Funded by
the Horizon 2020
Framework Programme
of the European Union

The First Summer School Portfolio -Deliverable-

This project has received funding from the European Union's Horizon 2020 research and innovation programme under **Grant Agreement No. 952464**

Deliverable number: D2.2

Due date: 31.10.2021

Nature: Report

Dissemination Level: Public

Work Package: WP2

Lead Beneficiary: West University of Timisoara (UVT or WUT)

Other contributors: UGENT and UNIMIB

Version 1

Florin-Alin Sava
Semnat digital de Florin-Alin Sava
Data: 2021.12.21 14:35:47 +02'00'

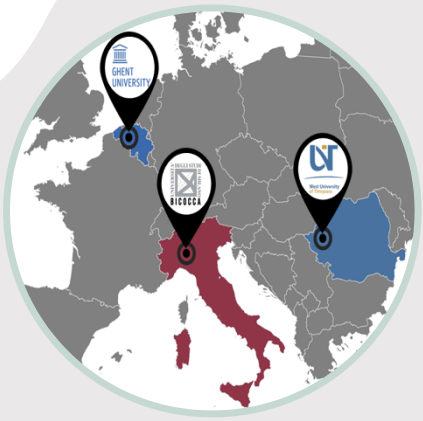
Abbreviations and Acronyms

WUT / UVT = West University of Timisoara

UGENT = Ghent University

UNIMIB = University of Milan-Bicocca

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.



LEARN VUL

ABOUT

The 2021 LEARNVUL Summer School organized as part of the Learning in emotionally vulnerable people project.
The project is in collaboration with University of Milano-Bicocca (UNIMIB) and Ghent University (UGENT) and has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 952464.

CONTACT

WEBSITE:
https://h2020learnvul.uvt.ro/

CONTACT:
Florin Alin Sava (Project Manager)

Address: Bd. Vasile Pârvan 4, Timișoara, 300223, RO

Phone: +40-(0)256-592311

E-mail: florin.sava@e-uvt.ro

The First Summer School Portfolio

The Summer School portfolio includes the following chunks of materials:

I. Before summer school materials

1) Invitation (call) posted on the website
2) The announcement (call) posted by the European Association of Personality Psychology and sent to their members.
3) Selection of the third-party participants (outside UVT, UNIMIB and UGENT)

II. During summer school materials

1. The summer school agenda (timetable)
2. The summer school instructors list
3. The summer school attendance list
4. Selection of pictures from the summer school
5. The summer school feedback form (results)

III. After summer school materials

1. Video lectures (publicly available)
2. Presentations and accompanying data (publicly available)

DISSEMINATION

All presentations and accompanying datasets are available in an Open Repository (Open Science Framework – OSF)

Citation: Sava, F. A. (2021, December 21). LEARNVUL Project Materials. https://doi.org/10.17605/OSF.IO/P48EW

All Video Lectures are available on YouTube at: https://cutt.ly/VUrPVdH

All materials are also available from our website.

West University of Timișoara
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
GHENT UNIVERSITY

LEARNING IN EMOTIONALLY VULNERABLE PEOPLE – Summer School 2021

Opened applications for 4 free Summer School attendees

19 – 24 September 2021, Timișoara, Romania

West University of Timisoara from Romania opened the call for applications as attendees at the 2021 LEARNVUL Summer School organized as part of the *Learning in emotionally vulnerable people* project. The project is in collaboration with University of Milano-Bicocca (UNIMIB) and Ghent University (UGENT) and has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 952464 (https://h2020learnvul.uvt.ro/).

Who cand apply?

The summer school is intended for mostly PhD and master students interested in empirical scientific research, and topics such as personality, cognitive psychology, or social psychology. However, as the program covers a great variety of topics with different levels of complexity, this summer school will also fit well the needs of more experienced researchers and postdoctoral researchers who want to improve their skills and expertise in the presented topics.

When and where?

The Summer School takes place in Romania and is organized by the West University of Timisoara (WUT) in collaboration with Ghent University (UGENT) and University of Milano-Bicocca (UNIMIB).

There will be **5 days** of workshops between **September 19 – September 24**.

All workshops will be held face-to-face in locations that will be announced later.

General information

Lecturers: Researchers from University of Milano-Bicocca (UNIMIB) and Ghent University (UGENT) are invited to provide lectures and hands-on activities during the 5 days of workshops.

Price: The costs related to travelling, accommodation and meals are supported by the organizer.

Free seats: 4

Status: Available

Deadline: 06 Aug 2021

Acquired skills

The Summer School activities will focus on a variety of topics meant to enhance participants' research skills.
Particularly, after attending this event, participants will be able to:
1) Use basic R for data manipulation;
2) Analyze data with basic frequentist and Bayesian statistic tests in R;
3) Analyze experimental data an ANOVA designs in R;
4) Understand multilevel analysis for experimental data in R, and more.

Besides topics regarding statistical data analysis, we will also discuss topics close to the project's scope:
5) Neuroticism as a risk factor for emotional disorders;
6) Evaluative learning;
7) Best practices in conducting an empirical study.
A complete version of the summer school topics and its related schedule is available here.

After the school, all materials will be made available online on the OSF repository.

Application

The applications are opened for all researchers (preferably early-stage researchers) interested in how personality traits (with a particular focus on Neuroticism) shape the evaluative learning processes. Researchers focusing on a broader topic – personality and information processing are also welcomed, particularly if they focus on information processing that make people predisposed to psychopathology (e.g. cognitive biases).
Requirements: The application will consist of: (i) Curriculum Vitae; and (ii) Motivation letter.
Contact: All applications will be sent via email to florin.sava@e-uvt.ro until Aug 6, 2021, with [Summer School Application] as subject.
The confirmation response will be provided within 3 working days.

Universitatea de Vest din Timișoara

Florin Sava <florin.sava@e-uvt.ro>

EAPP Member Mail - Announcements

1 mesaj

European Association of Personality Psychology <admin@membership.eapp.org> 23 iulie 2021, 18:23
Răspunde la: Senka Radovic <shadow.sombra.senka@gmail.com>
Către: Florin Sava <florin.sava@e-uvt.ro>

eapp European Association of Personality Psychology

Dear EAPP Members and Affiliates,

In this week's edition of the EAPP members' news, we would like to inform you about the following:

The first Summer School on Learning and Neuroticism

There are four available seats (full reimbursement for traveling, accommodation, and meals) for participants who will attend this summer school held in Timisoara, Romania (Sep 20 – Sep 24). The event is supported through an H2020 Twinning grant – LEARNVUL grant ID 952464 from the European Commission.

The topic of this summer school is mainly focused on technical issues using R, but conceptual aspects that make people scoring high on Neuroticism at risk to develop emotional disorders are also covered. Instructors will be provided by experts from Ghent University (Jan De Houwer and collaborators) and from the University of Milan-Bicocca (Marco Perugini and collaborators).

Further details regarding the application and the summer school schedule are provided by clicking here.

Kind regards,

The EAPP Executive Committee

Please be aware of possible criminal activities on the internet. EAPP will never request payments other than membership fees which are requested through the membership system (Wildapricot) by means of an invoice (always received by the address admin@membership.eapp.org). For more spoofing prevention measures, go to https://eapp.org/membership/spoofing-and-phishing-prevention/
If you have any doubt on a received email, do not hesitate to contact the Assistant Manager (senka.radovic@eapp.org) or the Secretary (secretary@eapp.org).

You receive this email because you are a member of the European Association for Personality Psychology (EAPP). Click below to unsubscribe.

Unsubscribe

Universitatea de Vest din Timișoara

Florin Sava <florin.sava@e-uvt.ro>

Application for LEARNVUL Summer School

9 mesaje

Kerli Ilves <kerli.ilves@ut.ee> 6 august 2021, 13:29
Către: "florin.sava@e-uvt.ro" <florin.sava@e-uvt.ro>

Hello,

My name is Kerli Ilves and I hereby will forward my CV and motivation letter to apply for a position for 2021 LEARNVUL Summer School in Timisoara, Romania.

Please let me know if you receive this letter or if you have any further questions for me.

Sincerely,
Kerli Ilves

2 atașamente

CV_academic_Ilves.pdf
469K

Ilves_Motivation_Letter_LEARNVUL_2021.pdf
90K

Kerli Ilves <kerli.ilves@ut.ee> 11 august 2021, 10:25
Către: "florin.sava@e-uvt.ro" <florin.sava@e-uvt.ro>

Hello,

I just wanted to check whether You have received all the necessary documents from me for this application process?

Sincerely,
Kerli Ilves

Saatja: Kerli Ilves
Saadetud: reede, 6. august 2021 13:29
Adressaat: florin.sava@e-uvt.ro <florin.sava@e-uvt.ro>
Teema: Application for LEARNVUL Summer School

[Textul citat a fost ascuns]

Florin Sava <florin.sava@e-uvt.ro> 12 august 2021, 18:23
Către: Kerli Ilves <kerli.ilves@ut.ee>

Dear Kerli Ilves,

I want to congratulate you for being selected to participate in the LEARNVUL summer school (which actually takes place on September 20-25, in Timisoara, Romania).
As you already discussed on the phone, please send us the following personal details:
Your full name, date of birth (YYYY/MM/DD) for purchasing the flight tickets (departure to Timisoara on Sunday - Sep 19, return from Timisoara to Tallin on Saturday - Sep 25). As I saw from google flight, there are no direct flights, but there is a single connection flight via Frankfurt or Munich. I hope these routes are ok for you.

The accommodation and meals are also covered, excepting for Sunday evening. I will let you know soon, such details.

Let me know if you have any questions.

Kind regards,
Florin Alin Sava

[Textul citat a fost ascuns]

--
Florin Alin Sava, Professor, Ph.D. in Psychology
Vice-Rector for Research, Development and Innovation
West University of Timisoara
phone: +40 256592311
mobile: +40 722510471
email: florin.sava@e-uvt.ro

Funded by
the Horizon 2020
Framework Programme
of the European Union

2021 LEARNVUL Summer School Schedule

Background information

Video-lectures concerning Basics of R will be prepared and made available online before the start of the summer school. These will include: *Using R-Studio, reading data, importing data, exporting data, examining data sets, transforming wide to long format or vice versa, examining variables, cleaning data, missing values, outliers, subsetting data, computing new variables (scales), labeling, reverse coding, collapsing variables in fewer categories, creating dummy variables, transforming variables, filter cases*. These will exemplify using some basic descriptive statistics and contain also exercises with solutions.

Day 1 (Monday, September 20)

Time	*Topic / Activity*
09.00 - 10.30	Introduction to data analysis and processing with R Part 3 (Part 1 and 2 will be made available *via* video lectures) with Q&A, additional data processing examples, and it will include simple examples for basic analyses (e.g., Pearson r, chi square, t student, basic visualization) **(Giulio Costantini)**
10.30 - 11.00	Coffee Break
11.00 - 12.30	Bayesian statistics and hypothesis testing. Conceptual issues and basics in R **(Daniele Romano)**
12.30 - 14.30	Lunch Break
14.30 - 16.00	Bayesian statistics. Testing competing hypothesizes vs. null hypothesis. The level of support for the null hypothesis **(Daniele Romano)**
16.00 - 16.30	Coffee Break
16.30 - 18.00	GLM 1. Simple ANOVA designs (between, within, mixed design, completely randomized, block design, visualization, with Bayesian equivalent) **(Giulio Costantini and Daniele Romano)**
19.00	Dinner

Day 2 (Tuesday, September 21)

Time	*Topic / Activity*
09.00 - 10.30	GLM 2. Generalizing ANOVA designs (ANCOVA, MANOVA, MANCOVA) **(Giulio Costantini)**
10.30 - 11.00	Coffee Break
11.00 - 12.30	Multilevel analysis for experimental data in R. Part 1 **(Marine Rougier and Jamie Cummins)**
12.30 - 14.30	Lunch Break
14.30 - 16.00	Multilevel analysis for experimental data in R. Part 2 **(Marine Rougier and Jamie Cummins)**
16.00 - 16.30	Coffee Break
16.30 - 18.00	Power analysis **(Marco Perugini and Giulio Costantini)**
19.00	Dinner

Day 3 (Wednesday, September 22 - Parallel sessions in the morning, outdoor activities in the afternoon)

Time	*Topic / Activity*
09.00 - 10.30	Visualizing data with R **(Giulio Costantini)**
10.30 - 11.00	Coffee Break
11.00 - 12.30	Evaluative Conditioning and Evaluative Learning **(Jan De Houwer)**
12.30 - 14.30	Lunch Break
14.30 - 16.00	Neuroticism and Emotional Vulnerability **(Emanuele Preti and Rossella Di Pierro)**
16.00 - 16.30	Coffee Break
16.30 - 17.45	Relating personality and learning (with a focus on Neuroticism) **(Florin Sava, Marco Perugini, Erica Casini and Jan De Houwer).**
19.00	Dinner

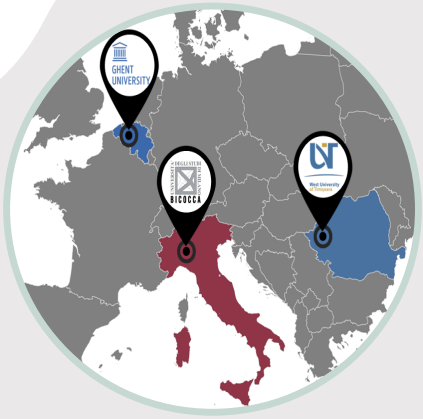
Day 4 (Thursday, September 23)

Time	*Topic / Activity*
09.00 - 10.00	Q & A regarding covered R topics. Exercises (Giulio, Marine, Daniele, Jamie, Marian)
10.30 - 11.00	Coffee Break
11.00 - 12.30	Q & A regarding covered R topics. Exercises. Part 2.
12.30 - 14.30	Lunch Break
14.30 - 18.00	Outdoor activity
19.00	Dinner

Day 5 (Friday, September 24)

Time	*Topic / Activity*
09.00 - 10.30	Data preprocessing from Inquisit and/or other data sources. From raw data to ready to analyze data (e.g., a standard evaluative conditioning study, linked with Inquisit) **(Jamie Cummins)**
10.30 - 11.00	Coffee Break
11.00 - 12.30	Best research practices in the Open Science and (post)reproducibility crisis era **(Marco Perugini)**
12.30 - 14.00	Lunch Break
14.00 - 15.30	**Putting it all together I. A research project from A to Z** A basic outline of a research project from design, implementation, data collection, to manuscript submission. It covers: (i) standards for open science, reproducibility (preregistration of design, data availability and instruction, data management plan); (ii) other aspects than open science (working with supervisors, gathering data tools, presenting research announcements, informed consent form, dealing with GDPR, etc.). (30/35 minutes each team) **(UNIMIB Team: Juliette Richetin and Cristina Zogmaister; Ghent Team: Jan De Houwer, Marine Rougier and Jamie Cummins)**
15.30 - 16.00	Coffee Break
16.00 - 18.00	**Putting it all together II. Practical aspects learned from the teams (exercises and feedback)** Exercises and feedback on the preparation of an outline of all steps for a research project of one's choice, from A to Z **(UNIMIB Team: Juliette Richetin and Cristina Zogmaister; Ghent Team: Jan De Houwer, Marine Rougier and Jamie Cummins)**
18.00 - 19.00	**Final round-up** Final round of questions about anything raised during the summer school, reflections about the summer school, implications for future studies, etc.
19.00	Dinner

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.



LEARN VUL

ABOUT

The 2021 LEARNVUL Summer School organized as part of the Learning in emotionally vulnerable people project.
The project is in collaboration with University of Milano-Bicocca (UNIMIB) and Ghent University (UGENT) and has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 952464.

CONTACT

WEBSITE:
https://h2020learnvul.uvt.ro/

CONTACT:
Florin Alin Sava (Project Manager)

Address: Bd. Vasile Pârvan 4, Timișoara, 300223, RO

Phone: +40-(0)256-592311

E-mail: florin.sava@e-uvt.ro

LEARNVUL SUMMER SCHOOL

This summer school was intended mostly for PhD students interested in empirical scientific research, and topics such as personality, cognitive psychology, learning psychology, and data analysis. However, as the program covered a great variety of topics with different levels of complexity, this summer school also fit well the needs of more experienced researchers who wanted to perfect their skills and expertise in the presented topics.

The Summer School took place onsite, in Timișoara, Romania and was organized by the West University of Timișoara in collaboration with Ghent University (UGENT) and University of Milano-Bicocca (UNIMIB).
Between September 20 – September 24, there were 5 days of workshops.

LECTURERS

Experienced researchers from the two leading institutions from the project, University of Milano-Bicocca (UNIMIB) and Ghent University (UGENT), were invited to provide lectures and hands-on activities during the 5 days of workshops.

University of Milan Bicocca	Marco Perugini, Professor Cristina Zogmaister, Associate Professor Emanuele Preti, Associate Professor Giulio Costantini, Assistant Professor Juliette Richetin, Assistant Professor Rossella Di Pierro, Assistant Professor Daniele Romano, Assistant Professor Erica Casini, Post--doc
Ghent University	Jan De Houwer, Professor Marine Rougier, Post--doc Jamie Cummins, Post--doc
West University of Timisoara	Florin Alin Sava, Professor

West University of Timișoara
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
GHENT UNIVERSITY

Funded by
the Horizon 2020
Framework Programme
of the European Union

2021 LEARNVUL Summer School Schedule

Attendance List

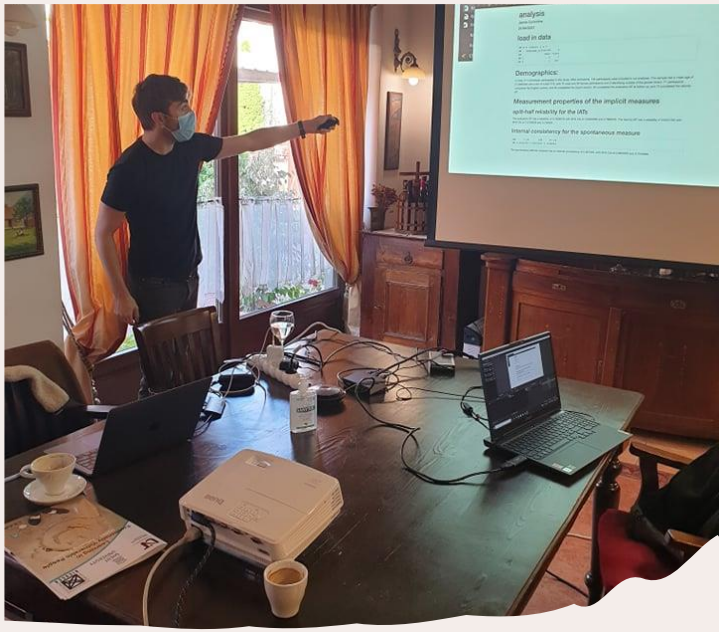
University	*Name*
University of Milan Bicocca	Marco Perugini, Professor (marco.perugini@unimib.it)
	Cristina Zogmaister, Associate Professor (cristina.zogmaister@unimib.it)
	Emanuele Preti, Associate Professor (emanuele.preti@unimib.it)
	Giulio Costantini, Assistant Professor (giulio.costantini@unimib.it)
	Juliette Richetin, Assistant Professor (juliette.richetin@unimib.it)
	Rossella Di Pierro, Assistant Professor (rossella.dipierro@unimib.it)
	Daniele Romano, Assistant Professor (daniele.romano@unimib.it)
	Erica Casini, Post-doc (erica.casini@unimib.it)
Ghent University	Jan De Houwer, Professor (jan.dehouwer@ugent.be)
	Marine Rougier, Post-doc (marine.rougier@ugent.be)
	Jamie Cummins, Post-doc (jamie.cummins@ugent.be)
University of Tartu	Kerli Ilves, Lecturer (kerli.ilves@ut.ee)
West University of Timisoara	Florin Alin Sava, Professor (florin.sava@e-uvt.ro)
	Laurențiu Maricuțoiu, Professor (laurentiu.maricutoiu@e-uvt.ro)
	Andrei Rusu, Associate Professor (andrei.rusu@e-uvt.ro)
	Cristina Măroiu, Post-doc (cristina.maroiu@e-uvt.ro)
	Luca Tisu, PhD Student (luca.tisu@e-uvt.ro)
	Cătălina Bunghez, PhD Student (catalina.bunghez@e-uvt.ro)
	Ștefan Marian, PhD Student (stefan.marian@e-uvt.ro)
	Florina Gabriela Huzoaica, PhD Student (florina.huzoaica@e-uvt.ro)
	Luisa Avram, PhD Student (luisa.avram@e-uvt.ro)
	Bianca Popescu, PhD Student (bianca.popescu95@e-uvt.ro)
	Anca Lazăr, PhD Student (anca.lazar89@e-uvt.ro)
	Eusebiu Ștefancu, PhD Student (eusebiu.stefancu98@e-uvt.ro)
	Dana Țurcan, Administrative Staff (dana.turcan@e-uvt.ro)

Photo Documentation

Twinning LEARNVUL Summer School

20 – 24 September 2021

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.



Workshops
(20-24th of September)

These activities were organized in Timișoara, at Casa del Sole.

Dinner at Cramele Recaș Vineyard
(22nd of September)

Outdoor activity - Paintball

(23rd of September)

2021 Twinning LEARNVUL Summer School - Feedback Form

	**Terrible**. Must be improved immediately!	**Poor**. Possible complaint, needs improving.	**Fair**. Not bad enough to merit complaining but ample room for improvement.	**Good**. Improvement can still be made.	**Fantastic**. No need for improvement!
Workshop activities					
The workshops meet my expectations.			10%	63%	27%
The structure of the program was well thought-out.			10%	72%	18%
The workload during the Summer School was appropriate.			10%	72%	18%
The content of the lectures was well illustrated with practical examples.			10%	45%	45%
Difficult tasks/concepts were explained in a comprehensive manner.			10%	63%	27%
The level of instructors' involvement was suitable to the group's interests.				27%	73%
The instructors took the time to answer to all the questions.				19%	81%
The participants had the opportunity to be actively involved in the activities.			18%	54%	28%
Staff and Accommodation					
The staff was well prepared.			10%	45%	45%
The staff was attentive.				27%	73%
The staff was enthusiastic and friendly.			10%	36%	54%
The facility was appropriate for the 2021 Summer School.				27%	73%
The facility was clean and well maintained.				45%	55%
The catering services were well-suited regarding the food and punctuality.				27%	73%

What lecture(s) did you find to be most useful for you at the 2021 Summer School?

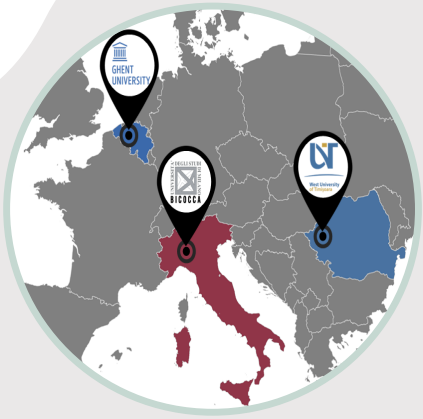
The lectures that participants considered to be the most useful were the workshops on Multilevel analysis for experimental data in R, Best research practices in the Open Science and (post)reproducibility crisis era, and Data processing, followed by the other lectures.

What lecture(s) did you find to be least useful for you at the 2021 Summer School?

Most of the lectures were considered useful, and only 2 participants pointed out that the lectures on statistics and R programming were a bit overwhelming, and the lectures on Relating personality and learning (with a focus on Neuroticism) were not very clear, because the results were still too much preliminary.

The overall feedback on the workshops was a positive one.

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.



LEARN VUL

ABOUT

The 2021 LEARNVUL Summer School organized as part of the Learning in emotionally vulnerable people project.
The project is in collaboration with University of Milano-Bicocca (UNIMIB) and Ghent University (UGENT) and has received funding from the European Union's Horizon 2020 Research and Innovation Programme under the Grant Agreement No. 952464.

CONTACT

WEBSITE:
https://h2020learnvul.uvt.ro/

CONTACT:
Florin Alin Sava (Project Manager)

Address: Bd. Vasile Pârvan 4, Timișoara, 300223, RO

Phone: +40-(0)256-592311

E-mail: florin.sava@e-uvt.ro

WORKSHOPS

The Summer School activities focused on a variety of topics meant to enhance participants research skills.
Particularly, after attending this school, participants will be able to:

1) Use basic R for data manipulation
2) Analyze data with basic frequentist and Bayesian statistic tests in R
3) Analyze experimental data an ANOVA design in R
4) Understand multilevel analysis for experimental data in R, and more.

Besides topics regarding statistical data analysis, there were also discussed topics close to the projects scope:

5) Neuroticism
6) Evaluative conditioning
7) Best practices in conducting an empirical study

DISSEMINATION

The video for each lecture presented during the Summer School, can be found on the 2021 LEARNVUL Summer School YouTube playlist, alongside a specific description of each activity.

Introduction to data analysis and processing with R: part I, part II, part III, part IV	Giulio Costantini
Bayesian statistics and hypothesis testing: part I, part II, and part III	Daniele Romano
General Linear models: part I and part II	Giulio Costantini and Daniele Romano
Multilevel analysis for experimental data in R: part I and part II	Marine Rougier and Jamie Cummins
Power analysis: part I and part II	Marco Perugini and Giulio Costantini
Evaluative Conditioning and Evaluative Learning	Jan De Houwer
Neuroticism and Emotional Vulnerability	Emanuele Preti and Rossella Di Pierro
Relating personality and learning (with a focus on Neuroticism)	Florin Sava, Marco Perugini and Jan De Houwer
Data preprocessing	Jamie Cummins
Best research practices in the Open Science and (post)reproducibility crisis era	Marco Perugini
Putting it all together. A research project from A to Z	Juliette Richetin and Cristina Zogmaister

All above presentations are included in the following pages (up to the end of this document)

West University of Timișoara
UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA
GHENT UNIVERSITY

2021 Twinning LEARNVUL Summer School

Workshop	Lecturers
Introduction to data analysis and processing with R: part I, part II, part III, part IV	Giulio Costantini
Bayesian statistics and hypothesis testing: part I, part II, and part III	Daniele Romano
General Linear models: part I and part II	Giulio Costantini and Daniele Romano
Multilevel analysis for experimental data in R: part I and part II	Marine Rougier and Jamie Cummins
Power analysis: part I and part II	Marco Perugini and Giulio Costantini
Evaluative Conditioning and Evaluative Learning	Jan De Houwer
Neuroticism and Emotional Vulnerability	Emanuele Preti and Rossella Di Pierro
Relating personality and learning (with a focus on Neuroticism)	Florin Sava, Marco Perugini and Jan De Houwer
Data preprocessing	Jamie Cummins
Best research practices in the Open Science and (post)reproducibility crisis era	Marco Perugini
Putting it all together. A research project from A to Z	Juliette Richetin and Cristina Zogmaister

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

Introduction to data analysis and processing with R (part I-III)

Giulio Costantini,

Università degli Studi di Milano-Bicocca

giulio.costantini@unimib.it

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

What is R?

What are R / Rstudio

- Open source version of S+
- A huge community of statisticians (and others, like us), writing software for statistical analyses.
- What is Rstudio: integrated development environment for R. Using Rstudio helps a lot
  - Great code editor
  - Import dataset (very useful at the beginning)
  - Some fundamental utilities, e.g., makes easier to create packages, integration with GitHub, R markdown for creating documents etc. (Most of these things will become clear when you will start mastering the R environment).

How to install R

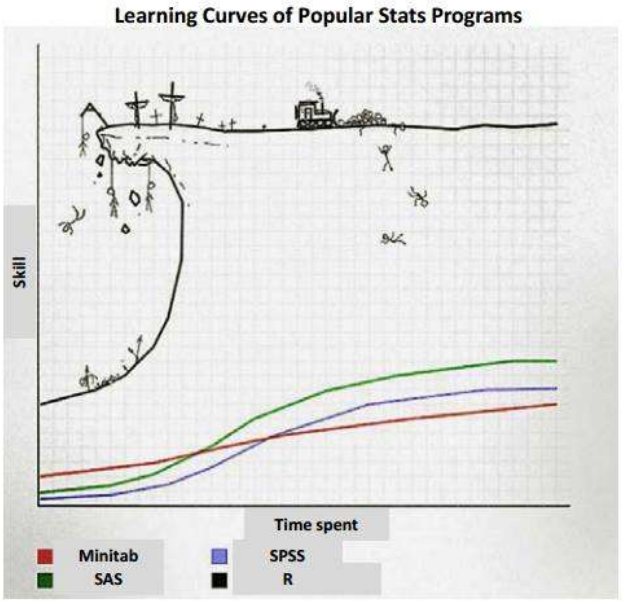
- Go to: https://cran.r-project.org/
- Select the version of R according to your OS.
- Install. Default options will fit in most cases. I suggest you to discard 32 bit files (if your OS is 64 bit) and message translations (annoying when you try googling them).
- In Windows: You may need to allow the system to read/write the installation folder.
  - Go to the R folder installation (e.g., «C:\Program Files\R\»)
  - Right click the R installation folder (e.g., «\R-4.1.1»)
  - Properties > Security > Modify
- You will need to re-install new version of R: It is important to keep R up to date

How to install Rstudio

- Go to: https://www.rstudio.com/
- Products > Rstudio > Desktop > Download
- Choose the free version

SPSS, R, Jamovi

	R	Jamovi	SPSS
Price	+	+	-
Functionalities	+	- (growing)	+
New functionalities	+	x	-
Customizability	+	x	-
Ease of use	-	+	+
Syntax	+	x	-
Stability	-	x	+
Diffusion	+	-	+



R Packages

- Packages are collections of functions, together with help files, sometimes datasets.
- Some basic statistical functions are already included in the basic installation of R, but most packages can be optionally installed.
- Packages are what makes R great: if you need to do something, it is very likely that someone else already implemented it in a package. There are packages for almost everything.

Installing/loading packages

«haven» is a package for reading SPSS data. Packages in R need to be first installed (i.e., downloaded from the internet) and then loded (i.e., made available in the current session).

A package needs to be installed only once, and loaded every time you use it

```
install.packages("haven")
```

```
library("haven")
```
or
```
require("haven")
```

Installing/loading packages (2)

A nice alternative is to use `pacman`, a package that manages packages

The only package you can manage with pacman, is pacman itself! The following code checks whether pacman is already installed, otherwise it installs it and loads it

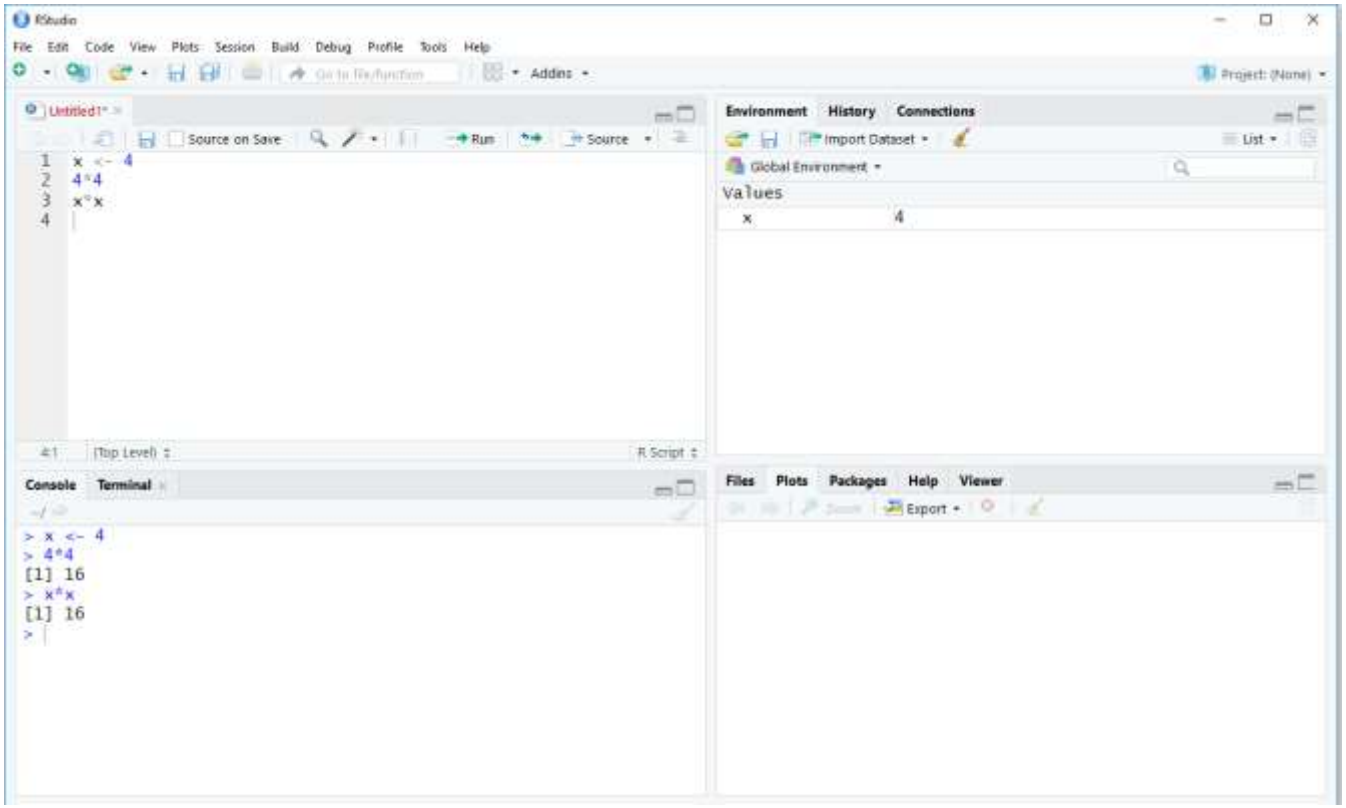
```
if(!require("pacman")) install.packages("pacman")
library("pacman")
```

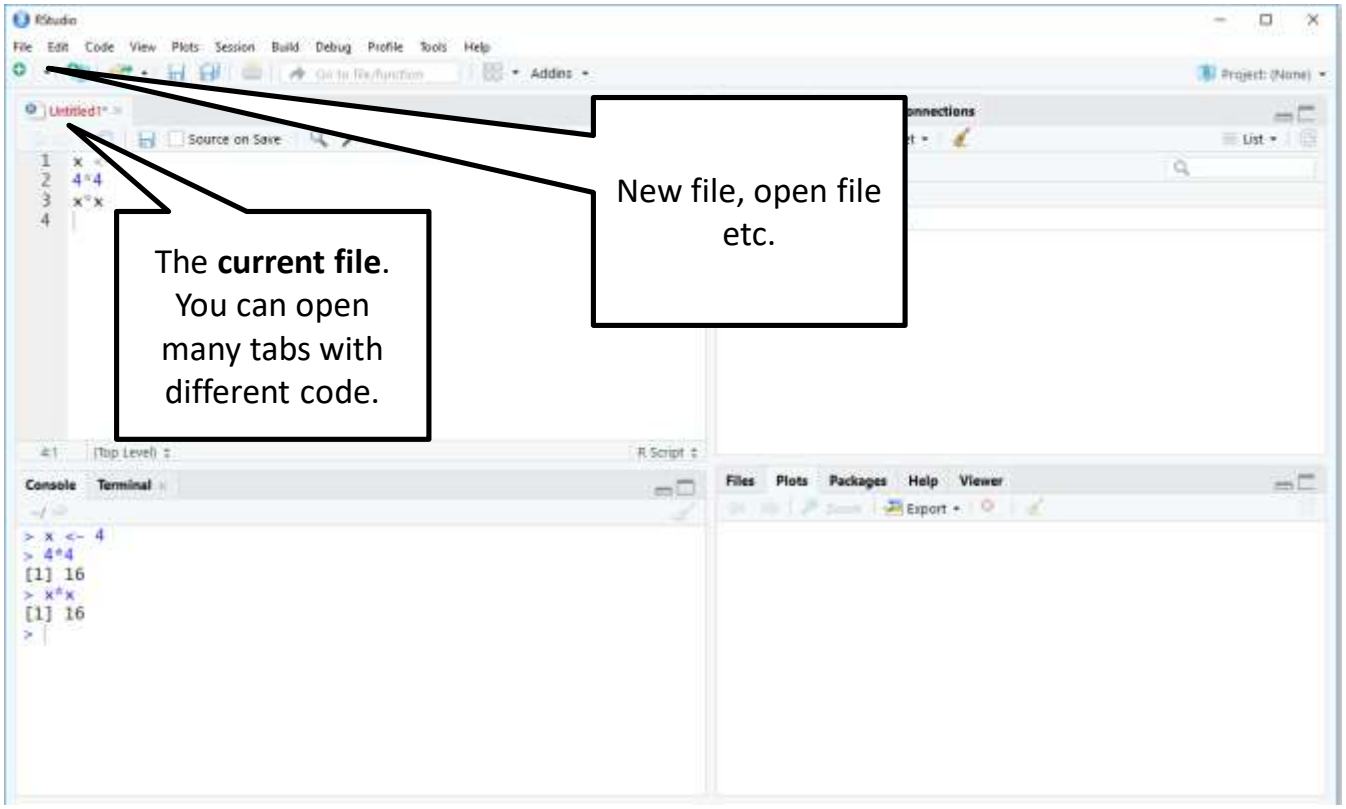
Installing/loading packages (2)

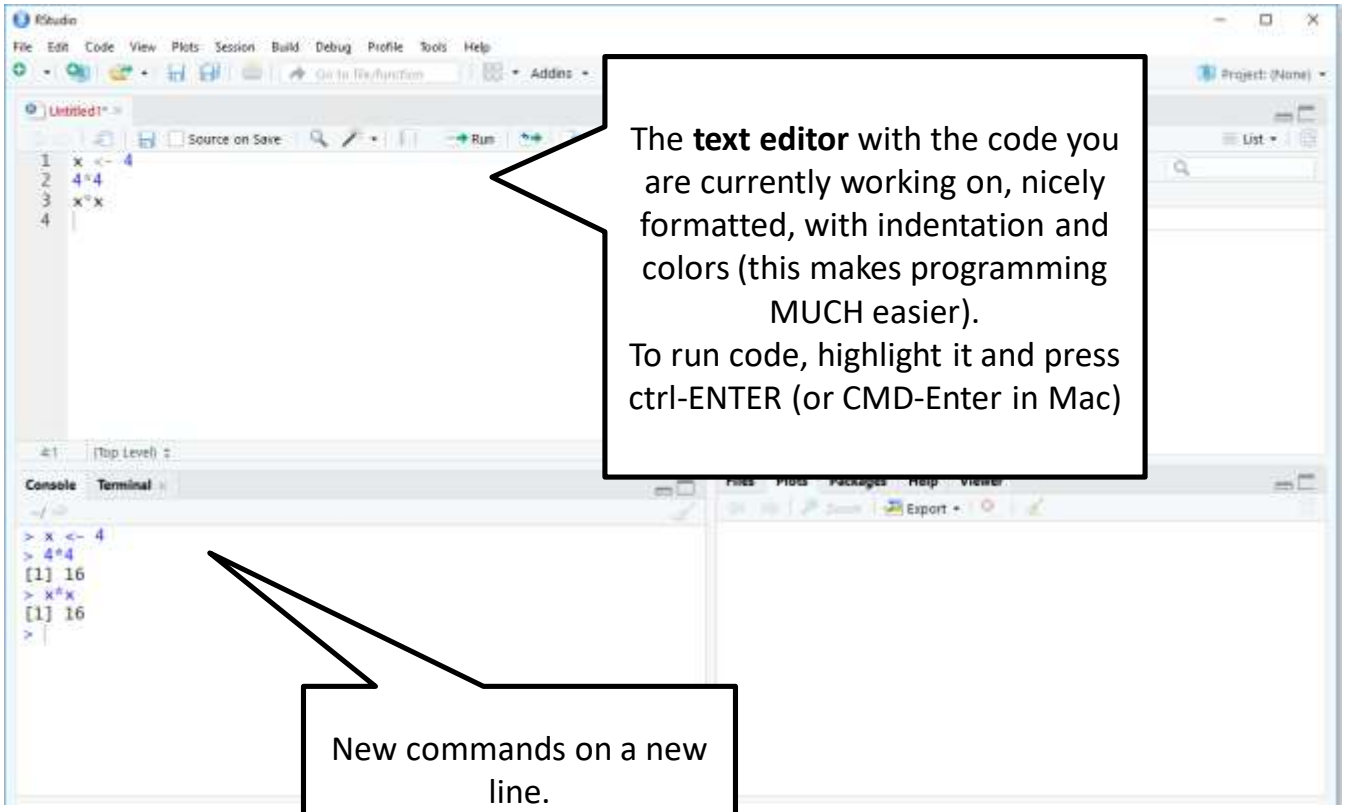
Then you can use pacman to install and load (if they are not installed alreay) or just load other packages, using pacman's function p_load

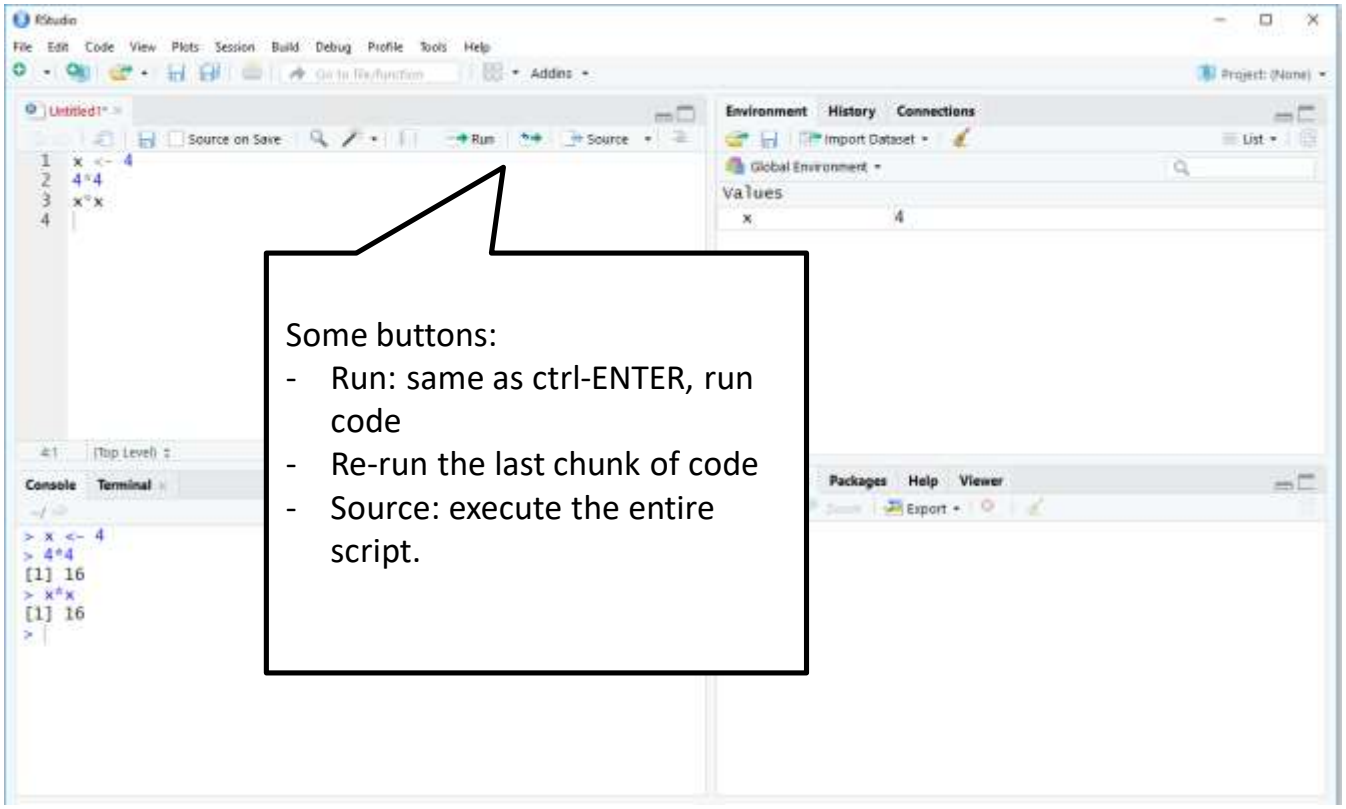
```
p_load("haven", "psych")
```

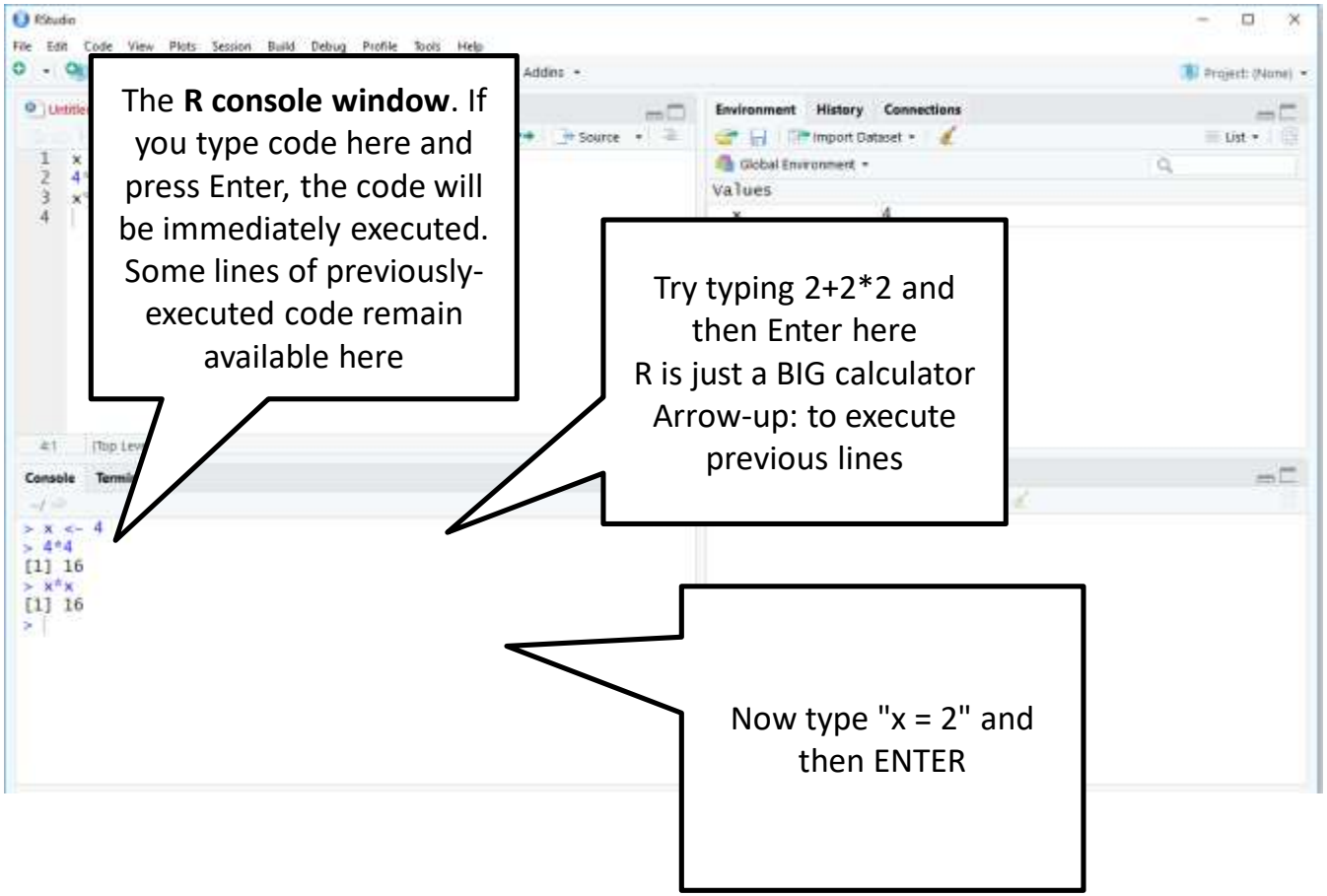
Rstudio and its windows

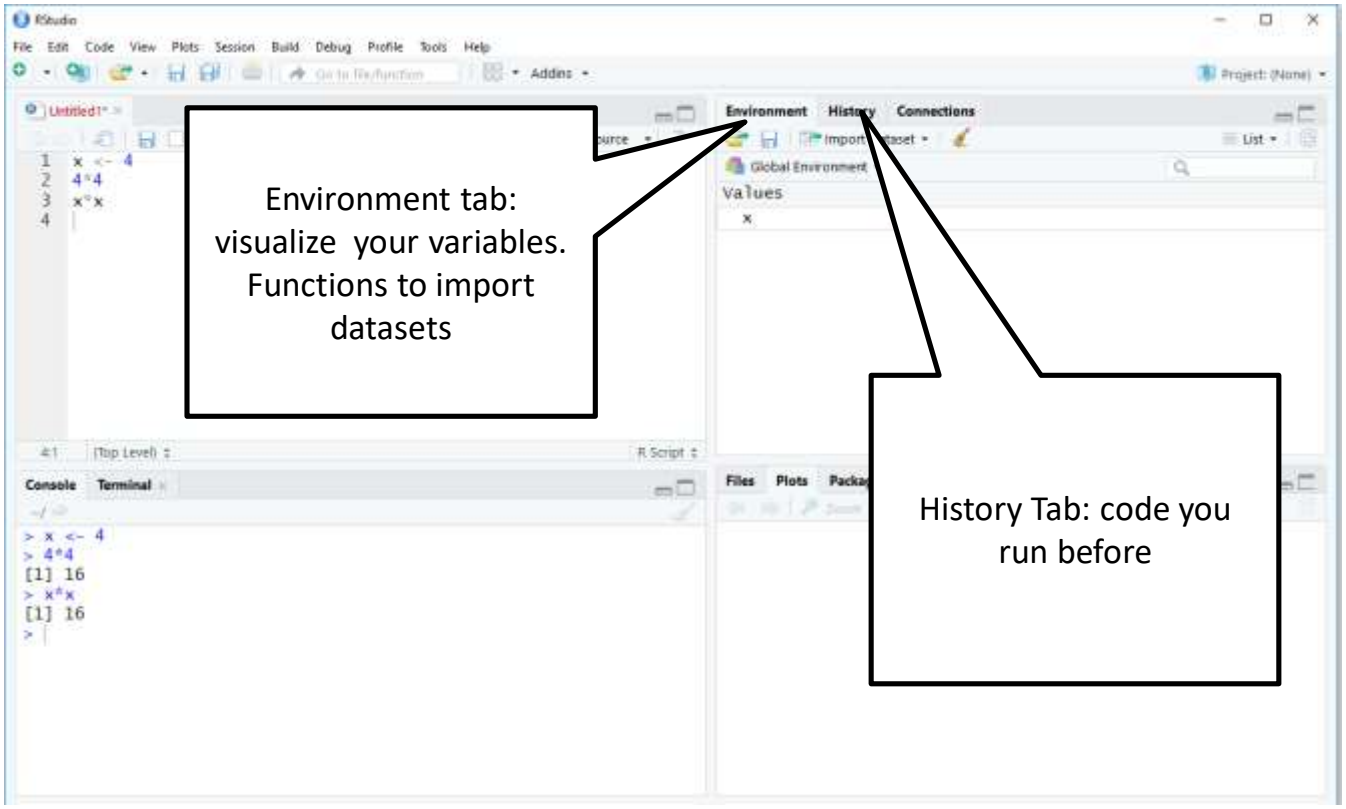

RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins
Project: (None)
Untitled1*
Source on Save
Run
Source
1 x <- 4
2 4^4
3 x^x
4
4:1 (Top Level)
R Script
Environment History Connections
Import Dataset
List
Global Environment
Values
x 4
Console Terminal
> x <- 4
> 4^4
[1] 16
> x^x
[1] 16
>
Files Plots Packages Help Viewer
Export

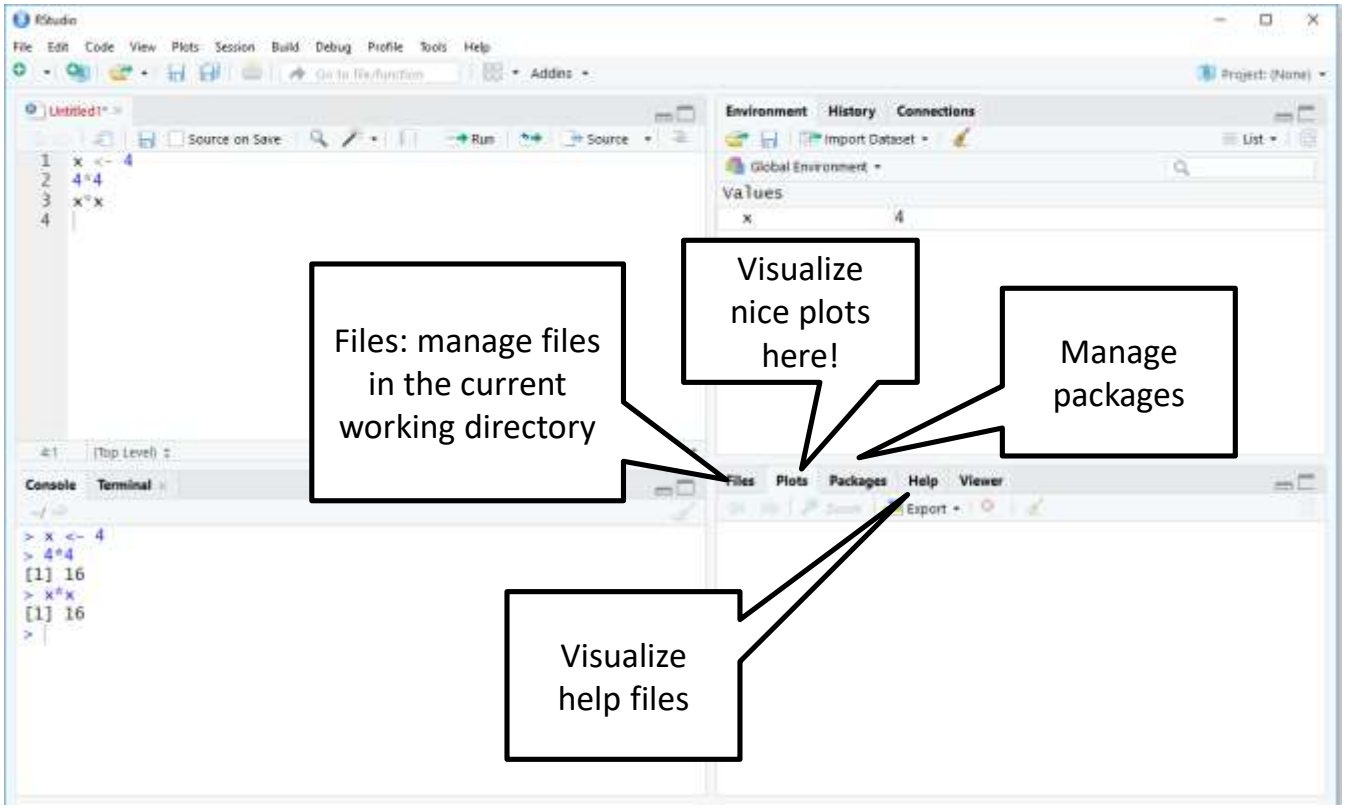

New file, open file etc.
The **current file**. You can open many tabs with different code.


