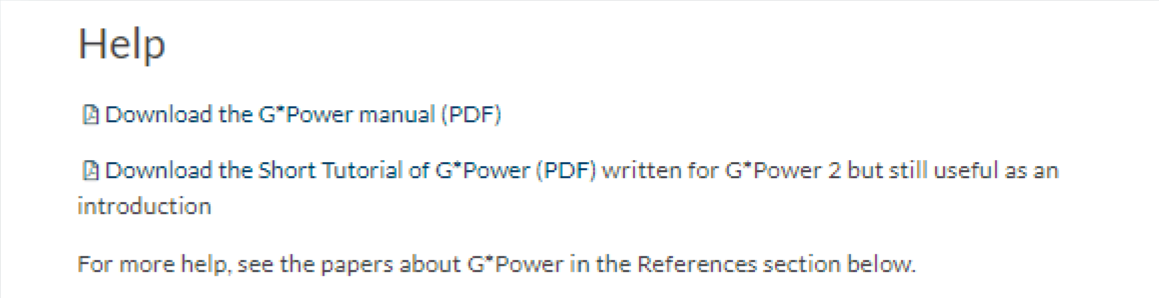
Download

By downloading G*Power you agree to these terms of use:

1. G*Power is free for everyone. Commercial distribution is strictly prohibited.
2. G*Power is distributed from this website. If you wish to distribute G*Power in some other way, then you need to seek permission from the authors. Please send us an e-mail in which you specify how and for what purpose you intend to distribute G*Power.
3. You may use screenshots of G*Power without asking for permission.
4. Considerable effort has been put into program development and evaluation, but there is no warranty whatsoever.

Download G*Power 3.1.9.7 for Windows XP, Vista, 7, 8, and 10 (about 20 MB). Please make sure to choose “unpack with folders” in your unzip tool.

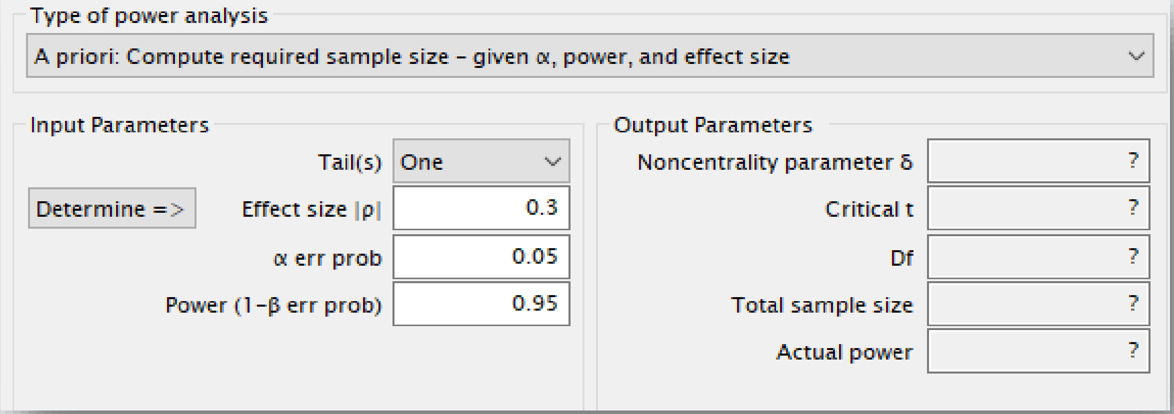
Download G*Power 3.1.9.6 for Mac OS X 10.7 to 10.15 (about 2 MB).