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins
Project: (None)
Untitled1*
Source on Save
Run
List
1 x <- 4
2 4*4
3 x^x
4
4:1 (Top Level)
Console Terminal
Files Plots Packages Help Viewer
Export
> x <- 4
> 4*4
[1] 16
> x^x
[1] 16
>
The **text editor** with the code you are currently working on, nicely formatted, with indentation and colors (this makes programming MUCH easier).
To run code, highlight it and press ctrl-ENTER (or CMD-Enter in Mac)
New commands on a new line.
Commands can also be on the same line, separated by a semicolon ";"


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins
Project: (None)
Untitled1*
Source on Save
Run
Source
1 x <- 4
2 4*4
3 x^x
4
4:1 (Top Level)
Console Terminal
> x <- 4
> 4*4
[1] 16
> x^x
[1] 16
>
Environment History Connections
Import Dataset
List
Global Environment
Values
x 4
Packages Help Viewer
Export
Some buttons:
- Run: same as ctrl-ENTER, run code
- Re-run the last chunk of code
- Source: execute the entire script.


The **R console window**. If you type code here and press Enter, the code will be immediately executed. Some lines of previously-executed code remain available here
Try typing 2+2*2 and then Enter here
R is just a BIG calculator
Arrow-up: to execute previous lines
Now type "x = 2" and then ENTER
> x <- 4
> 4*4
[1] 16
> x*x
[1] 16


Environment tab: visualize your variables. Functions to import datasets
History Tab: code you run before


Files: manage files in the current working directory
Visualize nice plots here!
Manage packages
Visualize help files

Some general info (1)

- Comments: Anything preceded by `#` will not be executed
- Text: `"use inverted commas for text"`
- R is **CaSe sENsItivE**.
- `TRUE` and `FALSE` are upper case.
- You can get help by typing `help(something)` or `?something`
- `help.search(something)` or `??something`.
- `NA` means missing value

Some general info (2)

- `ls()` – lists the variables that are currently in your **workspace**.
- `rm(list = ls())` clean workspace
- `getwd()` – to get the working directory
- To set the working directory (not good for code replicability!)
  `setwd("C:\\Somepath")` – for windows
  `setwd("/Somepath/")` – for MAC

Basic commands in R

Function	Commands	Function	commands
Check working directory	getwd()	Left assignment operator	x <- 2 or x = 2
Set working directory	setwd("C:/mydirectory/")	And, or, not	& | !
list all objects	ls()	Less than, less than or equal etc.	<, <=, >, >=
Remove an object	rm("myobject")	Equal to, different	==, !=
Get help on a function	help("function")	Exponentiation	^
Visualize a dataframe	head(dataset) or View(dataset)	Prioritizing operations	(2+3)*2
Variable names in a dataframe	names(dataset)	Absolute value	abs()
Add, subtract, multiply, divide	+, -, *, /	Square root	sqrt() or ^.5

Reading data

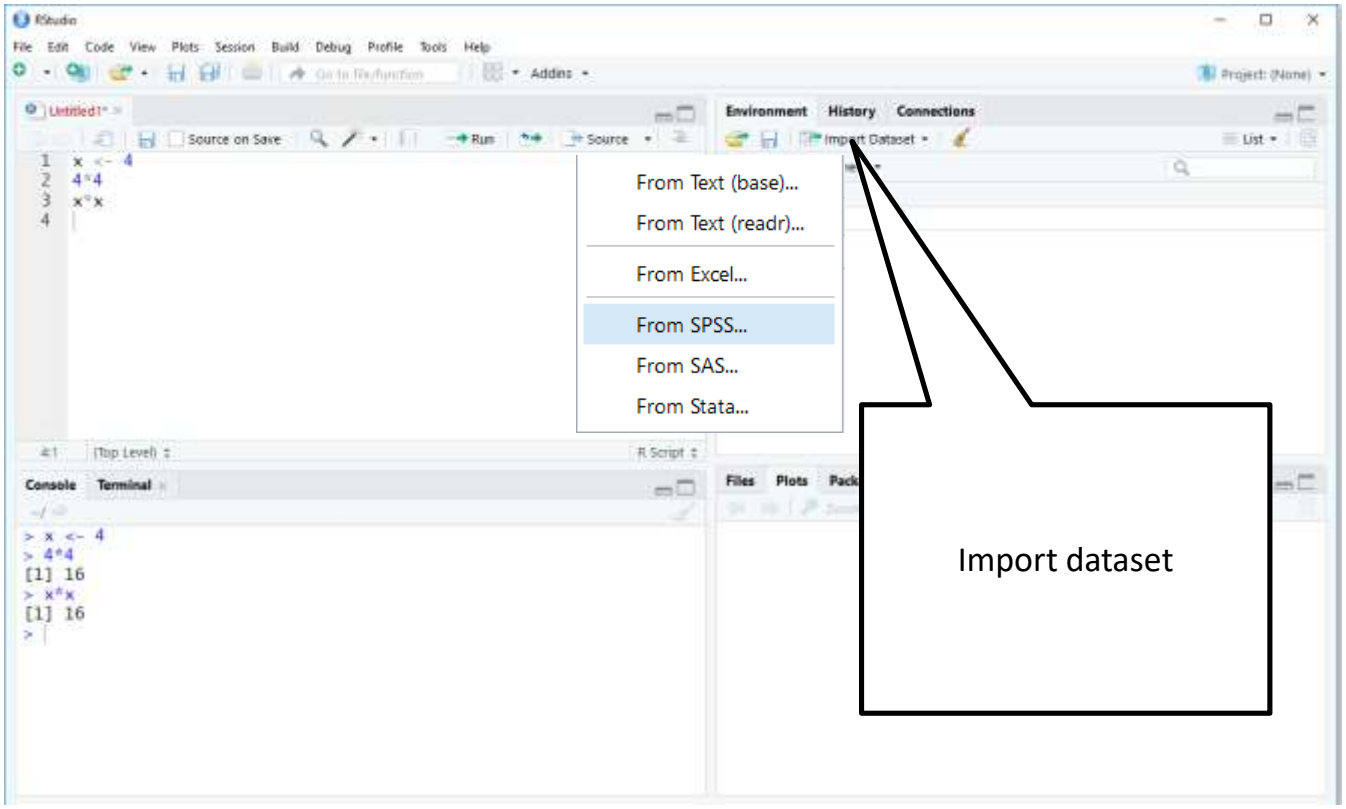
R scripts

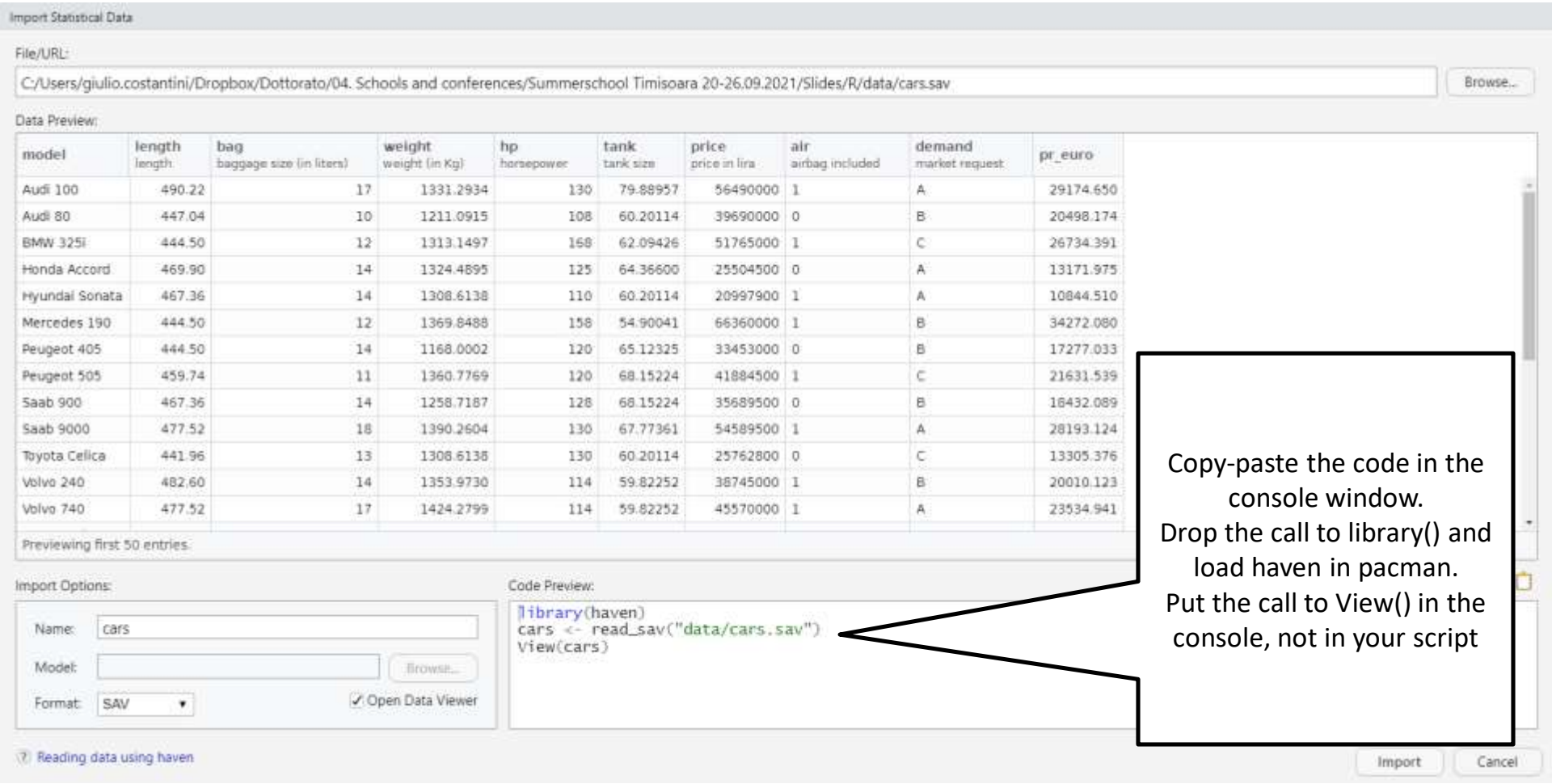
- R scripts are simple txt files, but with extension **.R** instead of **.txt**
- You can create an R file by just renaming a .txt file with a new extension.
- They are automatically opened by Rstudio if you have it installed.

Import data

- In R you can import data by writing code. Rstudio helps a bit by including the possibility to import data interactively. However, if you abuse this option (as other similar options in R studio), your code won't be reproducible (i.e., on another computer).
- The simplest pipeline to make your code as reproducible as possible, when you work on a project, is

1. Create an .R file
2. Save the data in the folder of your .R file (or in a subfolder, e.g., «yourpath/data»)
3. Open R-studio by double-clicking on your file. This will automatically set your working path to that of your R file
4. Open data from R-studio BUT instead of importing them directly, copy-paste the code in your file
5. Remove useless parts (e.g., View call, the library call if you already have the relevant packages loaded)


RStudio
File Edit Code View Plots Session Build Debug Profile Tools Help
Addins
Project: (None)
Untitled1*
Source on Save
Run
Source
1 x <- 4
2 4*4
3 x^x
4
Environment History Connections
Import Dataset
List
From Text (base)...
From Text (readr)...
From Excel...
From SPSS...
From SAS...
From Stata...
4:1 (Top Level)
R Script
Console Terminal
> x <- 4
> 4*4
[1] 16
> x^x
[1] 16
>
Files Plots
Import dataset



Copy-paste the code in the console window.
Drop the call to library() and load haven in pacman.
Put the call to View() in the console, not in your script

Rfile.R*

Source on Save

```
1  # load pacman
2  if(!require("pacman")) install.packages("pacman")
3  library(pacman)
4
5  # use pacman to load other packages
6  p_load("haven", "psych")
7
8  # load data using a function from package haven
9  cars <- read_sav("data/cars.sav")
10
```

10:1 (Top Level)

Console Terminal Jobs

R 4.1.0 · C:/Users/giulio.ca

```
> View(cars)
```

View() allows inspecting your data. You cannot modify data in here, you have to do it with the syntax (we will see how) or in excel before importing

Rfile.R* cars

Filter

	model	length length	bag baggage size (in liters)	weight weight (in Kg)	hp horsepower	tank tank size	price price in lira	air airbag included	demand market request	pr_euro
1	Audi 100	490.22	17	1331.2934	130	79.88957	56490000	1	A	29174.650
2	Audi 80	447.04	10	1211.0915	108	60.20114	39690000	0	B	20498.174
3	BMW 325i	444.50	12	1313.1497	168	62.09426	51765000	1	C	26734.391
4	Honda Accord	469.90	14	1324.4895	125	64.36600	25504500	0	A	13171.975
5	Hyundai Sonata	467.36	14	1308.6138	110	60.20114	20997900	1	A	10844.510
6	Mercedes 190	444.50	12	1369.8488	158	54.90041	66360000	1	B	34272.080
7	Peugeot 405	444.50	14	1168.0002	120	65.12325	33453000	0	B	17277.033
8	Peugeot 505	459.74	11	1360.7769	120	68.15224	41884500	1	C	21631.539
9	Saab 900	467.36	14	1258.7187	126	68.15224	35689500	0	B	18432.089
10	Saab 9000	477.52	18	1390.2604	130	67.77361	54589500	1	A	28193.124
11	Toyota Celica	441.96	13	1308.6138	130	60.20114	25762800	0	C	13305.376
12	Volvo 240	482.60	14	1353.9730	114	59.82252	38745000	1	B	20010.123
13	Volvo 740	477.52	17	1424.2799	114	59.82252	45570000	1	A	23534.941
14	BMW 535i	472.44	12	1651.0760	208	79.88957	69720000	1	C	36007.375

Showing 1 to 14 of 24 entries, 10 total columns

Some basic ways to inspect the dataframe

`summary(cars)`

```
> summary(cars)
    model               length           bag             weight           hp
 Length:24          Min.   :401.3   Min.   :10.00   Min.   : 993.4   Min.   : 81.0
 Class :character   1st Qu.:429.3   1st Qu.:12.00   1st Qu.:1079.0   1st Qu.:104.2
 Mode  :character   Median :444.5   Median :14.00   Median :1308.6   Median :120.0
                    Mean   :447.4   Mean   :13.96   Mean   :1252.8   Mean   :126.0
                    3rd Qu.:470.5   3rd Qu.:17.00   3rd Qu.:1355.7   3rd Qu.:137.0
                    Max.   :490.2   Max.   :19.00   Max.   :1651.1   Max.   :208.0
      tank            price              air             demand           pr_euro
 Min.   :45.06   Min.   :12387900   Min.   :0.0000   Length:24          Min.   : 6398
 1st Qu.:54.90   1st Qu.:20989500   1st Qu.:0.0000   Class :character   1st Qu.:10840
 Median :60.20   Median :36639750   Median :1.0000   Mode  :character   Median :18923
 Mean   :61.13   Mean   :38999538   Mean   :0.5417                      Mean   :20142
 3rd Qu.:67.87   3rd Qu.:52471125   3rd Qu.:1.0000                      3rd Qu.:27099
 Max.   :79.89   Max.   :88179000   Max.   :1.0000                      Max.   :45541
>
```

Descriptive statistics

Use the dollar syntax to select variables within a dataframe

```
mean(cars$price) # mean price
sd(cars$price) # sd of price
median(cars$price) # median price
scale(cars$price) # standardize the variable (mean 0, sd = 1)

mean(cars$price, na.rm = TRUE) # mean if you have missing values
```

Now you

- Import the ANT.csv data
  ANT = Attention Network Task, the data are from package ez.
- Visualize the data using View()
- Compute summary statistics using summary()

Tips:

- .csv is a textual format. Data are separated by a **colon**. You can use the readr package and select semicolon as the delimiter.
- Remember to load the readr package in the p_load call.

Solution

```
if(!require("pacman")) install.packages("pacman")
library(pacman)

# use pacman to load other packages
p_load("haven", "psych", "readr")

ANT <- read_delim("data/ANT.csv", delim = ";",
                  escape_double = FALSE, trim_ws = TRUE)

# I tipically comment calls to View() away, otherwise a
new window appears every time you run the script
#View(ANT)
summary(ANT)
```

Variables in R

This part is VERY boring, but you need to learn this stuff.

Variables

<- Left assignment (**=** works too, but not recommended)

-> Right assignment (almost never used)

```
X <- 4
4 -> X
```

When you create a variable, R does not visualize anything! You have to type that variable and hit enter to see its content

```
X
```

You can create and use the variable as you want

```
X*2
X^2
X >= 4
X > 4
```

Scalars

This creates a numeric scalar

```
X <- 10
```

String variables

```
Z <- "apple"
```

Logical variables

```
Q <- TRUE
```

(scalars in R are actually vectors of length 1)

Vectors

- A vector is a concatenation of elements.
- Variables in the datasets are vectors of different classes!
- The command head() allows you to inspect the first rows of each variable and its «type» (dbl = double precision numeric, chr = character). Here it is applied to ANT.

```
head(ANT)
```

```
# A tibble: 6 x 10
  subnum group     block trial cue     flank     location direction         rt error
   <dbl> <chr>     <dbl> <dbl> <chr>   <chr>     <chr>    <chr>          <dbl> <dbl>
1      1 Treatment     1     1 None    Neutral   up       left      3986773095     0
2      1 Treatment     1     2 Center  Neutral   up       left      3891821662     0
3      1 Treatment     1     3 Double  Neutral   up       left      3332185694     0
4      1 Treatment     1     4 Spatial Neutral   up       left      4197640401     0
5      1 Treatment     1     5 None    Congruent up       left      4464753886     0
6      1 Treatment     1     6 Center  Congruent up       left      3389765808     0
```

Creating vectors

Some ways for creating vectors

- `c("foo", "bar", "foo"), c(1, 2, 3), c(x, 1, 2, 3, x)` concatenate elements into a vector
- `1:10` simple numeric sequence
- `seq(from, to)` allows for more complex sequences, see ?seq
- `rep(x, times)` repeat element x
- `x <- runif(50)` Random data from a uniform distribution
- `y <- rnorm(50)` Random data from a normal distribution
- `z <- sample(1:5, 20, replace = TRUE)` Randomly sample from a vector

- `names(x) <- c("name1", "name2", …)` Give names to the elements of a vector

Vectors

```
q <- 1:10
x <- c(0, 1, 1, 2, 3, 5, 8)
y <- c("apple", "orange")
k <- c(TRUE, FALSE, TRUE, TRUE)
```

If you mix up things in a vector, R does not preserve the different types of data. One vector, one type of data

```
z <- c("apple", "orange", 1, 2, 3, 5, 8)
```

Vector elements can be named.

```
names(y) <- c("first fruit", "second fruit")
```

Length of a vector

```
length(x)
```

Factors

- Nominal variables (there is also a special type of factor for ordinal data)
- Variables are divided into levels
- The function `factor()` allows converting a numeric or string vector into a factor
- See ?factor for more options

```
gender <- c("male", "female", "female", "male")
gender2 <- factor(gender)
levels(gender2)

x <- sample(1:2, size = 10, replace = TRUE)
as.factor(x)
x <- factor(x, levels = c(1, 2), labels = c("male", "female"))
levels(x)
```

Indexing vectors

- `z[1]` first element of vector z
- `z[3]` third element of vector z
- `z[-3]` vector z except element 3
- `Z[1] <- "APPLE"` assignment
- `z[1:3] or z[c(1, 2, 3)]` elements 1, 2, and 3. Indexing a vector with a vector of positions.
- `x[c(TRUE, TRUE, TRUE, FALSE,…, FALSE)] or x[x<=1]` indexing a vector with a vector of logicals
- `y["first fruit"]` indexing a vector by element names
- `z[z != "APPLE"]`
- `z[z != "APPLE"] <- "PEACH"`

Now you

- Create a vector called “nms” with the names of six friends
- Create another vector called “age” with their age
- Select only the names of those friends that are younger than 30
- Create a new vector, in which those younger than 30 are now 32

Soloution

```
nms <- c("Mickey Mouse", "Donald Duck", "Daisy
Duck", "Goofy", "Pluto", "Gus Goose")
age <- c(26, 30, 22, 32, 8, 18)
nms[age < 30]
age2 <- age
age2[age < 30] <- 32
```

Some vector operators

- `%*%` Vector/matrix product (`*` performs elementwise product)

```
1:4 * 1:4
1:4 %*% 1:4
```

- `%in%` matching operator

```
"apple" %in% c("apple", "orange", "peach")
"Apple" %in% c("apple", "orange", "peach")
"Mary" %in% nms
```

Matrices

- Define a matrix from a vector

```
x <- matrix(1:9, nrow = 3, ncol = 3)
y <- matrix(1:9, nrow = 3, ncol = 3, byrow = TRUE)
```

- Define a matrix by concatenating vectors in columns…

```
x <- cbind(1:10, 11:20, 21:30)
```

- …Or rows

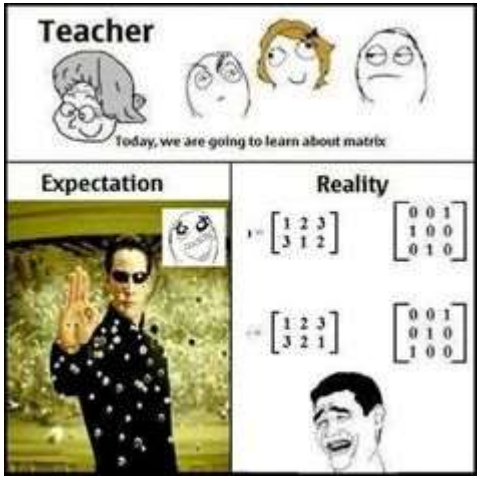
```
y <- rbind(1:10, 11:20, 21:30)
```

- Matrix columns and rows can be named

```
colnames(x) <- c("name1", "name2", …)
rownames(x) <- c("name1", "name2", …)
```

- Dimensions of a matrix

```
dim(x)
```



Indexing matrix elements

Same as for vectors, but now you can index separately both rows and columns

`x[1, 3]` Element in the first row, third column

`x[1, 3]` Element in the first row, third column

`x[1,]` first row of x

`x[1:2,]` first two rows of x

`x[c(TRUE, FALSE, TRUE), ]` using logical vectors

`x[, c("name1", "name2")]` using row and column names

The rows and columns of a matrix are vectors themselves

Matrix operators

`x+2, x-3, x*3, x/3, x^2` perform elementwise operations on a matrix

`x*y` elementwise product of two matrices

`x%*%y` matrix product

Apply allows applying a function to the rows or columns of a matrix

`apply(x, 1, sum)` sum by rows

`apply(x, 2, sum)` sum by column

There are shortcuts

`rowSums(x)`

`colSums(x)`

Arrays

- Like matrices, but with more than 2 dimensions.
- Data-cubes or hypercubes
- Same logic as matrices

```
x <- array(data = 1:27, dim = c(3, 3, 3))
x
x[1:2, 1, 1]
```

For arrays, you have to think in 3 or more dimensions

```
apply(x, 1, sum)
apply(x, 2, sum)
apply(x, 3, sum)
apply(x, 1:2, sum)
```

Lists

While vectors, matrices, and arrays in R can include only one type of element (e.g., numeric, string, etc.) Lists are flexible data structures whose elements can include different kinds of data

```
x <- c("cat", "dog")
y <- matrix(1:100, ncol = 10)
z <- 4
alist <- list(x, y, z)
names(alist) <- c("obj_x", "obj_y", "obj_z")
```

Double brackets for indexing single elements of lists

```
alist[[2]]
alist[["obj_x"]]
```

single brackets allow selecting multiple elements, the result being a list

```
alist[c(1, 3)]
```

Dollar sign: a convenient shortcut (very useful, see dataframes)

```
alist$obj_x
```

Applying operations to all elements of a list

```
lapply(alist, is.numeric)
```

Dataframes

- Look like matrices
- Are actually a special types of lists, in which all of the elements have the same length.
- Data imported in R are usually dataframes. If not, you can force them to be data-frames, using e.g., ANT <- data.frame(ANT)

```
Cod <- c("A","B","C","D","E","F","G","H","I","J")
Gender <- c("M","M","F","F","F","F","M","F","M","F")
Gender <- as.factor(Gender)
Age <- c(43,36,56,47,58,37,46,30,28,26)

mydata <- data.frame(Cod,  Gender,  Age)
```

See the difference between a dataframe and a matrix created with cbind?

```
cbind(Cod,  Gender,  Age)
```

Indexing dataframes

Elements of dataframes can be indexed like those of a matrix

```
mydata[, 2]
mydata[, "Gender"]
mydata[1, ]
```

And like those of a list, using the dollar sign

```
mydata$Gender
```

The same notation can be used for creating new variables

```
mydata$Years <- c(14,6,28,12,1,6,5,2,2,1)
mydata
```

Missing values

The special symbol NA is used in R for representing missing values

```
mydata[2, "Age"] <- NA

mean(mydata$Age)
mean(mydata$Age, na.rm = TRUE)
```

Other info about data classes

Get the class of an object

```
class(mydata)
class(mydata$Age)
class(mydata$Gender)
```

Ask about a specific class

```
is.matrix(), is.vector(), …
```

Ask about missing values

```
is.na(mydata$Age)
```

Try to coerce the class of an object

```
as.matrix(), as.vector(), as.data.frame(),
as.factor()
```

Now you! Fun with dataframes

Using the nms and age vector created just above

- Combine them into a dataframe called "neighbors"
  - Select only the rows of the dataframe corresponding to colleagues that are younger than 30
  - Create a new column in the dataframe, in which colleagues that are younger than 30 are now 32
  - Optional: inspect the class of the dataframe and of the varibles within the dataframe

Solution

```
neighbors <- data.frame(nms, age)
neighbors[neighbors$age < 30, ]
neighbors$age2 <- neighbors$age
neighbors$age2[neighbors$age < 30] <- 32
print(neighbors)
# optional, inspect the classes of elements
class(neighbors)
class(neighbors$nms)
class(neighbors$age)
class(neighbors$age2)
```

Saving data

Save the data in a .csv file. You can use function write.csv

```
write.csv(neighbors, file = "data\\neighbors.csv")
write.csv(neighbors, file = "data\\neighbors.csv", na = "",
row.names = FALSE)
```

Save and load using the R binary format, uses much less hard disk space, but you loose compatibility with other software.

```
save(neighbors, file = "data\\neighbors.RData")
load(file = "data\\neighbors.RData ")
```

Save/load everything in your current workspace

```
save.image(file = "data\\myimage.RData")
load("data\\myimage.RData")
```

Introduction to data analysis and processing with R – Part 2

Giulio Costantini -
giulio.costantini@unimib.it

Handling data with basic R

Import the cars data and ANT data. Using the cars data

1) Select only cars whose length is less than 450
2) Select only cars whose weight is more than 1100
3) Select only cars that comply with both conditions
4) Using the dataset in 3, select only the price of these cars
5) Calculate the mean of such prices, using colMeans()

Solution

```
# remember to load packages first!
if(!require("pacman")) install.packages("pacman")
require("pacman")
p_load("haven", "readr")

# load data
cars <- read_sav("data/cars.sav")
ANT <- read_csv("data/ANT.csv")

# complete the exercise
cars[cars$length < 450,]
cars[cars$weight > 1100,]
cars[cars$length < 450 & cars$weight > 1100,]
cars[cars$length < 450 & cars$weight > 1100, "price"]
colMeans(cars[cars$length < 450 & cars$weight > 1100,
"price"])
```

dplyr: handling data in an elegant way in R

dplyr aims to provide a function for each basic verb of data manipulation:

- %>% is a dplyr command that puts the output of the previous operation as the first input of what comes next
- select() to select variables based on their names.
- summarize() to condense multiple values to a single value.
- group_by() is used in conjunction to summarize to obtain summaries of groups of data
- filter() to select cases based on their values.
- mutate() and to add new variables that are functions of existing variables.
- arrange() to reorder the cases.
- rename() to rename variables

Select some columns

```
select(cars, length, bag, weight, hp)
select(cars, length:hp) # same thing
# can also change variable names in the process
select(cars, length, bag, weight, horsepower = hp)
```

«:» means from column `length` to column `hp`

```
> select(cars, length, bag, weight, hp)
   length bag    weight  hp
1  490.22  17 1331.2934 130
2  447.04  10 1211.0915 108
3  444.50  12 1313.1497 168
4  469.90  14 1324.4895 125
5  467.36  14 1308.6138 110
6  444.50  12 1369.8488 158
7  444.50  14 1168.0002 120
8  459.74  11 1360.7769 120
9  467.36  14 1258.7187 128
10 477.52  18 1390.2604 130
11 441.96  13 1308.6138 130
```

Filter observations

filter(yourdata, condition, condition, …) gives you only the cases that fulfill all conditions

e.g., filter only cars with airbag == 1

```
filter(cars, air == 1)
```

You can also use select, filter and other commands in conjunction, combining them with %>%

e.g.,

```
select(cars, length, air) %>% filter(air == 1)
```

Now you

Re-do this exercise, this time with dplyr

1) Select only cars whose length is less than 450
2) Select only cars whose weight is more than 1100
3) Select only cars that comply with both conditions
4) Using the dataset in 3, select only the price of these cars
5) Calculate the mean of such prices

Solution

```
filter(cars, length < 450)
filter(cars, weight > 1100)
filter(cars, length < 450, weight > 1100)
filter(cars, length < 450, weight > 1100) %>%
select(price)
filter(cars, length < 450, weight > 1100) %>%
select(price) %>% colMeans()
```

Excercise

Compute the correlation matrix among varaibles length, bag, weight and hp

Tip: The correlation matrix can be computed with command cor()

Solution

```
# Correct solution
cor( select(cars, length:hp) )

# Equivalent solution
select(cars, length:hp) %>% cor()
```

Summaries

summarize() allows you to obtain a new dataset including summary statistics you like

```
summarize(cars,
          Mlenght = mean(length),
          Mprice = mean(price))
```

Summaries by group

Summarize is very useful if preceded by group_by(), which allows obtaining summaries by one or more grouping variables

```
cars_summ <- group_by(cars, air) %>%
  summarize(Mlenght = mean(length),
            Mprice = mean(price))

cars_summ
```

Now you

Using the ANT data, create a new dataset "ANT_smm" that includes

- Mrt = the average reaction-time
- Merr = the average error rate,

by subject and by block

Tip. In group_by, separate grouping conditions using a comma

Solution

```
ANT_smm <- group_by(ANT, subnum, block) %>%
  summarize(Mrt = mean(rt), Merr = mean(error))
ANT_smm
```

Compute new variables

You can use mutate() to create new variables.

e.g., in cars you can create a new variable, «space», that is just the sum of of bag and tank. The output is a new dataset, including the new variable

```
mutate(cars, space = bag + tank)
```

Now you

Using the ANT_smm dataset created above, compute the so-called inverse-efficienty score for each subject and each block

$$IES = \frac{Mrt}{1 - Merr}$$

Solution

```
ANT_smm %>%
        mutate(IES = Mrt /(1-Merr))
```

Instead of using the saved dataset, you can also keep adding functions to the above code using the %>% command, the results is the same

```
group_by(ANT, subnum, block) %>%
  summarize(Mrt = mean(rt), Merr = mean(error)) %>%
  mutate(IES = Mrt /(1-Merr))
```

Computing scale scores

By combining the commands you learned so far, you should be able to compute scale scores.

The BPD data (simluated) include a questionnaire (that I made up!) for borderline personality disorder, items are on a scale from 1 to 6. Higher scores correspond to more BPD.

Create a mean score for a scale including items BPD_1 to BPD_6 and call it BPD_score.

Solution

First, let's import the data

BPD <- read_csv("data/BPD.csv")

There are a few possible solutions

```
# using mutate and summing variables
BPD <- mutate(BPD,
BPD_score = (BPD_1 + BPD_2 + BPD_3 + BPD_4 + BPD_5 +
BPD_6)/6)

# using rowSums
BPD$BPD_score <- select(BPD, BPD_1:BPD_6) %>% rowMeans()
```

Reverse-coding items

The item BPD_7_r is reverse coded, meaning that higher scores correspond to lower agreement.

1) Reverse-score it, by creating a new varaible (BPD_7) which is equal to 7-BPD_7_r
2) Include it in the scale that you just computed

Transforming variables

With mutate you can also transform variables as you like. For example, in the ANT data you can create a log_RT variable, which is the logarithm of reaction times

Tip. The function for logarithm is log()

Solution

```
ANT <- mutate(ANT, log_rt = log(rt))
```

Collapse variables into fewer categories

Function cut() allows collapsing variables into categories.

Imagine that we want to create a new variable telling us if RTs are fast, medium or slow.

First, we can inspect the quartiles with function

```
quantile(ANT$rt)
```

```
> quantile(ANT$rt)
      0%      25%      50%      75%     100%
179.5972 356.2357 404.9896 454.4902 657.6986
```

Collapse variables into fewer categories

Then we can convert the rt variable into a new variable, rt_categories, telling us whether a rt is fast (<= 356 ms), medium (356 < rt < 454) or slow (>= 657 ms).

See ?cut to check for all options.

To include all values smaller than 356 ms or larger than 657 ms, you can include –Inf and Inf as the first and last extreme in the breaks argument

Collapse variables into fewer categories

```
ANT <- mutate(ANT,
rt_categories = cut(rt,
breaks = c(-Inf, 356, 454, Inf), labels = FALSE))
ANT$rt_categories
```

Creating dummy variables

Sometimes you want to dichotomize a variable according to its value.

Try creating a new variable in ANT, rt_dummy, which separates RTs larger and smaller than the median (405 ms).

You should be able figure it out!

Solution

```
ANT <- mutate(ANT, rt_dummy =
          cut(rt, breaks = c(-Inf, 405, Inf),
          labels = FALSE))
```

Outliers

Outlier detection is not a simple topic and would require several lesson. A good resource is

Wilcox, R. (2012). *Modern Statistics for the Social and Behavioral Sciences: A practical Introduction.* CRC press.

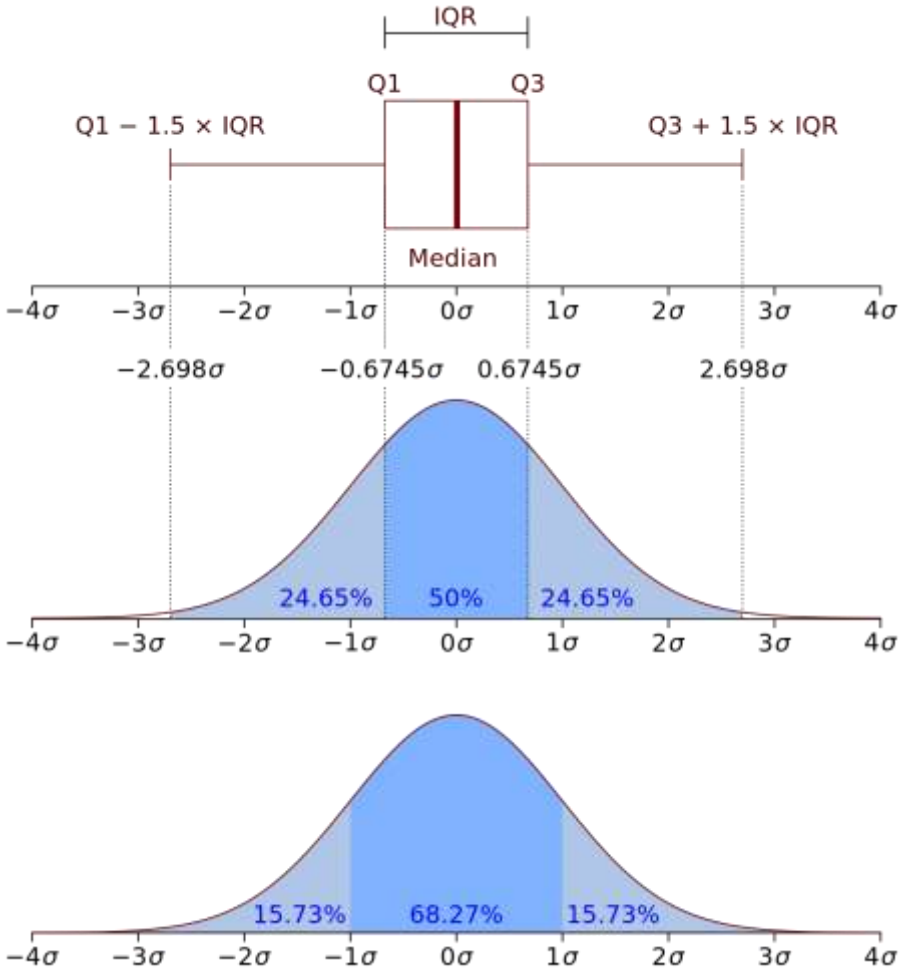
We will only focus on a simple-but-effective strategy, the boxplot.

Boxplot

The boxplot is a quite «robust» method (does not suffer from masking, like the $M \pm 2SD$ and similar methods).

Image from Wikipedia.

See ?boxplot.stats for more information



Boxplot in R

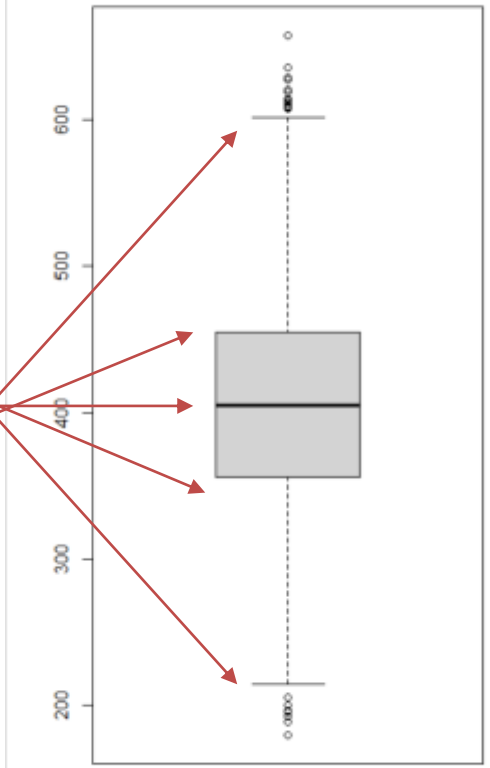
box <- boxplot(ANT$rt)

box$stats

You can then choose to filter out cases whose rt is lower than 214.6342 or larger than 601.0676.

Try it yourself

```
> box$stats
         [,1]
[1,] 214.6342
[2,] 356.2310
[3,] 404.9896
[4,] 454.4974
[5,] 601.0676
```



Solution

```
ANT <- filter(ANT, rt >= 214.6342, rt <= 601.0676)
```

Handling repeated-measures data

Wide format

One row by subject, but as many columns as repeated measures.

```
select(BPD, ID, Sex, Age, selfHurtPRE:SelfHurt3Months)
```

	ID	Sex	Age	SelfHurtPRE	SelfHurtPOST	SelfHurt1Month	SelfHurt2Months	SelfHurt3Months
1	1	M	15	3.574414	5.369372	4.545787	3.010593	4.161612
2	2	M	19	4.550380	4.557880	4.371093	5.358189	4.171579
3	3	M	15	6.040344	6.050010	5.786008	5.670039	5.508103
4	4	F	15	4.674558	6.677108	6.021034	4.256716	4.762845
5	5	F	19	4.734384	2.986224	2.736148	4.099973	3.757674
6	6	M	17	6.111710	3.728257	4.954271	4.448321	4.852450
7	7	F	18	4.722836	3.875594	5.390017	3.910833	6.565479
8	8	M	16	5.438501	4.911779	5.207618	4.424561	4.979036
9	9	F	19	4.917107	6.432360	5.223686	6.213898	5.501442
10	10	M	19	5.427812	5.563453	2.204504	1.778136	3.451040

Long format

- Many rows by subject, but one column for each variable of interest.
- Rows indicating stable properties of the subject are repeated.
- Most used format in R.

	ID	Sex	Age	variable	score
1	1	M	15	SelfHurtPRE	3.574414
2	1	M	15	SelfHurtPOST	5.369372
3	1	M	15	SelfHurt1Month	4.545787
4	1	M	15	SelfHurt2Months	3.010593
5	1	M	15	SelfHurt3Months	4.161612
6	2	M	19	SelfHurtPRE	4.550380
7	2	M	19	SelfHurtPOST	4.557880
8	2	M	19	SelfHurt1Month	4.371093
9	2	M	19	SelfHurt2Months	5.358189
10	2	M	19	SelfHurt3Months	4.171579
11	3	M	15	SelfHurtPRE	6.040344
12	3	M	15	SelfHurtPOST	6.050010

From wide to long: function melt

```
melt(

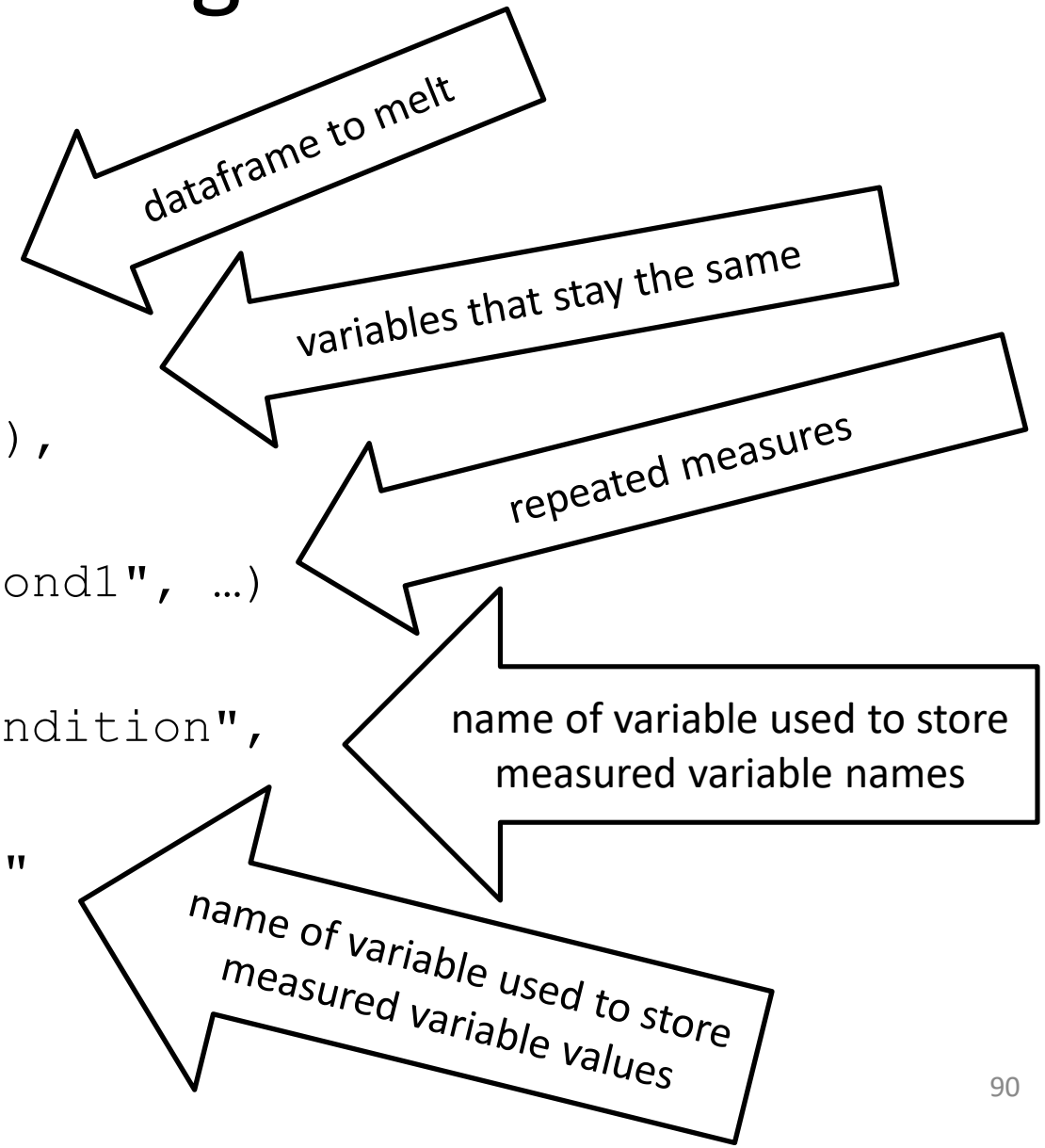
data = yourdata,

id.vars = c("ID", …),

measure.vars = c("cond1", …)

variable.name = "Condition",

value.name = "score"
)
```



From wide to long with melt()

You need to load the package **reshape2**

Example with those BPD variables

```
select(BPD, ID, Sex, Age, SelfHurtPRE:SelfHurt3Months) %>%
  melt(id.vars = c("ID", "Sex", "Age"), value.name = "score")
```

Now you

The file also includes an evaluation of Anger before the therapy, after the therapy, and at 1, 2, and 3 months.

Create a dataset in long-format with that information and call it BPD_long.

Solution

```
BPD_long <- select(BPD, ID, Sex, Age,
                        AngerPRE:Anger3Months) %>%
  melt(id.vars = c("ID", "Sex", "Age"),
               value.name = "score")
```

From long to wide – function dcast

Function dcast from package reshape2

```
ew <- dcast(data,
       formula =  subject_variables ~ condition,
       value.var = "variable of interest")
```

- data = the data to reshape
- left hand side of formula: Variables that should be in separate rows
- right hand side of formula: Variables that should be in separate columns
- value.var = The variable in which the values of interest are stored

Example

```
BPD_wide <- dcast(BPD_long,
       ID + Sex + Age ~ variable,
       value.var = "score")
```

Now you

Convert again BPD_wide to long format (call it BPD_long2) and again to wide format (call it BPD_wide2)

Solution

```
BPD_long2 <- melt(BPD_wide, id.vars = c("ID", "Sex", "Age"),
                  value.name = "score")
```

```
BPD_wide2 <- dcast(BPD_long2, ID + Sex + Age ~ variable,
                   value.var = "score")
```

Introduction to data analysis and processing with R – Part 3

Q&A

Any question regarding parts 1 and 2?

https://bit.ly/3ArFMVR

Recap parts 1 and 2

1. Install/load packages dplyr, psych, haven, readr
2. Import the file BPD.csv and the file cars.sav
3. Calculate summary statistics for the BPD sample
    1. mean and SD of Age
    2. Number of male and female participants
4. Compute the same summary statistics separately by Therapy condition (vaiable «Therapy»)

Tip. To count the number of males and females, you should, within the summarize function:

A) Create a vector of logical values, in which each observation with the desired gender is marked as TRUE and other observations are marked as FALSE

B) Sum the truth values (TRUE is = 1, FALSE = 0)

Solution

```
if(!require("pacman")) install.packages("pacman")
p_load("psych", "dplyr", "haven", "readr")

BPD <- read_csv("data/BPD.csv")
cars <- read_sav("data/cars.sav")

summarize(BPD,
          Mage = mean(Age),
          SDage = sd(Age),
          nM = sum(Sex == "M"),
          nF = sum(Sex == "F"))

group_by(BPD, Therapy) %>%
  summarize(Mage = mean(Age),
          SDage = sd(Age),
          nM = sum(Sex == "M"),
          nF = sum(Sex == "F"))
```

Importing dataset in R packages

Some R packages also include datasets, that are used for examples. For instance, the ANT data are from the R package ez.

Importing in your environment data from an R package is easy:

- Install/load the package
- Use the command data(dataset1, dataset2, …) to import the data of interest

Try importing:

- The ANT data from pacakge ez
- The bfi data from package psych

Solution

```
if(!require("pacman")) install.packages("pacman")
p_load("ez", "psych")

data(ANT, bfi)

# you can also get information about the data using
?ANT
?bfi
```

Cross-tabulation

- Funciton **table()** allows you to cross-tabulate two (or more) categorical variables
  (i.e., factors, but also numeric/character variables that can be interpreted as factors work).
- **Now you**: Cross-tabulate the Therapy condition and Sex in BPD data

Solution

```
select(BPD, Sex, Therapy) %>%
table()
```

Chi-square test for independence

Test the null hypothesis H_0 that cell probablities are equal to probabilities that would be expected if the two variables are independent.

In R, the function is **chisq.test()**

Most relevant input options:

- x. a crosstabulation, e.g., the output of table()
- simulate.p.value. If TRUE, the p-value is estimated via a Monte Carlo simulation.
- B. Number of Monte Carlo replicates.

Now you. Test the hypothesis that the Therapy condition is independent of gender.

Solution

```
select(BPD, Sex, Therapy) %>%
  table() %>%
  chisq.test()
```

Now you

The bfi data in package psych include 25 Big Five items from 2800 subjects, plus their gender (1 = M, 2 = F), age, and level of education (1 = highschool studies – 5 = graduate degree).

1. Cross-tabulate gender and education
2. Test the hypothesis of independence

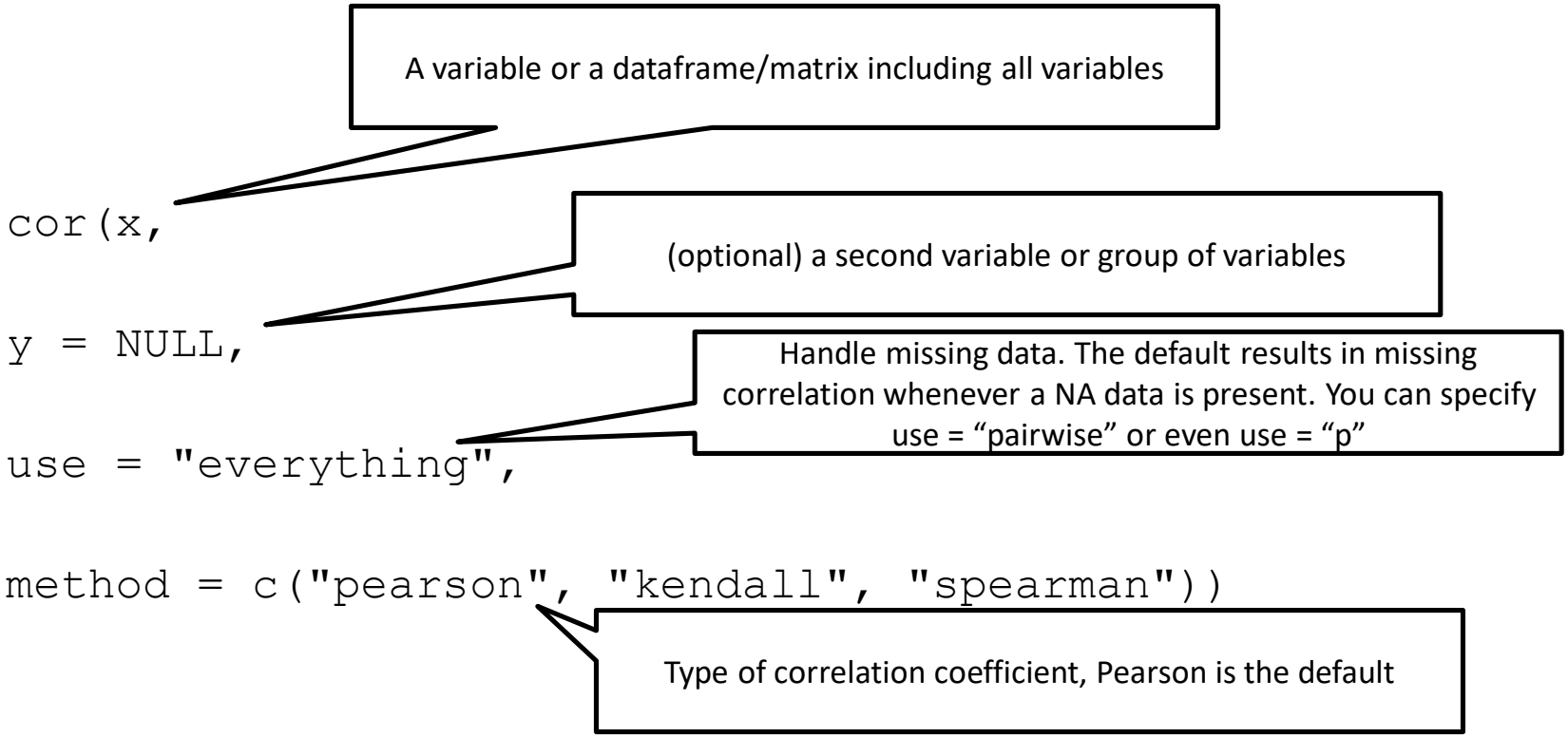
Solution

```
select(bfi, gender, education) %>%
  table()

select(bfi, gender, education) %>%
  table() %>%
  chisq.test()
```

Correlation

Function cor



Correlation exercise #1

Compute the correlations between the following pairs of variables within the cars dataset

1. length and weight
2. hp and price
3. correlation matrix of length, bag, weight, hp, and tank
4. Same as previous point, but Spearman's rank correlation matrix

Solution

```
select(cars, length, weight) %>% cor()

select(cars, hp, price) %>% cor()

select(cars, length:tank) %>% cor()

select(cars, length:tank) %>% cor(method =
"spearman")
```

Correlation exercise #2

If you want also p-values and possibly confidence intervals, you can use function corr.test from package psych.

1. Quickly check the help file of function corr.test() to see how it works

Tip. It works very much like function cor() , with some additional for correcting for multiple comparisons

2. Compute the correlation matrix of length, bag, weight, hp, and tank using corr.test and inspect the p-values (both above and below the diagonal – they are different! Why?)

Solution

```
select(cars, length:tank) %>% corr.test()
```

```
Call:corr.test(x = .)
Correlation matrix
       length  bag weight   hp tank
length   1.00 0.00   0.76 0.31 0.61
bag      0.00 1.00   0.00 0.04 0.12
weight   0.76 0.00   1.00 0.79 0.74
hp       0.31 0.04   0.79 1.00 0.69
tank     0.61 0.12   0.74 0.69 1.00
Sample Size
[1] 24
Probability values (Entries above the diagonal are adjusted for multiple tests.)
       length  bag weight   hp tank
length   0.00 1.00      0 0.68 0.01
bag      0.99 0.00      1 1.00 1.00
weight   0.00 0.99      0 0.00 0.00
hp       0.14 0.84      0 0.00 0.00
tank     0.00 0.58      0 0.00 0.00

To see confidence intervals of the correlations, print with the short=FALSE option
```

Solution (2)

If you also want CIs, you need to print the output of corr.test with the short = FALSE option

```
select(cars, length:tank) %>% corr.test() %>%
print(short = FALSE)
```

```
 Confidence intervals based upon normal theory.  To get bootstrapped values, try cor.ci
            raw.lower raw.r raw.upper raw.p lower.adj upper.adj
lngth-bag       -0.41  0.00      0.40  0.99     -0.41      0.40
lngth-weght      0.52  0.76      0.89  0.00      0.38      0.92
lngth-hp        -0.10  0.31      0.64  0.14     -0.23      0.71
lngth-tank       0.27  0.61      0.81  0.00      0.13      0.86
bag-weght       -0.41  0.00      0.40  0.99     -0.46      0.45
bag-hp          -0.37  0.04      0.44  0.84     -0.45      0.51
bag-tank        -0.30  0.12      0.50  0.58     -0.40      0.58
weght-hp         0.56  0.79      0.90  0.00      0.43      0.93
weght-tank       0.48  0.74      0.88  0.00      0.34      0.91
hp-tank          0.40  0.69      0.86  0.00      0.26      0.89
```

Scatterplot with regression line

pairs.panels() is a convenient function in package psych that visualizes a SPLOM, a scatterplot with smoother (or regression line) between all pairs of variables

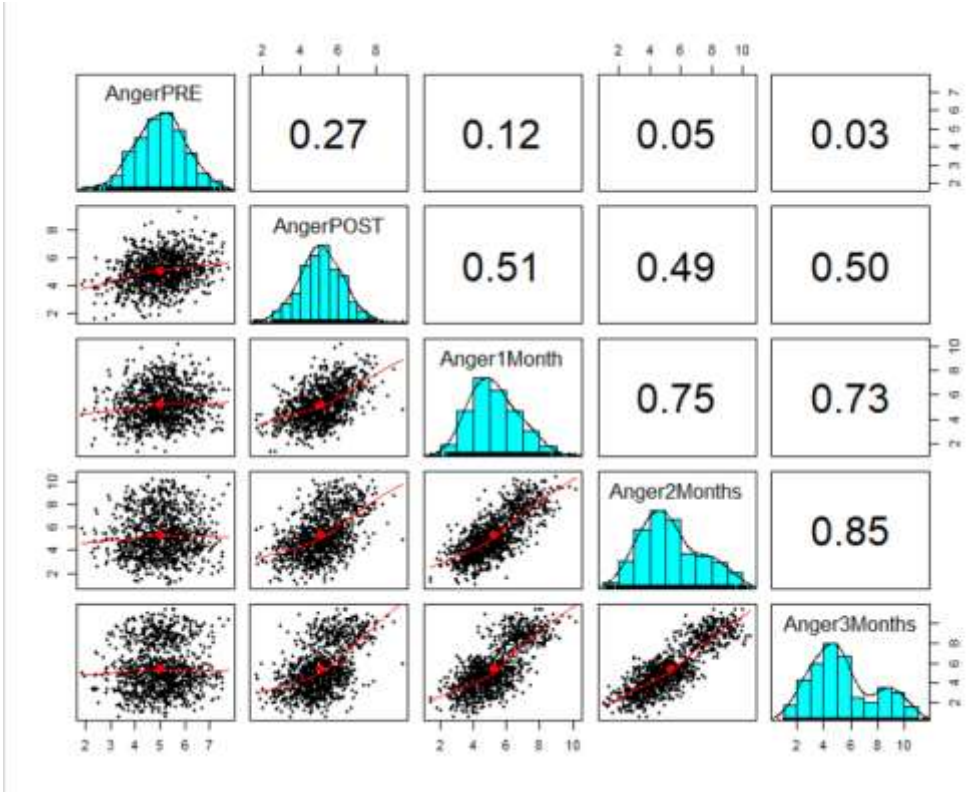
```
select(cars, length:tank) %>% pairs.panels()
select(cars, length:tank) %>% pairs.panels(lm = TRUE)
```

Exercise SPLOM

1. In the BPD data, visualize a SPLOM of variables AngerPRE, AngerPOST, …, Anger3Months
2. Compute also the Spearman correlation matrix of the same variables, including also p-values

Solution

```
select(BPD, AngerPRE:Anger3Months) %>%
pairs.panels()
select(BPD, AngerPRE:Anger3Months) %>% corr.test()
```



Independent-samples T-test

Function t-test in R allows performing independent-samples and paired samples t-tests.

In case of independent samples t-test, you typically want to use it like this:

t.test(DV ~ IV, data = yourdata)

t.test(price ~ air, data = cars)

Paired samples t-test

In case of paired sample, data in wide format are expected and the type of input changes.

t.test(data$X1, data$X2, paired = TRUE)

e.g.,

```
t.test(x = BPD$AngerPRE,
       y = BPD$AngerPOST,
       paired = TRUE)
```

Now you

1) Calculate a mean score for agreeableness in the bfi data (mean of items between A1 and A5).
   Bonus: Consider the presence of missing values: by default, listwise elimination will be applied. If you use rowMeans with na.rm = TRUE, in case of missing values, you can calculate the mean of the available items.
2) Perform an independent-samples t-test predicting Agreeableness from Gender.
3) Perform a paired samples t-test, to test the null hypothesis that the mean of items A1 and A2 are the same

Solution

```
bfi$Agreeableness <- select(bfi, A1:A5) %>%
rowMeans(na.rm = TRUE)
t.test(Agreeableness ~ gender, data = bfi)
t.test(bfi$A1, bfi$A2)
```

Visualizing data with R

Giulio Costantini,

Università degli Studi di Milano-Bicocca

giulio.costantini@unimib.it

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

Options we have seen so far

- pairs.panels() -> SPLOM
- ezPlot -> ANOVA visualization
- boxplot

Other functions

- plot(y ~ x) -> simple scatterplot
- hist(x) -> histogram
- dev.off() -> close the comunication with the current plotting device

ggplot2

A general way to get wonderful plots in R is ggplot2

ggplot2 has its own logic, different types of layers that compose the plot.

- ggplot: basic functions specifying the data and the aesthethics (aes)
- aes: mapping between a variable and a characteristic of the plot (e.g., shape corresponds to gender, size corresponds to age).
- geom: layers containing elements (e.g., bars, points, lines, text)
- facets: subset plots in different panels (e.g., different subplots by condition)

Other options

- Scales: define options for the axes.
- themes: global options – appearence of the graph
- There are A LOT of options in ggplot2 and we will not see them all. You can find much more information here: https://ggplot2.tidyverse.org/

Define aesthetics

- Load the ggplot2, readr and the dplyr package
- Import the BPD dataset

Let's first tell ggplot which is the dataset that we want to visualize, and how are the graphical elements associated to vairables (aes).

```
ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM))
```

Some aesthetics

- x: x-axis
- y: y-axis
- size: size
- color: color (of the outline)
- fill: color (the filling of a shape)
- shape: distinct shapes of points (to distinguish different groups)
- ... and many more.

Getting started with ggplot2

```
ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM))
```

- The above code alone does not visualize anything! That's because we need to tell ggplot which «geoms» we want to use.
- Geoms can be points, lines, etc., and are added to the ggplot() using the «+» symbol.
- We can specify as many geoms we want to!
- For example geom_point() visualizes points, and geom_smooth() visualizes regression lines.
- Each geom has several options
  - Look ?geom_point()
  - Google what you want to plot, that's often the easiest way

Adding geoms

```
ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point()
```

```
ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point() +
  geom_smooth()
```

Adding options to geoms

```
ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point() +
  geom_smooth(method = "lm")
```

```
ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE)
```

Facets

facet_grid() and facet_wrap() allow drawing multiple plots according to one or more variables

- facet_grid() allows specifying variables that define rows and columns
  - facet_grid(rows ~.)
  - facet_grid(.~cols)
  - facet_grid(rows ~ cols)
- facet_wrap(~var1 + var2…) just places the plots one after the other

Try visualizing separate scatterplots by Sex, by Therapy or both

Solution(s)

```
# with facet_grid
ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  facet_grid(.~ Therapy)

ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  facet_grid(Sex~ .)

ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  facet_grid(Sex~ Therapy)

# with facet_wrap
ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  facet_wrap(.~ Therapy+Sex)
```

Themes

There is plenty of themes for ggplot, some examples that you can try adding to your plot:

- theme_bw()
- theme_classic()
- theme_line()

Furthermore, theme() lets you set all options manually (try this when you have a lot of time to experiment)

Themes

```
ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  facet_wrap(.~ Therapy+Sex) +
  theme_bw()

ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  facet_wrap(.~ Therapy+Sex) +
  theme_classic()

ggplot(data = BPD, aes(x = BPDCL, y = SCL_SOM)) +
  geom_point() +
  geom_smooth(method = "lm", se = FALSE) +
  facet_wrap(.~ Therapy+Sex) +
  theme_light()
```

Now you

Draw a histogram of variable BPDCL, considering that

1) The geom for histograms is geom_histogram()
2) geom_histogram only requires either the x or the y aesthetic, but not both

Also draw separate histograms by Therapy

Solution

```
ggplot(data = BPD, aes(x = BPDCL)) +
  geom_histogram()
```

```
ggplot(data = BPD, aes(x = BPDCL)) +
  geom_histogram() +
  facet_wrap(~Therapy)
```

Boxplot, violin plot

1. Create a series of boxplots, in which on the X axis you have the Therapy condition and on the Y axis you have SelfHurtPOST, knowing that geom_boxplot() is what you need for boxplots.
2. Do the same with violinplots – geom_violin()
3. Also add datapoints to the violinplot
4. Make points transparent, setting the «alpha» option within geom_point to .1

Solution

```
ggplot(data = BPD, aes(x = Therapy, y = SelfHurtPOST)) +
  geom_boxplot()

ggplot(data = BPD, aes(x = Therapy, y = SelfHurtPOST)) +
  geom_violin()

ggplot(data = BPD, aes(x = Therapy, y = SelfHurtPOST)) +
  geom_violin() +
  geom_point()

ggplot(data = BPD, aes(x = Therapy, y = SelfHurtPOST)) +
  geom_violin() +
  geom_point(alpha = .1)
```

Let's add some color

Color is just another aesthetic, you can associate it to a variable and the different levels of that vairable (continuous or discrete) will be automatically associated to different colors

Advanced: You can customize the colors, using commands scale_color_...

For example, **associate different levels of Therapy to colors**

Solution

```
plot(data = BPD, aes(x = Therapy, y = SelfHurtPOST,
color = Therapy)) +
  geom_violin() +
  geom_point(alpha = .1)
```

Bring R with you!

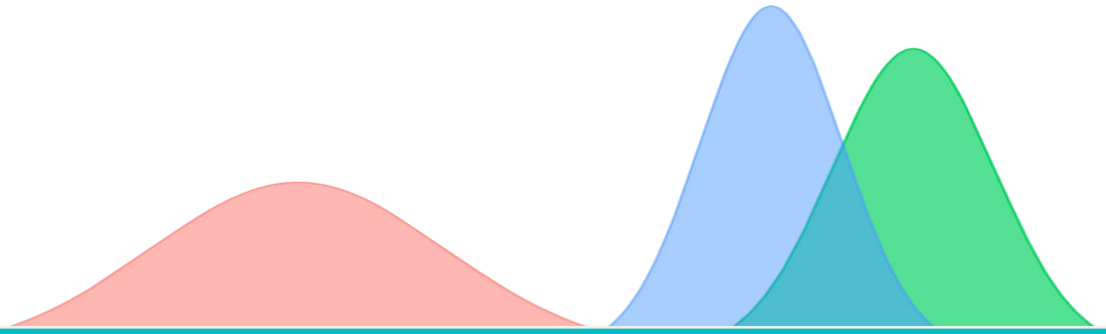
- During this lectures, we only scratched the surface of what you can do with R and ggplot.
- For each of the topics we have seen, there are plenty of additional possibilities. No human being can know them all.
- At the beginning, R is scary and you may think that it is not for you! If you do not have previous R programming experience, it is absolutely normal that you feel that you have been struggling so far.
- The best way to learn R is to keep using it despite failure. Try and try again and do not give up!
- **It takes time to learn R**, but it is also extremely rewarding, I promise.

Bayesian statistics and hypothesis testing.
Conceptual issues and basics

Daniele Romano,

Università degli Studi di Milano-Bicocca

daniele.romano@unimib.it



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

Who is speaking

daniele.romano@unimib.it

daniele.romano84 / danromano

UNIVERSITA' DEGLI STUDI DI MILANO BICOCCA

UNIVERSITAS STUDIORUM LUPIENSIS

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

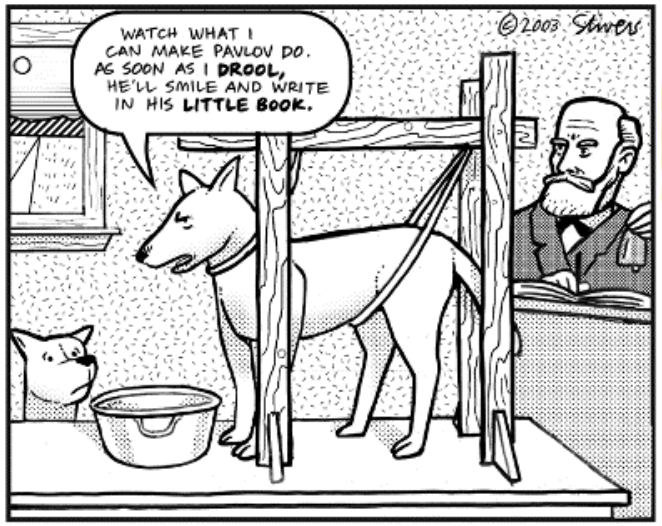
Birkbeck UNIVERSITY OF LONDON

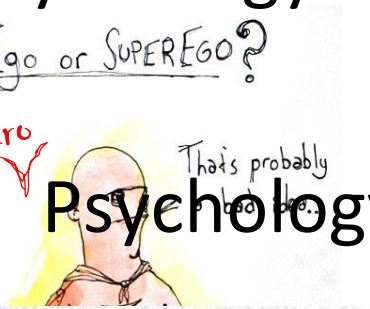
UNIVERSITA' CAMPUS BIO-MEDICO DI ROMA

UCL Université catholique de Louvain

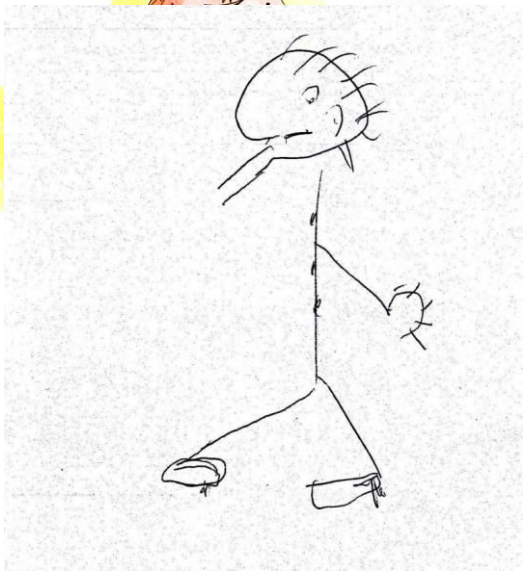
Psychology

Experimental Psychology

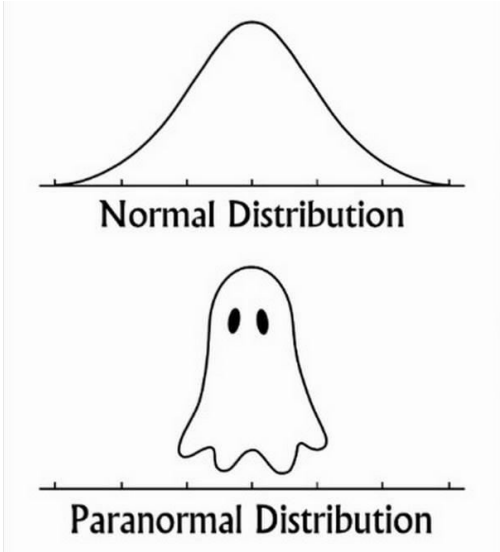




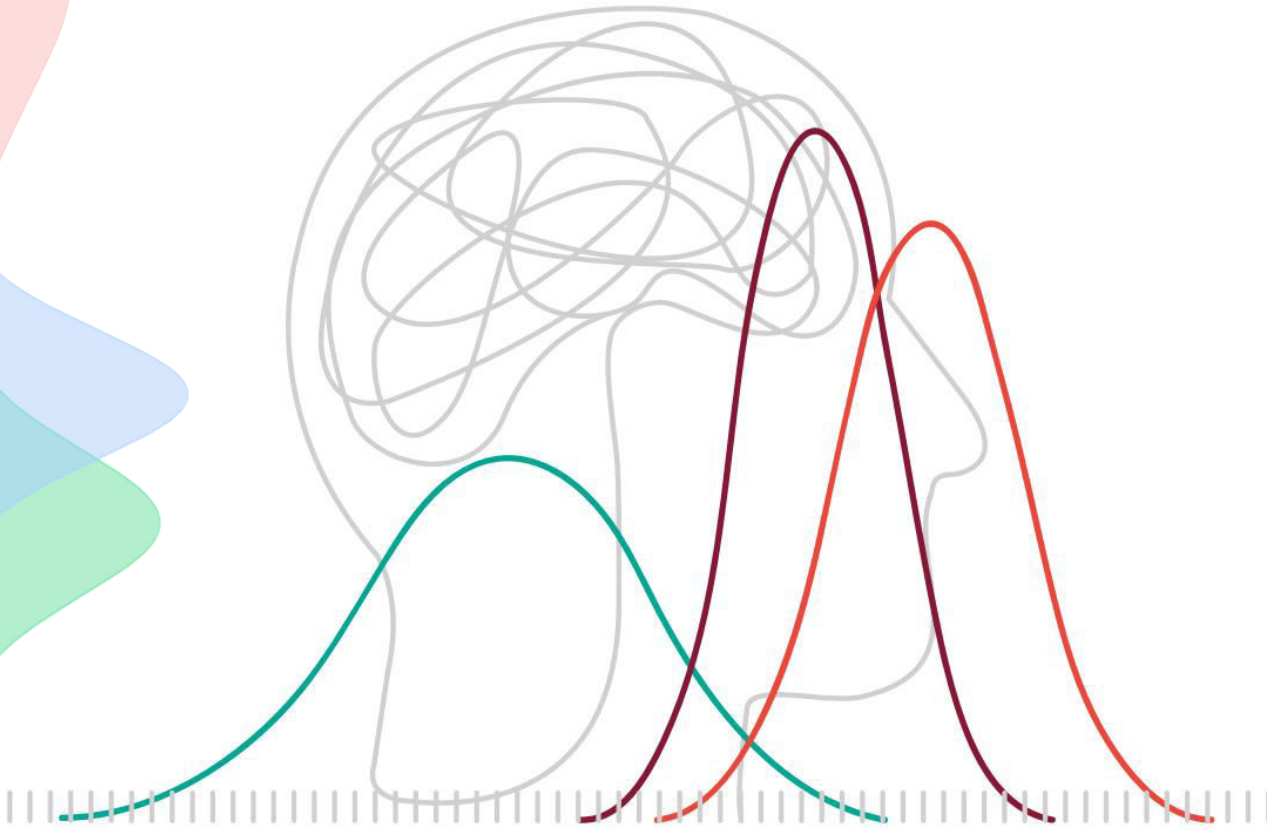
Neuro Psychology



Psychology ~~Metric~~



UNIVERSITÀ di VERONA
Dipartimento di SCIENZE UMANE



Monday May 31st to Saturday June 5th, 2021, University of Verona, Verona, Italy.

LECTURERS

Rosalyn Moran *(King's College London – UK)*
Richard Morey *(Cardiff University – UK)*
Mel Slater *(University of Barcelona – Spain)*
Massimiliano Badino *(University of Verona – Italy)*
Fabio Spizzichino *(University of Rome "La Sapienza" – Italy)*
Marco Tullio Liuzza *("Magna Graecia" University of Catanzaro – Italy)*
Daniele Romano *(University of Milano-Bicocca – Italy)*
Michele Scandola *(University of Verona – Italy)*

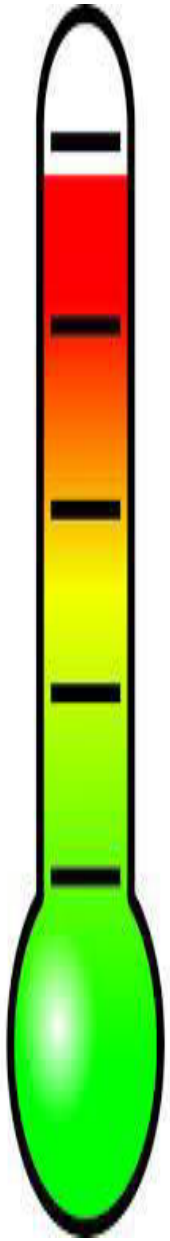
sites.hss.univr.it/**bayeshsc**
#BayesHSC2020
Deadline for applications is **March 15, 2020**

2nd BAYESIAN STATISTICAL ANALYSES
FOR THE HUMAN, SOCIAL AND COGNITIVE SCIENCES

DIRECTORS

Michele Scandola *(University of Verona – Italy)*
Valentina Moro *(University of Verona – Italy)*

https://sites.hss.univr.it/bayeshsc/



Creator

Expert

User-friendly

User

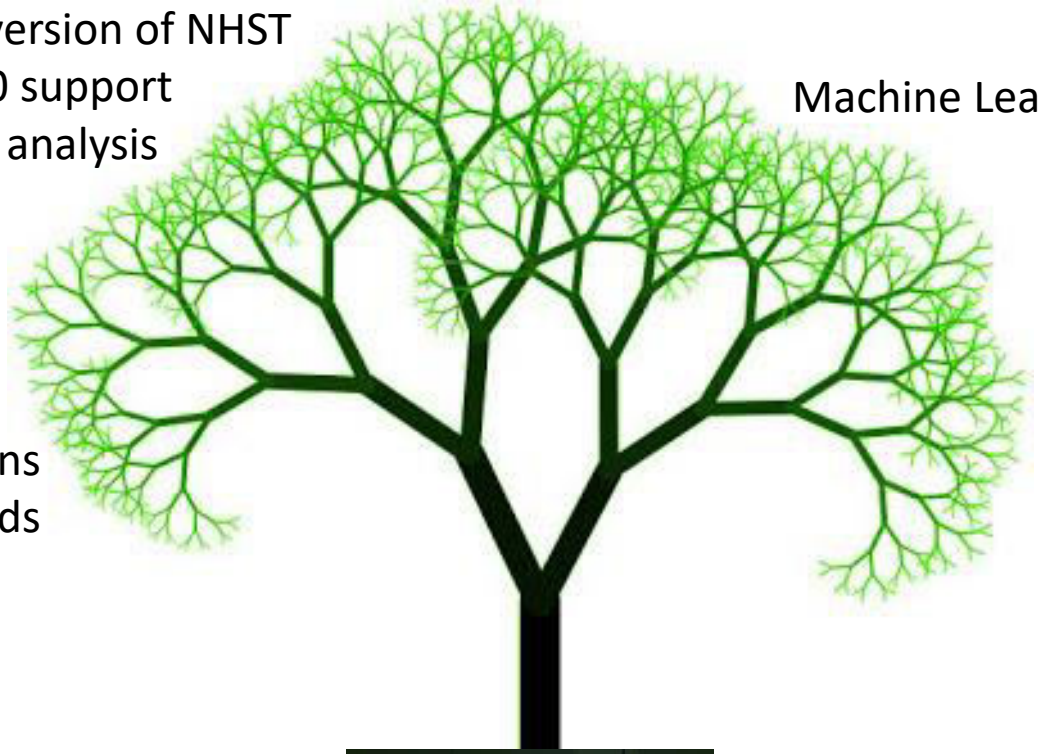
Naive

What can you do with Bayesian Basics

Imaging
(DCM)

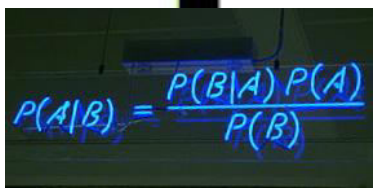
Inferential stats
- Bayesian version of NHST
- Test for H0 support
- Multilevel analysis

Machine Learning



Efficient Experimental designs
- Adaptive methods

Theoretical Models
(The Bayesian Brain)



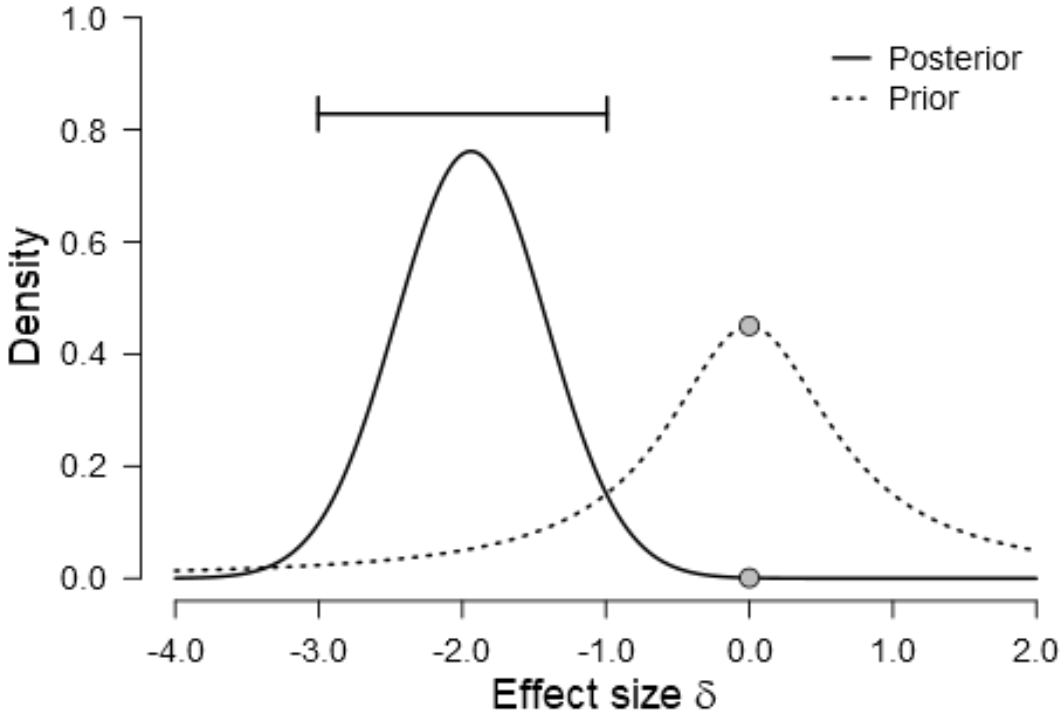
Inferential Statistics

Bayesian version of NHST

It is always necessary?
NO

When we should use it?
ALWAYS

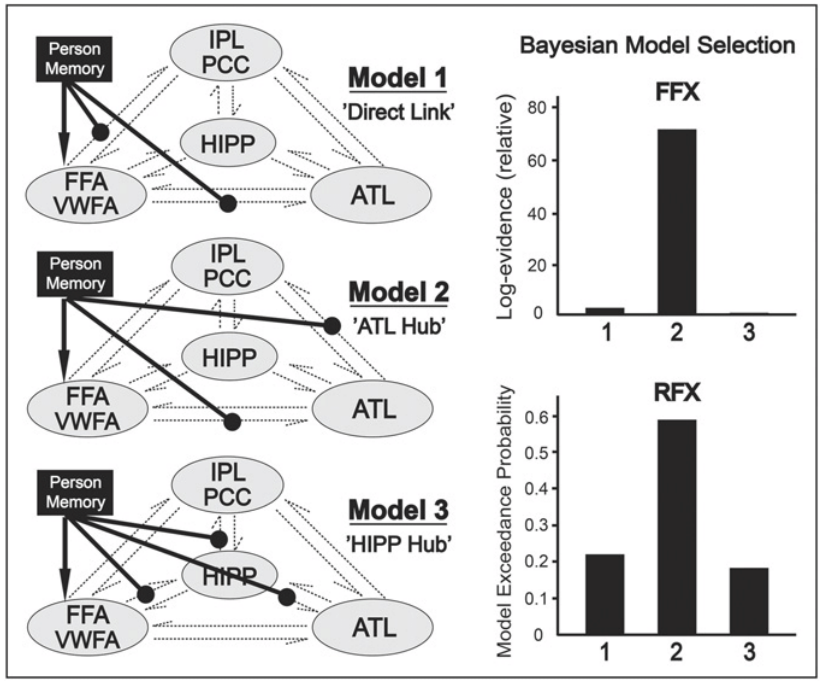
When is it advantageous?
POTENTIALLY ALWAYS

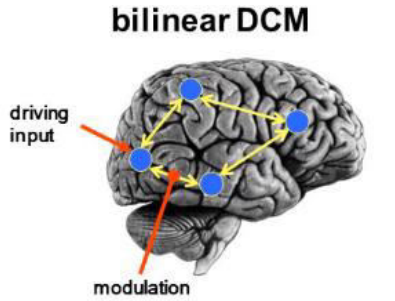


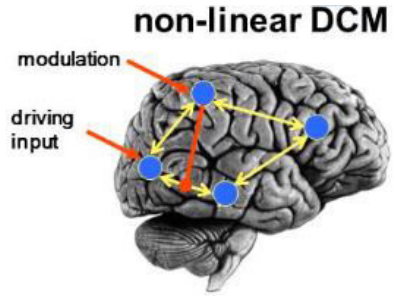
Brain Imaging

Is it always necessary?
No, it depends on the sceintific question.

However, modern questions and experiments tend to go in the direction of
Dynamic Causal Modelling (DCM)
Which gives a better comprhension of dynamic brain processes, surpassing the mere localisationism.







Efficient Experiments

PERGAMON

Vision Research 39 (1999) 2729–2737

Vision Research

Bayesian adaptive estimation of psychometric slope and threshold

Leonid L. Kontsevich *, Christopher W. Tyler

Smith-Kettlewell Eye Research Institute, 2232 Webster Street, San Francisco CA 94115, USA

Received 1 October 1997; received in revised form 10 August 1998

Bayesian adaptive method for acquisition of both threshold and slope of the psychometric function.

i.e., Threshold estimation within 2 dB (23%) precision requires less than 30 trials for a typical 2-Alternative Forced-Choice detection task. To get the slope estimate with the same precision takes about 300 trials.

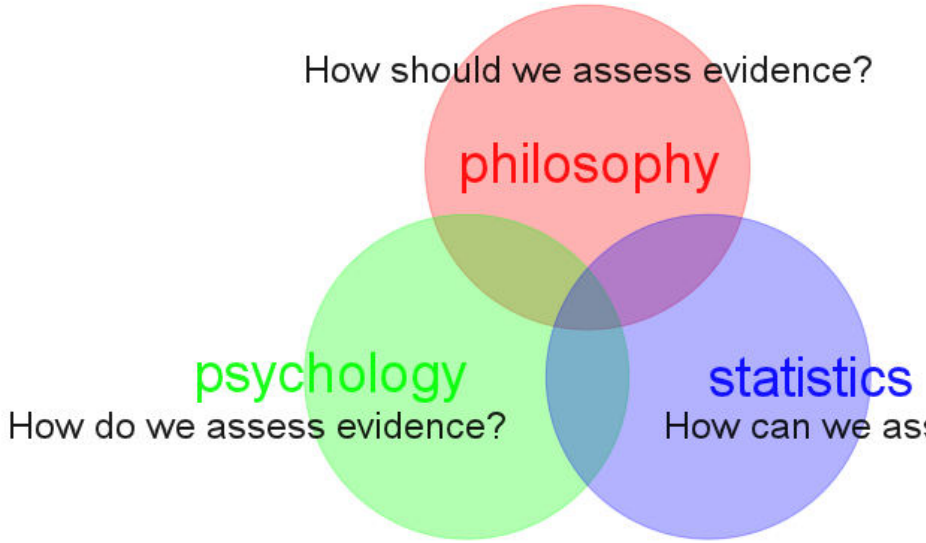
Before going Bayesian

Go to www.menti.com and enter the following code 5883 0613

Jerzy Neyman (1957): Statistics is about decision and action

Ronald A. Fisher (1955): Statistics is about knowledge and rationality

Gigerenzer et al., 2004: Statistics has become an unthinking "ritual"



"Tests should only be regarded as tools which must be used with discretion and understanding, and not as instruments which in themselves give the final verdict."

(Neyman, J., & Pearson, E. S. (1967). Joint statistical papers. Univ of California Press)

Probability

Classical definition: the probability of an event is the ratio between the number of favorable cases and the number of possible cases.

Frequentist definition: The probability of an event is defined as the relative frequency that it assumes on a large number of tests all performed under the same conditions.

Subjectivist or Bayesian theory: probability is the measure of the degree of confidence that a coherent individual assigns to the occurrence of a given event based on his knowledge.

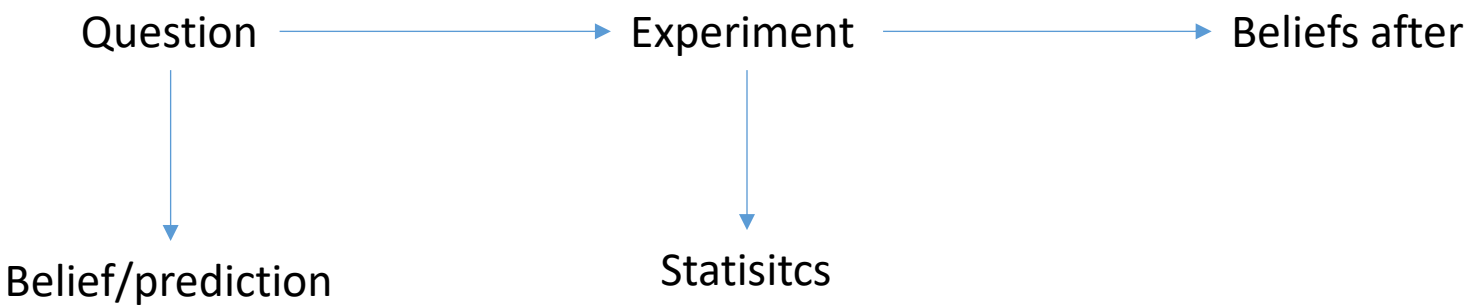
"Scientific knowledge is a body
of statements of varying degrees
of certainty --
some most unsure,
some nearly sure,
none absolutely certain."

Bayesian Statistics is about Uncertainty and how we deal with that uncertainty.

A few questions:

- How strong is the evidence for anthropogenic global warming?
- Is someone who rejects the theory of relativity rational?
- Should we believe in subliminal priming, given the literature on it?

These are all perfectly scientific questions, concerning evidence, reason, and belief.



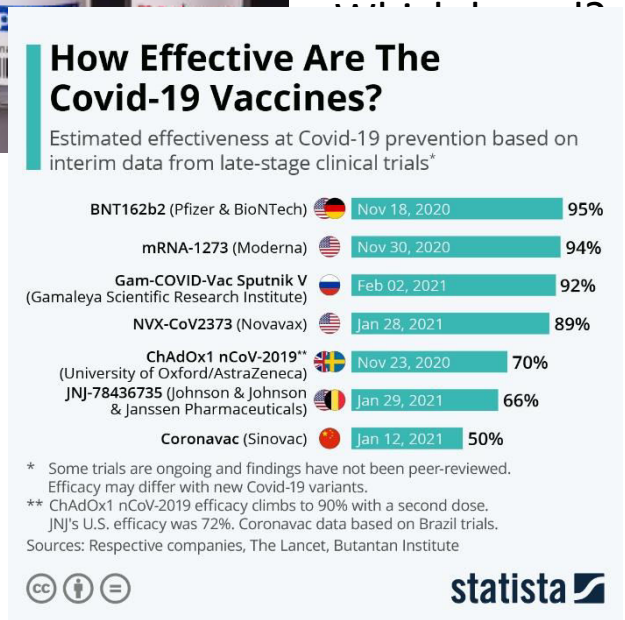
Inferential statistic is useful taking «informed» decision

Take the best **decision**, given the available information

I'm the president of Notsowealthland

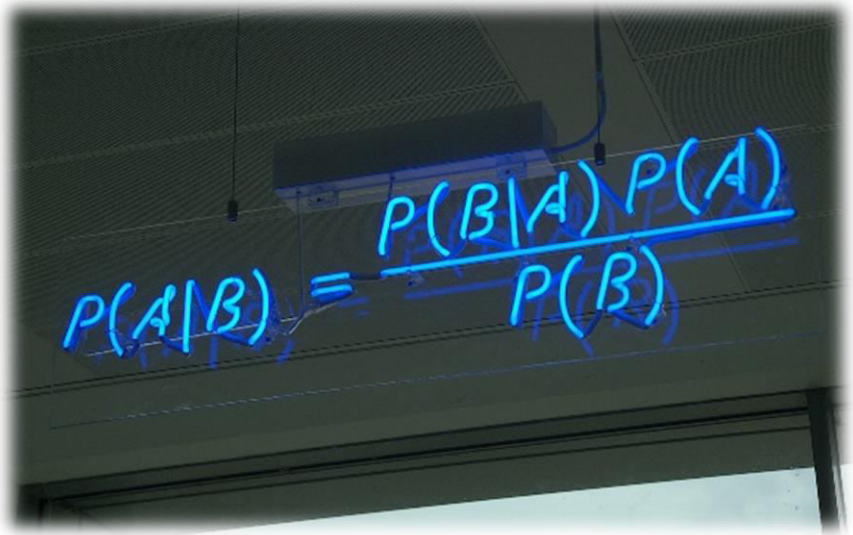
Covid spread out

Where do I invest Notsowealthland money to buy vaccine before their realization?
mRNA vaccine
Adenovirus vaccine

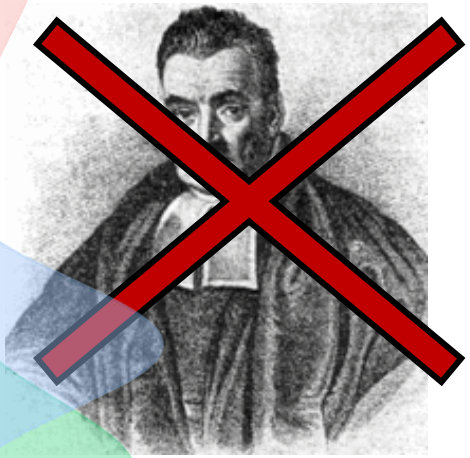


After a year of new evidence, should I revise my budget?

Let's go Bayesian



Why are we going Bayesian NOW



XVIII Century

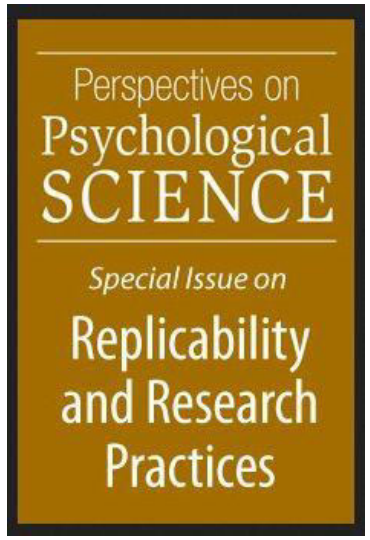
Development of Bayes Theorem

XIX Century

Development of Psychological Sciences

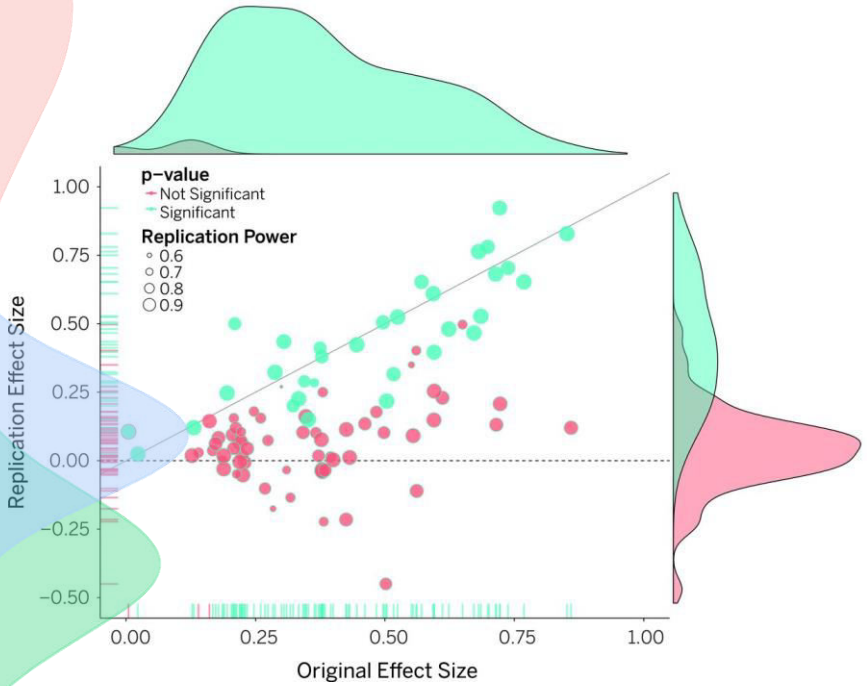
XX Century

Development of NHST

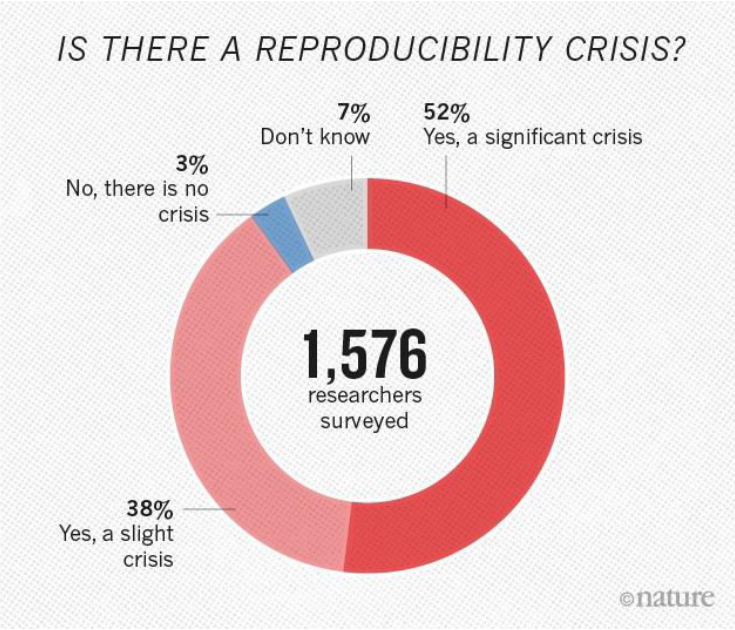


2012

Psychological Science replicability crisis


p−value
Not Significant
Significant
Replication Power
0.6
0.7
0.8
0.9
Replication Effect Size
Original Effect Size
1.00
0.75
0.50
0.25
0.00
−0.25
−0.50
0.00
0.25
0.50
0.75
1.00


Many Psychology Findings Not as Strong as Claimed, Study Says
The New York Times
Science
AAAS
nature
Over half of psychology studies fail reproducibility test
Vox
SCIENCE & HEALTH
Scientists replicated 100 recent psychology experiments. More than half of them failed.

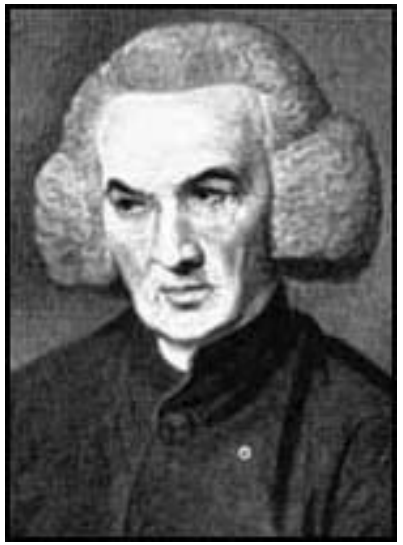

IS THERE A REPRODUCIBILITY CRISIS?
7%
Don't know
52%
Yes, a significant crisis
3%
No, there is no crisis
1,576
researchers surveyed
38%
Yes, a slight crisis
©nature

Lack of confidence in the meaning of p-value
Development of computers
Development of algorithm to calculate the posterior
New theoretical framework

What does it mean «Bayesian»

It’s a set of methods originating from a single theorem (Bayes rule) which is entangled with the idea of **Conditional probability**.

Theoretical Framework
(Bayes)



Mathematical Translation
(Laplace and Price)

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

What is the core idea of Bayes Theorem

Calculate the probability that an event occur, given the data and while considering the previous knowledge

OR

Given a set of information, that «bias» my decision, I observe data and update my believes according to the new information

OR

How new information changes what I believe

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

How it works

Basics

Conditional probability IS NOT Joint probability

If I know that B is the case, what is the probability that A is also the case?

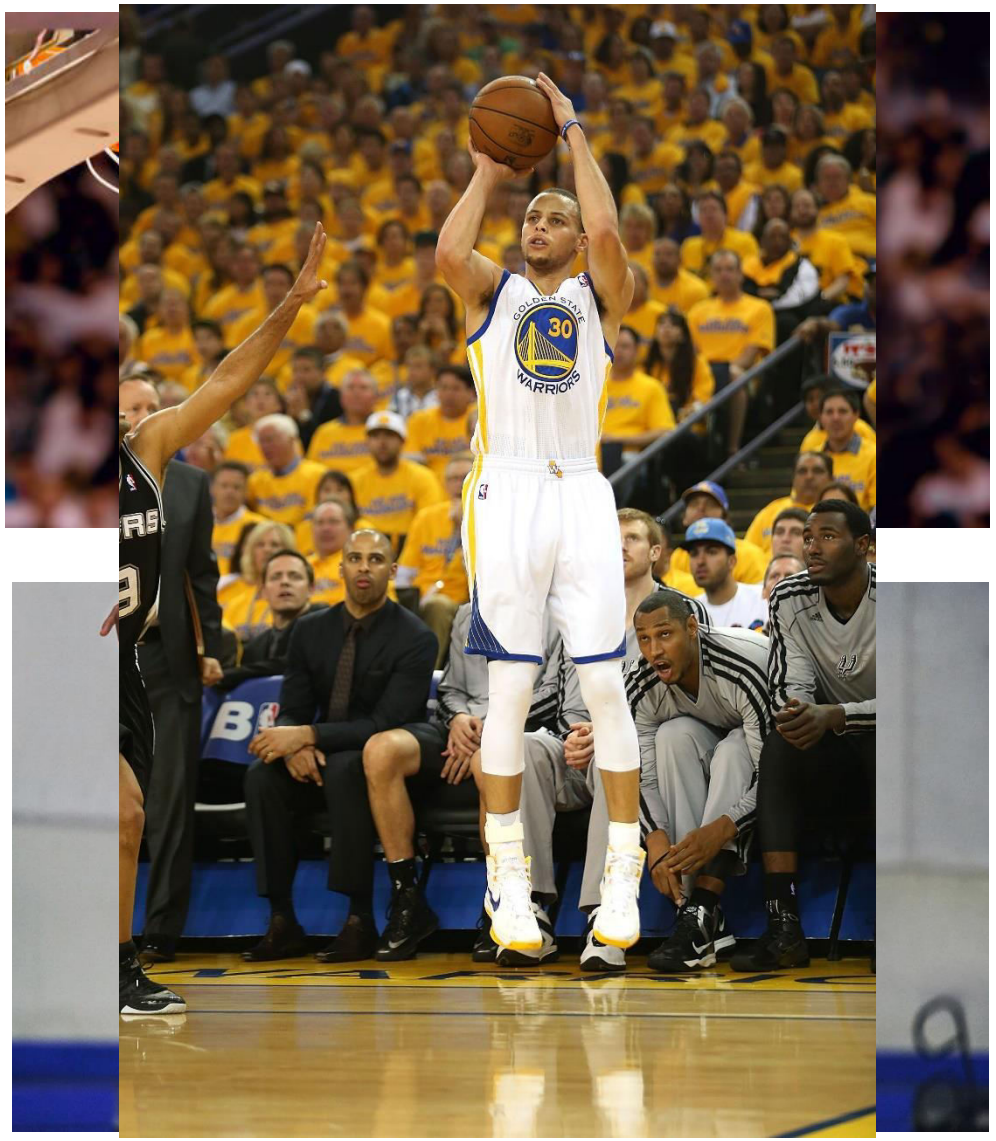
P(A|B) is not the same as P(B|A)

How likely is this a 3 pointer? P(3 pointer)

P(3 pointer|Steph Curry) is not the same as P(Steph Curry|3 pointer)

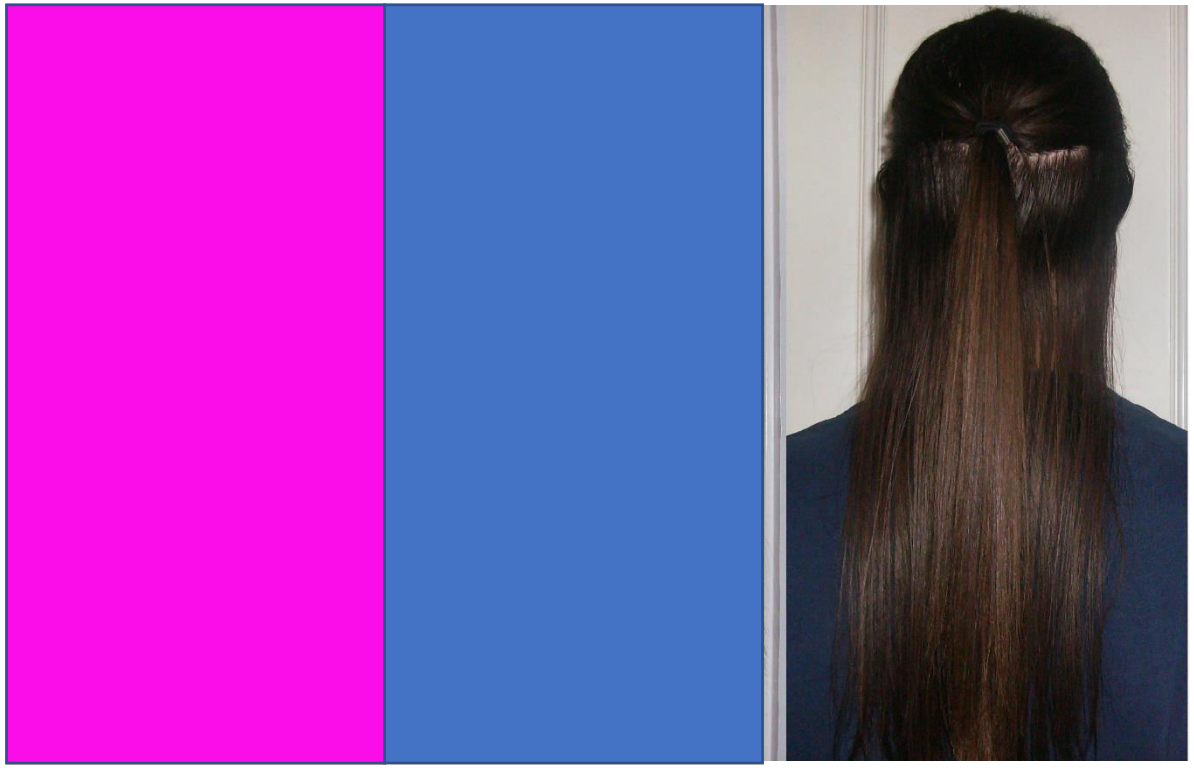
If I know that Steph is shooting,
what is the probability that it is a 3-pointer?