Help

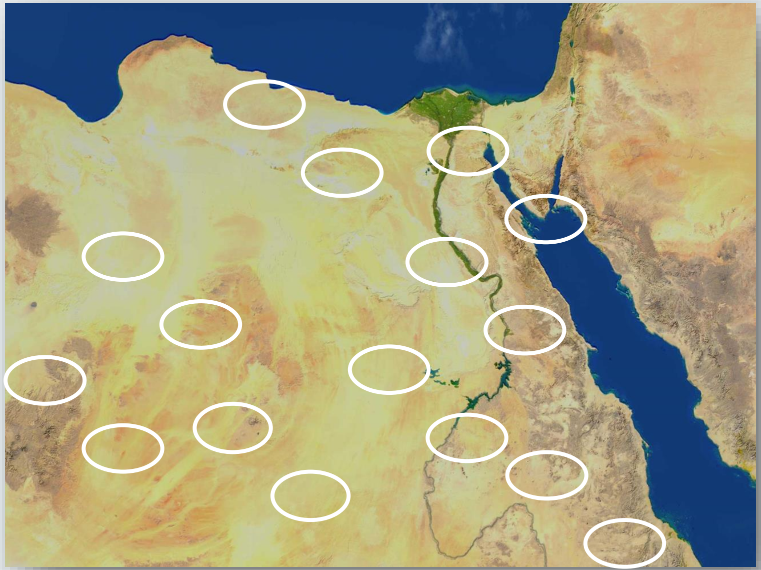
Download the G*Power manual (PDF)

Download the Short Tutorial of G*Power (PDF) written for G*Power 2 but still useful as an introduction

For more help, see the papers about G*Power in the References section below.

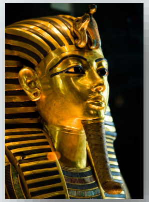


Summarise



H_0: There is no treasure

H_a: There is a treasure



		Sample	
		H_0 rejected	H_0 not rejected
Population	H_0 was true	Type I error **α**	Correct inference **1-α**
	H_0 was false	Correct inference **1-β**	Type II error **β**

Knowledge Check

Please write the null and the alternative hypothesis for the research scenarios below

a) A researcher wants to test if students spend time to watch the independent learning assignment videos.

H_0: The mean hours the students spend watching the videos equal **zero**.
H_a: The mean hours the students spend watching the videos are different than zero.
H_a: The mean hours the students spend is larger than zero (one sided).

b) A researcher wants to test if the proportion of women with PhDs in 2021 has **changed** compared to 2020.

H_0: The proportion of women with PhD in 2020 is equal to the proportion of women with PhD in 2021. The difference between the two proportions is zero.
H_a: The difference between the two proportions is different than zero (two sided).

c) A researcher wants to test if men consume **more** alcohol (units per week) than women.

H_0: The average units of alcohol that men consume is the same as the average units of alcohol than women consume. The difference between the two averages is zero.
H_a: The average units men consume is larger than the average units women consume (one sided).

Knowledge Check Solutions

A research team tested a drug for anxiety against placebo. At the end of the study, they observed that the people who took the drug had the same levels of anxiety as those who used placebo. What should they conclude?

The drug works equally well with the placebo

There is no evidence to suggest that the drug works better than the placebo

Because we cannot accept the null, we simply cannot reject it!

H_0: The drug and the placebo work equally well (zero difference)

H_a: The drug works better than the placebo (there is a difference)

Reflection

Imagine being judges in the court of law ruling about a crime.

What is the null and what is the alternative hypothesis?

Reference List

Kevin R. Murphy, Brett Myors, Allen H. Wolach. Statistical Power Analysis: A Simple and General Model for Traditional and Modern Hypothesis Tests. Routledge, 2009

Thank you

Please contact your module leader or the course lecturer of your programme, or visit the module's forum for any questions you may have.

If you have comments on the materials (spotted typos or missing points) please contact Dr Vitoratou:

Silia Vitoratou, PhD
Psychometrics & Measurement Lab,
Department of Biostatistics and Health Informatics
IoPPN, King's College London, SE5 8AF, London, UK
silia.vitoratou@kcl.ac.uk

For any other comments or remarks on the module structure, please contact one of the three module leaders of the Biostatistics and Health Informatics department:

Zahra Abdula: **zahra.abdulla@kcl.ac.uk**

Raquel Iniesta: **raquel.iniesta@kcl.ac.uk**

Silia Vitoratou: **silia.vitoratou@kcl.ac.uk**