If I know that the shoot is a 3-pointer,
what is the probability that it was Steph shooting?



You are at the cinema, you have this person in front and you need to pass.
How do you refer to this person? Sir or Lady?

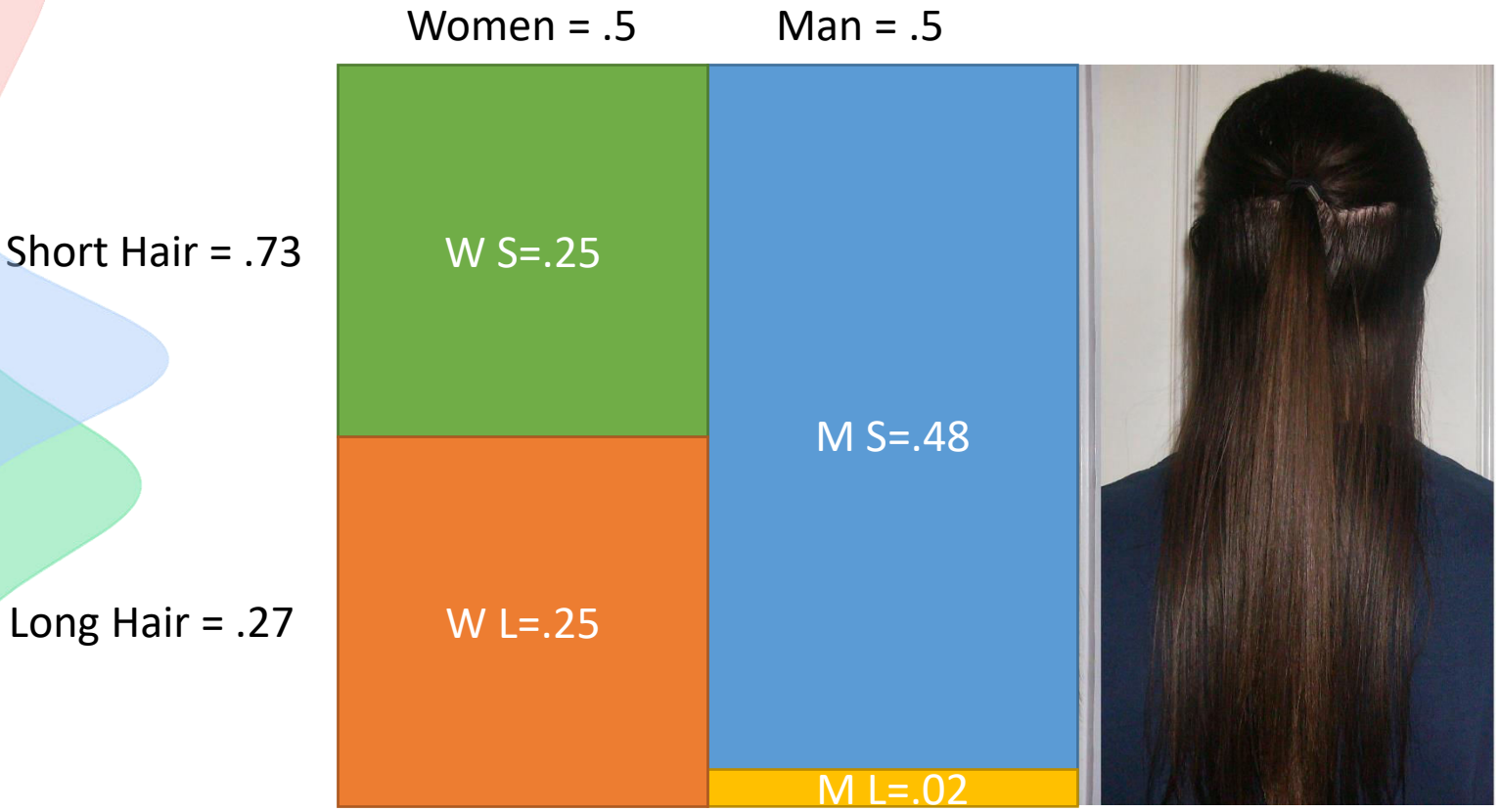
P(Woman)= .5
P(long hair|is a woman)=.5
P(long hair|is a man)=.04



Joint Probability

P(woman)*P(short hair|woman)=.5*.5=.25
P(man)*P(long hair|man)=.5*.04=.02

P(woman)*P(long hair|woman)=.5*.5=.25
P(man)*P(short hair|man)=.5*.96=.48



$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

P(man|long hair)=.04*.5/.27=.07

93% chance that is a woman.

7% chance that is a man.

«Excusme Lady»

Same cinema, same person. But this is the man restroom line.
Do you call this person Sir or Lady?

P(Woman)= .02
P(long hair|is a woman)=.5
P(long hair|is a man)=.04

Joint Probability

P(woman)*P(short hair|woman)=.02*.5=.01
P(man)*P(long hair|man)=.98*.04=.02

P(woman)*P(long hair|woman)=.02*.5=.01
P(man)*P(short hair|man)=.98*.96=.48

Women = .5 Man = .5

Short Hair = .95

Long Hair = .05

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

P(man|long hair)=.04*.98/.05=.80

20% chance that is a woman.

80% chance that is a man.

«Excusme Sir»

Exists a Rare disesease affecting 1‰ population.
A person is positive at the test which is correct in 99% of cases (False Positive 1%)

What's the probability that the person has the disease?

Probability that the test is positive
assuming that is true that I have the disease

Probability of actually having the disease
Given that the test is positive

Overall probability of having the disease
net of any other information (Prior Probability)
i.e., Frequency of the disease in the population

$$P(H|T) = \frac{P(T|H) * P(H)}{P(T)}$$

Probability that the test is positive

Exists a Rare disesease affecting 1‰ population.
A person is positive at the test which is correct in 99% of cases (False Positive 1%)

What's the probability that the person has the disease?

Probability that the test is positive
assuming that is true that I have the disease

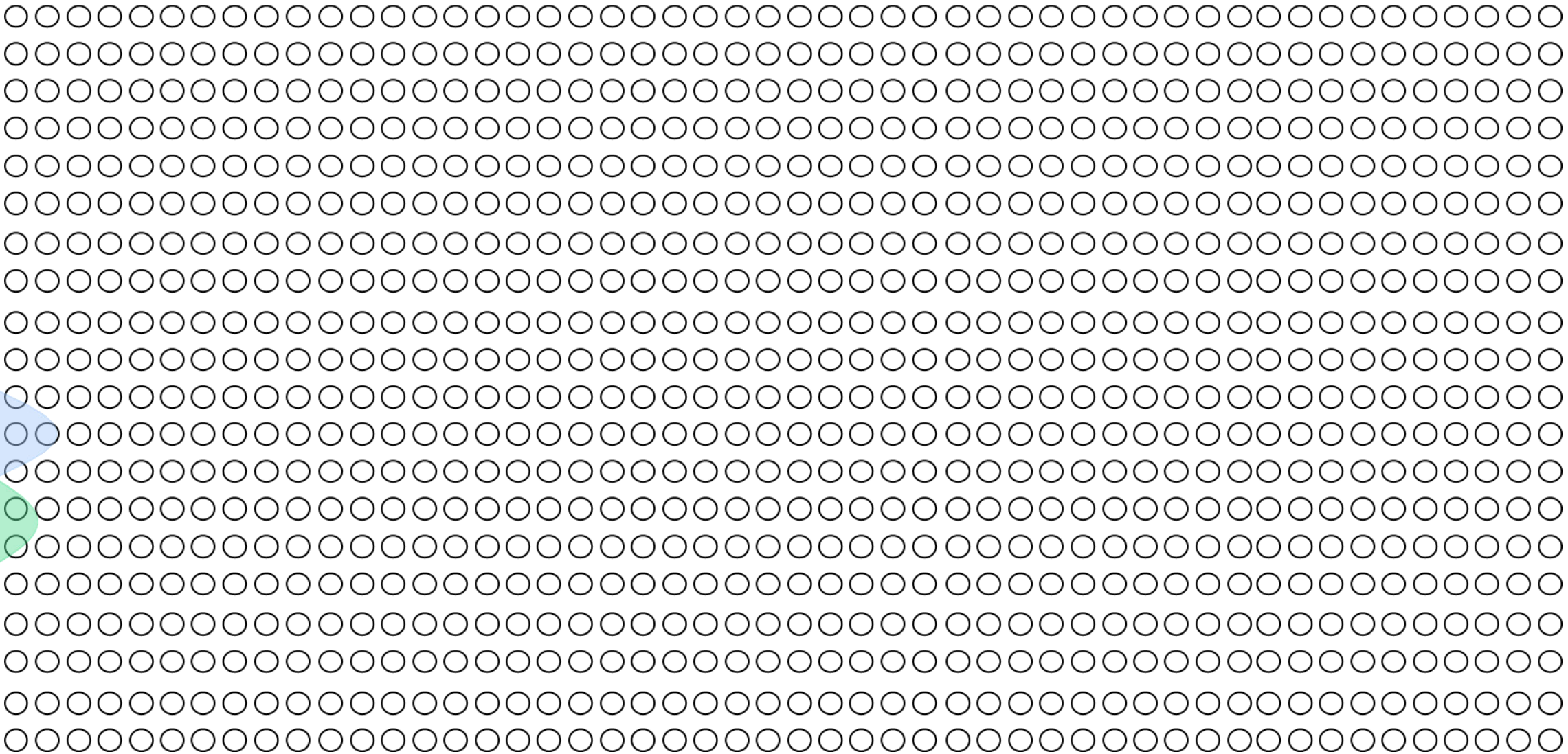
Probability of actually having the disease
Given that the test is positive

Overall probability of having the disease
net of any other information (Prior Probability)
i.e., Frequency of the disease in the population

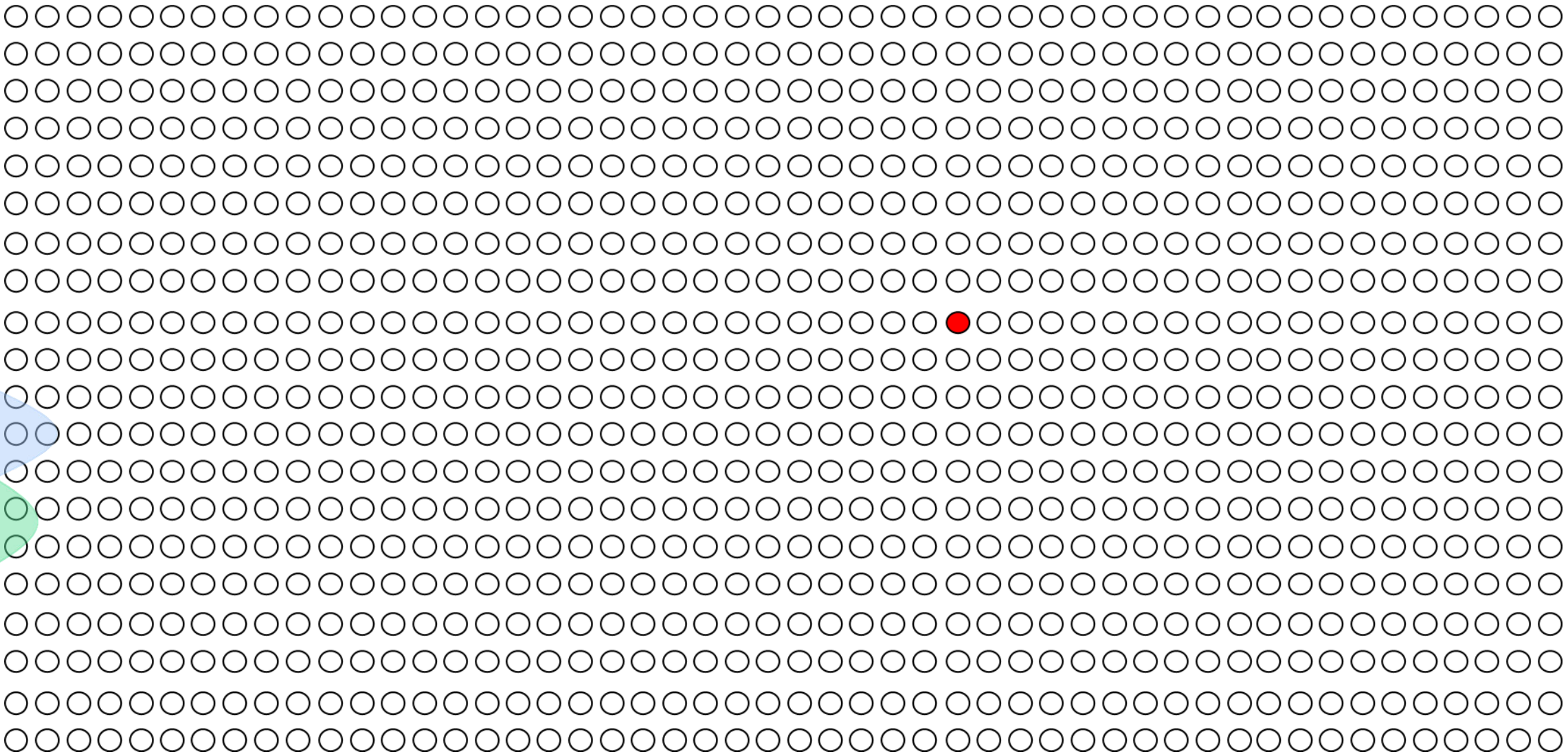
$$P(H|E) = \frac{.99 * .001}{.001*.99 + .999*.01}$$

Probability that the test is positive

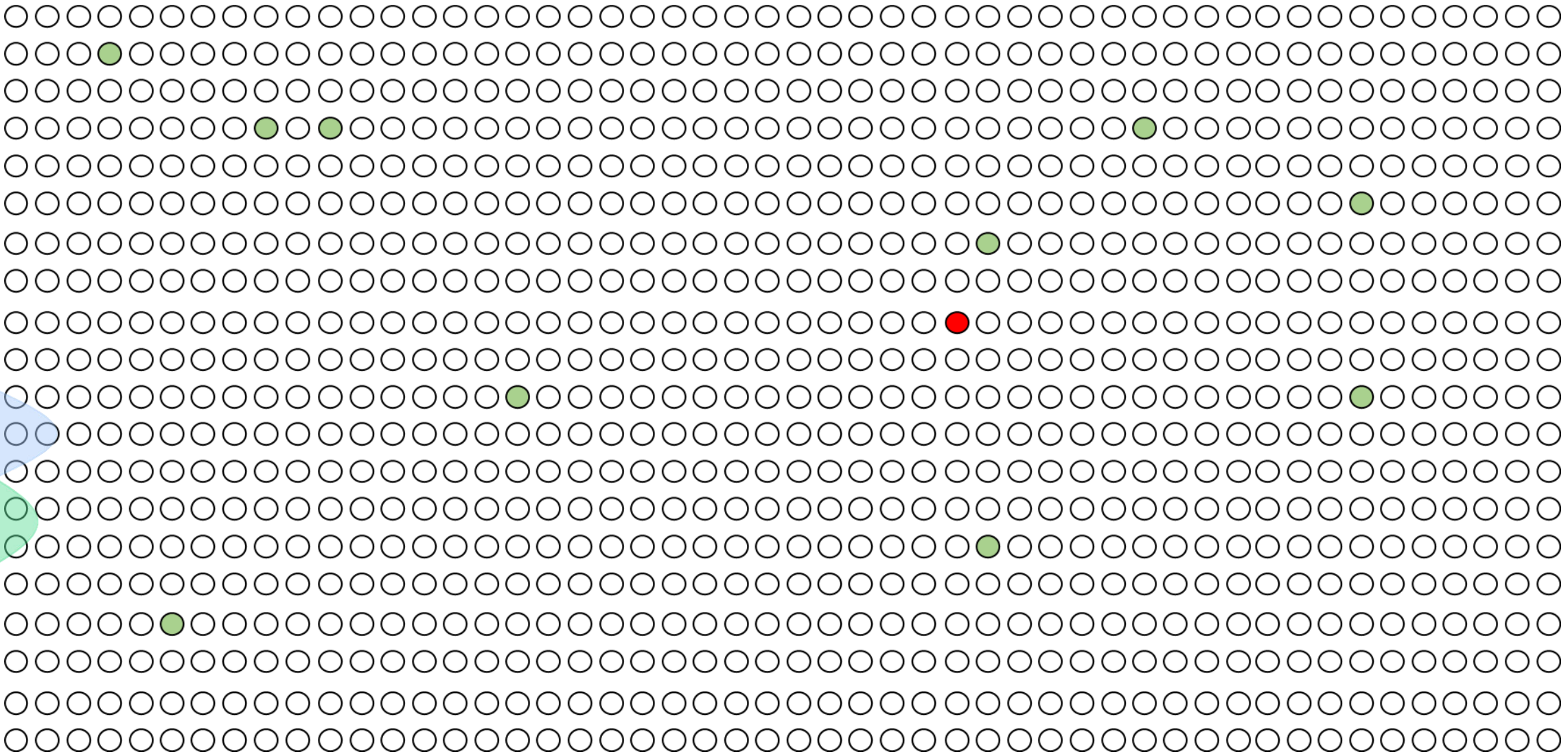
9%



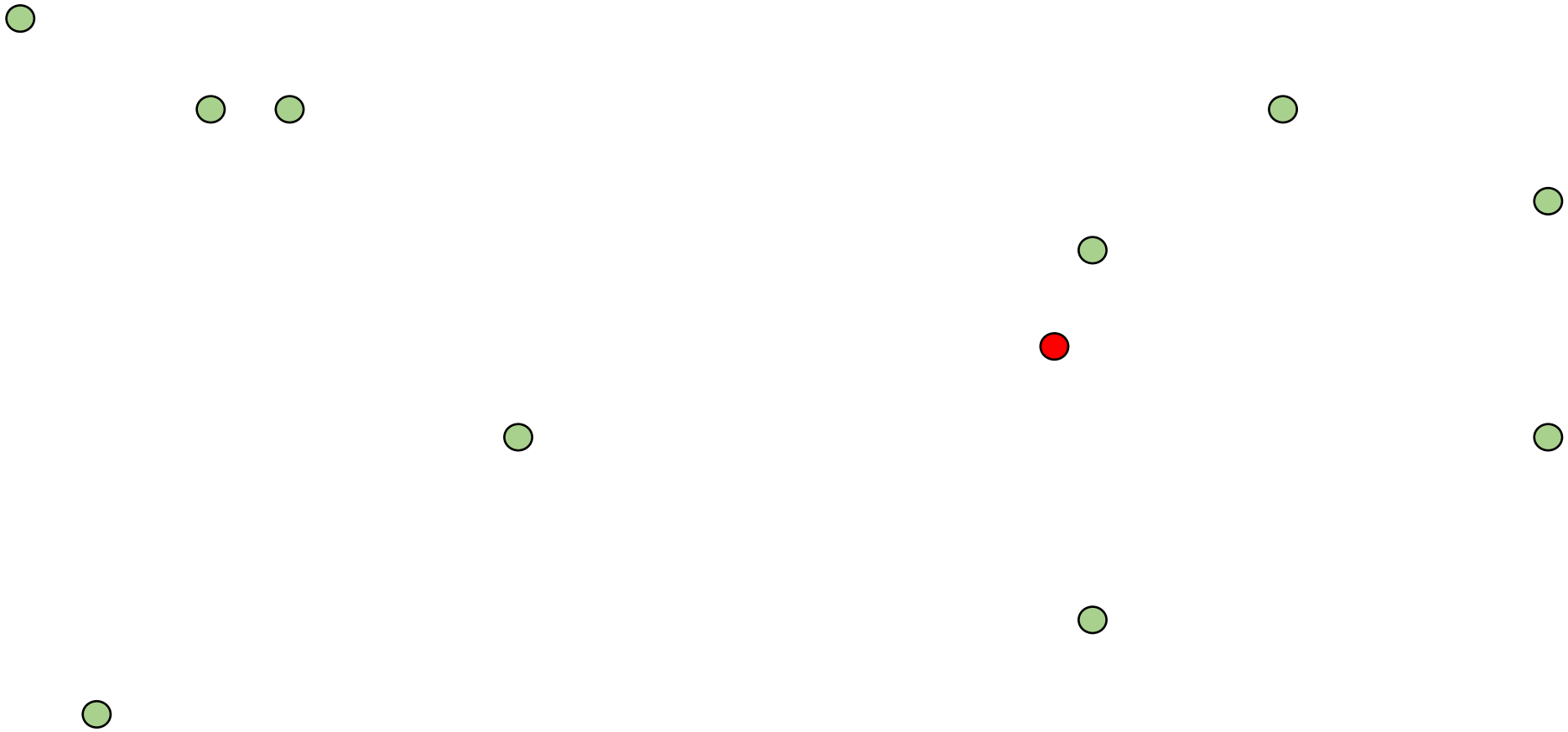
A Rare disesease 1‰ population. The test is positive. The Test is correct in 99% of cases. False positive 1%



1 out of 1000 actually has the disease



1% of false positive, means that 10 out of 1000 are positive to the test without having the disease



If your **test is positive**, you are one of those 11 people => 9% probability of actually having the disease

If you do a second (indipendent) test which is positive again

Probability that the test is positive
assuming that is true that I have the disease

Probability of actually having the disease
Given that the test is positive

Prior Probability of having the disease
(i.e. Previous posterior -> UPDATING)

$$P(H|E) = \frac{.99 * .09}{.09*.99 + .91*.01}$$

Probability that the test is positive

91%

Using probability theory to change beliefs

The central inferential principle in Bayesian statistics is Bayes' theorem.

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

deceptively simple consequence of the definition of conditional probability that achieves great significance when we apply it to beliefs:

Plausibility(A given data)=Plausibility(data given A)/Plausibility(data)*Plausibility(A)

If uncertainty is represented by probability, then can use Bayes' theorem:

$$p(\theta \mid y) = \frac{p(y \mid \theta)}{p(y)} \times p(\theta)$$

$$\text{Beliefs after (posterior)} = \text{Evidence (likelihood)} + \text{Beliefs before (prior)}$$

$p(y \mid \theta)$ is our model of the data, called the likelihood. What is $p(y)$?

$$p(y) = \int_{\Theta} p(y \mid \theta) p(\theta)\, d\theta = \text{Average likelihood}$$

- If a particular θ is better than average in explaining the data, then our belief in that θ grows; otherwise, it shrinks.

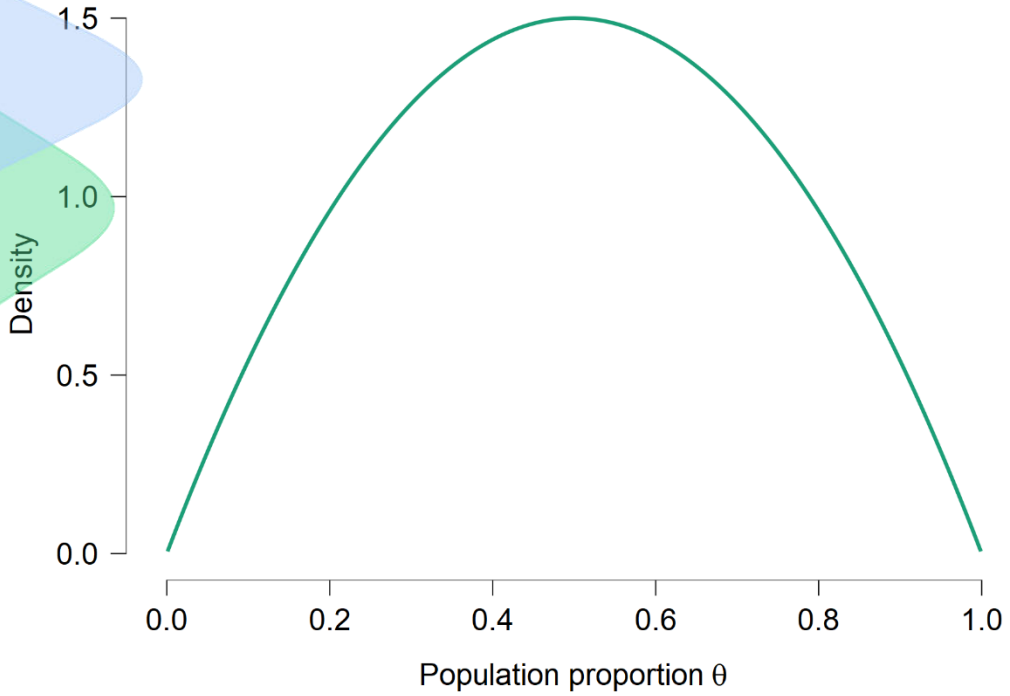
An example

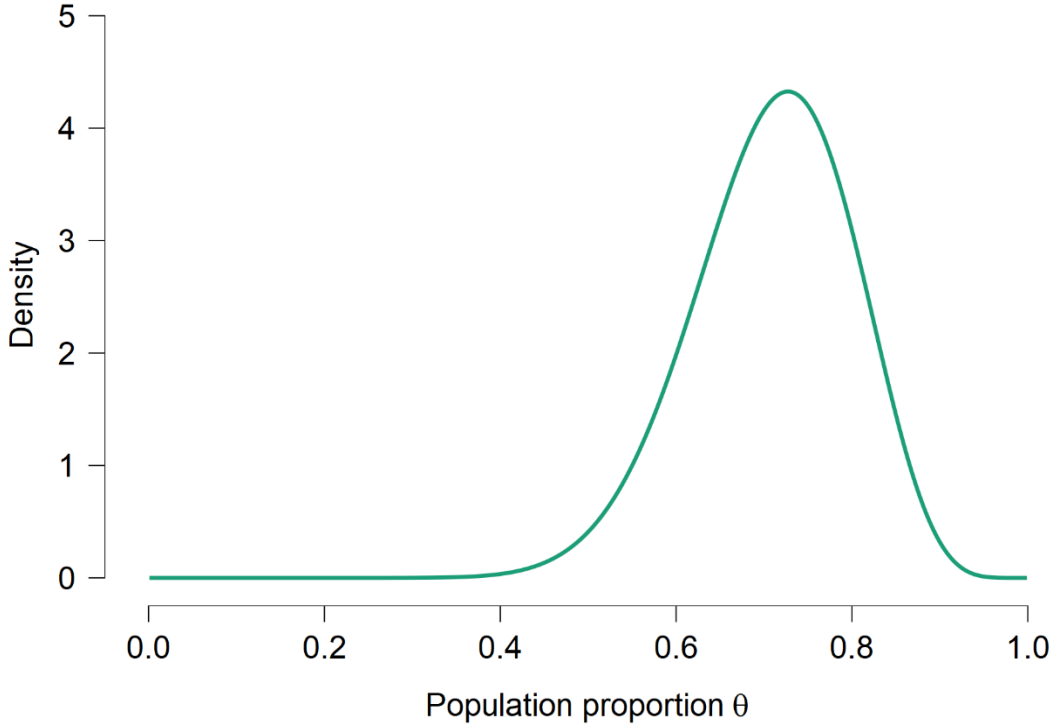
Suppose we are interested in determining the probability that an unfair coin comes up "heads". We model this as a Binomial:

$$p(y \mid \theta) = \binom{N}{y} \theta^y (1-\theta)^{N-y}$$

We have a prior belief distribution $p(\theta)$, and we observe 15 heads out of 20 flips.

We learned y=15; we want to know how the data should change our beliefs about θ.





How did this work, exactly?

Step by step, using the Beta and Binomial distributions:

Prior:

$$\theta \sim \text{Beta}(2, 2)$$

$$p(\theta) = \frac{1}{\text{Be}(2, 2)} \theta^{2-1}(1-\theta)^{2-1}$$

Likelihood:

$$y \sim \text{Binomial}(20, \theta)$$

$$p(15 \mid \theta) = \binom{20}{15} \theta^{15}(1-\theta)^{20-15}$$

Combining the prior (beliefs) and likelihood (evidence)

Likelihood times the prior

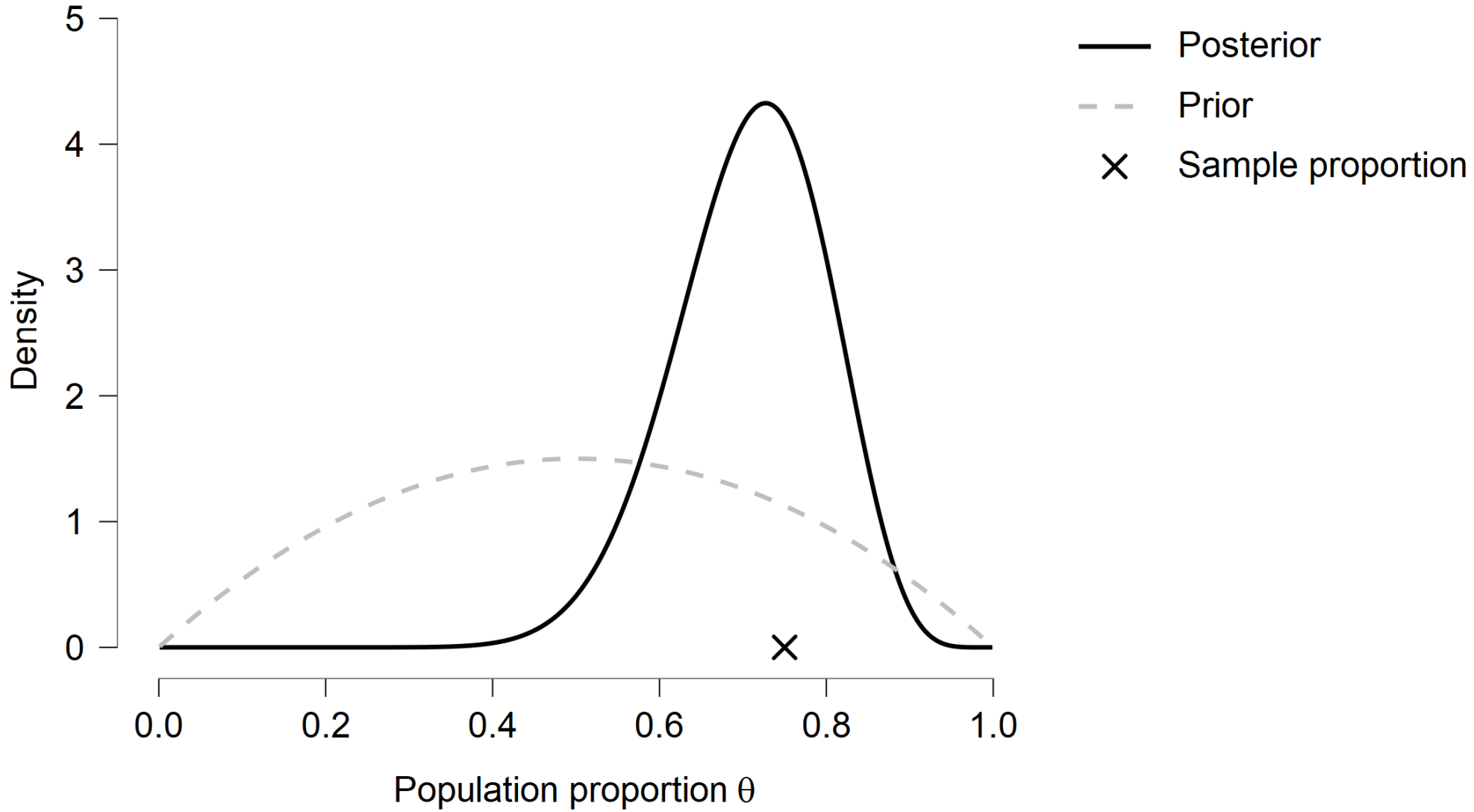
$$p(15 \mid \theta)p(\theta) \propto \theta^{15}(1-\theta)^{20-15} \times \theta^{2-1}(1-\theta)^{2-1}$$
$$\propto \theta^{17-1}(1-\theta)^{7-1}$$

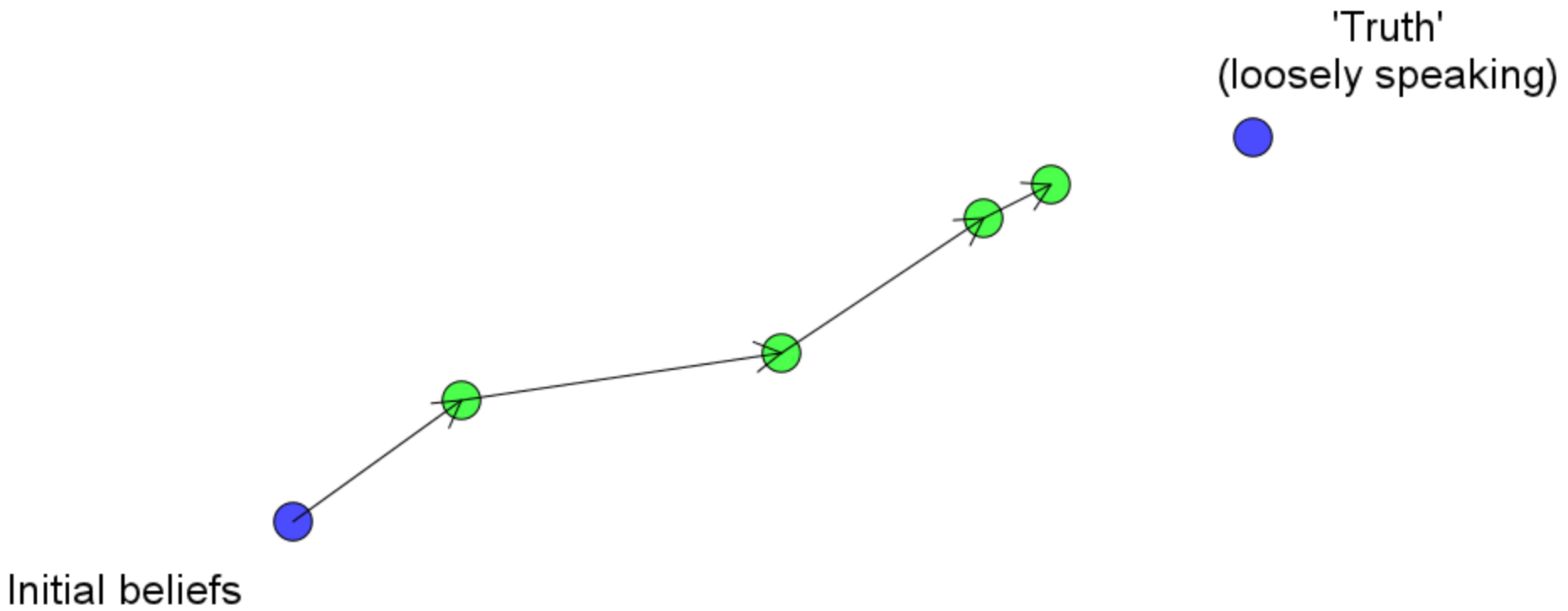
This is a $\text{Beta}(17, 7)$ distribution! The data moved our beliefs from a $\text{Beta}(2, 2)$ to a $\text{Beta}(17, 7)$.

The algebra will not always be this easy, but the principle is always the same:

$$\text{Posterior beliefs} \propto \text{Likelihood} \times \text{Prior beliefs}$$

This is what "rational" statistical learning looks like.


Posterior
Prior
Sample proportion
Density
0
1
2
3
4
5
0.0
0.2
0.4
0.6
0.8
1.0
Population proportion θ

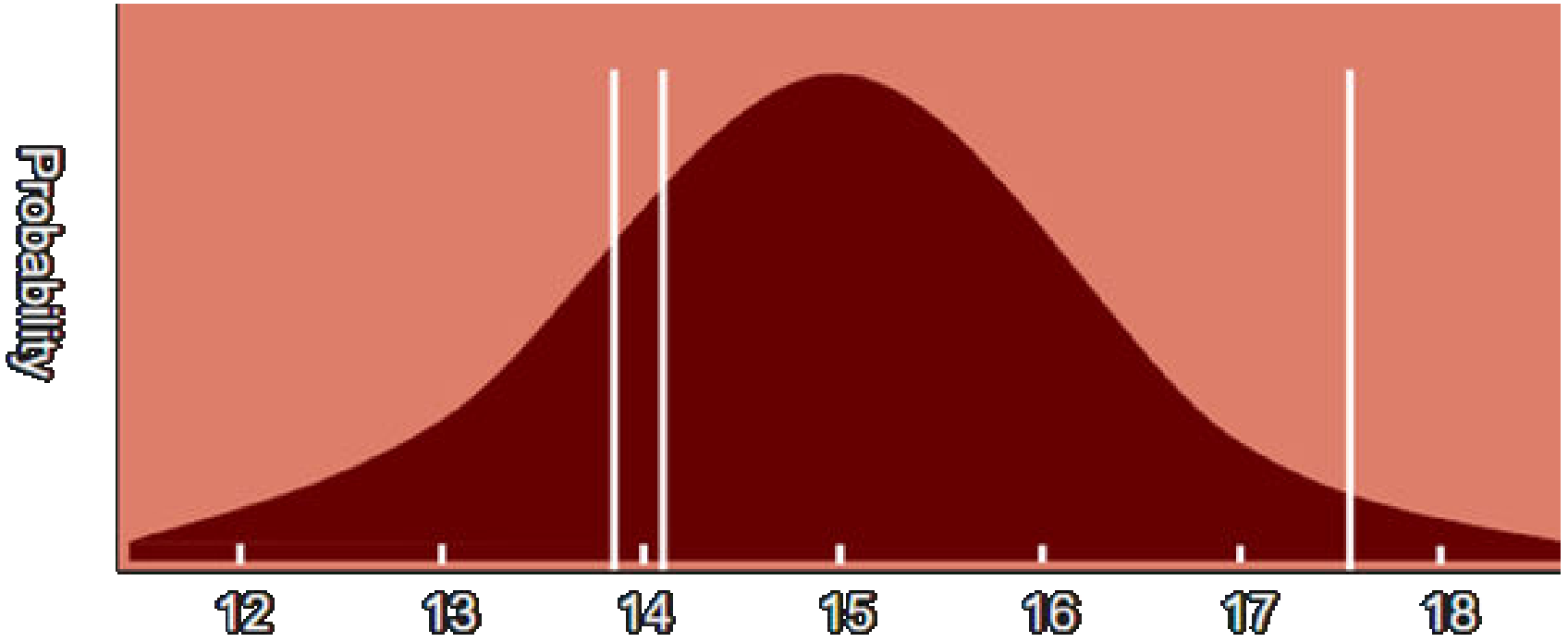

'Truth'
(loosely speaking)
Initial beliefs

How much does it weigh?

We have three Observations: 13.9, 17.5, 14.1

Non-Bayesian Estimate of the real weigh

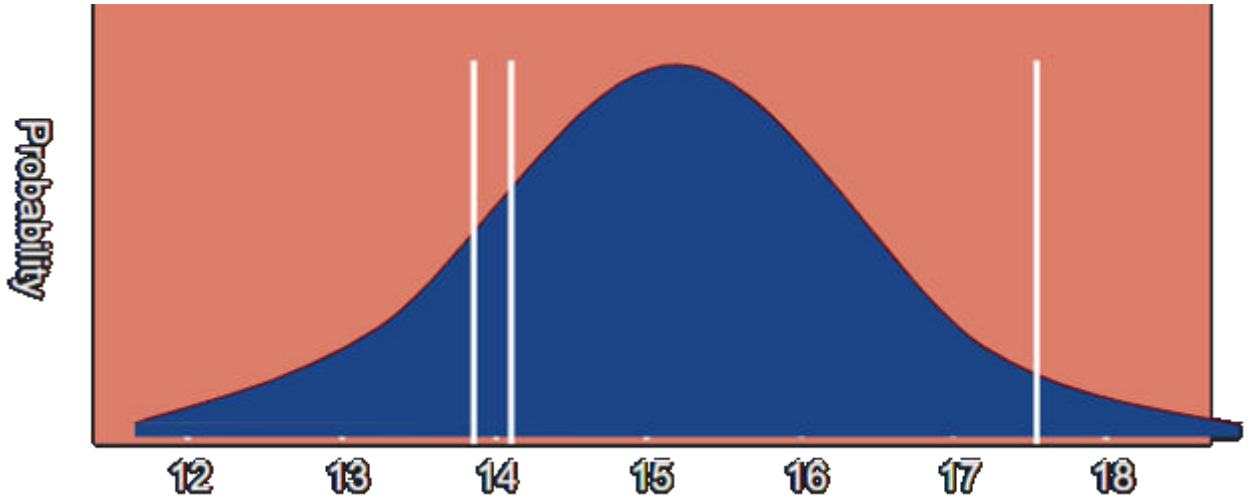
Normal distribution centred on the average of 15.2 with a Standard Error of 1.6



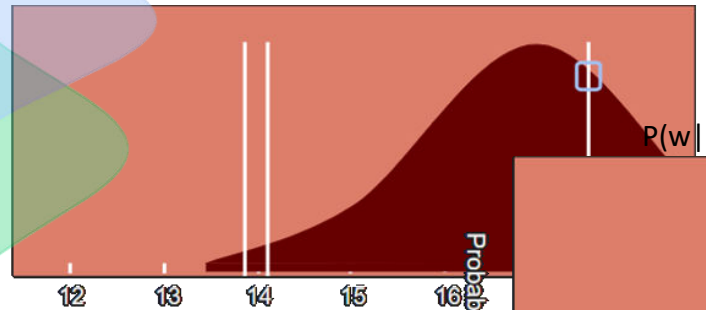
Apply Bayes Theorem: assuming a uniform prior

$$P(w \mid m) = P(m \mid w)$$

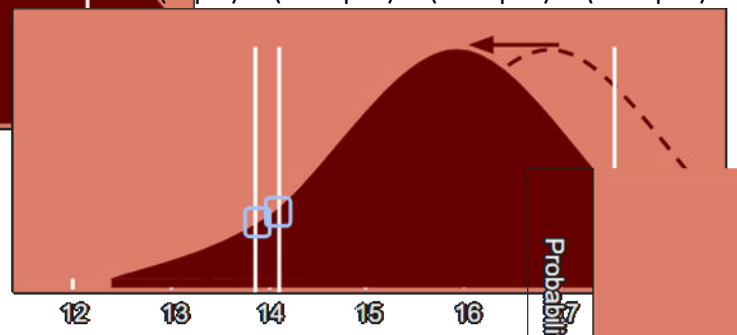
Mean= 15.2 - Maximul Likelihood Estimate (MLE)



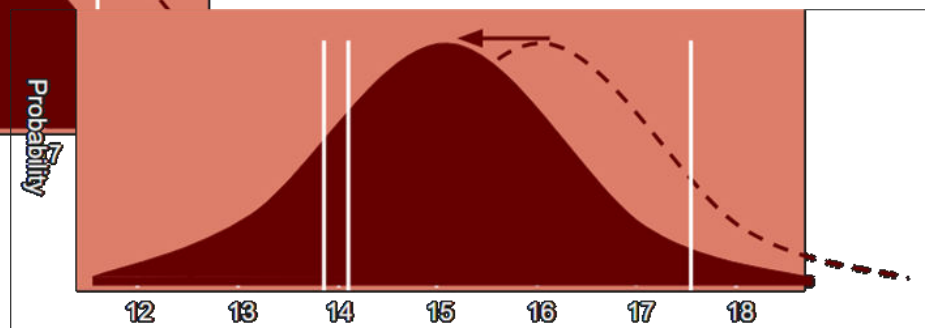
P(w|m)=P(13.9|17)*P(14.1|17)*P(17.5|17)



P(w|m)=P(13.9|16)*P(14.1|16)*P(17.5|16)

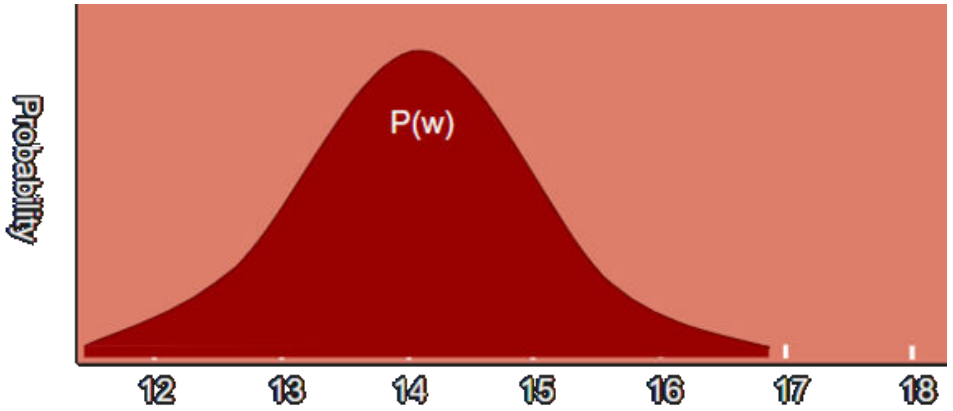


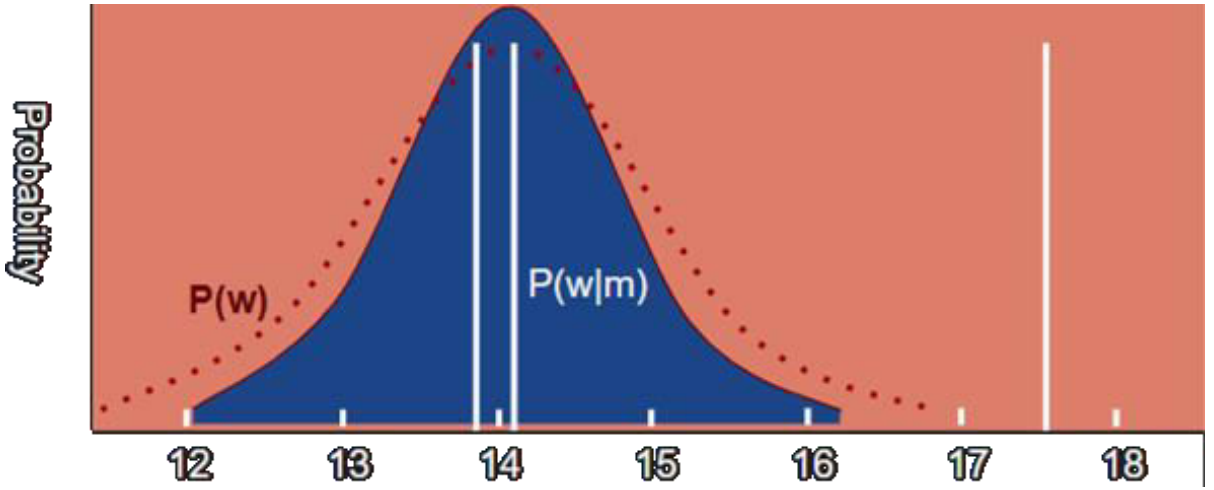
P(w|m)=P(13.9|15)*P(14.1|15)*P(17.5|15)



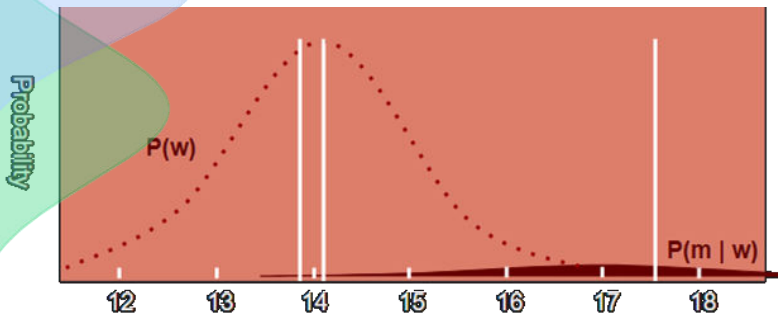
Now let's apply a prior:

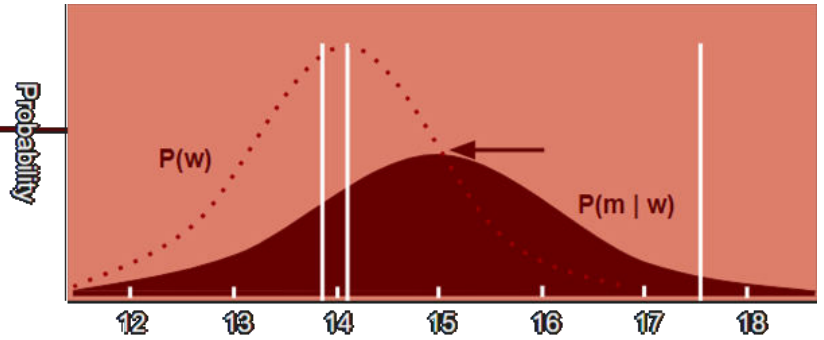
Normal curve with a mean at 14.2 and a StErr of .5

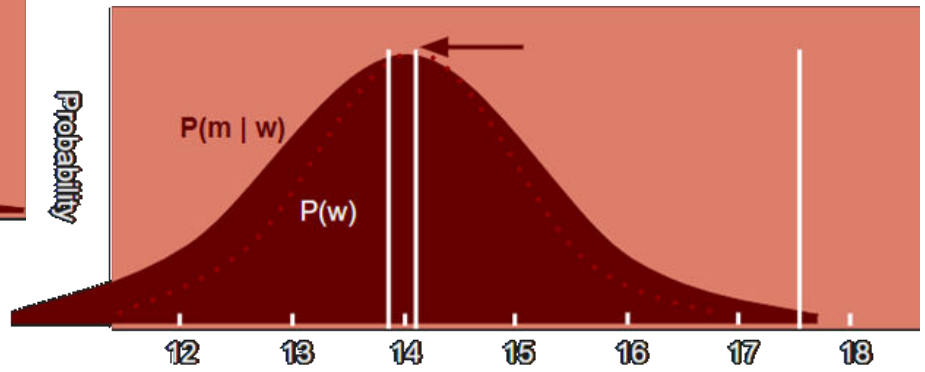




The new estimate is a normal distribution with a mean at 14.1 and a StErr of .4

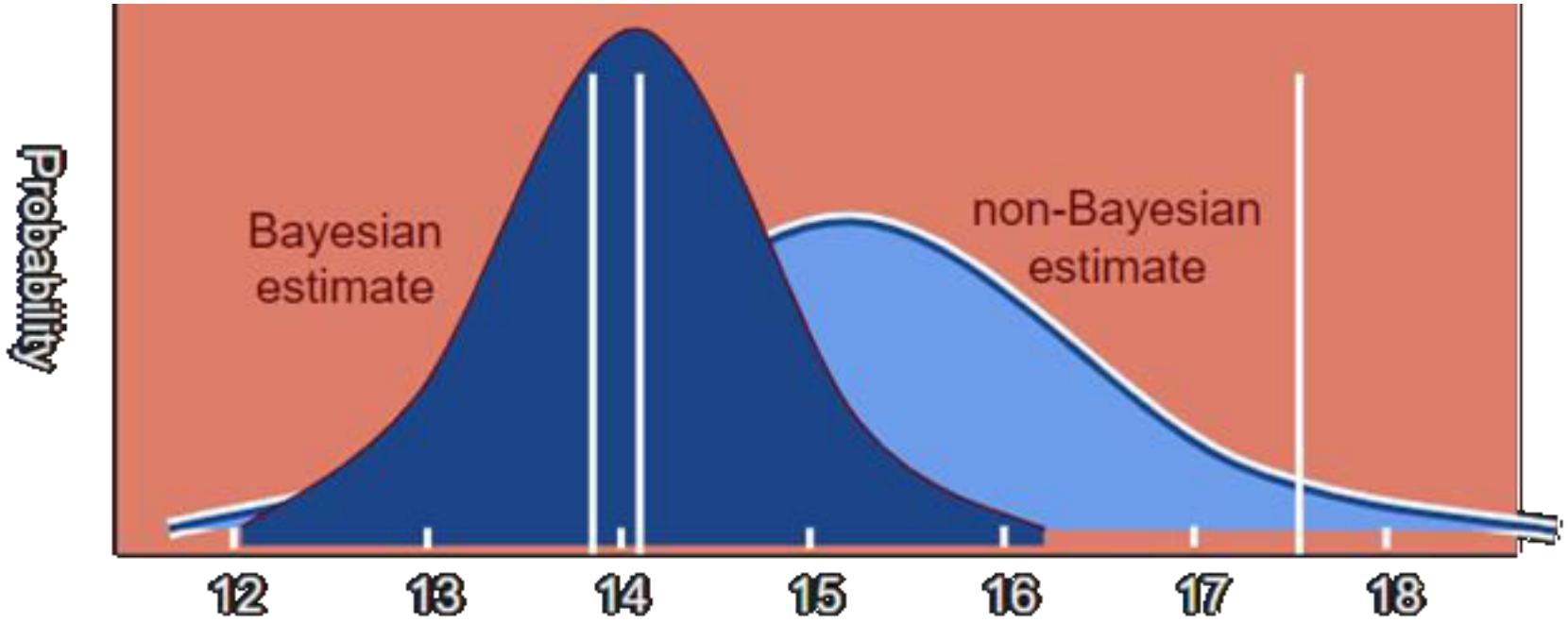






The distribution is narrower => Confidence is greater

Is this better?



Data are the data, but the Prior?

$$p(\theta \mid y) = \frac{p(y \mid \theta)}{p(y)} \times p(\theta)$$

$$\text{Beliefs after (posterior)} = \text{Evidence (likelihood)} + \text{Beliefs before (prior)}$$

DATA

????

Bayes factors and posterior model probabilities are generally sensitive to the choices of prior parameters, and thus one cannot simply select vague proper priors to get around the elicitation issue.

Standard Priors:

- normal distribution with mean 0 and standard deviation (σ) 10 (Balanced)
- Student's t distribution with mean 0, standard deviation (σ) 10; 3 df (Best for parameter estimation)
- Cauchy distribution with mean 0 and scale $\sqrt{2}/2$. (best for Bayes Factor)

Personalised Priors (According to the recommendations by Dienes 2019)

- normal distribution with the standard deviation coincident to the clinically relevant difference divided by 2,
- Cauchy distribution with the scale parameter divided by 7.

Elicited priors (expert and non expert) (Stefan et al., 2021).

- Different procedures available

Notably, it has been recently shown that the adoption of different prior distribution may lead to substantial changes in the Bayes factor quantification, but these changes rarely affect the qualitative conclusions of a hypothesis test (Stefan 2021)

Risks

We are not always aware of what we believe

Putting what we believe into a distribution correctly is tricky

We want to be able to be surprised by our data
Inaccurate beliefs can make it hard or impossible to learn

Bayesian approach may trap people interpreting ripetitive events as inevitable events.

P ≠ 1

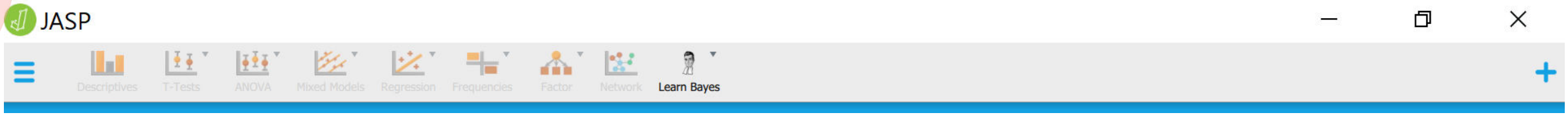
$$p(\theta \mid y) = \frac{p(y \mid \theta)}{p(y)} \times p(\theta)$$

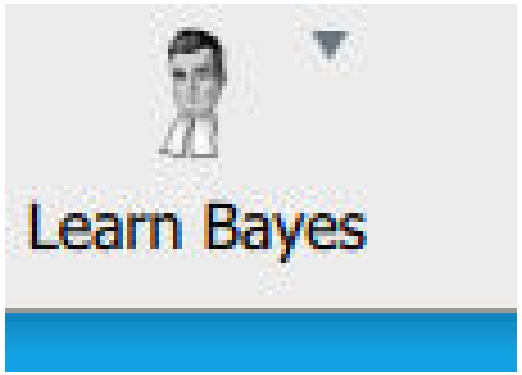
Beliefs after (posterior) = Evidence (likelihood) + Beliefs before (prior)

P ≠ 0

Let’s Play a bit with numbers

If you have installed JASP: go open it and open the tab Learn Bayes



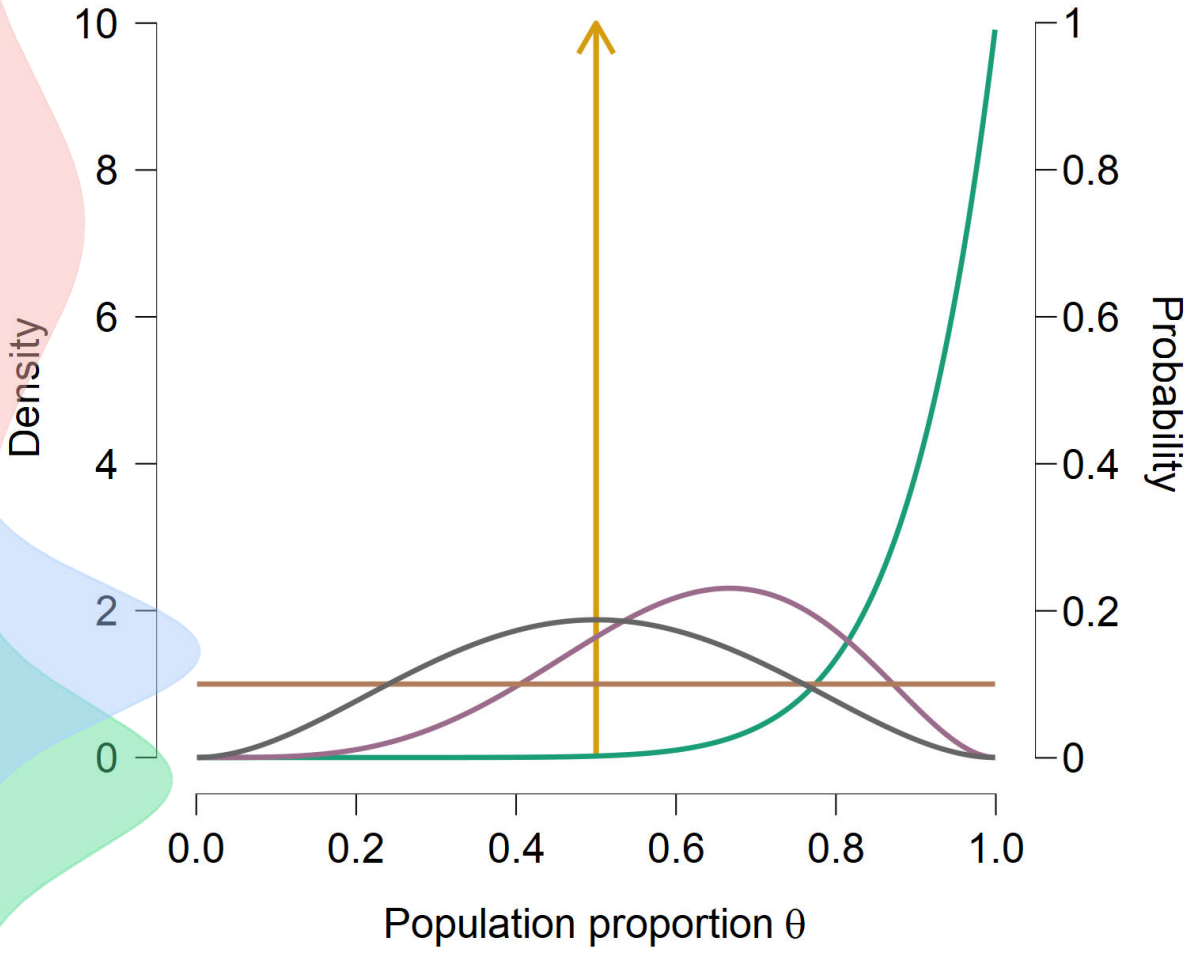


If you don't find it, it may stand under this + button

0,0,1,0,0,1,1,1,1,0,1,1,1,1

Data Summary

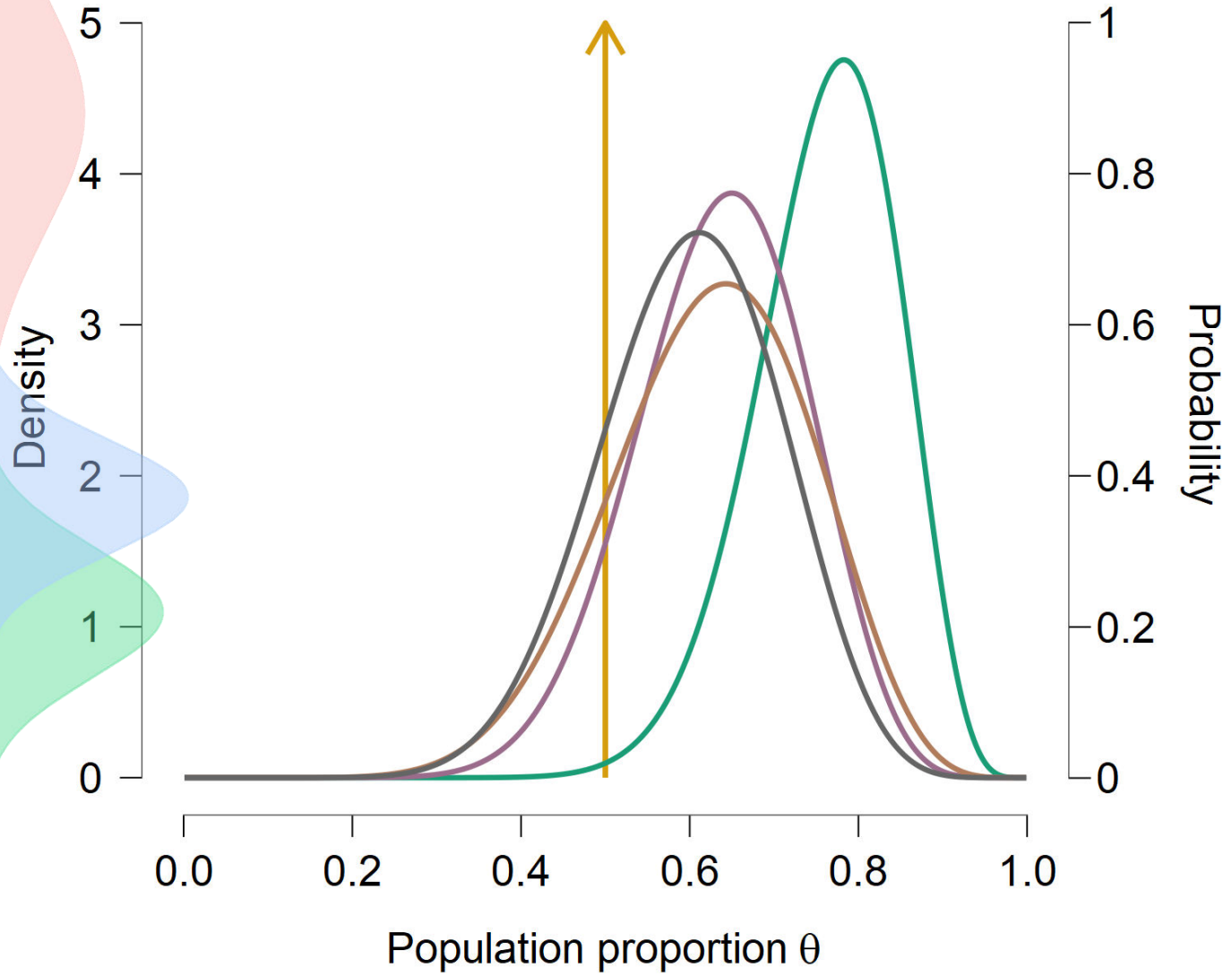
	Counts	Proportion
Successes	9	0.643
Failures	5	0.357
Total	14	



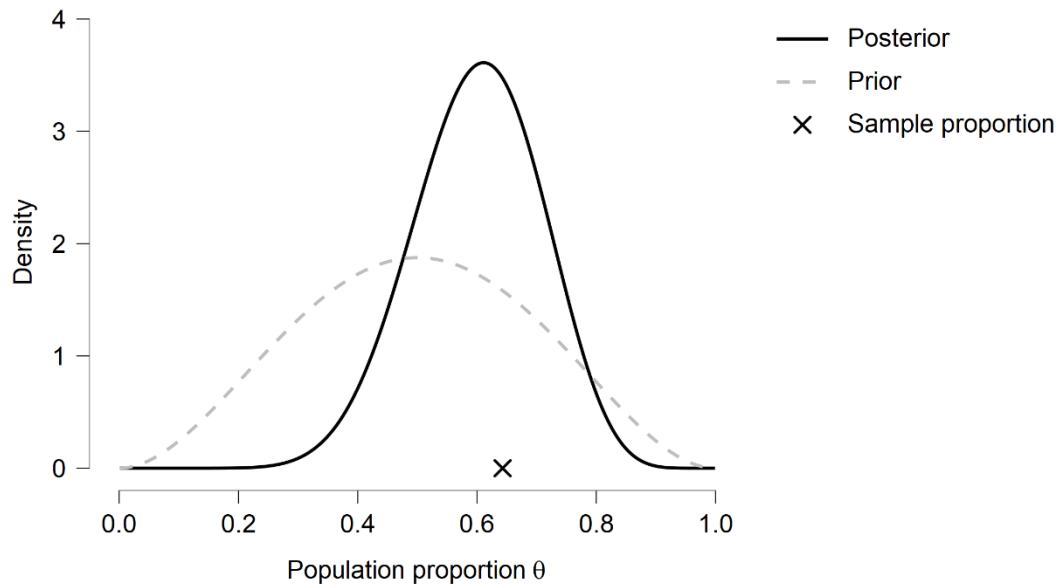
Estimation Summary

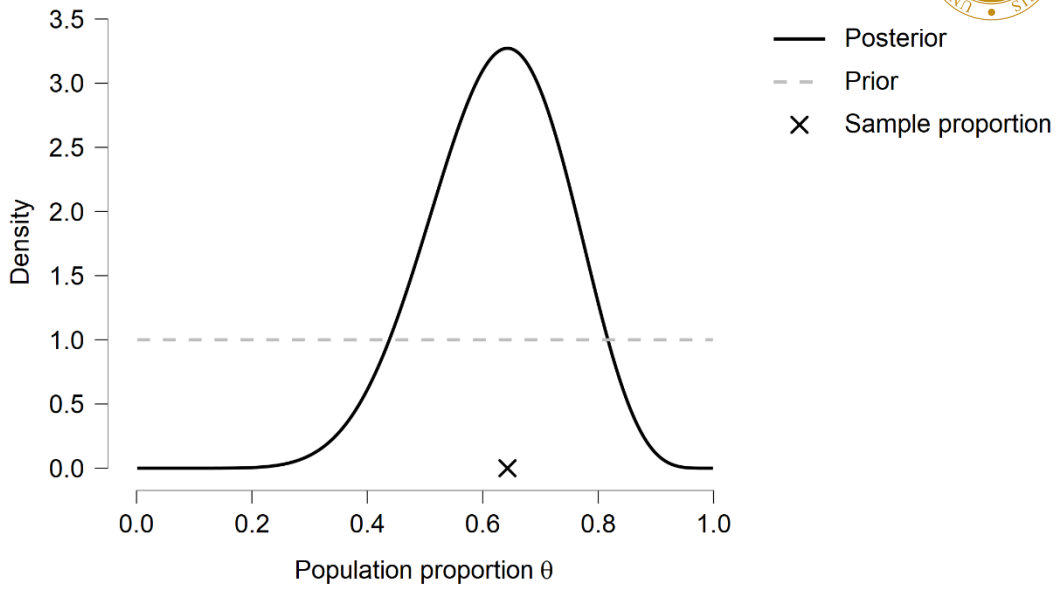
Model	Prior (θ)	Prior Mean
Spike	spike at 0.5	0.500
non-informative	beta (1, 1)	0.500
Symmetric	beta (3, 3)	0.500
Asymmetric Weak	beta (5, 3)	0.625
Asymmetric Strong	beta (10, 1)	0.909

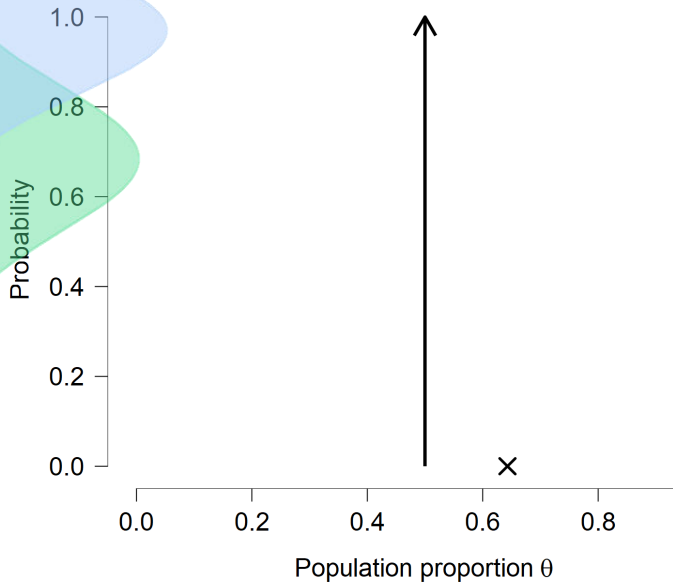
Note. These results are based on 9 successes and 5 failures.

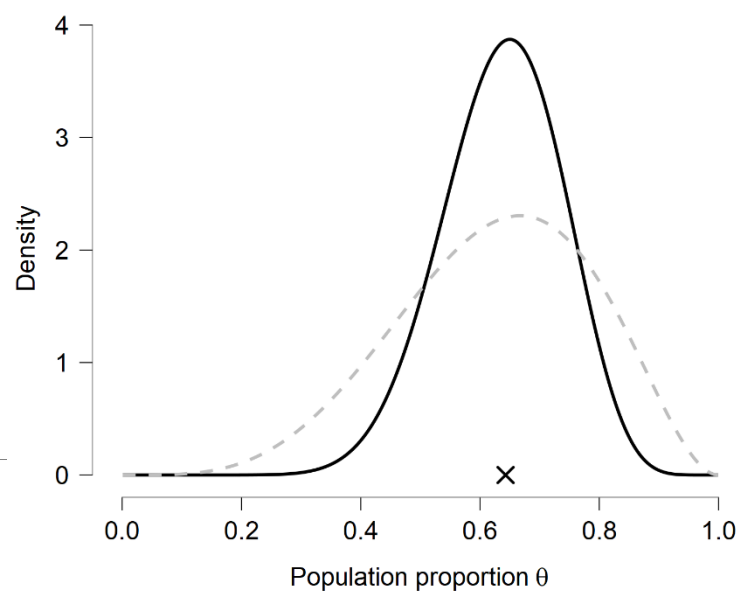


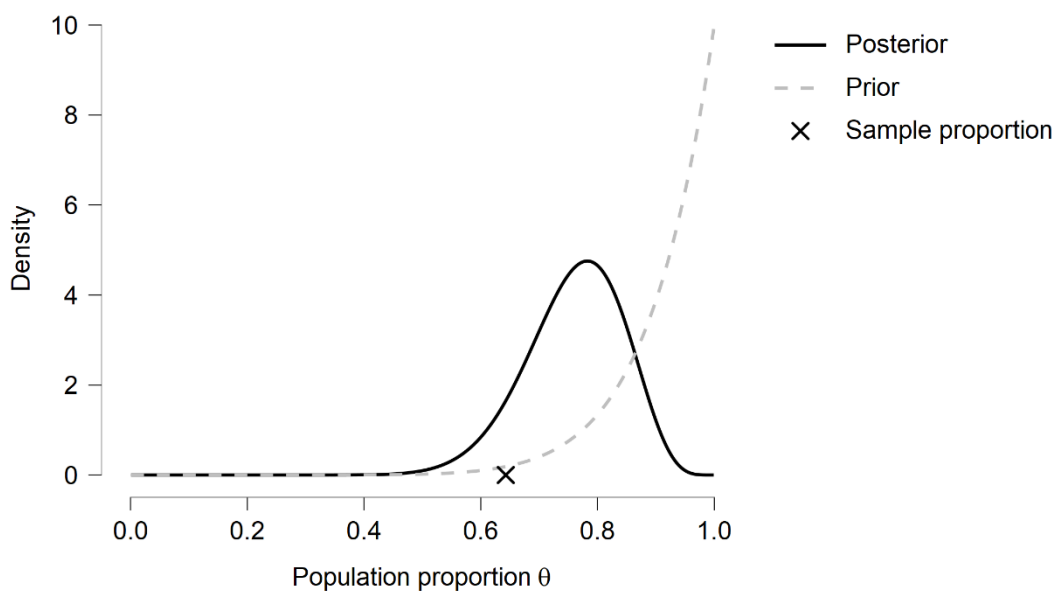
Posterior (θ)	Posterior Mean
spike at 0.5	0.500
beta (10, 6)	0.625
beta (12, 8)	0.600
beta (14, 8)	0.636
beta (19, 6)	0.760

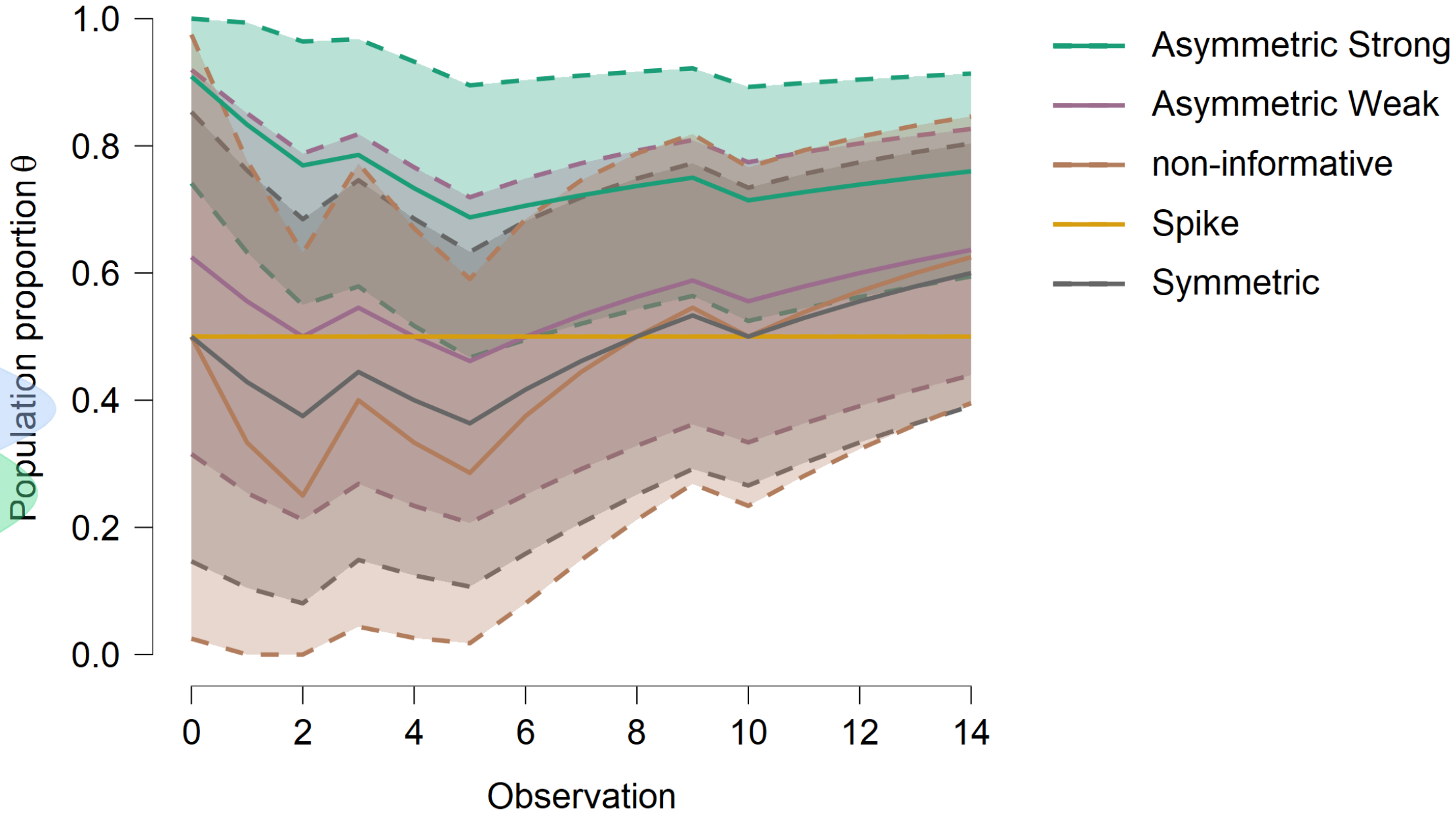

Posterior
Prior
Sample proportion
Density
Population proportion θ


Posterior
Prior
Sample proportion
Density
Population proportion θ


Probability
Population proportion θ

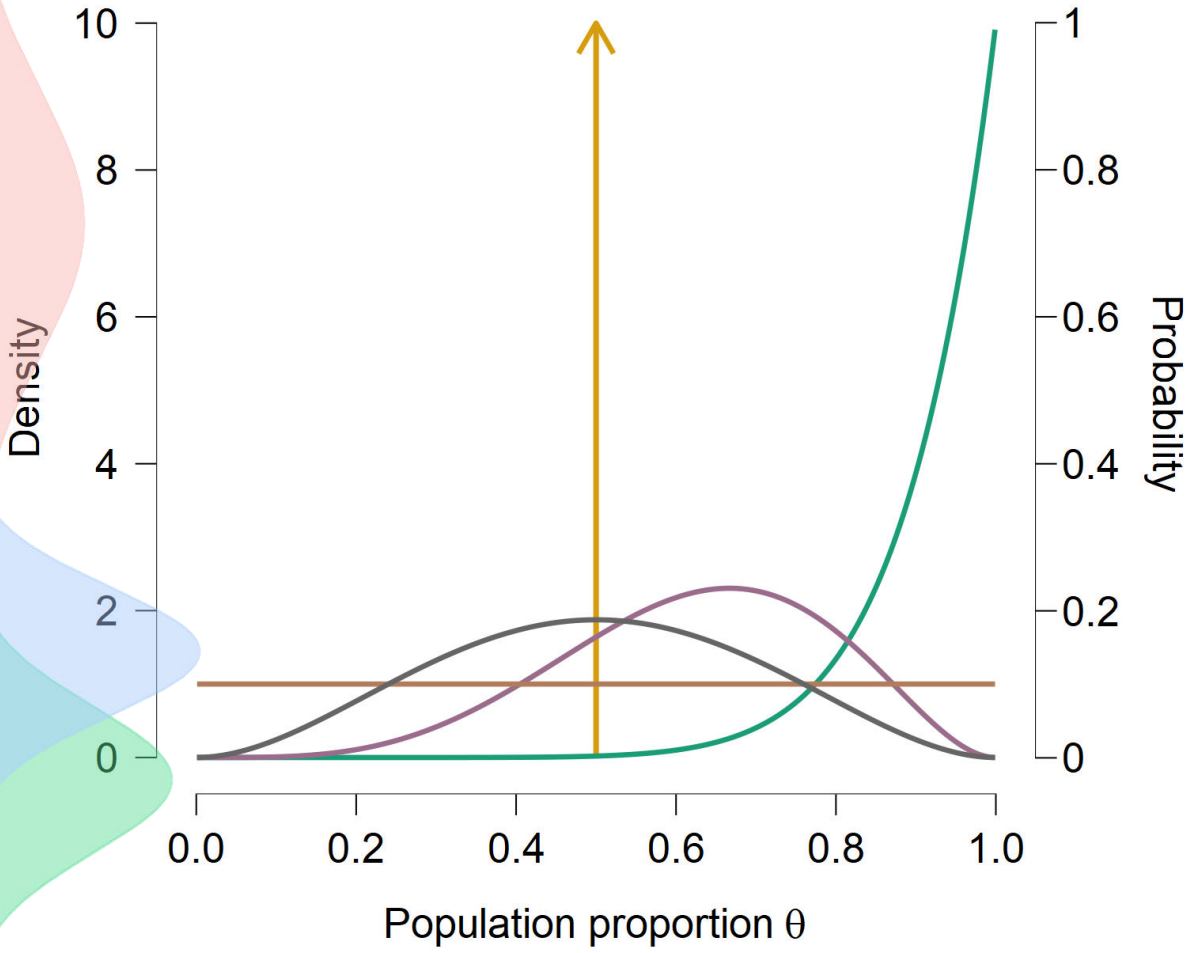

Density
Population proportion θ


Posterior
Prior
Sample proportion
Density
Population proportion θ


1.0
0.8
0.6
0.4
0.2
0.0
Population proportion θ
0
2
4
6
8
10
12
14
Observation
Asymmetric Strong
Asymmetric Weak
non-informative
Spike
Symmetric

Data Summary

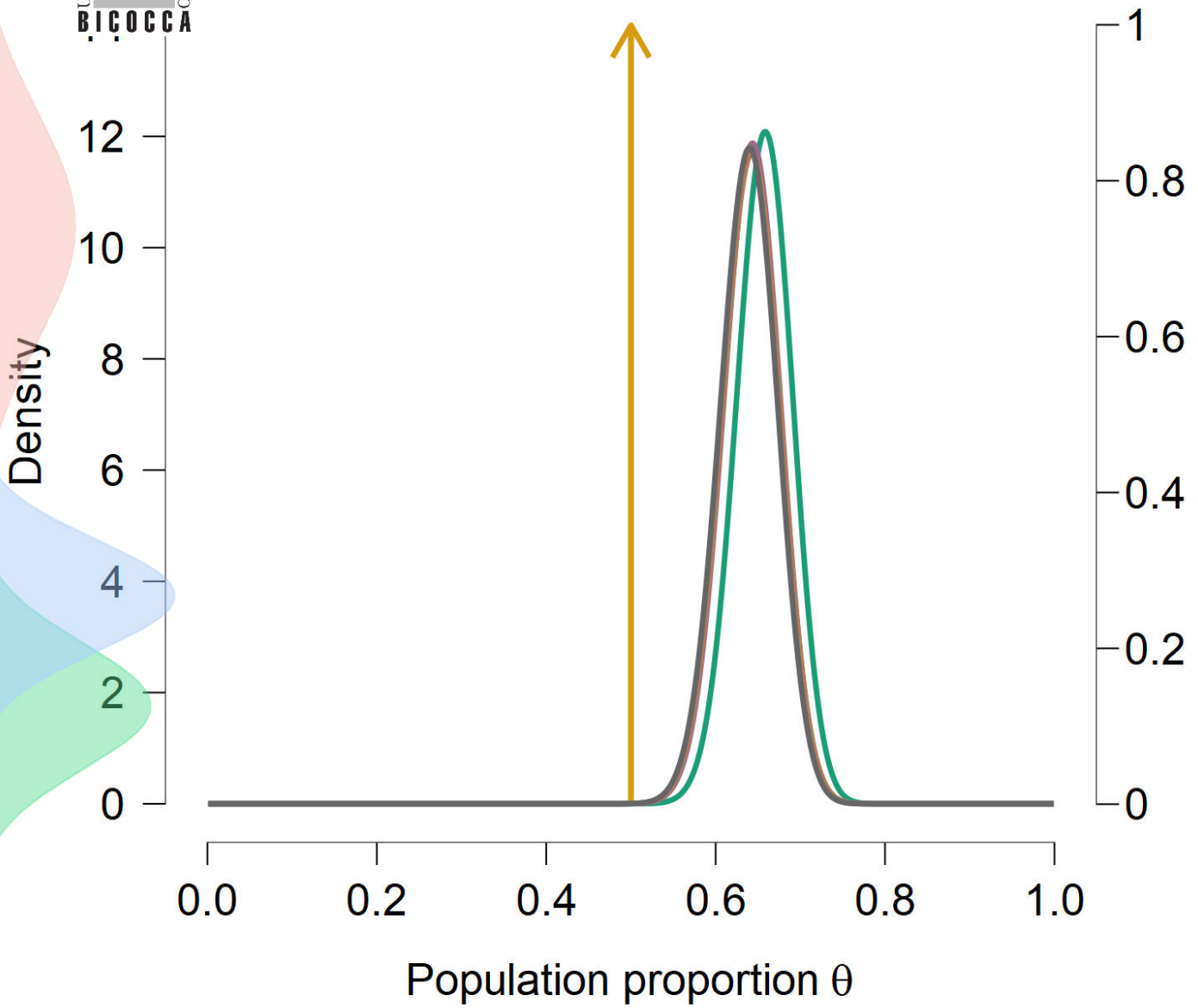
	Counts	Proportion
Successes	126	0.643
Failures	70	0.357
Total	196	



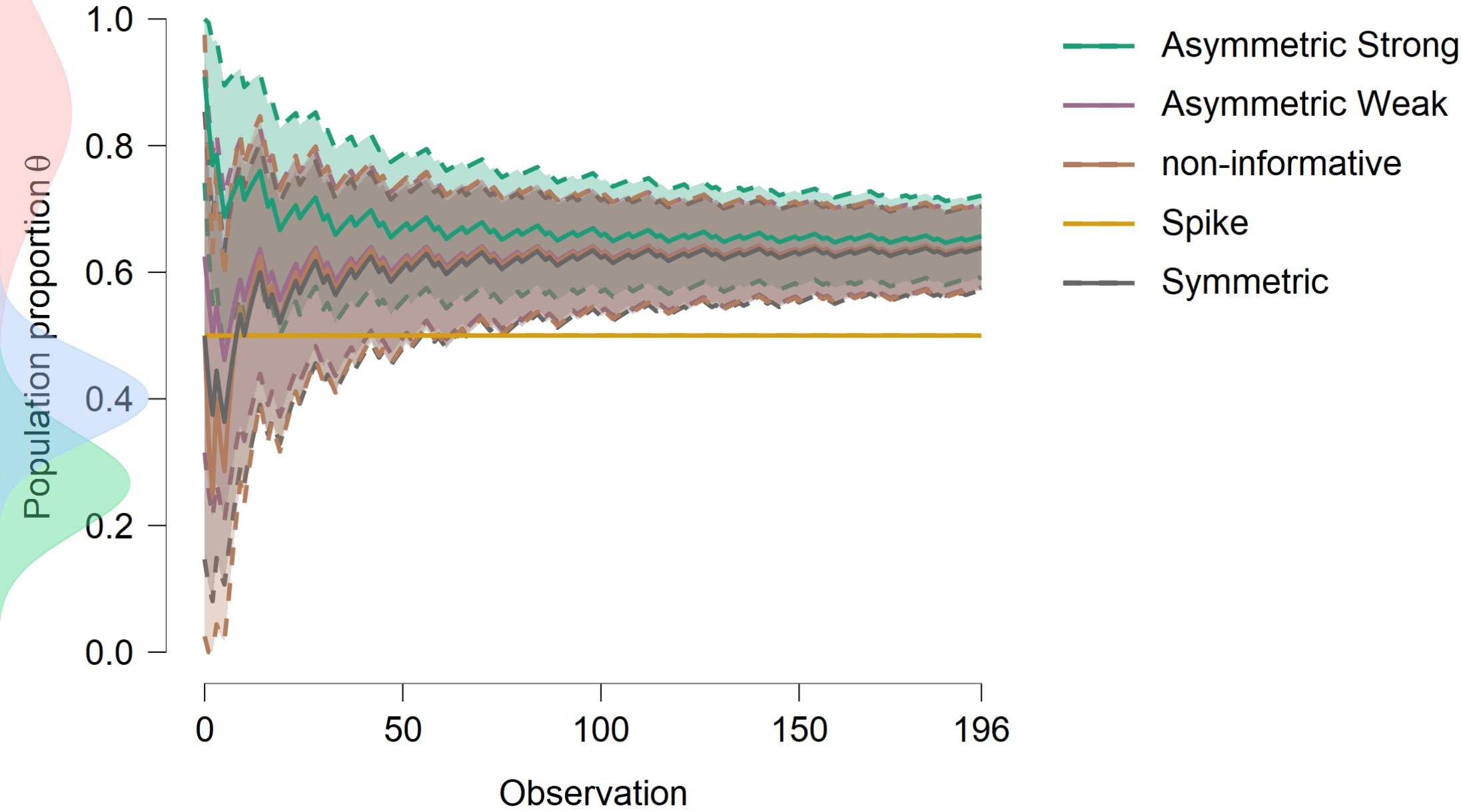
Estimation Summary

Model	Prior (θ)	Prior Mean
Spike	spike at 0.5	0.500
non-informative	beta (1, 1)	0.500
Symmetric	beta (3, 3)	0.500
Asymmetric Weak	beta (5, 3)	0.625
Asymmetric Strong	beta (10, 1)	0.909

Note. These results are based on 9 successes and 5 failures.



Posterior (θ)	Posterior Mean
spike at 0.5	0.500
beta (127, 71)	0.641
beta (129, 73)	0.639
beta (131, 73)	0.642
beta (136, 71)	0.657


1.0
0.8
0.6
0.4
0.2
0.0
Population proportion θ
0
50
100
150
196
Observation
Asymmetric Strong
Asymmetric Weak
non-informative
Spike
Symmetric

End of the first part

daniele.romano@unimib.it

daniele.romano84 / danromano

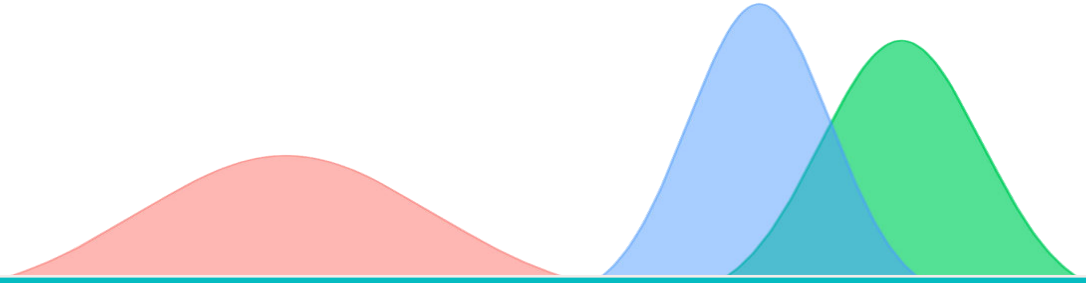
Bayesian statistics and hypothesis testing.

Testing competing hypothesizes vs. null hypothesis. The level of support for the null hypothesis

Daniele Romano,

Università degli Studi di Milano-Bicocca

daniele.romano@unimib.it



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

Reminder

In the frequentist conceptualization, probability is defined as a limit.
It is the proportion of events in infinitely many repeated samples.

Bayesians, however, define probability as a measure of subjective belief.

- **Statistical Evidence**
- "Statistical evidence" is data (in the context of a probability model) that would affect a rational person's belief regarding a statistical question of interest.

- We will only consider relative evidence
- Beliefs will be modelled using probability distributions
- "Rationality" for our purposes is probability theory/conditionalization
- **These constraints are the foundation of Bayesian statistics.**

The effect of data...

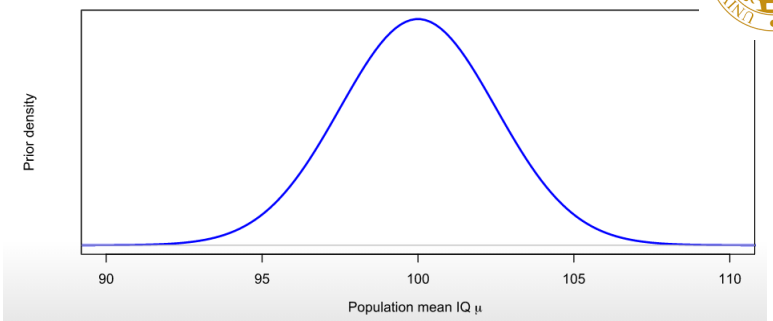
Suppose we observe $\bar{y} = 103$ with $N = 15$, and suppose the population variance $\sigma^2 = 15^2$.

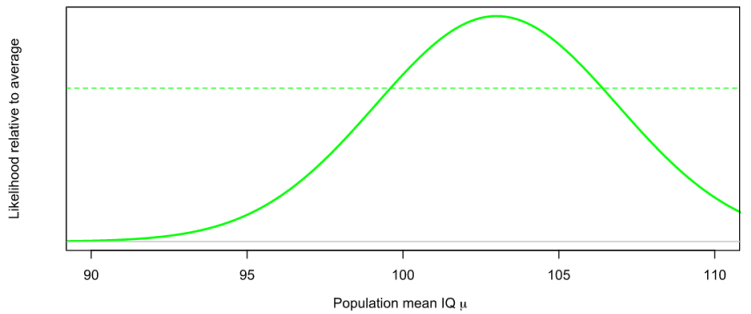
$$\text{Posterior} \propto \text{Likelihood} \times \text{Prior}$$

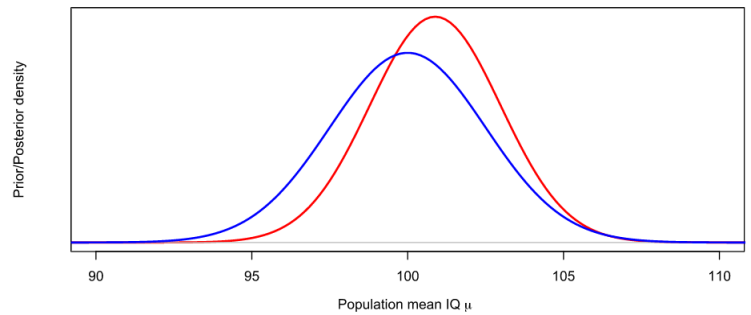
Multiplying our prior and likelihood together,

$$\text{Posterior} \propto \left[\prod_{i=1}^{15} \exp\left\{-\frac{1}{2 \times 15^2}(y_i - \mu)^2\right\}\right] \times \exp\left\{-\frac{1}{2 \times 2.5^2}(\mu - 100)^2\right\}$$

...looks complicated, but is just the function of μ produced by multiplying the likelihood and prior.







From prior to posterior

A few points...

- Posteriors converge to the "true value" - if the prior allows it
- Likelihood "swamps" the prior with sufficient data
- Various attempts at "noninformative" priors exist (but beware!)

But how do we deal with more than one parameter?

Extending to multiple parameters

Suppose we don't know μ or $\sigma 2$. Bayes theorem still applies:

$p(\mu,\sigma 2|\boldsymbol{y}) \propto p(\boldsymbol{y}|\mu,\sigma 2)p(\mu,\sigma 2)$

Now $p(\mu,\sigma 2)$ is a joint prior; $p(\mu,\sigma 2|\boldsymbol{y})$ is a joint posterior.

Joint priors/posteriors are multivariate probability distributions.

Marginalizing

Integration is hard!

- Analytic integration is only possible in very simple circumstances
- Large numbers of parameters pose a problem (integration over 1000 dimensions?)
- Luckily solutions exist...

What if we could sample from a joint posterior distribution?

Posterior sampling

Posteriors are the target of inference for many Bayesians.

- Yield estimates of parameters in light of observed data
- Transforming samples enables reparametrization
- Fairly easy to compute, even for complex models (WinBUGS, JAGS, etc.)
- Often robust to various prior specifications
- BUT: posteriors are conclusions, not evidence!

- **The Bayes factor**

- **The Bayes factor: Weight of statistical evidence**
- $\frac{p(M0|y)}{p(M1|y)} = \frac{p(y|M0)}{p(y|M1)} \times \frac{p(M0)}{p(M1)}$
- Posterior odds=Bayes factor×Prior odds

- In words: "The relative probability of the data under the hypotheses is exactly the strength of the relative statistical evidence between the two hypotheses."

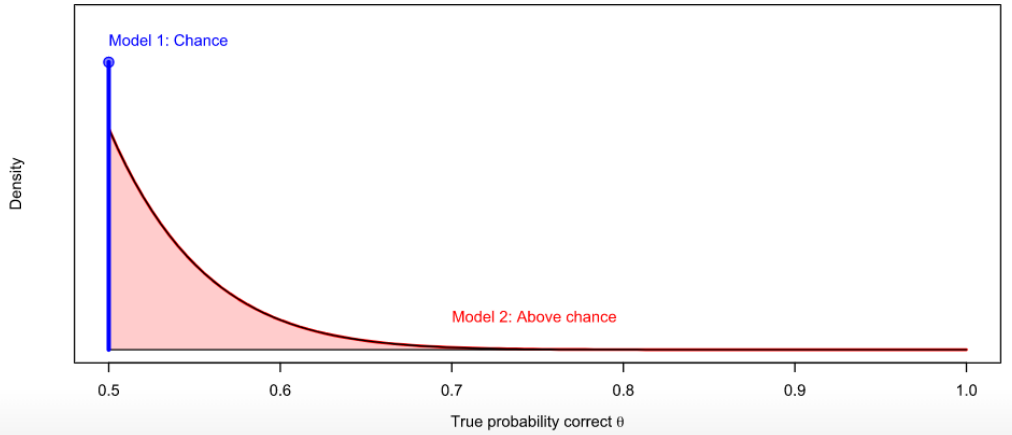
Convincing evidence?

- **A Bayes factor of 2.1908276 means...**
- The data were 2.1908276 times more likely under the alternative model M2
- We should shift our beliefs by a factor of 2.1908276 toward the above-chance model M2
- If we were evenly split between the two before (prior odds of 1) we would favor M2 by 2.1908276 after (posterior odds)

- **The "weight of evidence" is 2.1908276. Our beliefs should barely change.**

An example: ESP

Suppose we are interested in true accuracy θ in a two-alternative forced choice task.



An example: ESP

In a large experiment, we observe 528 out of 1000 correct responses $(p = 0.036)$. A frequentist rejects $\mathcal{M}_1$. How much evidence is this for $\mathcal{M}_2$, against $\mathcal{M}_1$?

$$p(y \mid \mathcal{M}_1) = \binom{1000}{528} 0.5^{528}(1-.5)^{1000-528} = 0.0052624$$

For $\mathcal{M}_2$, we have to average over the prior:

$$p(y \mid \mathcal{M}_2) = \int_{.5}^{1} \binom{1000}{528} \theta^{528}(1-\theta)^{1000-528} p(\theta)\, d\theta = 0.011529$$

Thus...

$$\text{Bayes factor} = \frac{p(\boldsymbol{y} \mid \mathcal{M}_2)}{p(\boldsymbol{y} \mid \mathcal{M}_1)} = \frac{0.011529}{0.0052624} = 2.1908276$$

Bayes factors describe the relative probability of data under competing positions

Computing Bayes factors

Integration is hard

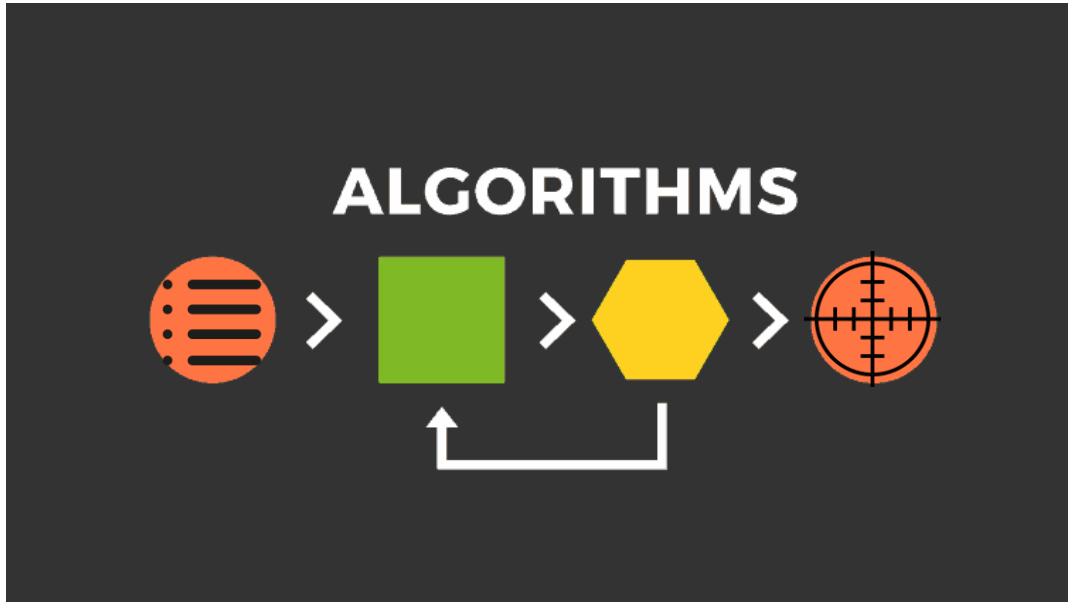
- How to compute Bayes factors:
  1. Monte Carlo Markov Chains
  2. From Gibbs sampler
  3. Approximations (no sampling)

$$B_{01} = \frac{\int_{\theta \in \Theta_0} \Pr(\text{Data}|\mathcal{M}_0, \boldsymbol{\theta})\pi_0(\boldsymbol{\theta})d\boldsymbol{\theta}}{\int_{\theta \in \Theta_1} \Pr(\text{Data}|\mathcal{M}_1, \boldsymbol{\theta})\pi_1(\boldsymbol{\theta})d\boldsymbol{\theta}}$$

Monte Carlo Markov Chain (MCMC)

MCMCs are part of the family of Monte Carlo simulations developped to approximate a complicated system with a statistical sample.

MCMC is a family of algorithms that can approximate the posterior estimation:
- Precisesly
- Efficiently



Markov Chain (MC-)

- it's a walk (or a chain of events/values) where, the actual one is directly related to, and dependent from, only the previous value

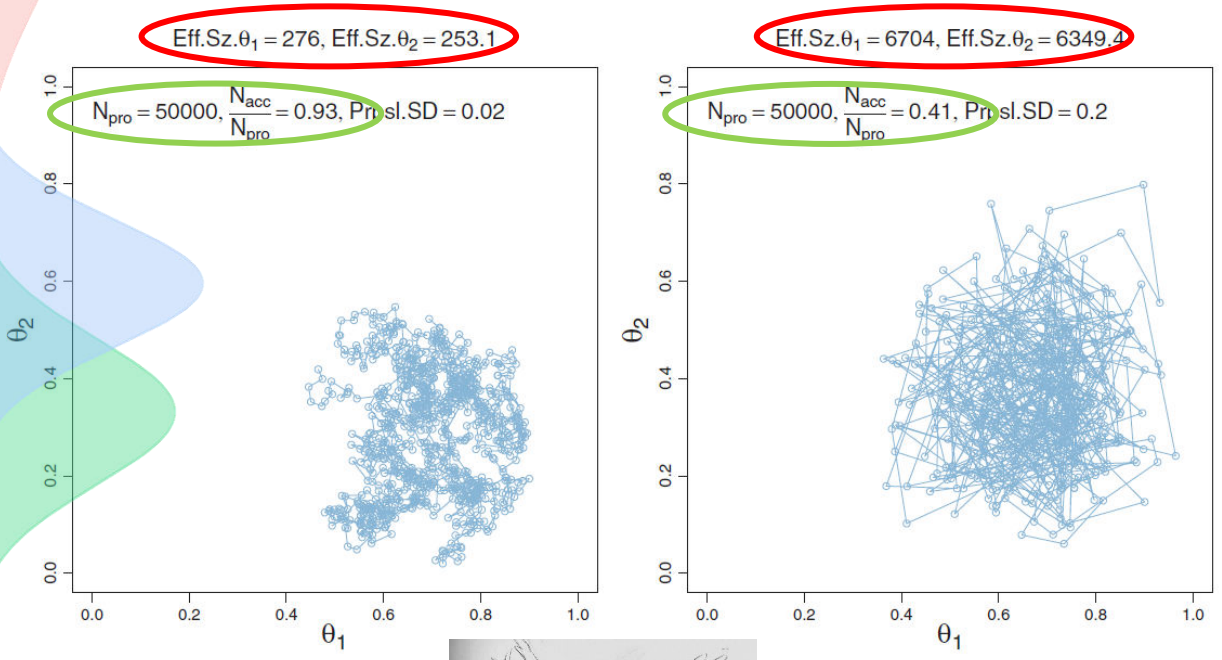
Monte Carlo (-MC)

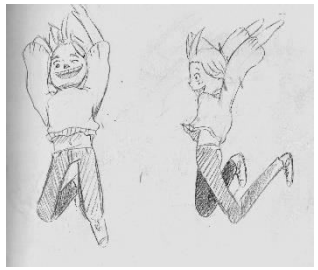
- The steps are sampled randomly (or to a probabilistic rule)

Comparing the Gibbs and Metropolis methods

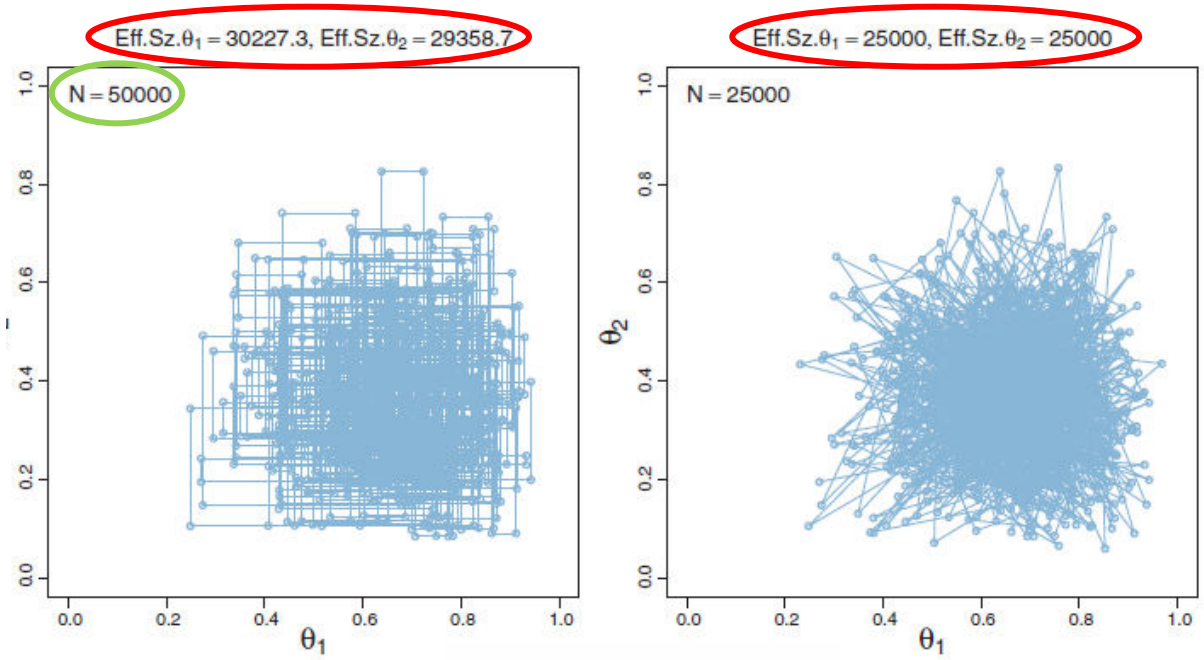
In the infinitely long run, the Gibbs and Metropolis methods converge to the same distribution.
What differs is the efficiency of getting to any desired degree of approximation accuracy in a finite run.

Metropolis





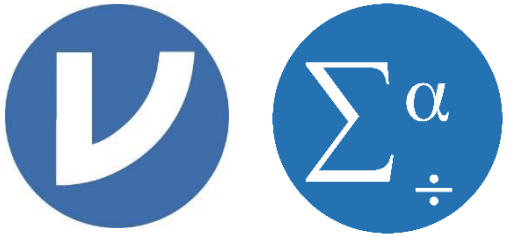
Gibbs



How to run Bayesian Statistics

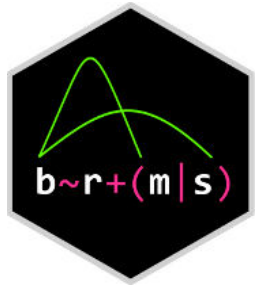
Graphic User Interface

JASP, JAMOVI, SPSS (?!)
User-friendly, but not very flexible

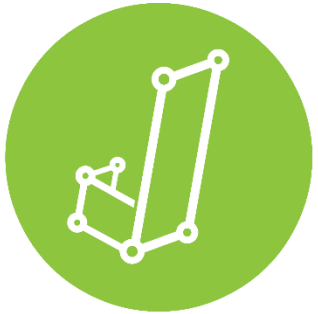


R environment

BayesFactor BMRS JAGS STAN
harder implementation
free to draw your own specific model



An Intorduction to



JASP

A Fresh Way to Do Statistics

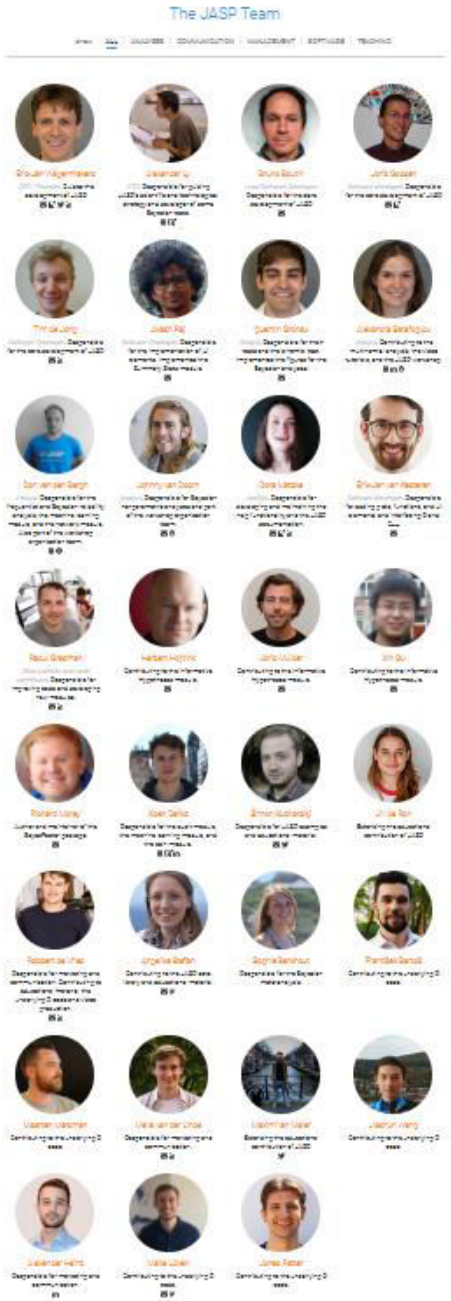
Daniele Romano

daniele.romano@unimib.it

Credit

Eric-Jan Wagenmakers

CEO / Founder. Guides the development of JASP.

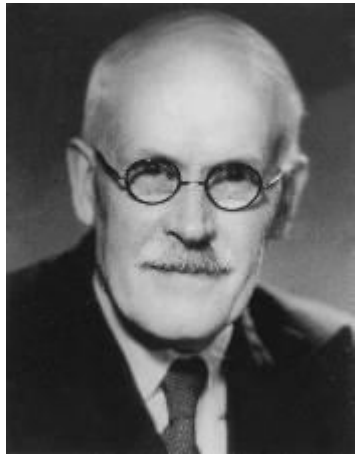


“Teachers are free to use the guides and share it with their students.”

https://jasp-stats.org/team/

What is JASP

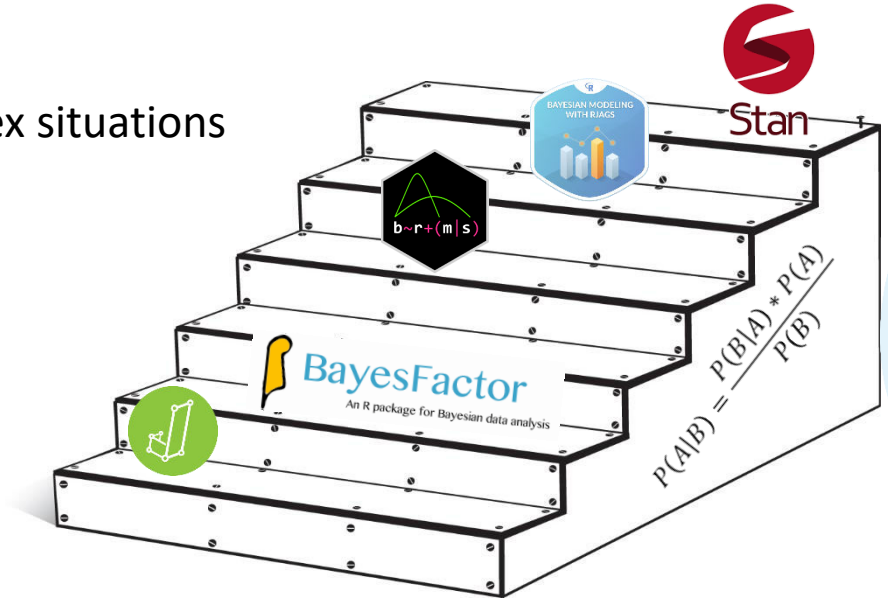
JASP stands for Jeffrey's Amazing Statistics Program



Why JASP

Probably the most used software for Bayesian analysis, setting a sort of a standard.
For basic analysis is easy, fast, and «enough»
In contrast to many statistical packages**, JASP provides a simple drag and drop interface,** easy access menus, intuitive analysis with real-time computation and display of all results.
All tables and graphs are presented in APA format and can be copied directly and/or saved independently.

- It is very intuitive for the basics, and not flexible enough for the complex situations
- A lot of free (valuable) materials online
- The JASP team organizes in depth courses on a yearly basis.



Scopes of this introduction

- Getting informed about JASP
- Familiarize with the environment
- See the potential and the limit of the software.

- Run Bayesian t-tests
- Run Bayesian ANOVA (maybe)

Installation

Invite JASP to your PC: Installation

- Click https://jasp-stats.org/download/ to start downloading.
- Consider your operating system (Windows, Mac, Linux).

Good to Know

JASP is released under a GNU Affero GPL v3 license, which is an open-source license that guarantees that JASP will always be (for) free.

The GNU Affero General Public License is a free, copyleft license for software and other kinds of works, specifically designed to ensure cooperation with the community in the case of network server software.

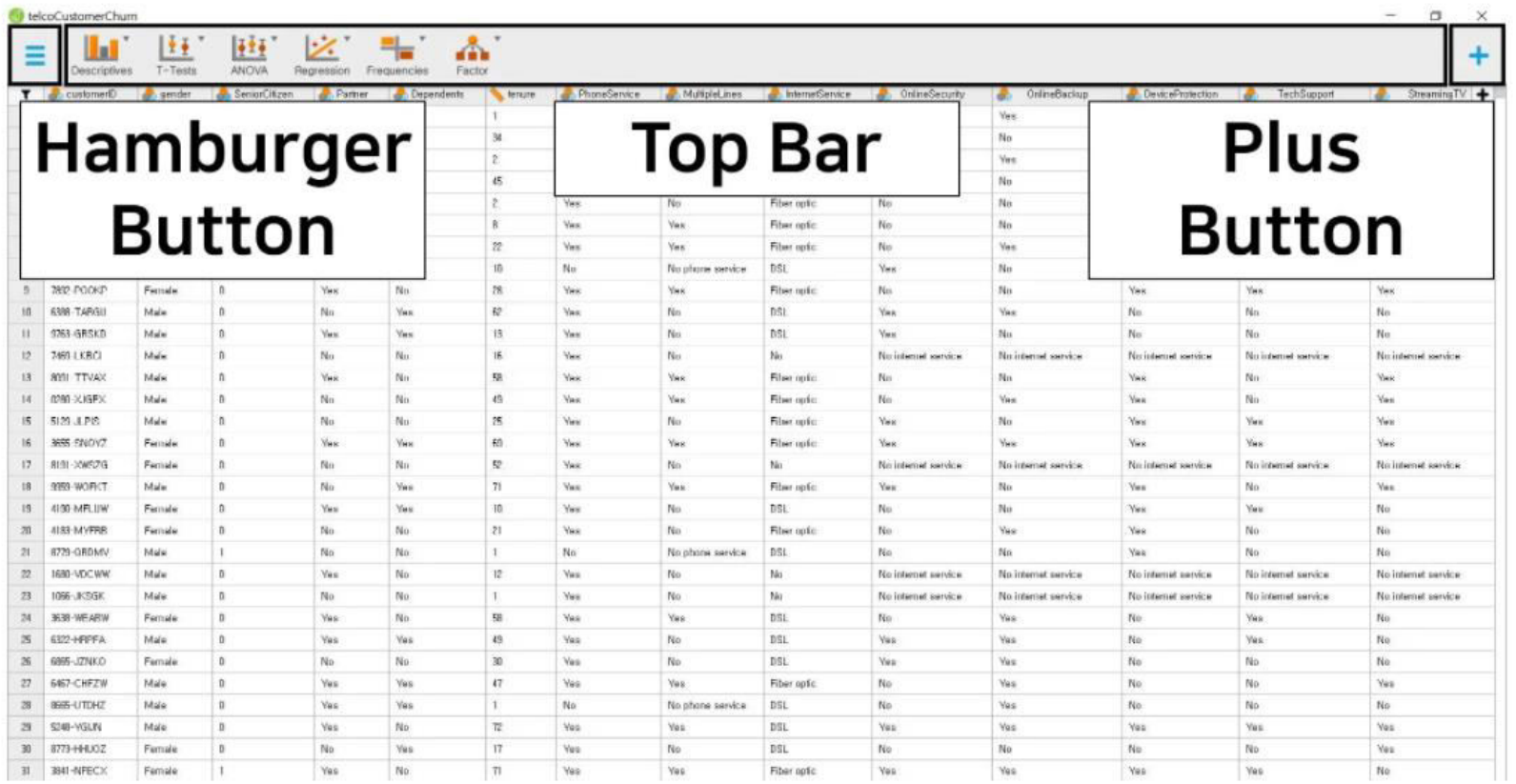
The environnement

Hamburger button (top left)

•With the hamburger button, you can open, save, or export data.

Top bar (top middle)

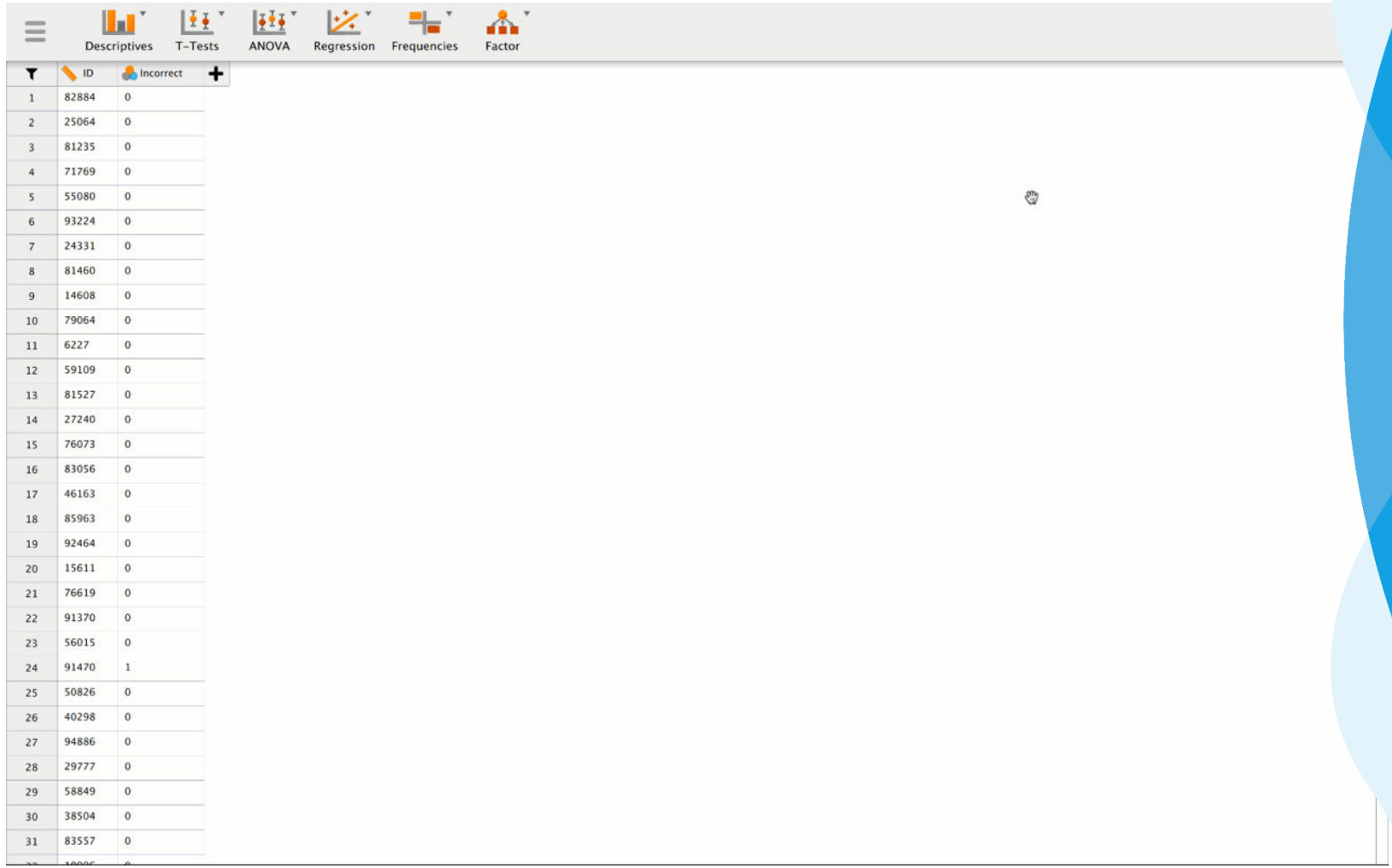
•As you might have guessed, the top bar is a group of primary analyses. You can simply st any analysis you want by clicking the analysis option. For example, if you want to perform regression analysis, simply click the 'Regression' button. Do you want to know the next step We will guide you through the steps in 'IV. Data analysis and interpretation'. So please follow our guidance until that section.



The environnement

Plus button (top right)

- As the 'plus' sign implies, the plus button is an add-on button for advanced analytic techniques.
- When you click the plus button, you can encounter various analysis options such as JAGS, Machine Learning, Meta-Analysis, Network, SEM (Structural Equation Modeling), and more. They will help you answer the complex research questions.



	ID	Incorrect
1	82884	0
2	25064	0
3	81235	0
4	71769	0
5	55080	0
6	93224	0
7	24331	0
8	81460	0
9	14608	0
10	79064	0
11	6227	0
12	59109	0
13	81527	0
14	27240	0
15	76073	0
16	83056	0
17	46163	0
18	85963	0
19	92464	0
20	15611	0
21	76619	0
22	91370	0
23	56015	0
24	91470	1
25	50826	0
26	40298	0
27	94886	0
28	29777	0
29	58849	0
30	38504	0
31	83557	0

File format

JASP has its own **.jasp** format but can open a variety of different dataset formats such as:
.csv / .txt / .tsv (tab-separated values) / **.sav** (IBM SPSS data file) / **.ods** (Open Document spreadsheet)

Variables identification

	ParticipantNumber	Condition	q1_check
1	1	1	2
2	2	2	3
3	3	3	7
4	4	4	4
5	5	1	3
6	6	2	3
7	7	3	4
8	8	4	7
9	9	1	5
10	10	2	4
11	11	3	4
12	12	4	7
13	13	1	6
14	14	2	3
15	15	3	4
16	16	4	7

Automatic detection of the variable level of measurement.

Watchout, sometimes it's wrong!

You can change it manually

Database handling

All files must have a header label in the first row. Once loaded, the dataset appears in the window:

	ParticipantNumber	Condition	q1_check
1	1	1	2
2	2	2	3
3	3	3	7
4	4	4	4
5	5	1	3
6	6	2	3
7	7	3	4
8	8	4	7
9	9	1	5
10	10	2	4
11	11	3	4
12	12	4	7
13	13	1	6
14	14	2	3
15	15	3	4
16	16	4	7

JASP is quite inefficient managing databases.

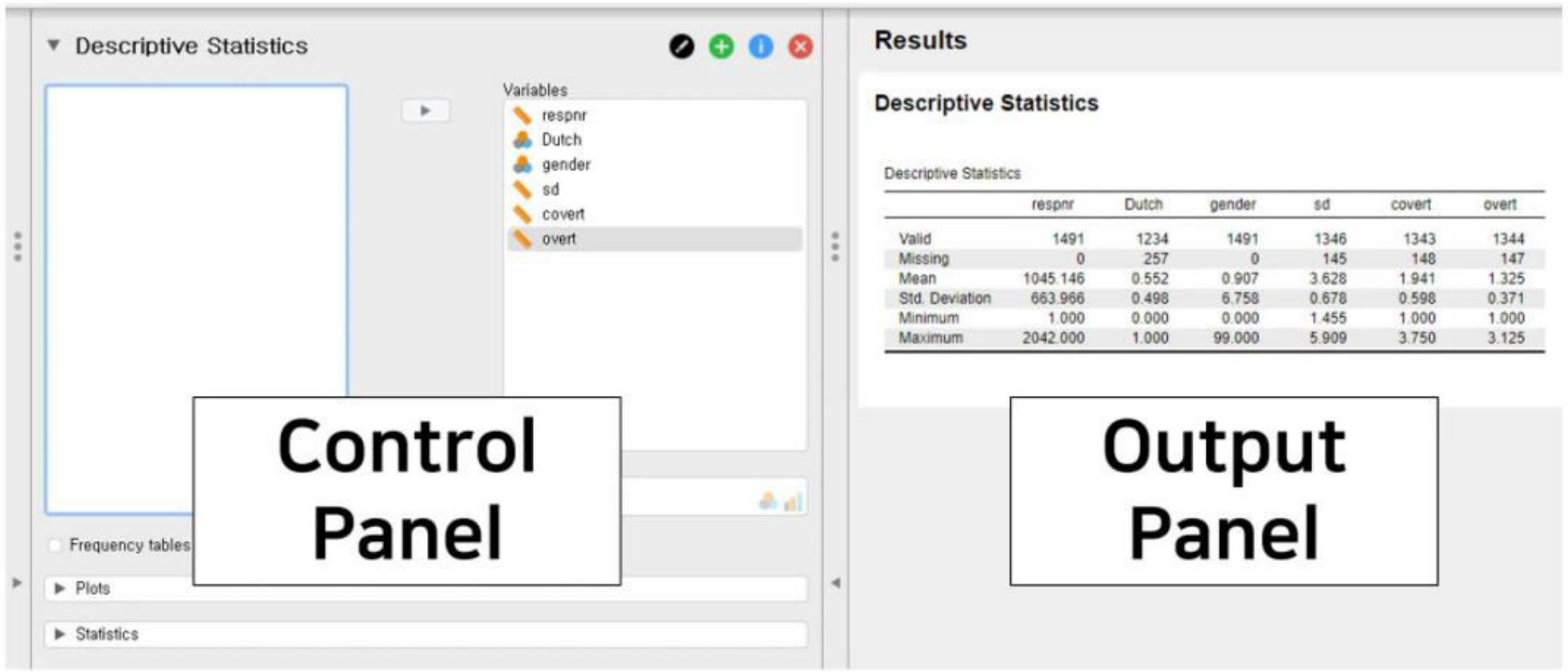
You can add filters, or calculat variables, but it is done very inefficiently (and sometime cause crashes)

The suggestion is to do all thes operations outside (like in Exce and then update the data file.

The database can be open externally by double-clicking data panel

Panels

There are three panels: Data, Control, Output.

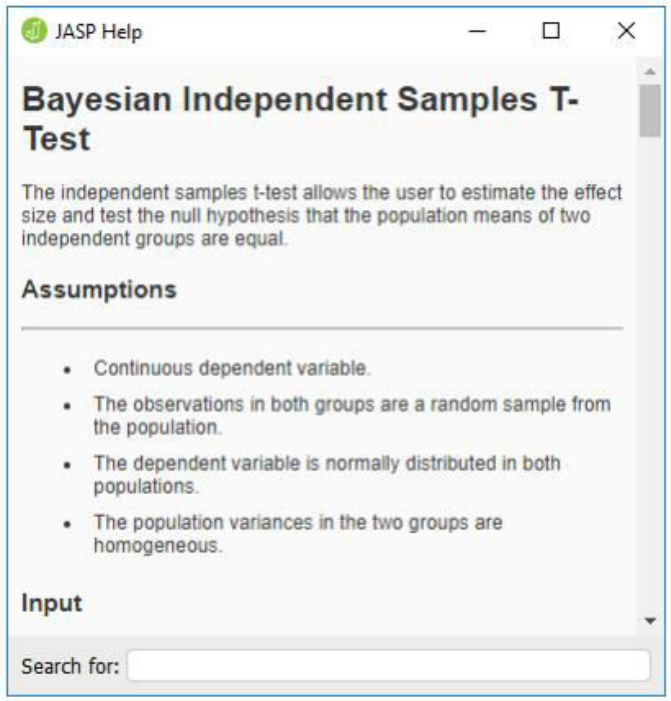


Top Bar Analysis Menu

Descriptives • Descriptive stats	**Bayesian Correlation & Regression** • Correlation • Linear regression
Bayesian T-Tests • Independent • Paired • One sample	**Bayesian Frequencies** • Binomial test • Multinomial test • Contingency tables
Bayesian ANOVA • Independent • Repeated measures • Mixed factor	**BAIN** • Bayesian informative hypotheses evaluation

Information

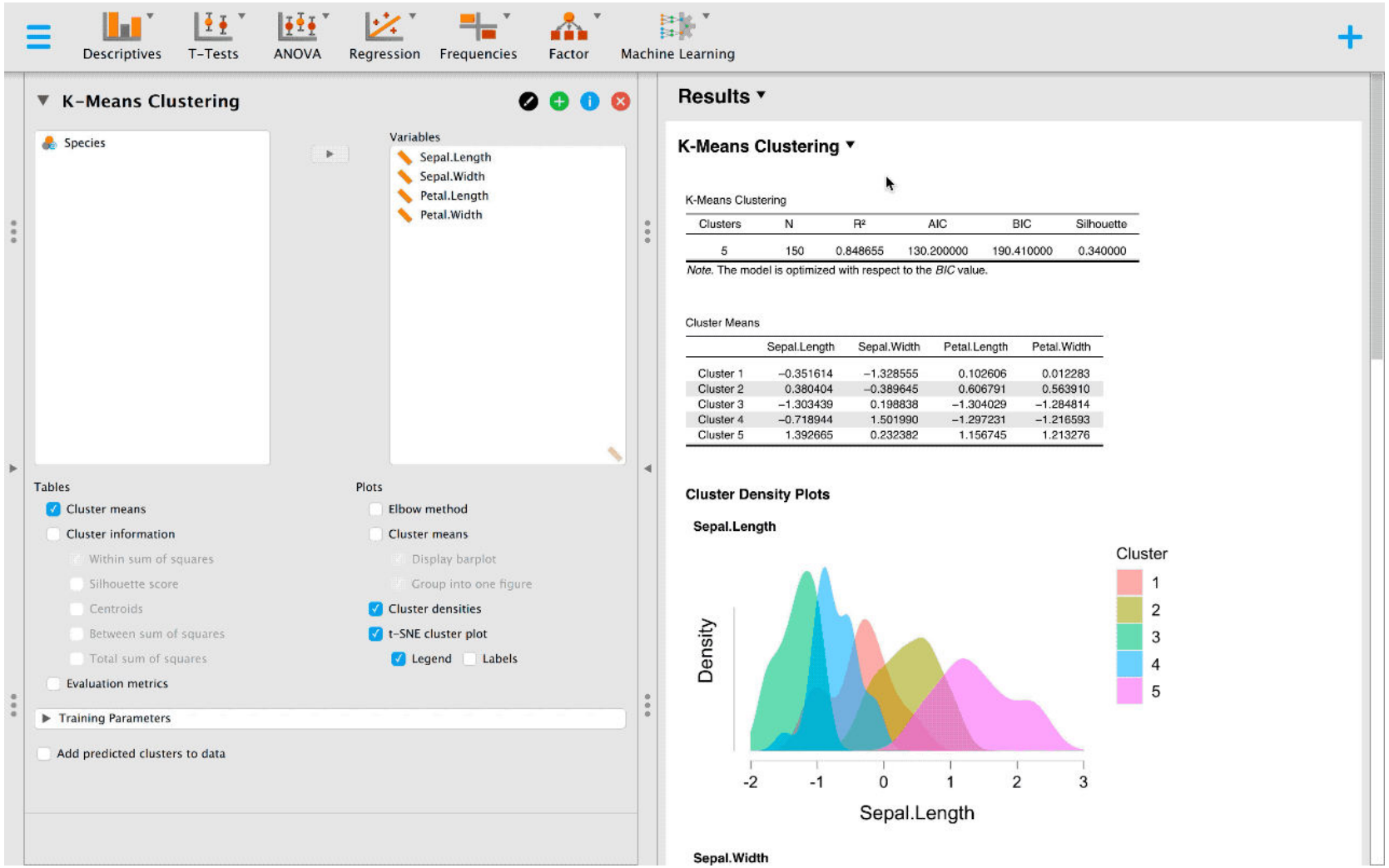
The blue information icon provides detailed information on each of the statistical procedures used and a search option.



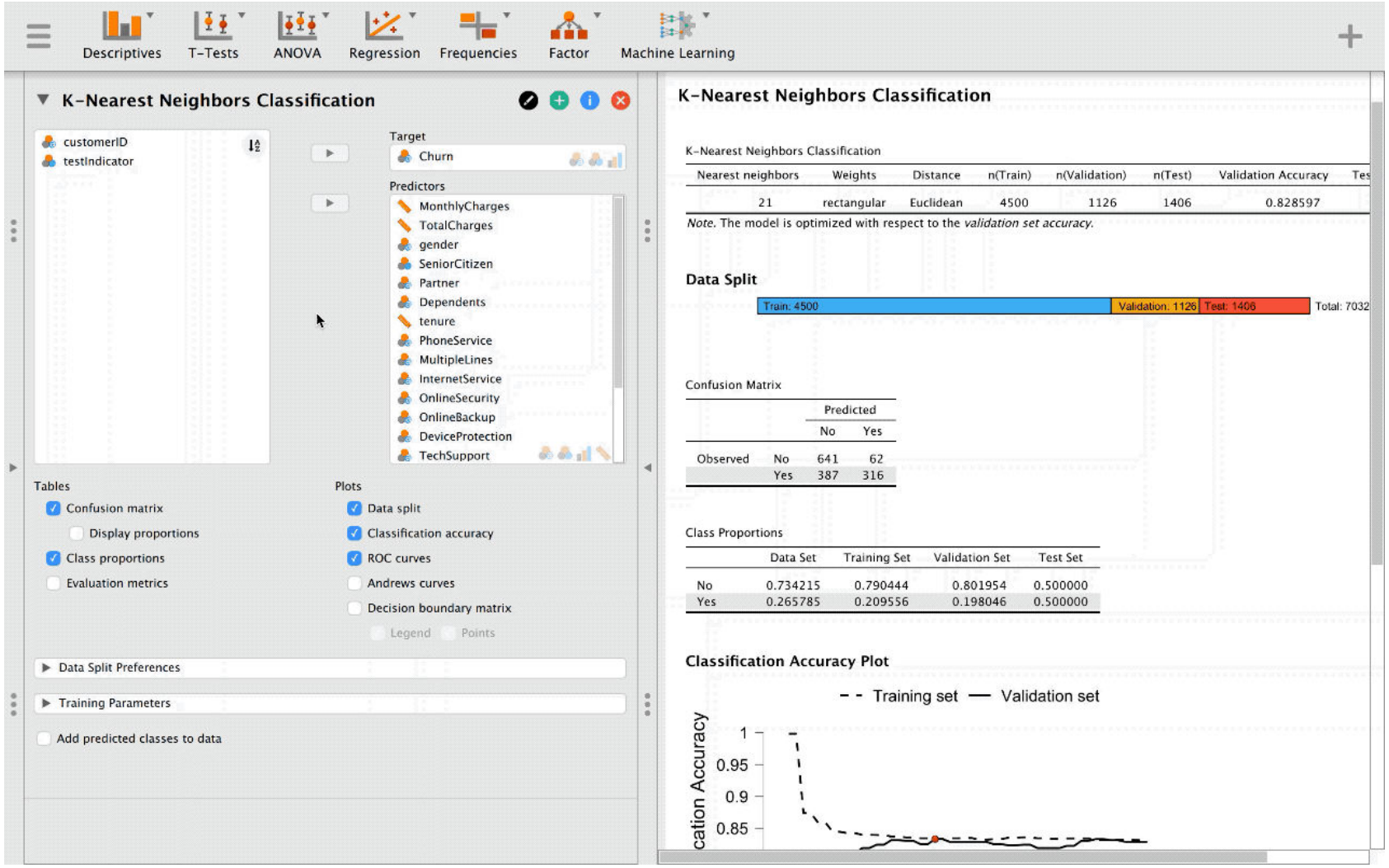
Info reported:

- Assumptions
- Available input options
- Available Priors
- Given Output information
- References for the analysis
- R packages involved

Annotations



Export Results



An (easy) example. Bayesian t-test

- **Open** Kitchen Rolls for independent sample t-test

Kitchen Rolls: Independent Samples T-Test

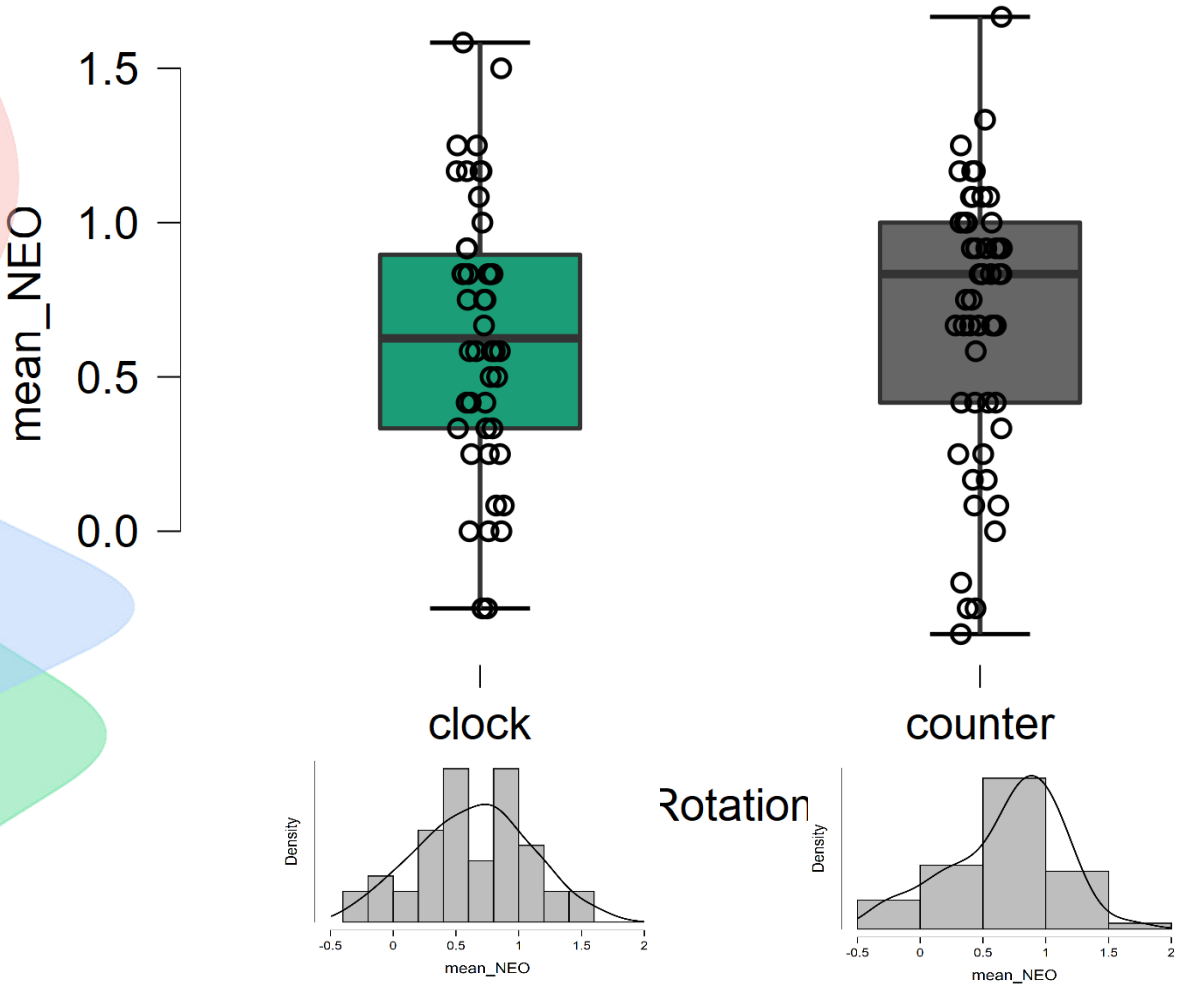
Description:

This data set, "Kitchen Rolls", provides Openness to Experience scores for two groups of students - while filling out the personality questionnaire, both groups rotated a kitchen roll with their hands (one group clockwise, the other group counterclockwise).

Variables:

- **mean_NEO** - Mean score on the NEO PI-R that consists of 12 questions (q1_NEO, ..., q12_NEO).
- **Rotation -** Experimental group of the direction the kitchen roll was turned (clock = clockwise, counter = counterclockwise).

Topolinski, S. and Sparenberg, P (2012). Turning the hands of time: Clockwise movements increase preference for novelty. *Social Psychological and Personality Science,* 3:308-314.
Wagenmakers, E.-J., Beek, T. F., Rotteveel, M., Gierholz, A., Matzke, D., Steingroever, H., … Pinto, Y. (2015). Turning the hands of time again: A purely confirmatory replication study and a Bayesian analysis. *Frontiers in Psychology*, *6*.



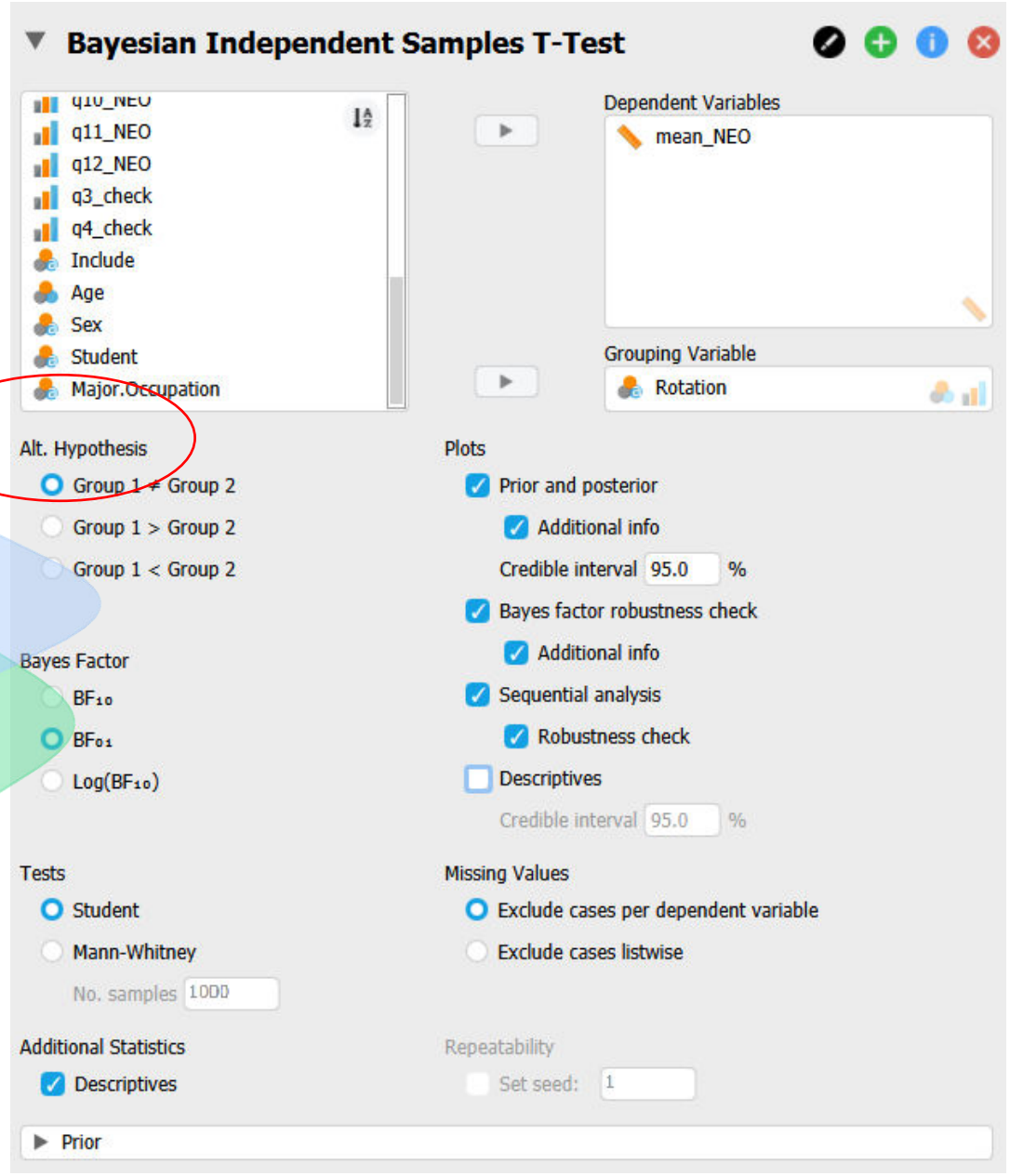
Independent Samples T-Test

	t	df	p	Cohen's d
mean_NEO	-0.534	97	0.595	-0.108

Note. Student's t-test.

An (easy) example. Bayesian t-test

- **Open** Kitchen Rolls for independent sample t-test
- Explore the environment
- Navigate the t-test options
  - Plots
  - Robust analysis
- Observe and comment the output
- Explore the 



Bayesian Independent Samples T-Test

	BF_{01}	error %
mean_NEO	4.151	0.033

Bayesian Independent Samples T-Test

	BF_{0+}	error %
mean_NEO	6.736	~ 0.007

Note. For all tests, the alternative hypothesis specifies that group *clock* is greater than group *counter* .

Bayesian Independent Samples T-Test

Variables: q10_NEO, q11_NEO, q12_NEO, q3_check, q4_check, Include, Age, Sex, Student, Major.Occupation

Dependent Variables: mean_NEO

Grouping Variable: Rotation

Alt. Hypothesis
- (•) Group 1 ≠ Group 2
- ( ) Group 1 > Group 2
- ( ) Group 1 < Group 2

Plots
- [x] Prior and posterior
  - [x] Additional info
  - Credible interval 95.0 %
- [x] Bayes factor robustness check
  - [x] Additional info
- [x] Sequential analysis
  - [x] Robustness check
- [ ] Descriptives
  - Credible interval 95.0 %

Bayes Factor
- ( ) BF_{10}
- (•) BF_{01}
- ( ) $Log(BF_{10})$

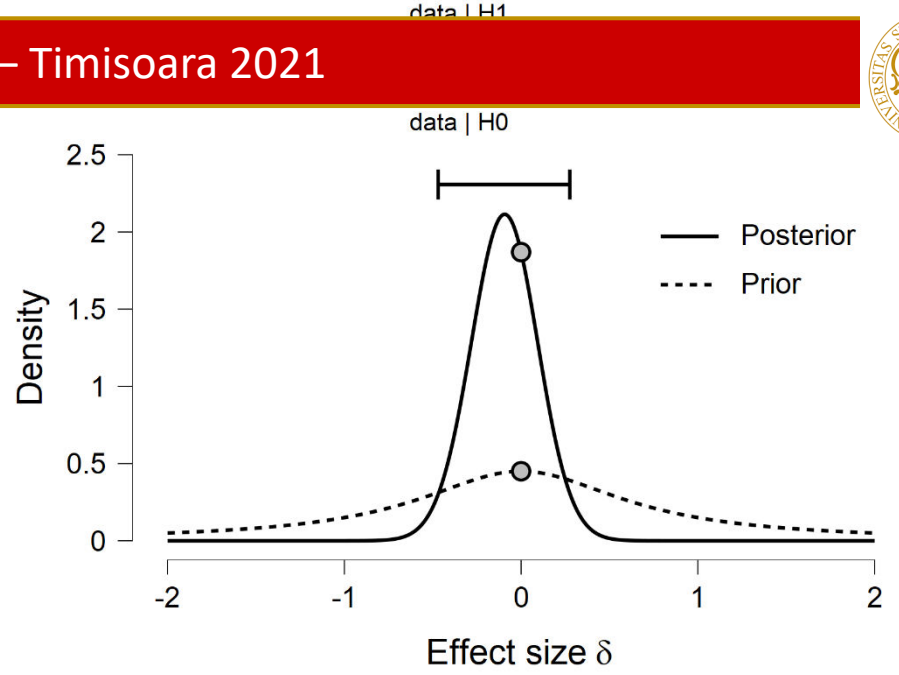
Tests
- (•) Student
- ( ) Mann-Whitney
  - No. samples 1000

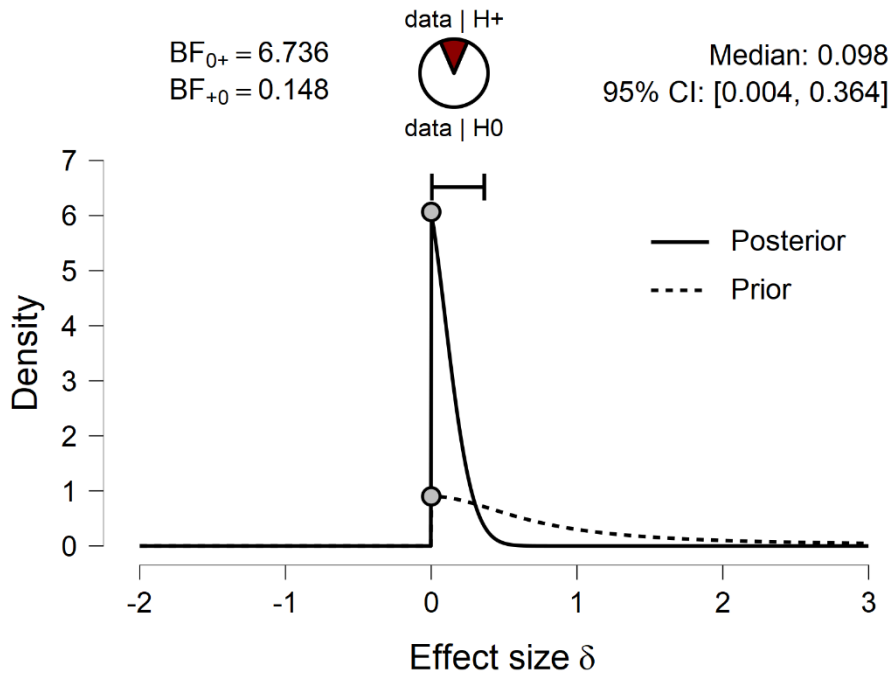
Missing Values
- (•) Exclude cases per dependent variable
- ( ) Exclude cases listwise

Additional Statistics
- [x] Descriptives

Repeatability
- [ ] Set seed: 1

▶ Prior





q11_NEO
q12_NEO
q3_check
q4_check
Include
Age
Sex
Student
Major.Occupation

Dependent Variables
mean_NEO

Grouping Variable
Rotation

Alt. Hypothesis
- Group 1 ≠ Group 2
- Group 1 > Group 2
- Group 1 < Group 2

Bayes Factor
- BF_{10}
- BF_{01}
- $Log(BF_{10})$

Plots
- Prior and posterior
  - Additional info
  - Credible interval 95.0 %
- Bayes factor robustness check
  - Additional info
- Sequential analysis
  - Robustness check
- Descriptives
  - Credible interval 95.0 %

Tests
- Student
- Mann-Whitney

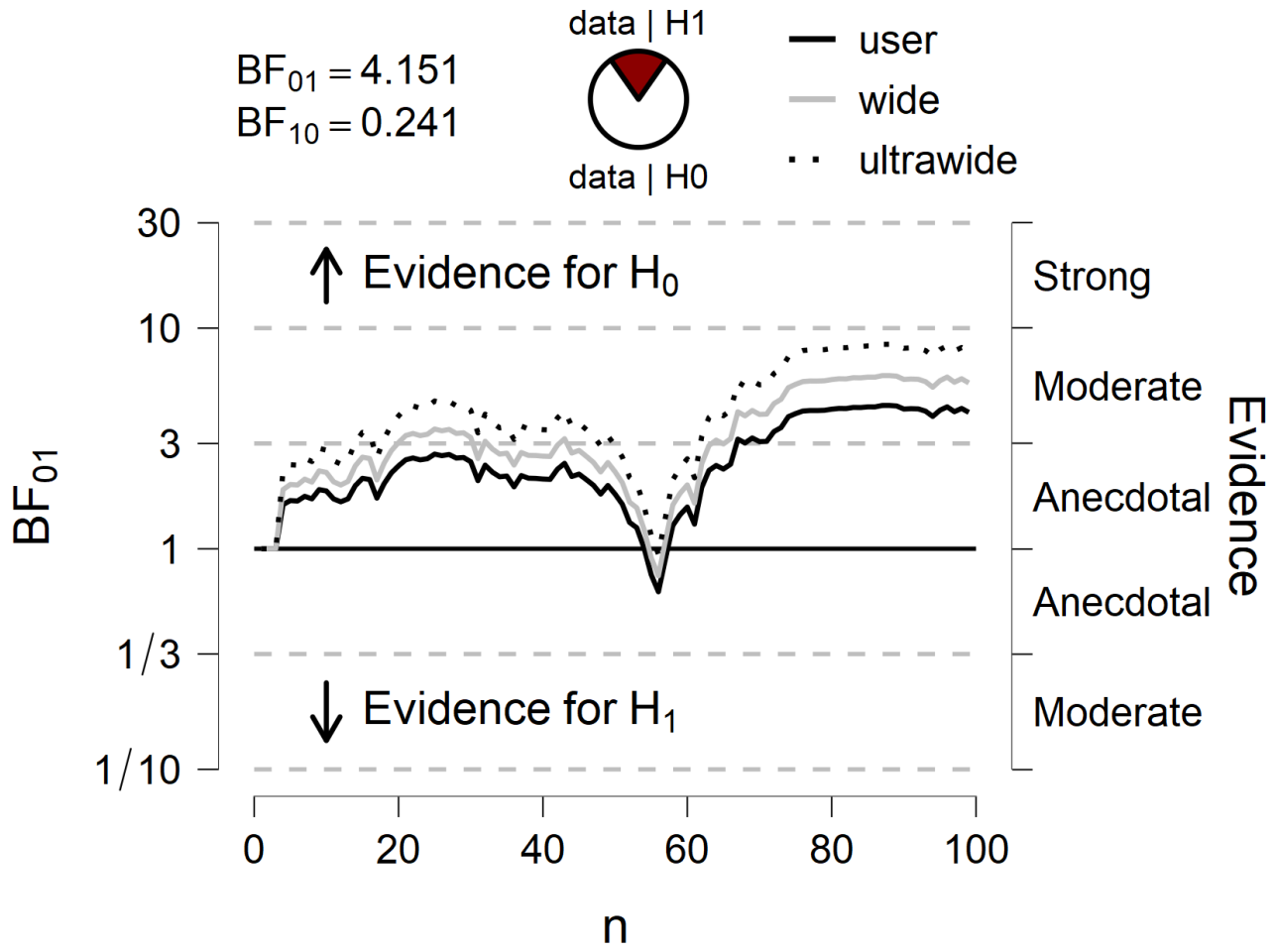
No. samples 1000

Missing Values
- Exclude cases per dependent variable
- Exclude cases listwise

Additional Statistics
- Descriptives

Repeatability
- Set seed: 1

Prior



q11_NEO
q12_NEO
q3_check
q4_check
Include
Age
Sex
Student
Major.Occupation

Dependent Variables
mean_NEO

Grouping Variable
Rotation

Alt. Hypothesis
- Group 1 ≠ Group 2
- Group 1 > Group 2
- Group 1 < Group 2

Bayes Factor
- BF_{10}
- BF_{01}
- $Log(BF_{10})$

Plots
- Prior and posterior
  - Additional info
  - Credible interval 95.0 %
- Bayes factor robustness check
  - Additional info
- Sequential analysis
  - Robustness check
- Descriptives
  - Credible interval 95.0 %

Tests
- Student
- Mann-Whitney
  - No. samples 1000

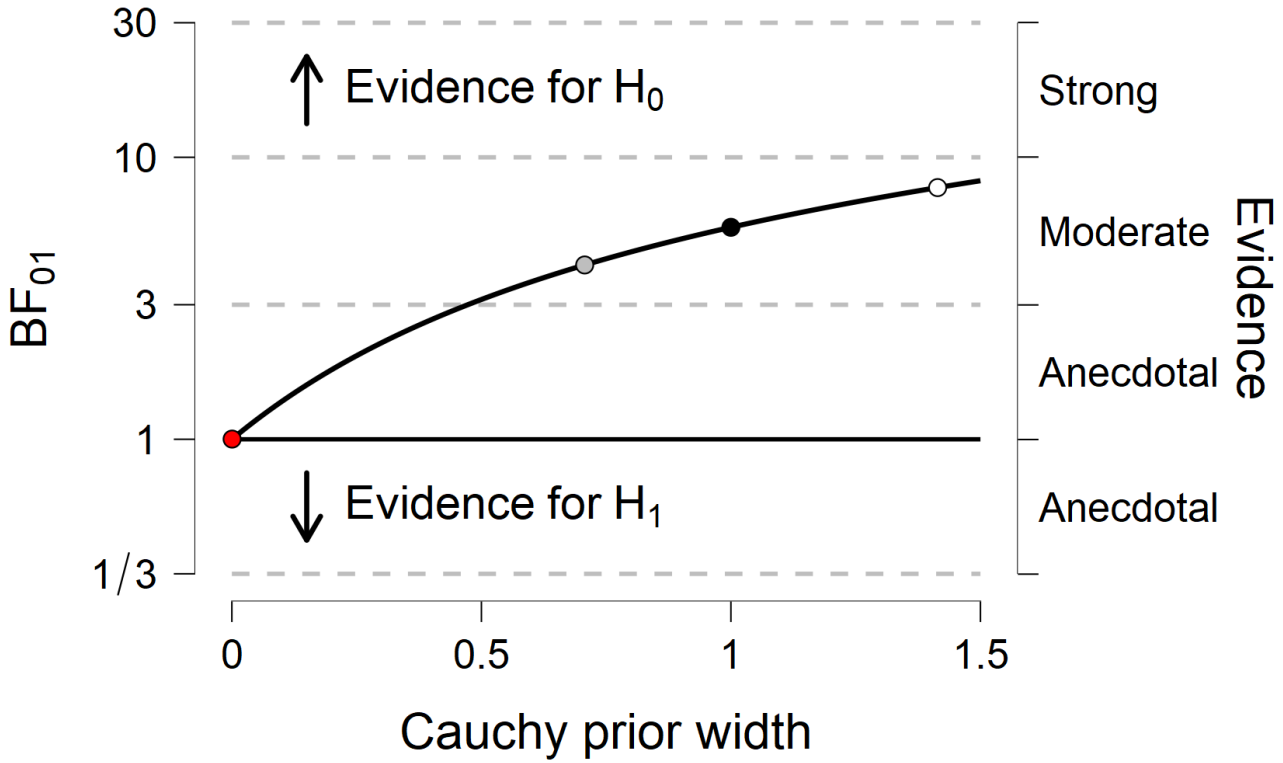
Missing Values
- Exclude cases per dependent variable
- Exclude cases listwise

Additional Statistics
- Descriptives

Repeatability
- Set seed: 1

Prior

- max BF_{01}: 1.002 at r = 5e-04
- user prior: $BF_{01} = 4.151$
- wide prior: $BF_{01} = 5.645$
- ultrawide prior: $BF_{01} = 7.808$



Prior

Default

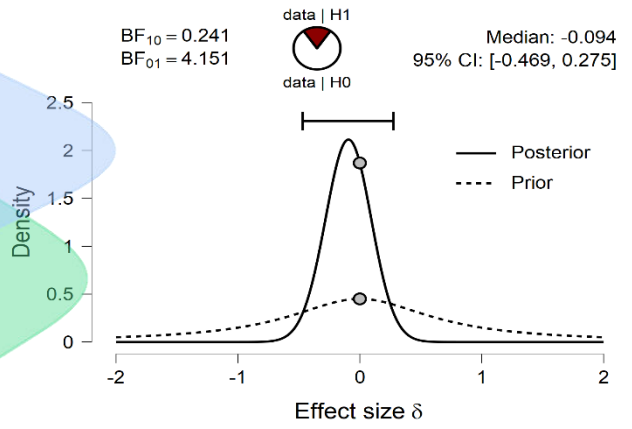
Cauchy scale 0.707

Informed

Cauchy location: 0 scale: 0.707

Normal

t

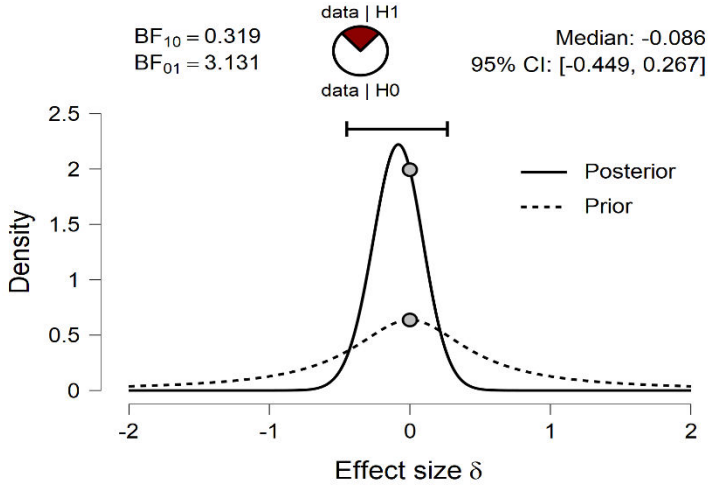


Informed

Cauchy location: 0 scale: 0.5

Normal

t

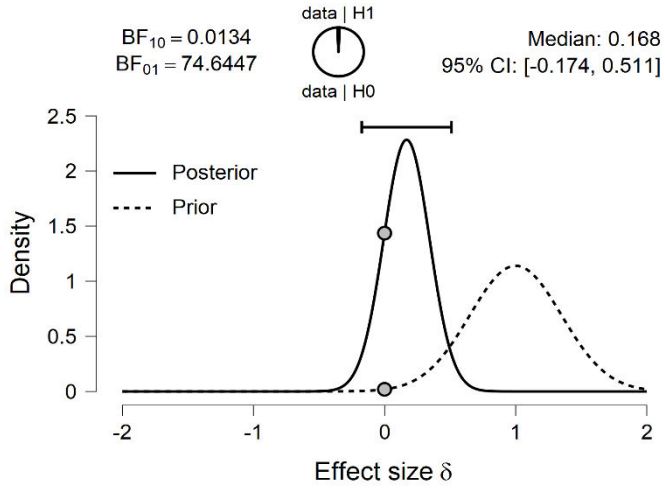


Informed

Cauchy

Normal mean: 1 std: 0.35

t



Standard Priors:

- normal distribution with mean 0 and standard deviation (σ) 10 (Balanced)
- Student's t distribution with mean 0, standard deviation (σ) 10; 3 df (Best for parameter estimation)
- Cauchy distribution with mean 0 and scale $\frac{\sqrt{2}}{2}(.707)$. (best for Bayes Factor)

Personalised Priors (According to the recommendations by Dienes 2019)

- normal distribution with the standard deviation coincident to the clinically relevant difference divided by 2,
- Cauchy distribution with the scale parameter divided by 7.

Elicited priors (expert and non expert) (Stefan et al., 2021).

- Different procedures available

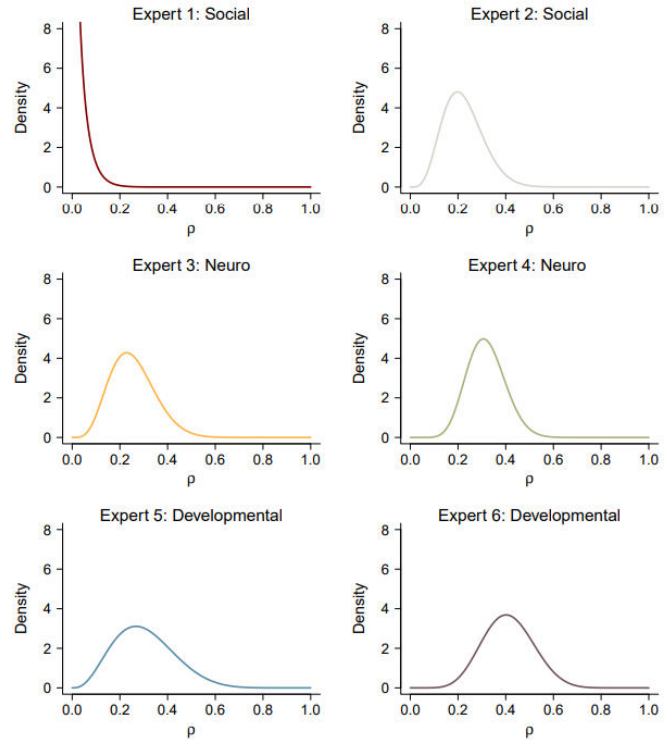
Expert Agreement in Prior Elicitation and its Effects on Bayesian Inference

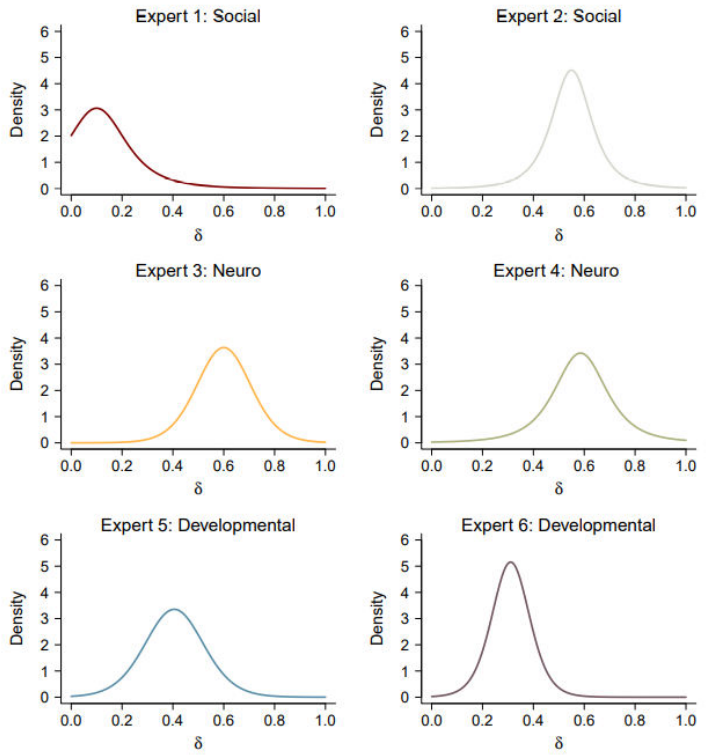
Angelika M. Stefan*[a], Dimitris Katsimpokis*[b], Quentin F. Gronau[a]
Eric-Jan Wagenmakers[a]

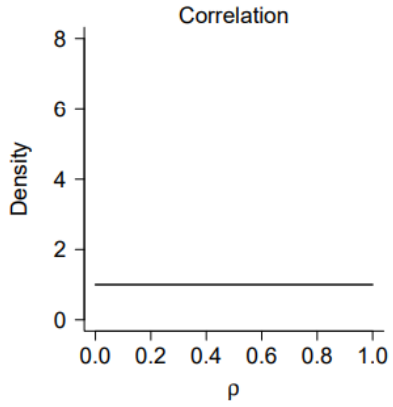
[a] Department of Psychology, University of Amsterdam, The Netherlands

[b] Department of Psychology, University of Basel, Switzerland

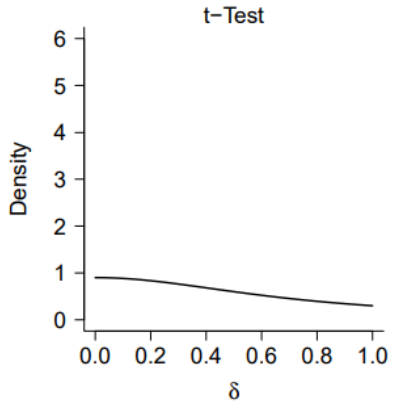
* AMS and DK contributed equally to the paper and share first authorship.



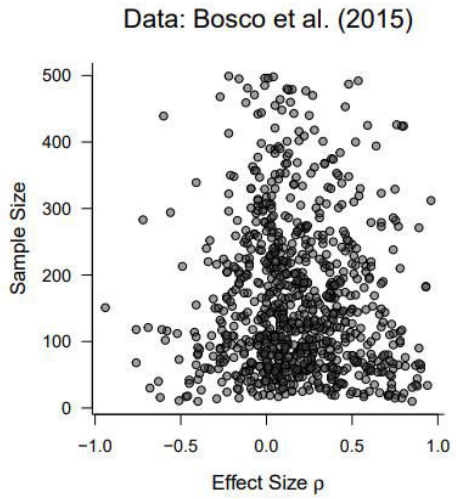


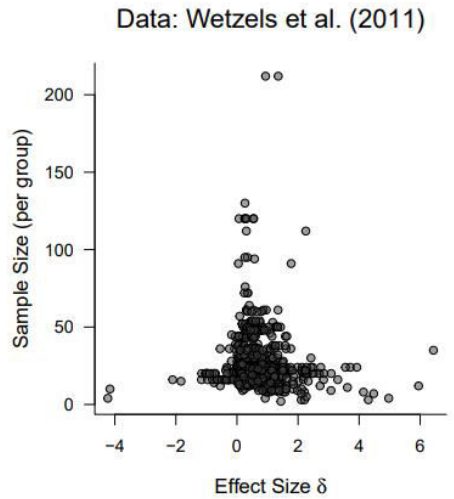


Uniform



Cauchy scale.707





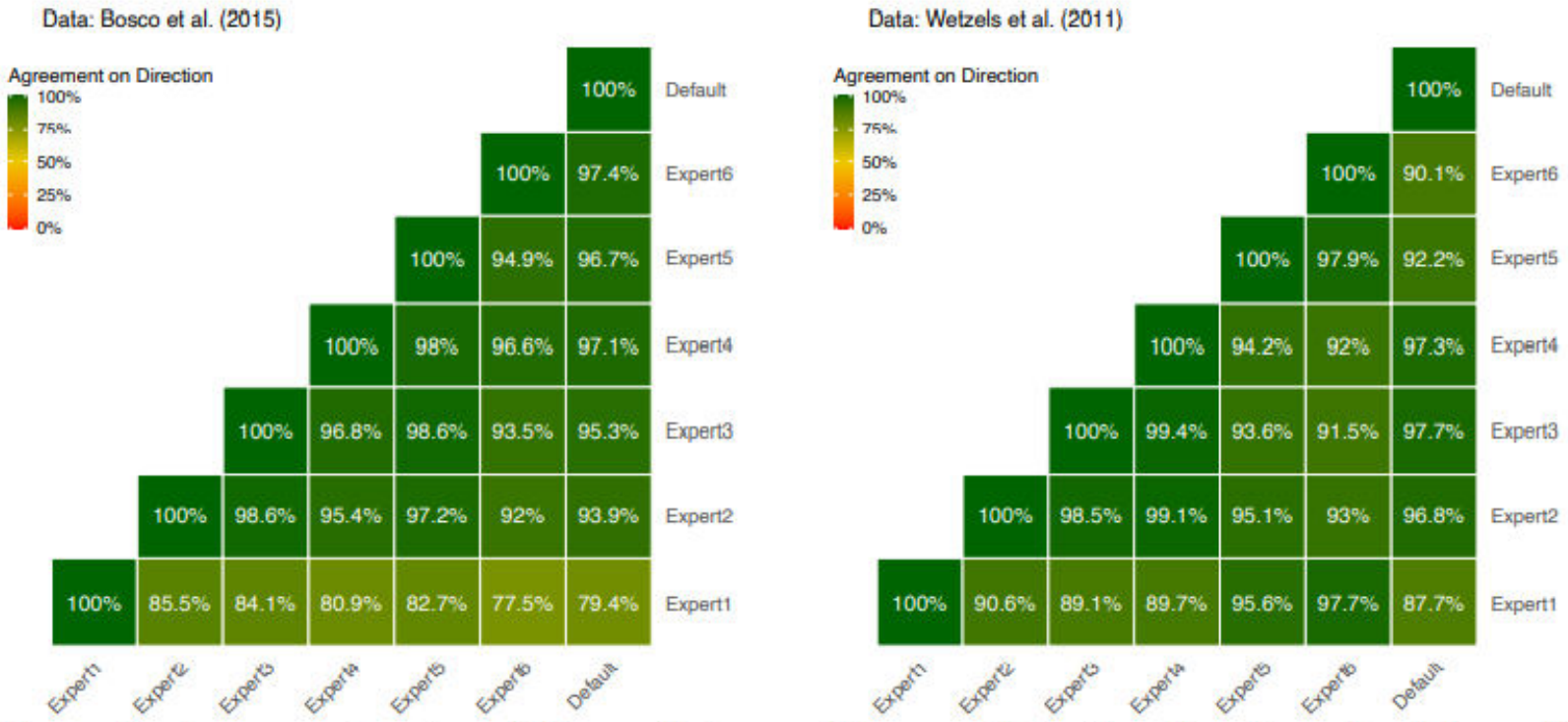


Figure 7. Agreement rates of Bayes factors with regard to the direction of evidence for all combinations of prior distributions. Agreement criterion: Both Bayes factors are either larger than 1 or smaller than 1.

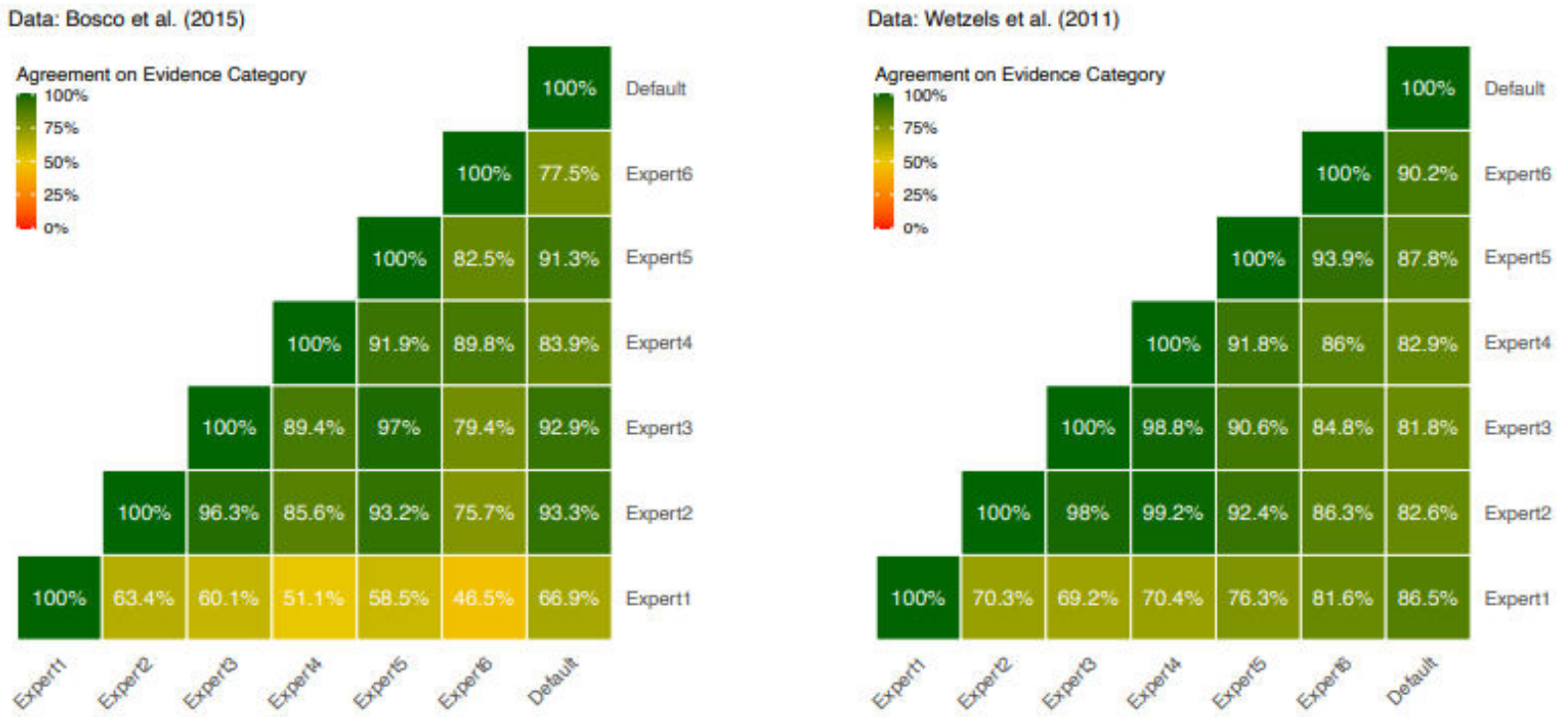


Figure 8. Agreement rates of Bayes factors with regard to the evidence category for all combinations of prior distributions. Here, strong evidence is defined as $BF_{10} > 10$ or $BF_{10} < 1/10$. Bayes factors are considered to possess the same strength of evidence if both Bayes factors show strong evidence for the same hypothesis or if both Bayes factors show inconclusive evidence.

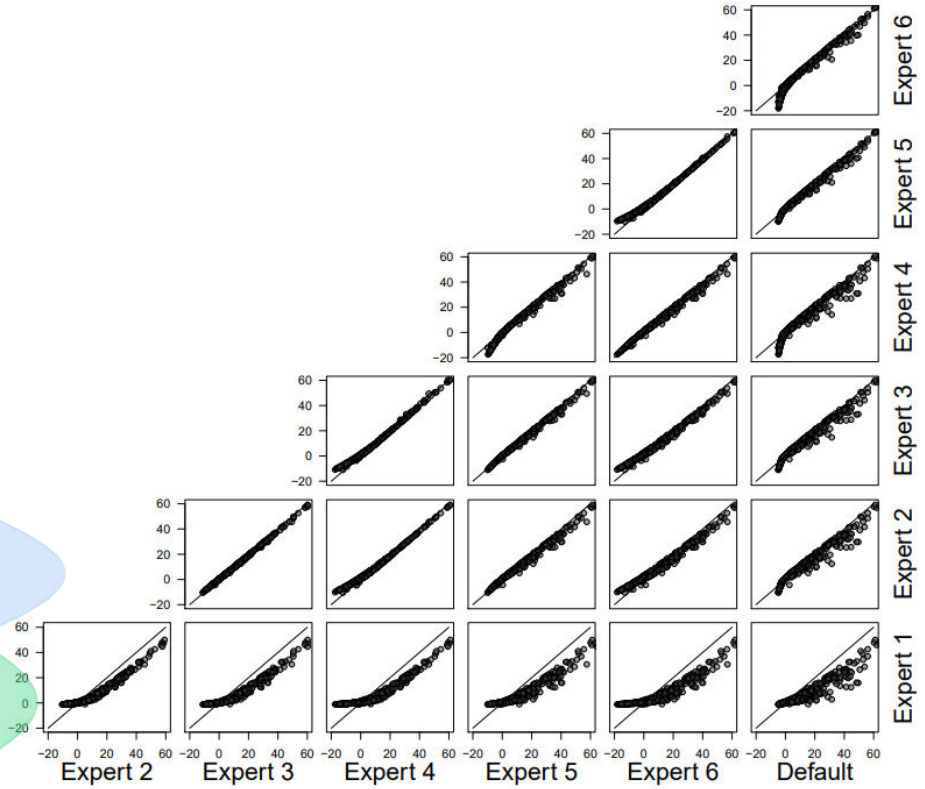


Figure 9. Correspondence between log Bayes factors for all prior distributions in the Bosco et al. (2015) database. The diagonal line marks equal values.

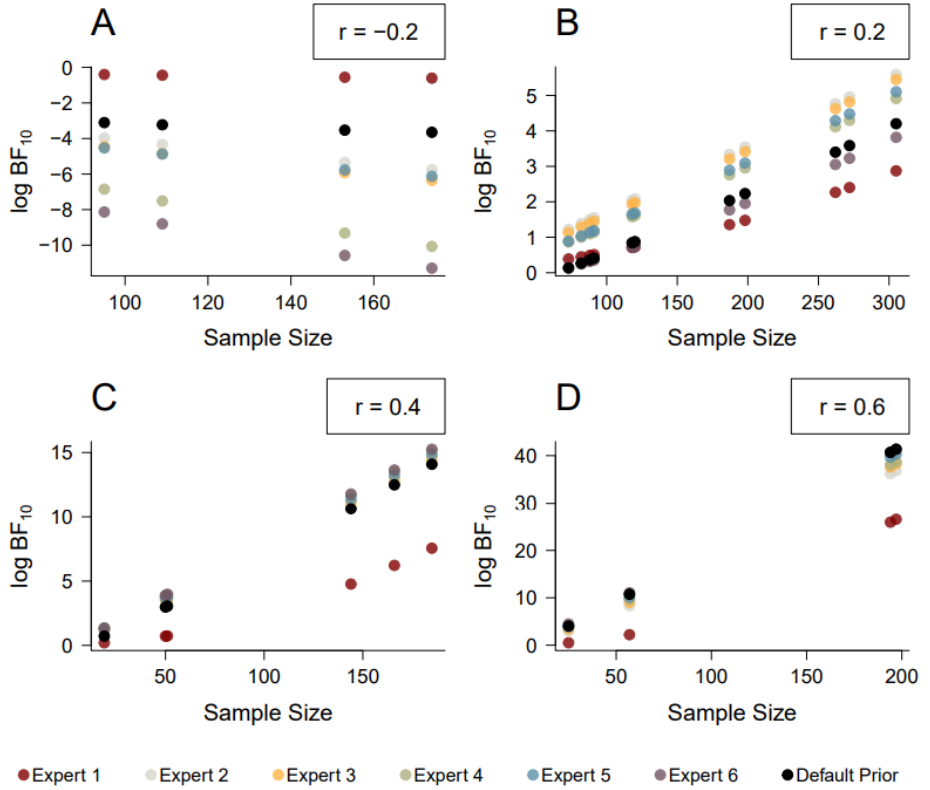


Figure 11. Variation in log Bayes factors for four different observed effect sizes (panel A-D) in the Bosco et al. (2015) database depending on different priors (color coded) and sample sizes.

Modest differences in elicited expert knowledge are still visible in the statistical results, but rarely change the qualitative conclusions of the model comparison. Concerns that idiosyncrasies between experts might jeopardize the objectivity of their statistical analyses are easily overstated. We hope that this insight will lead more researchers to embrace informed Bayesian inference in the future.

A (sligthly) more complex example. ANOVA

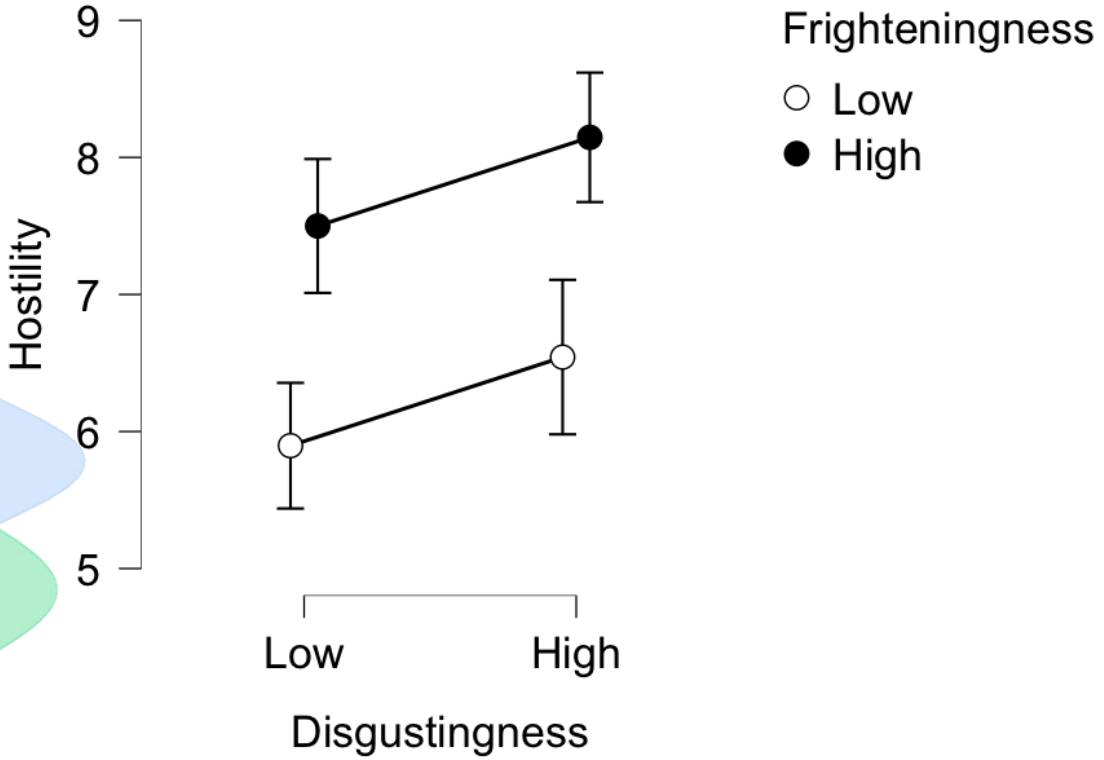
- **Open «Bugs»** in ANOVA folder

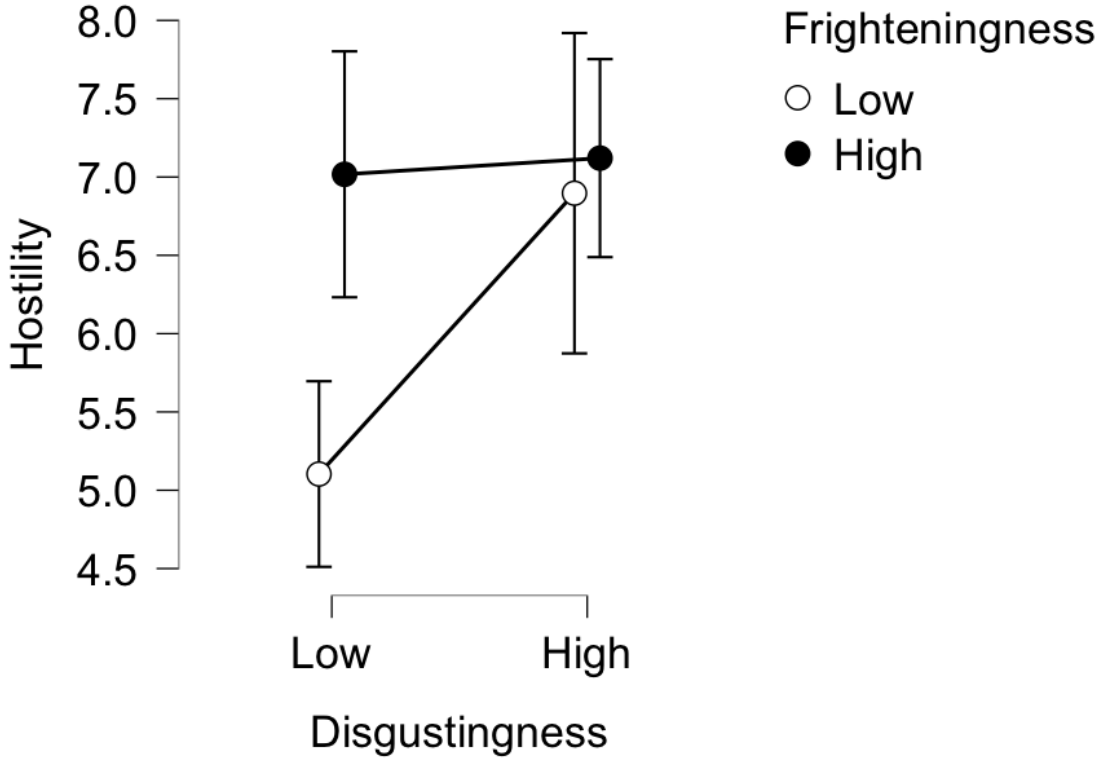
Within Subjects Effects

Cases	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2
Disgustingness	None	49.174	1.000	49.174	12.061	8.122e -4	0.018
Disgustingness ✻ Gender	None	1.760	1.000	1.760	0.432	0.513	6.358e -4
Residuals	None	346.552	85.000	4.077			
Frighteningness	None	138.075	1.000	138.075	32.122	1.939e -7	0.050
Frighteningness ✻ Gender	None	5.523	1.000	5.523	1.285	0.260	0.002
Residuals	None	365.366	85.000	4.298			
Disgustingness ✻ Frighteningness	None	13.799	1.000	13.799	4.688	0.033	0.005
Disgustingness ✻ Frighteningness ✻ Gender	None	13.799	1.000	13.799	4.688	0.033	0.005
Residuals	None	250.177	85.000	2.943			

Note. Sphericity corrections not available for factors with 2 levels.

Note. Type III Sum of Squares


9
8
7
6
5
Hostility
Low
High
Disgustingness
Frighteningness
Low
High

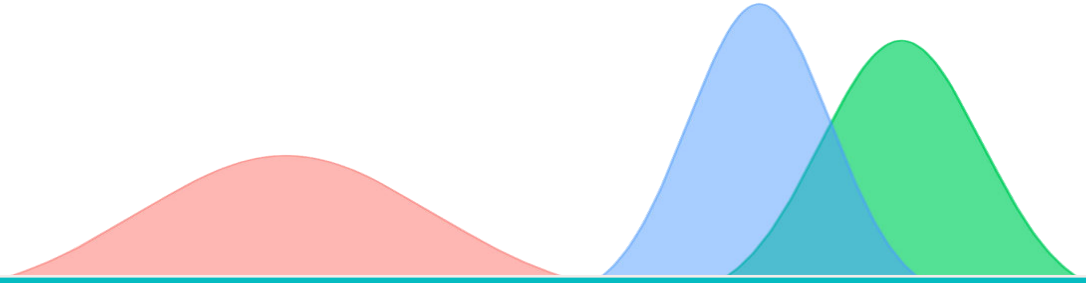

8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
Hostility
Low
High
Disgustingness
Frighteningness
Low
High

daniele.romano@unimib.it

daniele.romano84 / danromano

ANOVA with Bayesian statistics

Daniele Romano,
Università degli Studi di Milano-Bicocca
daniele.romano@unimib.it



This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

A (sligthly) more complex example. ANOVA

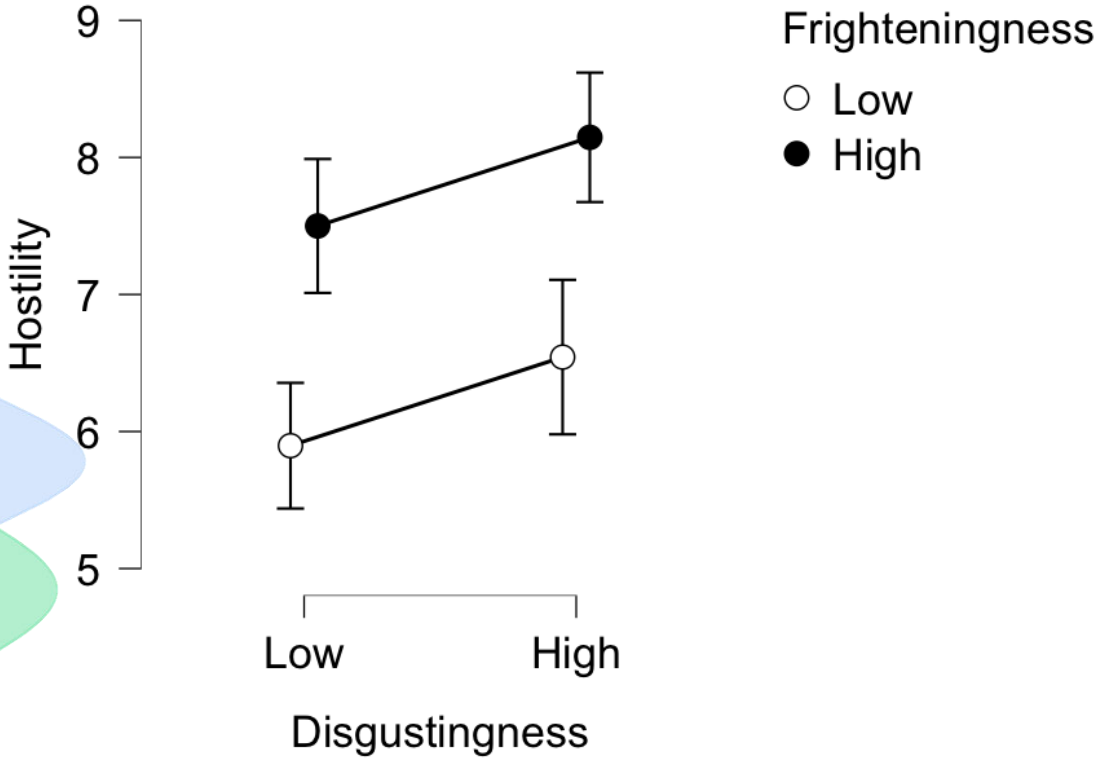
- **Open «Bugs»** in ANOVA folder

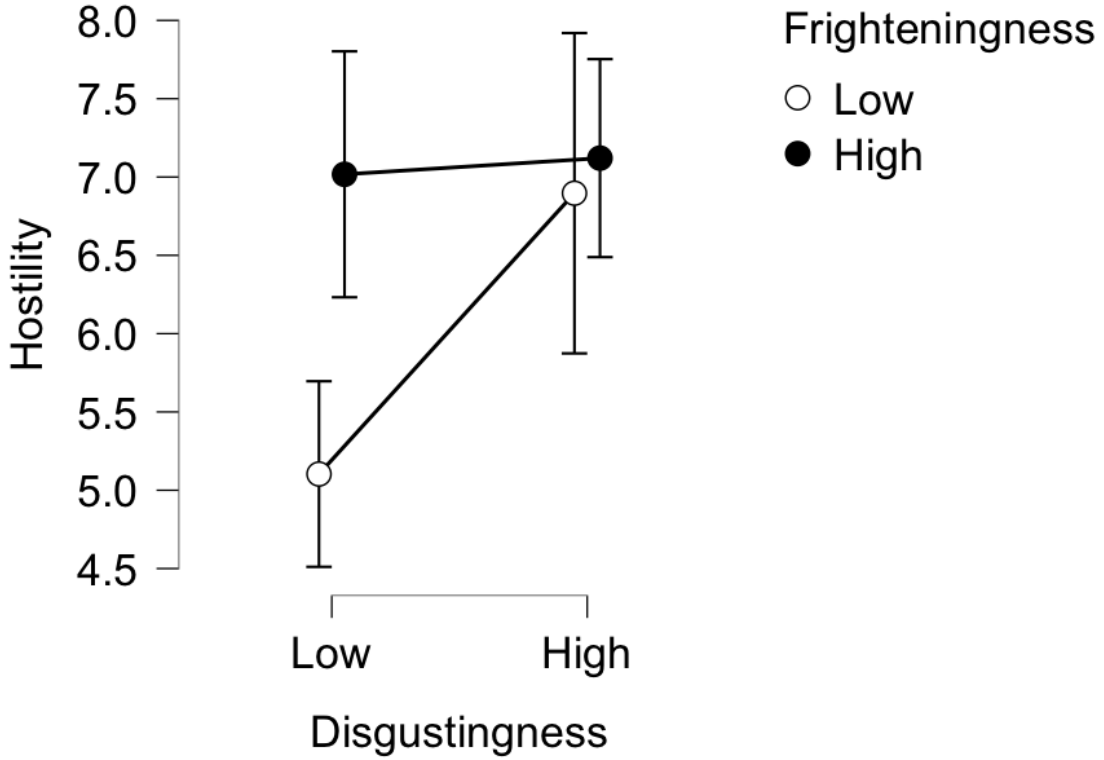
Within Subjects Effects

Cases	Sphericity Correction	Sum of Squares	df	Mean Square	F	p	η^2
Disgustingness	None	49.174	1.000	49.174	12.061	8.122e -4	0.018
Disgustingness ✻ Gender	None	1.760	1.000	1.760	0.432	0.513	6.358e -4
Residuals	None	346.552	85.000	4.077			
Frighteningness	None	138.075	1.000	138.075	32.122	1.939e -7	0.050
Frighteningness ✻ Gender	None	5.523	1.000	5.523	1.285	0.260	0.002
Residuals	None	365.366	85.000	4.298			
Disgustingness ✻ Frighteningness	None	13.799	1.000	13.799	4.688	0.033	0.005
Disgustingness ✻ Frighteningness ✻ Gender	None	13.799	1.000	13.799	4.688	0.033	0.005
Residuals	None	250.177	85.000	2.943			

Note. Sphericity corrections not available for factors with 2 levels.

Note. Type III Sum of Squares


9
8
7
6
5
Hostility
Low
High
Disgustingness
Frighteningness
Low
High


8.0
7.5
7.0
6.5
6.0
5.5
5.0
4.5
Hostility
Low
High
Disgustingness
Frighteningness
Low
High

A (sligthly) more complex example. Bayesian ANOVA

- **Open «Bugs»** in ANOVA folder
- Explore the environment of Bayesian ANOVA

Journal of Mathematical Psychology 56 (2012) 356–374

Contents lists available at SciVerse ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Default Bayes factors for ANOVA designs

Jeffrey N. Rouder [a,*], Richard D. Morey [b], Paul L. Speckman [c], Jordan M. Province [a]

[a] Department of Psychological Sciences, University of Missouri, United States
[b] Faculty of Behavioural and Social Sciences, University of Groningen, The Netherlands
[c] Department of Statistics, University of Missouri, United States

Rouder (2012) modeling framework is influenced by Gelman (2005), who conceptualizes ANOVA as a hierarchical model in which effects are clustered within factors.

Predictions for competing models are obtained by marginalizing over the parameters, and the evidence for one model relative to another is the ratio of the densities of these predictions evaluated at the observed data

A (sligthly) more complex example. Bayesian ANOVA

- **Open «Bugs»** in ANOVA folder
- Explore the environment of Bayesian ANOVA
- Navigate the options
  - Bayes Factor
  - Effects
  - Single model Inference
  - Post-Hoc
- Observe and comment the output
- Explore the 

Bayesian ANOVA

Model Comparison - Dependent Variable:

- Models: The first column contains all the models included in the analysis.
  - Null model: This model contains the grand mean and the random factors.
  - Independent Variable model: This model adds the effect of the independent variable.
- P(M): This column contains the prior model probability.
- P(M|data): This column contains the updated probability of the model given the data. This is called the posterior model probability.
- BF_M : This column contains the posterior model odds. This is the change from the prior odds to the posterior odds for the model.

- BF_{10} : This column contains the Bayes factor that quantifies evidence for the alternative hypothesis relative to the null hypothesis/null model. However, when the option Compare to best model is selected, the column will contain the Bayes factor that quantifies evidence for this model relative to the best model.
- BF_{01} : This column contains the Bayes factor that quantifies evidence for the null hypothesis/null model relative to the alternative hypothesis. However, when the option Compare to best model is selected, the column will contain the Bayes factor that quantifies evidence for the best model relative to this model.

- error % : The error of the Gaussian quadrature integration routine used by the BayesFactor package for the computation of the Bayes factor.

Analysis of Effects - Dependent Variable:

•Effects: This column contains the components included in the models, such as independent variables and their interactions.

•P(incl): This column contains the prior inclusion probability. This is the prior probability summed across all models that include the component.

•P(incl|data): This column contains the posterior inclusion probability. This is the summed posterior probability over all models that include the component.

•**$BF_{inclusion}$:** This column contains the change from prior inclusion odds to posterior inclusion odds for each component averaged by all the models that includes the component.

Model Averaged Posterior Summary:

•Variable: This column contains all the fixed factors and their interactions included in the models. The first row contains information about the intercept.

•Level: Each level of the factor and combination of levels of the interactions that are included in the model.

•Mean: The model averaged mean. For the factors, this is the deviation from the intercept for each level of the factor. The level means for a factor sum to zero.

•SD: The standard deviation of the model averaged mean.

•% Credible interval: The credible interval of the mean. By default, this is set to 95%.

- Lower: The lower bound of the credible interval of the mean.
- Upper: The upper bound of the credible interval of the mean.

This example JASP file demonstrates the use of mixed design ANOVA. We will test the relation between hostility towards insects and their disgustingness and frighteningness for males and females separately.

This data set, "Bugs", provides the extent to which men and women want to kill arthropods that vary in freighteningness (low, high) and disgustingness (low, high). Each participant rates their attitudes towards all anthropods. Subset of the data reported by Ryan et al. (2013).

Variables:

- **Gender -** Participant's gender (Female, Male).
- **Lo D, Lo F -** Desire to kill an artrhopod with low freighteningness and low disgustingness.
- **Lo D, Hi F -** Desire to kill an artrhopod with low freighteningness and high disgustingness.
- **Hi D, Lo F -** Desire to kill an artrhopod with high freighteningness and low disgustingness.
- **Hi D, Hi F -** Desire to kill an artrhopod with high freighteningness and high disgustingness.
- For a description of the remaining variables see Ryan et al. (2013).

The desire to kill an arthropod was indicated on a scale from 0 to 10.

Ryan, R. S., Wilde, M., & Crist, S. (2013). Compared to a small, supervised lab experiment, a large, unsupervised web-based experiment on a previously unknown effect has benefits that outweigh its potential costs. *Computers in Human Behavior, 29*, 1295-1301.

Bayesian Repeated Measures ANOVA

Subject
Region
Education

Repeated Measures Factors

Frighteningness
H-F
L-F
Level 3

Repeated Measures Cells

Hi D, Hi F	H-F,H-D
Lo D, Hi F	H-F,L-D
Hi D, Lo F	L-F,H-D
Lo D, Lo F	L-F,L-D

Between Subject Factors

Gender

Covariates

Bayes Factor

Tables

Bayes Factor
- (•) BF_{10}
- ( ) BF_{01}
- ( ) $Log(BF_{10})$

Order
- (•) Compare to best model
- ( ) Compare to null model

Tables
- [x] Effects
  - (•) Across all models
  - ( ) Across matched models
- [x] Estimates
- [x] Model averaged R^2
- [ ] Descriptives

Credible interval 95.0 %

Plots
- [x] Model averaged posteriors
  - (•) Group levels in single plot
  - ( ) Individual plot per level
- [ ] Q-Q plot of residuals
- [ ] Posterior R^2

Additional Options

Prior
- r scale fixed effects 0.5
- r scale random effects 1
- r scale covariates 0.354

Posterior Samples
- (•) Auto
- ( ) Manual
  - No. samples 1000

Numerical Accuracy
- (•) Auto
- ( ) Manual
  - No. samples 10000

Repeatability
- [ ] Set seed: 1

Model Comparison

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Frighteningness + Disgustingness	0.053	0.431	13.642	1.000	
Frighteningness + Disgustingness + Gender	0.053	0.168	3.630	0.389	2.780
Frighteningness + Disgustingness + Frighteningness ✻ Disgustingness	0.053	0.167	3.602	0.387	3.434
Frighteningness + Disgustingness + Gender + Frighteningness ✻ Disgustingness	0.053	0.063	1.202	0.145	2.664
Frighteningness + Disgustingness + Gender + Frighteningness ✻ Gender	0.053	0.060	1.146	0.139	3.905
Frighteningness + Disgustingness + Gender + Disgustingness ✻ Gender	0.053	0.037	0.682	0.085	5.921
Frighteningness + Disgustingness + Gender + Frighteningness ✻ Disgustingness + Frighteningness ✻ Gender	0.053	0.023	0.416	0.052	3.628
Frighteningness + Disgustingness + Gender + Frighteningness ✻ Disgustingness + Disgustingness ✻ Gender	0.053	0.015	0.282	0.036	14.943
Frighteningness + Disgustingness + Gender + Frighteningness ✻ Gender + Disgustingness ✻ Gender	0.053	0.012	0.226	0.029	3.277
Frighteningness	0.053	0.010	0.174	0.022	2.036
Frighteningness + Disgustingness + Gender + Frighteningness ✻ Disgustingness + Frighteningness ✻ Gender + Disgustingness ✻ Gender	0.053	0.005	0.097	0.012	5.978
Frighteningness + Disgustingness + Gender + Frighteningness ✻ Disgustingness + Frighteningness ✻ Gender + Disgustingness ✻ Gender + Frighteningness ✻ Disgustingness ✻ Gender	0.053	0.005	0.094	0.012	5.869
Frighteningness + Gender	0.053	0.004	0.064	0.008	2.174
Frighteningness + Gender + Frighteningness ✻ Gender	0.053	0.001	0.023	0.003	2.473
Disgustingness	0.053	3.935e -9	7.083e -8	9.128e -9	2.259
Disgustingness + Gender	0.053	1.406e -9	2.532e -8	3.262e -9	2.192
Disgustingness + Gender + Disgustingness ✻ Gender	0.053	3.324e -10	5.983e -9	7.709e -10	12.465
Null model (incl. subject)	0.053	1.998e -10	3.597e -9	4.635e -10	1.511
Gender	0.053	7.306e -11	1.315e -9	1.695e -10	2.072

Note. All models include subject

Bayesian Repeated Measures ANOVA

Subject
Region
Education

Repeated Measures Factors

Frighteningness
H-F
L-F
Level 3

Repeated Measures Cells

Hi D, Hi F	H-F,H-D
Lo D, Hi F	H-F,L-D
Hi D, Lo F	L-F,H-D
Lo D, Lo F	L-F,L-D

Between Subject Factors

Gender

Covariates

Bayes Factor

Tables

Bayes Factor
- BF10
- BF01
- Log(BF10)

Tables
- Effects
  - Across all models
  - Across matched models
- Estimates
- Model averaged R^2
- Descriptives

Credible interval 95.0 %

Order
- Compare to best model
- Compare to null model

Plots
- Model averaged posteriors
  - Group levels in single plot
  - Individual plot per level
- Q-Q plot of residuals
- Posterior R^2

Additional Options

Prior
- r scale fixed effects 0.5
- r scale random effects 1
- r scale covariates 0.354

Posterior Samples
- Auto
- Manual
  - No. samples 1000

Numerical Accuracy
- Auto
- Manual
  - No. samples 10000

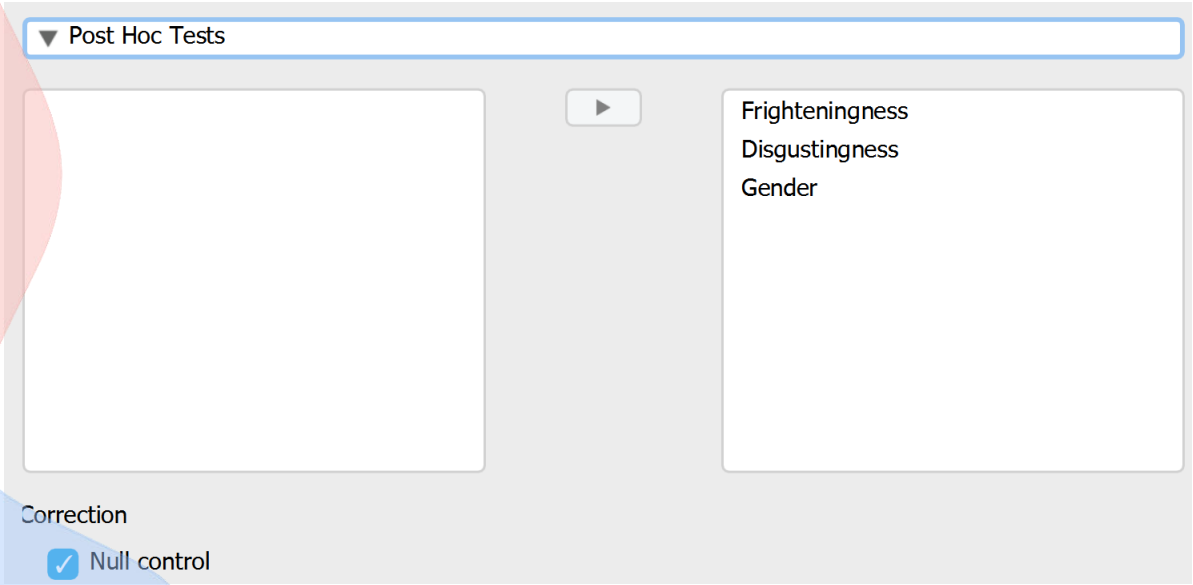
Repeatability
- Set seed: 1

Analysis of Effects

Effects	P(incl)	P(excl)	P(incl|data)	P(excl|data)	BF_{incl}
Frighteningness	0.737	0.263	1.000	5.947e -9	6.005e +7
Disgustingness	0.737	0.263	0.986	0.014	24.488
Gender	0.737	0.263	0.393	0.607	0.231
Frighteningness ✻ Disgustingness	0.316	0.684	0.278	0.722	0.834
Frighteningness ✻ Gender	0.316	0.684	0.107	0.893	0.259
Disgustingness ✻ Gender	0.316	0.684	0.075	0.925	0.175
Frighteningness ✻ Disgustingness ✻ Gender	0.053	0.947	0.005	0.995	0.094

Model Averaged Posterior Summary

Variable	Level	Mean	SD	95% Credible Interval Lower	Upper
Intercept		6.834	0.240	6.346	7.296
Frighteningness	H-F	0.693	0.106	0.476	0.902
	L-F	-0.693	0.106	-0.906	-0.480
Disgustingness	H-D	0.363	0.105	0.135	0.559
	L-D	-0.363	0.105	-0.580	-0.164
Gender	Female	0.207	0.227	-0.256	0.648
	Male	-0.207	0.227	-0.669	0.234
Frighteningness ✻ Disgustingness	H-F & H-D	-0.131	0.101	-0.335	0.067
	H-F & L-D	0.131	0.101	-0.070	0.333
	L-F & H-D	0.131	0.101	-0.071	0.333
	L-F & L-D	-0.131	0.101	-0.336	0.068
Frighteningness ✻ Gender	H-F & Female	0.128	0.107	-0.085	0.339
	H-F & Male	-0.128	0.107	-0.342	0.083
	L-F & Female	-0.128	0.107	-0.342	0.083
	L-F & Male	0.128	0.107	-0.085	0.339
Disgustingness ✻ Gender	H-D & Female	-0.065	0.106	-0.278	0.147
	H-D & Male	0.065	0.106	-0.149	0.276
	L-D & Female	0.065	0.106	-0.149	0.276
	L-D & Male	-0.065	0.106	-0.278	0.147
Frighteningness ✻ Disgustingness ✻ Gender	H-F & H-D & Female	0.181	0.104	-0.022	0.393
	H-F & H-D & Male	-0.181	0.104	-0.395	0.020
	H-F & L-D & Female	-0.181	0.104	-0.395	0.020
	H-F & L-D & Male	0.181	0.104	-0.022	0.393
	L-F & H-D & Female	-0.181	0.104	-0.395	0.020
	L-F & H-D & Male	0.181	0.104	-0.022	0.393
	L-F & L-D & Female	0.181	0.104	-0.022	0.393
	L-F & L-D & Male	-0.181	0.104	-0.395	0.020



Post Hoc Tests

Post Hoc Comparisons - Frighteningness

		Prior Odds	Posterior Odds	$BF_{10, U}$	error %
H-F	L-F	1.000	8.599e +7	8.599e +7	1.171e −12

Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy (0, r = 1/sqrt(2)) prior. The "U" in the Bayes factor denotes that it is uncorrected.

Post Hoc Comparisons - Disgustingness

		Prior Odds	Posterior Odds	$BF_{10, U}$	error %
H-D	L-D	1.000	52.467	52.467	3.302e −6

Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy (0, r = 1/sqrt(2)) prior. The "U" in the Bayes factor denotes that it is uncorrected.

Post Hoc Comparisons - Gender

		Prior Odds	Posterior Odds	$BF_{10, U}$	error %
Female	Male	1.000	0.372	0.372	2.829e −6

Note. The posterior odds have been corrected for multiple testing by fixing to 0.5 the prior probability that the null hypothesis holds across all comparisons (Westfall, Johnson, & Utts, 1997). Individual comparisons are based on the default t-test with a Cauchy (0, r = 1/sqrt(2)) prior. The "U" in the Bayes factor denotes that it is uncorrected.

Pros and cons

Pros

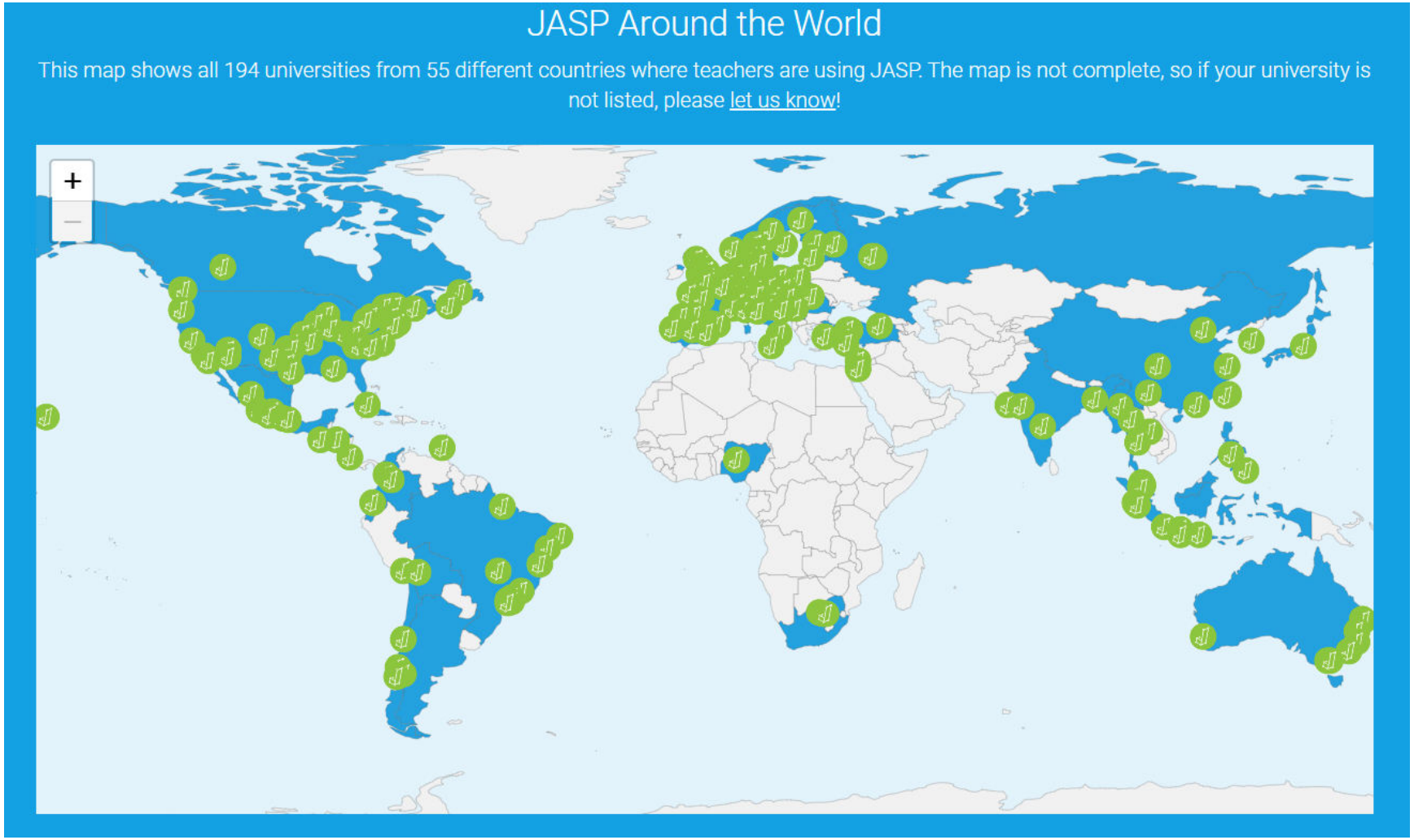
- Intuitive, Easy to use
- (Usually) FAST
- Worldwide used and accepted
- Replicable
- Gives a sort of «Standard»
- Constantly updated
- FREE and Open Science Inspired (e.g., connected to OSF)
- Inclusive: geared for both frequentist and Bayesian statistics
- Extensive: available from basic (e.g., t-test, regression, ANOVA) to advanced analytic techniques (e.g., machine learning, structural equation modeling, meta-analysis, network analysis)

Cons

- No direct (easy) access to the code
- Limited options in analysis
- Limited graphical display
- The immediate automatic update can slow the process
- Can crash if you add many layers to your analysis
- Graphical representations not modifiable
- Not so manageable data handling.

Should we use JASP?

- Yes we can.
- Seriously: **YES**!



Conclusions. Should we use JASP?

- Yes we can.
- Seriously: **YES**!
- But we **do not HAVE TO** use it.
- Sometimes is useful and time saving (standard parameters are fine, useful in simple experimental design). The use of standards save the researchers from inappropriate or hard to justify unusual parameters or solutions (if Uncertain… Go with the standards, at least they will be replicable)
- However this limit the flexibility, which is one of the main plus of Bayesian stats as well as «subjectivity» which is also a potential plus from certain perspectives.

Byes Factor is the underlying engine of JASP and can be used in r with (a little) more flexibility

BayesFactor's model notation: similar to aov's and lmer's

- tilde (~) means "predicted from"; separates DV (RT) from IVs
- asterisk (*) means "expand to all main effects and interactions"
- colon (:) indicates an interaction

Thus `RT ~ shape*color + ID` is the same as `RT ~ shape + color + shape:color + ID`.

```
anovaBF(RT ~ shape + color + shape:color + ID, whichRandom="ID", data=puzzles)
```

```
## Bayes factor analysis
## --------------
## [1] shape + ID                      : 2.815 ±0.85%
## [2] color + ID                      : 2.833 ±1.06%
## [3] shape + color + ID              : 11.66 ±1.62%
## [4] shape + color + shape:color + ID : 4.449 ±4.88%
##
## Against denominator:
##   RT ~ ID
## ---
## Bayes factor type: BFlinearModel, JZS
```

BRMS can fit complex hierarchical models with the flexibility over priors and family distributions

---

brm *Fit Bayesian Generalized (Non-)Linear Multivariate Multilevel Models*

```
## Not run:
# Poisson regression for the number of seizures in epileptic patients
# using normal priors for population-level effects
# and half-cauchy priors for standard deviations of group-level effects
prior1 <- prior(normal(0,10), class = b) +
  prior(cauchy(0,2), class = sd)
fit1 <- brm(count ~ zAge + zBase * Trt + (1|patient),
            data = epilepsy, family = poisson(), prior = prior1)

# generate a summary of the results
summary(fit1)
```

Authors:
Paul-Christian Bürkner [aut, cre],
Jonah Gabry [ctb],
Sebastian Weber [ctb],
Andrew Johnson [ctb],
Martin Modrak [ctb],
Hamada S. Badr [ctb],
Frank Weber [ctb],
Mattan S. Ben-Shachar [ctb]

Critiques of Bayesian statistics

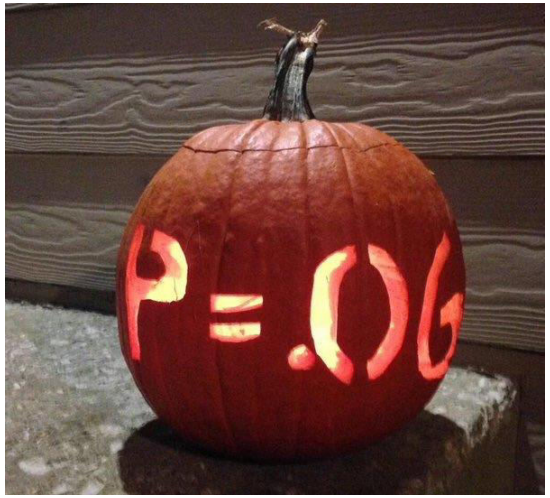
- Statistics should be objective, not subjective
- Inference is dependent on priors
- Beliefs as probability distributions?
- Gives the same inferences as frequentism anyway

Possible (basic) scenarios

- I colelcted 25 participants, run the analysis (NHST) and found a p-value of .07
  - **NHST** - I collect more participants, but watchout to inflate alpha
  - **Bayesian** – I collect more participants (does not affect alpha)

- I improve my design, run the power analysis a-priori, test the suggested 30 participants and found again p-value of .07
  - **NHST** - Potential (wrong) Answer: It's not significant, conditions are equivalent
  - **Bayesian** – Under conditions of uncertainity, adding information (more participants), improve the precision of esimates without inflating Type 1 error

- My 30 participants are associated with a p-value of .18
  - **NHST** - Potential (wrong) Answer: It's not significant, conditions are equivalent
  - **Bayesian** – Do I have evidnece in favour of H0?



Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications

Eric-Jan Wagenmakers[1] · Maarten Marsman[1] · Tahira Jamil[1] · Alexander Ly[1] · Josine Verhagen[1] · Jonathon Love[1] · Ravi Selker[1] · Quentin F. Gronau[1] · Martin Šmíra[2] · Sacha Epskamp[1] · Dora Matzke[1] · Jeffrey N. Rouder[3] · Richard D. Morey[4]

	Bayesian Inference	Classical Inference
Desiderata for Parameter Estimation		
1. To incorporate prior knowledge	✔	✗
2. To quantify confidence that θ lies in a specific interval	✔	✗
3. To condition on what is known (i.e., the data)	✔	✗
4. To be coherent (i.e., not internally inconsistent)	✔	✗
5. To extend naturally to complicated models	✔	✗
Desiderata for Hypothesis Testing		
1. To quantify evidence that the data provide for $\mathcal{H}_0$ vs. $\mathcal{H}_1$	✔	✗
2. To quantify evidence in favor of $\mathcal{H}_0$	✔	✗
3. To allow evidence to be monitored as data accumulate	✔	✗
4. To not depend on unknown or absent sampling plans	✔	✗
5. To not be "violently biased" against $\mathcal{H}_0$	✔	✗

PROS and CONS

NHST

- Rigid assumptions (!)
- Very diffuse (+)
- Easy to understand (+)
- Easy to misunderstand (-)
- Standards for reporting analysis and results (easier to test replication) (+)
- Directional testing (test one hypothesis) (-)

Bayesian

- Flexible (+)
- Powerful and sensitive (+)
- Drive solid conclusions either way (H0/H1) (+)
- Not always certain conclusions (more transparent, not always intuitive) (!)
- Complex, time demanding, calculations (-)

(!) = Pay attention
(+) = PROS
(-) = CONS

Go to www.menti.com and enter the following code 5883 0613

Resources

https://jasp-stats.org/download/ Download website

https://forum.cogsci.nl/index.php?p=/categories/jasp-bayesfactor Forum for requests and doubts

https://jasp-stats.org/jasp-materials/ Additional material from the JASP team

https://jasp-stats.org/r-package-list/ List of all the packages used by JASP

References

Basic reading:

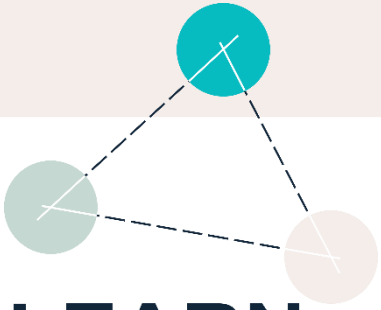
- Rouder, J. N., Morey, R. D., Speckman, P. L., & Province, J. M. (2012). Default Bayes factors for ANOVA designs. *Journal of mathematical psychology*, *56*(5), 356-374.
- Wagenmakers, E. J., Marsman, M., Jamil, T., Ly, A., Verhagen, J., Love, J., ... & Morey, R. D. (2018). Bayesian inference for psychology. Part I: Theoretical advantages and practical ramifications. *Psychonomic bulletin & review*, *25*(1), 35-57.
- Keysers, C., Gazzola, V., & Wagenmakers, E. J. (2020). Using Bayes factor hypothesis testing in neuroscience to establish evidence of absence. *Nature neuroscience*, *23*(7), 788-799.
- Van de Schoot, R., Kaplan, D., Denissen, J., Asendorpf, J. B., Neyer, F. J., & Van Aken, M. A. (2014). A gentle introduction to Bayesian analysis: Applications to developmental research. *Child development*, *85*(3), 842-860.

Additional Reading:

- Schönbrodt, F. D., & Wagenmakers, E. J. (2018). Bayes factor design analysis: Planning for compelling evidence. *Psychonomic bulletin & review*, *25*(1), 128-142.
- Stefan, A. M., Evans, N. J., & Wagenmakers, E. J. (2020). Practical challenges and methodological flexibility in prior elicitation. *Psychological Methods*.
- Scandola, M., & Romano, D. (2021). Bayesian multilevel single case models using 'Stan'. A new tool to study single cases in neuropsychology. *Neuropsychologia*, *156*, 107834.

daniele.romano@unimib.it

daniele.romano84 / danromano



General linear models

Giulio Costantini,

Università degli Studi di Milano-Bicocca

giulio.costantini@unimib.it

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

Linear Regression

Some examples in this section are taken from James et al.'s, 2013, An Introduction to Statistical Learning, with Applications in R, which is freely available online at

http://www-bcf.usc.edu/~gareth/ISL/

(a 2021 edition is also available there)

and from the wonderful Stanford online courses by Hastie and Tibshirani,

https://online.stanford.edu/courses/sohs-ystatslearning-statistical-learning

Loading Boston data

Package MASS includes a dataset called Boston. Load the package and the dataset in R.

Load also the following packages:

- dplyr
- lm.beta
- ez
- psych
- haven
- readr

Solution

```
if(!require("pacman")) install.packages("pacman")
p_load("MASS", "dplyr", "lm.beta", "ez", "psych",
"haven")

data(Boston)
```

Linear regression

`lm` is the function for fitting *linear models*, it takes as input a formula. If the variables in the fromula are taken from a dataset, it is necessary to indicate in the field «data» the dataset from which the variables should be taken.

```
fit1 <- lm(DV ~ X1+X2…, data=yourdata)
```

`summary` of the fitted object gives you coefficients, R^2 and p-values

```
summary(fit1)
```

The formula syntax

An expression of the form `y ~ model` is interpreted as a specification that the response `y` is modelled by a predictor specified symbolically by `model`

Y is predicted by x

```
Y ~ x
```

predictors are separated by «+»

```
Y ~ x1 + x2
```

«:» denotes interaction

```
Y ~ x1:x2
```

multiple regression with interaction

```
Y ~ x1 + x2 + x1:x2
```

Same as above, the «*» symbol indicates interaction plus all main effects. **This is useful for moderation analysis**

```
Y ~ x1*x2
```

The I() operator specifies that what is inside the parentheses should be interpreted as an arithmetic transformation of the variables, and not as formula syntax. The following code specified that there is only one predictor, which is a variable given by the sum of x1 and x2.

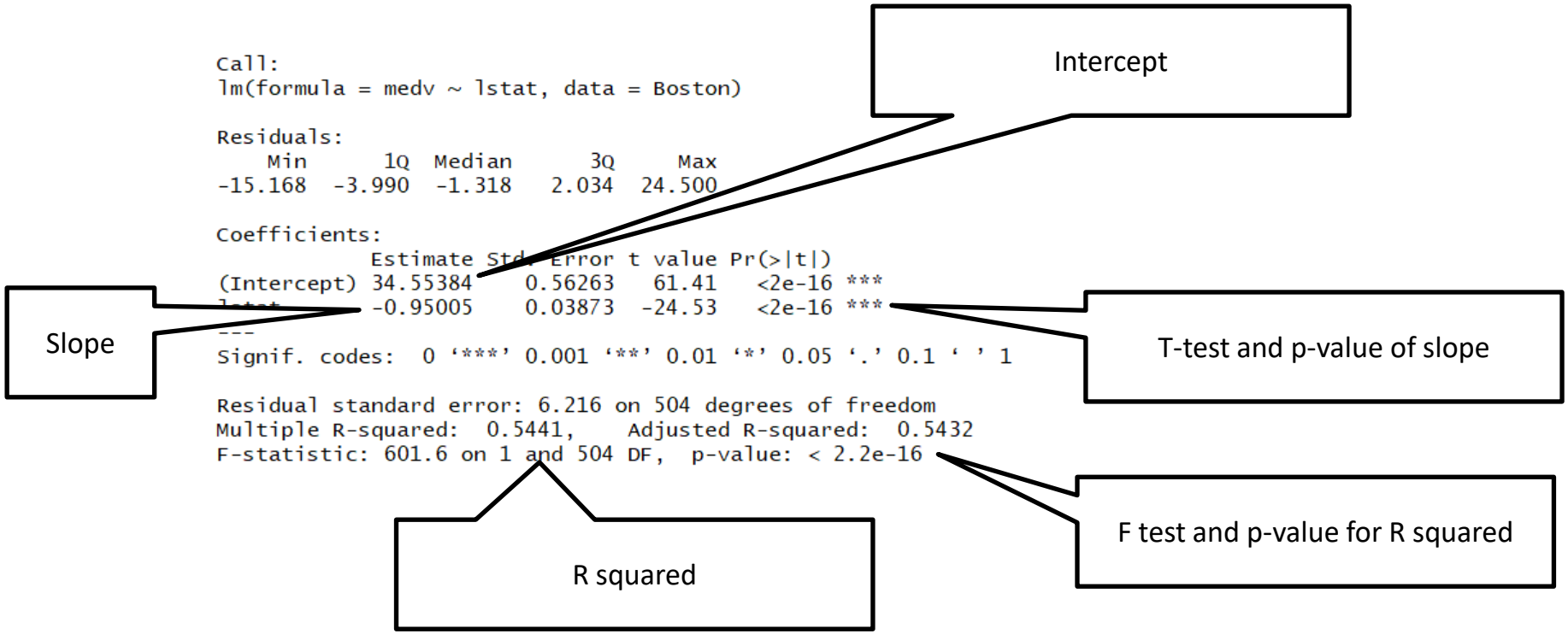
```
y ~ I(x1+x2)
```

Linear regression - example

Let’s regress medv = median household value on lstat = % lower status population.

```
fit1 <- lm(medv ~ lstat, data = Boston)
summary(fit1)
```

Reading the outptut



Getting betas

Function lm.beta allows also getting standardized regression coefficients.

```
fit1_beta <- lm.beta(fit1)
summary(fit1_beta)
```

```
Coefficients:
             Estimate Standardized Std. Error t value Pr(>|t|)
(Intercept) 34.55384      0.00000    0.56263   61.41   <2e-16 ***
lstat       -0.95005     -0.73766    0.03873  -24.53   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 6.216 on 504 degrees of freedom
Multiple R-squared:  0.5441,    Adjusted R-squared:  0.5432
F-statistic: 601.6 on 1 and 504 DF,  p-value: < 2.2e-16
```

Now you

1. Perform a simple linear regression predicting `medv` from `crim`, the per capita crime rate by town. Inspect the b coefficient, the p-value and the $R^{2.}$

2. Inspect also the beta coefficients

3. Visualize the variables medv, crim, and lstat in a scatterplot with regression line

Solution

```
fit2 <- lm(medv ~ crim, data = Boston)

fit2_beta <- lm.beta(fit2)
summary(fit2_beta)

select(Boston, medv, crim, lstat) %>% pairs.panels()
```

Multiple linear regression

You can just add predictors separated by the “+” symbol

Predict medv from crim, nox, and rm

Solution

```
# an alternative code to do the same as above
fit3 <- lm(medv ~ crim + nox + rm, data = Boston)
%>% lm.beta

summary(fit3)
```

ANOVA

ANOVA using package ez

- There are several ways to perform ANOVAs in R.
- The R package ez includes several convenient functions that make between-ss, within-ss, and mixed ANOVA very easy.
- By default, ezANOVA uses type-2 sum of squares, whereas, e.g., SPSS use type 3. This can sometimes lead to different results.
  For more detail about types of sum of squares in ANOVA, see e.g., https://mcfromnz.wordpress.com/2011/03/02/anova-type-iiiiii-ss-explained/
- Furthermore, ezANOVA treats covariates in a specific way: Covariates are partialled out from the dependent variable before the analyses. This can also lead to differences

```
ezANOVA(
data,# the dataset
dv, # dependent variable
wid, # Subject identifier
within, # Within subject predictors
within_full, # ALL within subject variables (also levels not used for the analysis!)
within_covariates, # within-subject covariates
between, # between-subject predictors
between_covariates, # between-subject covariates
type = 2, # type of sum of squares. Type 3 is the same as the SPSS and Jamovi default
detailed, # if TRUE, returns more information
return_aov # if TRUE, returns "classic R" output
)
```

The independent variable(s) in ez can be specified as a c() list, including the names of the vairables separated by a comma

function **ezPlot** for getting easy anova plots

```
ezPlot(
# the following variables are just like ezANOVA
data, dv, wid, within, within_full, within_covariates, between,
between_covariates, type,

x, # variable on the x axis?
split, # v. that define different lines
row, # v. that splits the plot into different plots next to each other.
col) # v. that splits the plot into different plots one above another?
```

Open the data

- Open the “Canova.sav” dataset
- Use the function as_factor on Canova, to tell R that nominal variables should be imported as factors, instead of as “haven labelled”, as format that is specific to package haven and that unfortunately does not work well with package ez

Solution

```
Canova <- read_sav("data/Canova.sav")

Canova <- as_factor(Canova)
```

Between-ss ANOVA

1. One-way Anova: Examine whether education affects verbal argumentation score (argum).
2. Generate a plot with ezPlot

Tip. For generating the plot, copy paste ezANOVA code and put education on the x axis.

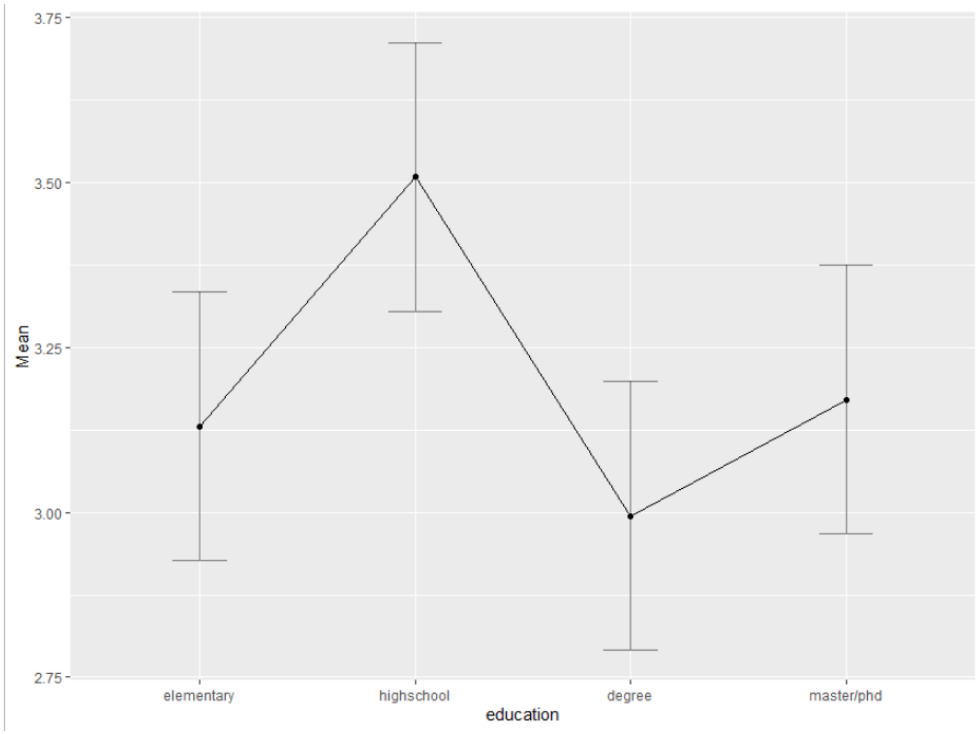
Solution

```
ezANOVA(data = Canova,
        dv = argum,
        wid = id,
        between = education)
```

```
$ANOVA
     Effect DFn DFd        F          p p<.05        ges
1 education   3  96 2.974093 0.03549958     * 0.08503704

$`Levene's Test for Homogeneity of Variance`
  DFn DFd       SSn      SSd         F         p p<.05
1   3  96 0.6532879 24.70736 0.8461129 0.4719885
```

```
ezPlot(data = Canova,
       dv = argum,
       wid = id,
       between = education,
       x = education)
```



How to get estimated marginal means?

Here is a trick! Just save the output of ezPlot to an object and inspect the «data» content

```
mns <- ezPlot(data = Canova,
              dv = argum,
              wid = id,
              between = education,
              x = education)
mns$data
```

	education	N	Mean	SD	FLSD	lo	hi
1	elementary	21	3.130476	0.7647449	0.4068334	2.927059	3.333893
2	highschool	41	3.508780	0.6506888	0.4068334	3.305364	3.712197
3	degree	24	2.995000	0.7747033	0.4068334	2.791583	3.198417
4	master/phd	14	3.171429	0.7830793	0.4068334	2.968012	3.374845

Effect size: no partial eta squared?

- ezANOVA reports the **generalized eta squared**
- This effect size measure has been developed as a generalization of partial eta squared that is invariant across designs (e.g., if the effect of a factor is estimated in a within-subject vs. between-subject design, if it is observed vs. manipulated).
- If all factors are manipulated between-subjects, it is equal to partial eta squared. Otherwise, you should specify the argument “observed”.
- For details, see *Bakeman, R. (2005). Recommended effect size statistics for repeated measures designs. Behavior Research Methods, 37 (3), 379-384.*

How to get post-hoc tests?

You need to specify return_aov = TRUE in ezANOVA

```
fit <- ezANOVA(data = Canova,
               dv = argum,
               wid = id,
               between = education,
               return_aov = TRUE)
```

We will just see Tukey HSD post-hoc tests. As argument “which”, you can specify which factor you want to examine (the default is to examine all factors)

```
TukeyHSD(fit$aov, which = "education")
```

How to get post-hoc tests?

```
TukeyHSD(fit$aov)
```

```
Fit: aov(formula = formula(aov_formula), data = data)

$education
                           diff        lwr        upr     p adj
highschool-elementary  0.37830430 -0.1301070  0.8867156 0.2162155
degree-elementary     -0.13547619 -0.7016007  0.4306483 0.9235844
master/phd-elementary  0.04095238 -0.6127519  0.6946566 0.9984257
degree-highschool     -0.51378049 -1.0007257 -0.0268353 0.0344942
master/phd-highschool -0.33735192 -0.9238227  0.2491188 0.4391356
master/phd-degree      0.17642857 -0.4607235  0.8135807 0.8872911
```

Factorial ANOVA

- In ezANOVA, just add more variables to the «between» argument.
- In ezPlot, you may want to specify how to represent additional predictors (split, row, col)

For instance, **include also the variable gender as a predictor in the previous analysis**

Solution

```
fit <- ezANOVA(data = Canova,
               dv = argum,
               wid = id,
               between = c(education, gender),
               return_aov = TRUE)
fit

TukeyHSD(fit$aov)

mns <- ezPlot(data = Canova,
               dv = argum,
               wid = id,
               between = c(education, gender),
              x = education,
              split = gender)
mns
```

ANCOVA

- Also control for the linear effect of participants' age (birth)

Tip. Consider argument between_covariates

Solution

```
# ANCOVA
fit <- ezANOVA(data = Canova,
          dv = argum,
          wid = id,
          between = c(education, gender),
          between_covariates = birth,
          return_aov = TRUE)
fit

TukeyHSD(fit$aov)

mns <- ezPlot(data = Canova,
          dv = argum,
          wid = id,
          between = c(education, gender),
          between_covariates = birth,
          x = education,
          split = gender)
mns
```

Within-ss and Mixed ANOVA

- Data in long-format are expected: You are now able to melt them if necessary!
- You just need to specify the follwing additional arguments in ezANOVA:

- within: within-subject variables
- within_full: in case of unbalanced data, within-subject variables that you do not want to analyze, but you want to consider when computing cell means.

In Mixed ANOVA, you just specify both between and within-subject variables.

Note. We won't see post-hoc for repeated-measures anova in R, as they are quite complex to get. In general, it is better and easier to use mixed models instead of within-ss anova.

Open the ANT dataset

You can just import it from package ez

```
data(ANT)
```

Within-ss ANOVA exercise

1. Predict the RTs in the ANT data from flank, cue (and their interaction, which is included by default).
2. Get the corresponding plot and marginal means

Solution

```
fit <- ezANOVA(ANT,
               dv = rt,
               wid = subnum,
               within = c(flank, cue))
fit

mns <- ezPlot(ANT,
              dv = rt,
              wid = subnum,
              within = c(flank, cue),
              x = cue,
              split = flank)
mns
mns$data
```

Mixed ANOVA

Include also group as a between-subject predictor.

Tip. In ezPlot, you can represent groups on separate rows, using the argument «row»

Solution

```
fit <- ezANOVA(ANT,
               dv = rt,
               wid = subnum,
               between = group,
               within = c(flank, cue))
fit
mns <- ezPlot(ANT,
              dv = rt,
              wid = subnum,
              within = c(flank, cue),
              between = group,
              x = cue,
              split = flank,
              row = group)
mns
mns$data
```

A recap exercise (at home)

1. Load packages readr, ez, reshape2, dplyr
2. Load the BPD data
3. Convert the SelfHurtPRE,…, SelfHurt3Months to long format (name the variable as «time» and the values as «selfharm»), retaining the following information regarding each subect: ID, Sex, Age, and Therapy
4. Perform a mixed ANOVA, predicting self harm from
    1. Therapy (between-subject)
    2. Time (within-subject)
    3. Age (covariate)
5. Obtain a plot, in which time is on the x axis and therapy defines different lines.

Solution (1)

```
if(!require("pacman")) install.packages("pacman")
p_load("readr", "ez", "reshape2", "dplyr")

BPD <- read_csv("data/BPD.csv")

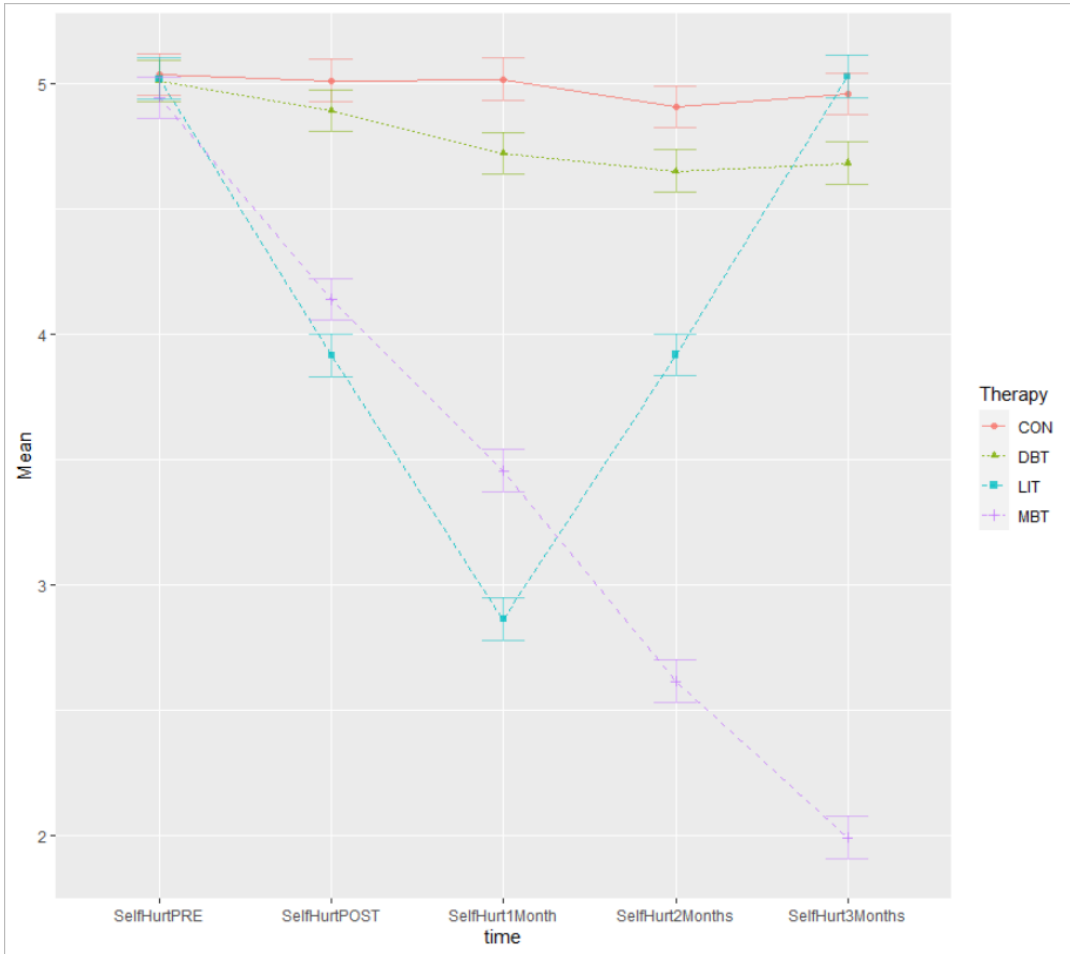
BPD_long <- melt(BPD, id.vars = c("ID", "Sex", "Age", "Therapy"),
                 measure.vars = c("SelfHurtPRE", "SelfHurtPOST",
                                  "SelfHurt1Month", "SelfHurt2Months",
                                  "SelfHurt3Months"),
                 variable.name = "time",
                 value.name = "selfharm")
```

Solution (2)

```
fit <- ezANOVA(BPD_long,
        wid = ID,
        dv = selfharm,
        within = time,
        between = Therapy,
        between_covariates = Age)
fit

mns <- ezPlot(BPD_long,
               wid = ID,
               dv = selfharm,
               within = time,
               between = Therapy,
               between_covariates = Age,
              x = time,
              split = Therapy)
mns
```

Solution (3)


Mean
5
4
3
2
SelfHurtPRE
SelfHurtPOST
SelfHurt1Month
SelfHurt2Months
SelfHurt3Months
time
Therapy
CON
DBT
LIT
MBT

General linear model (Part 2)

Giulio Costantini -
giulio.costantini@unimib.it

Post-hoc for repeated measures ANOVA

Post-hoc for repeated measures ANOVA

First, we need to collapse data, taking the mean of the DV across conditions of interest. Assuming that the data are balanced, that's very easy in dplyr

```
ANT_collapsed <- group_by(ANT, subnum, flank, cue) %>%
  summarize(rt = mean(rt))
```

You need to do ANOVA with aov

I did not show that to you explicitly, but that's how ws ANOVA can be specified in vanilla R (without ez)

```
fit2 <- aov(rt~flank*cue +
Error(subnum/(flank*cue)),
      data = ANT_collapsed)
```

Use emmeans to get post-hoc

```
p_load("emmeans")

emmeans(fit2, specs = ~ flank) %>%
pairs(adjust = "tukey")

emmeans(fit2, specs = ~ cue) %>%
pairs(adjust = "tukey")

emmeans(fit2, specs = ~ cue:flank)
%>% pairs(adjust = "tukey")
```

GLM assumptions

GLM assumptions

Let’s do a multiple linear regression

Using the dataset Boston in package MASS, let’s predict medv from all variables in the dataset

The “dot” in the syntax formula means “add all other variables in the dataset as predictors”

Solution

```
library(MASS)
data(Boston)
fit1 <- lm(medv~., data = Boston)
summary(fit1)
```

How do we know if the assumptions for running a linear model are fulfilled?

Assumptions of GLM

- *Linearity*: If the relationships are not linear, the linear model may not be the best way to test them.
- *Homoscedasticity*: the variance of the residuals is the same for all predicted values.
- Normality: *Residuals are normally distributed*.

Other assumptions:

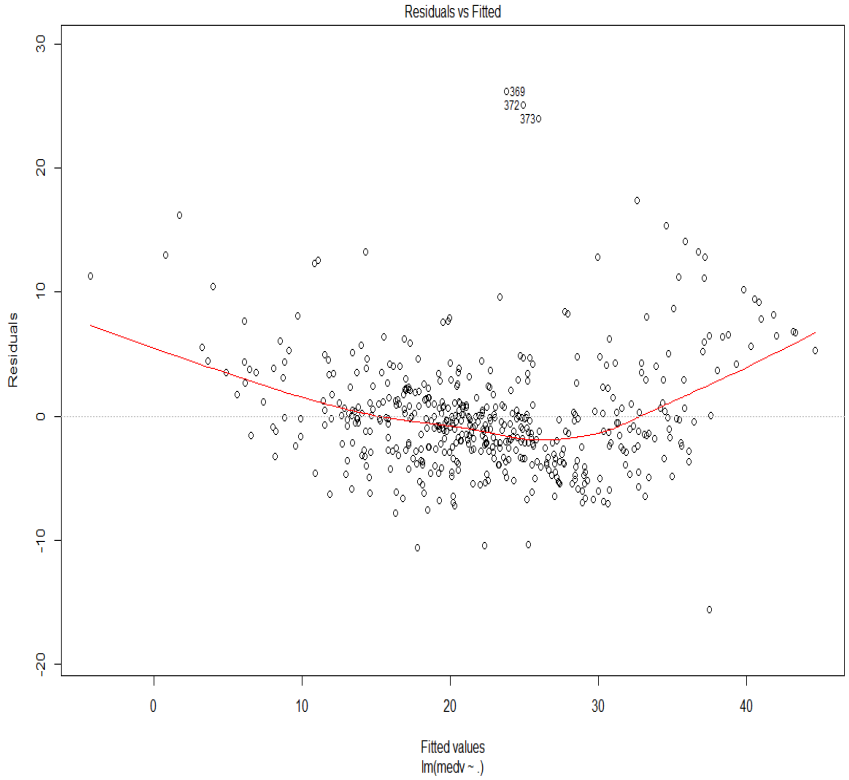
- *Outliers*: the results should not be due to the presence of outliers, i.e., observations that come from a population different from the one of interest.
- *Independence of observations*: This should be known by design (e.g., repeated measures), and can be solved by fitting a Mixed Model instead of a GLM

Checking assumptions in R (1)

`plot(fit1)` returns 4 plots by default.

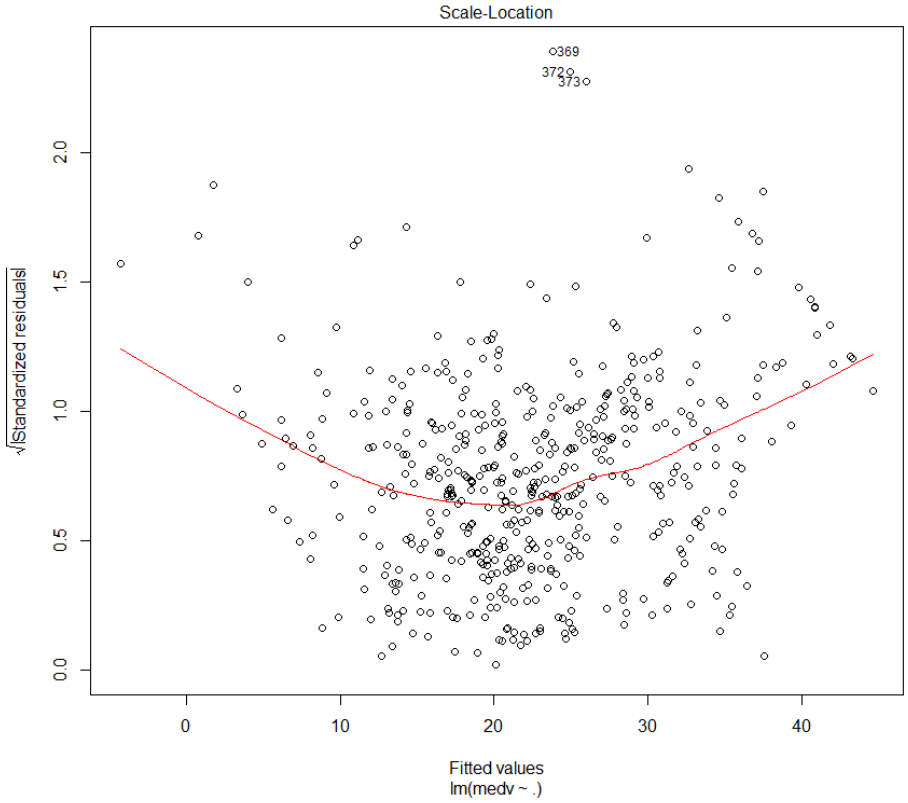
(1) “residuals vs. fitted”: gives the fitted values on the x-axis and the residuals on the y-axis. This allows checking for

- Homoschedasticity: If the variance of the residuals is the same at the different levels of the fitted values (i.e., the dots form a homogeneous band), the assumption is met.
- Linearity: if you see a clear pattern, it may be due to nonlinear relationships.



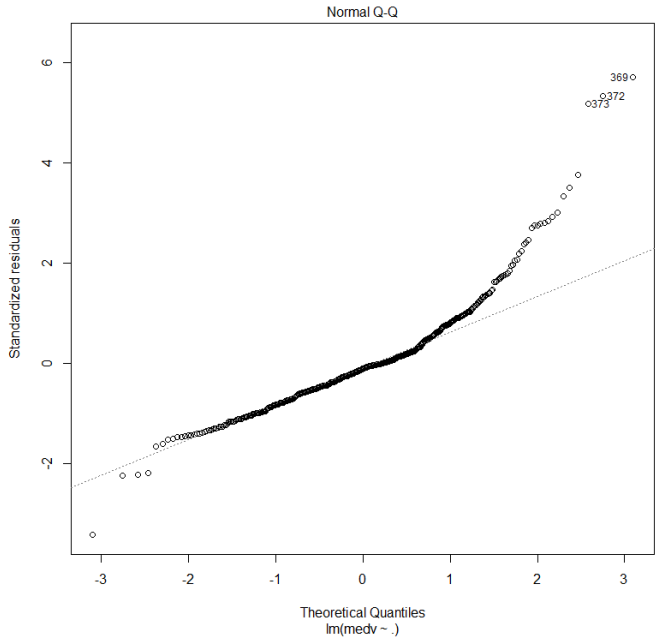
Checking assumptions (2)

(2) Scale-location: Same as the first plot, but the residuals on the y-axis are now standardized (z-scores).



Checking assumptions (3)

(3) “Normal q-q”. QQ means quantile-quantile. A point (*x*, *y*) on the plot corresponds to one of the quantiles of the second distribution (*y*-coordinate) plotted against the same quantile of the first distribution (*x*-coordinate). The normal qq plot compares the quantiles that would be expected if a variable was normally distributed, against the empirically observed values. **If the residuals are normally distributed, the points lay on a straight line.**



Checking assumptions (4)

You can also check normality of residuals using the Shapiro-Wilk test. A significant results reveals significant deviations from normality

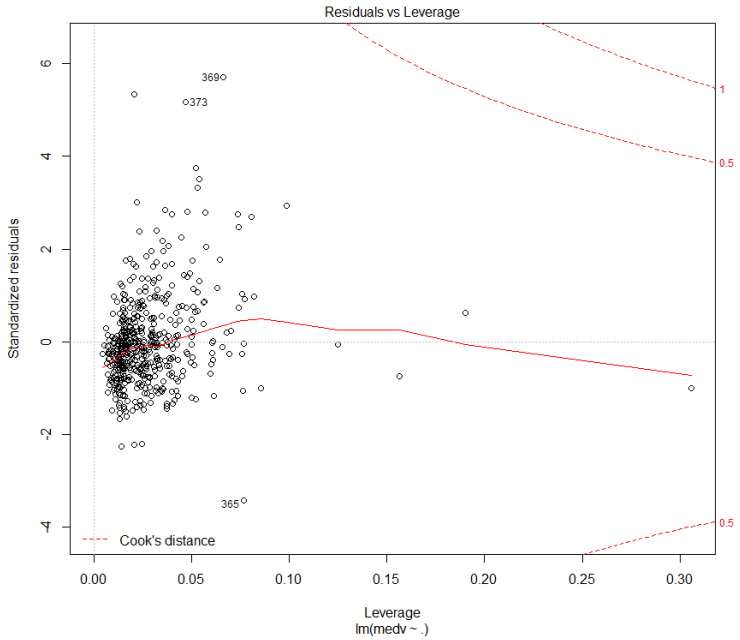
```
res <- residuals(fit1) # save residuals
shapiro.test(res) # run test
```

Checking assumptions (5)

(4) “Residuals vs. leverage”: it allows checking for outliers.

- x-axis = leverage, a measure of how much each data point influences the regression.
- y-axis = residuals.
- Contour values = Cook’s distance, a measure of how much the regression would change if a point was deleted.

If the red smoothed line stays close to the horizontal gray dashed line and if no point has a large Cook’s distance (i.e., > 1), you do not have outliers.



Checking assumptions for between-subject ANOVA

Simply set the return_aov = TRUE option in ezANOVA and apply what we have seen to that fitted object ($aov).

Moderation / interaction

- Interaction effects are defined by the **multiplicative** effect of predictors. The equation is

$$y = b_o + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2$$

- The dependent variable y must be continuous. Predictors can be continuous or dichotomous.
- Interactions must be always estimated together with main **effects.** In R formula syntax, this is achieved with any of the following
  Y ~ X1*X2
  Y ~ X1 + X2 + X1:X2
- Centered (or standardized) predictors to reduce multicollinearity between the interaction term and main effects

Package pequod: a simpler way to do the same thing

This package has a few useful function for testing and visualizing interaction/moderation

- **lmres** is like *lm*, with an embedded centering option
- **simpleSlope** does a simple slope analysis, i.e., computes the regression for M+1SD and M-1SD of moderator (age).
- **PlotSlope** visualizes the simple slope plot

Interaction with pequod

1. Load packages "readr", "dplyr", "car", "pequod", "MASS", "psych"
2. Open the dataset *Boston* in package *MASS*
3. Let’s predict the median value of the houses (medv) from the lstat and age. Use the standardized varibles.

Same model with package pequod

```
if(!require("pacman")) install.packages("pacman")
require("pacman")
p_load("readr", "dplyr", "car", "pequod", "MASS")

fit <- lmres(medv ~ lstat * age,
             centered = c("lstat", "age"),
             data = Boston)


summary(fit)
```

Output

```
Models
          R      R^2   Adj. R^2      F      df1  df2  p.value
Model   0.745   0.556      0.553 209.312   3.000  502  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-15.806  -4.045  -1.333   0.000   2.085  27.552

Coefficients
             Estimate    StdErr   t.value    beta p.value
(Intercept)  22.03061   0.35325  62.36587         < 2e-16 ***
lstat        -1.10712   0.05850 -18.92517 -0.8596 < 2e-16 ***
age           0.05186   0.01442   3.59762  0.1587 0.00035 ***
lstat.XX.age  0.00416   0.00185   2.24428  0.0833 0.02525 *
---
```

Proportion of explained variance (R^2), F test and p-value

Output

```
Models
          R      R^2   Adj. R^2      F      df1  df2  p.value
Model   0.745   0.556      0.553 209.312   3.000  502  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residuals
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-15.806  -4.045  -1.333   0.000   2.085  27.552

Coefficients
              Estimate    StdErr   t.value    beta p.value
(Intercept)   22.03061   0.35325  62.36587         < 2e-16 ***
lstat         -1.10712   0.05850 -18.92517 -0.8596 < 2e-16 ***
age            0.05186   0.01442   3.59762  0.1587 0.00035 ***
lstat.XX.age   0.00416   0.00185   2.24428  0.0833 0.02525 *
```

Multiple regression coefficients with corresponding p-values, including also the interaction term given by the product of lstat x age, which is significant.

Standardized coefficients are reported in column «beta». However, the value for the product is obtained by standardizing the interaction term after it is computed, which is slightly different than manually standardizing the two variables before computing the product.

Simple slope analysis

To further explore the results, we can run a simple-slope analysis, i.e., compute the regression for values of the moderator +1SD and -1SD.

To do this, we have to choose which between lstat and age is the moderator. We chose lstat as predictor and age as moderator.

```
sslope <- simpleSlope(fit,
                      pred = "lstat_sc",
                      mod1 = "age_sc")

sslope
```

The fitted object «fit» computed above is given then as input of function simpleSlope.

Simple slope analysis

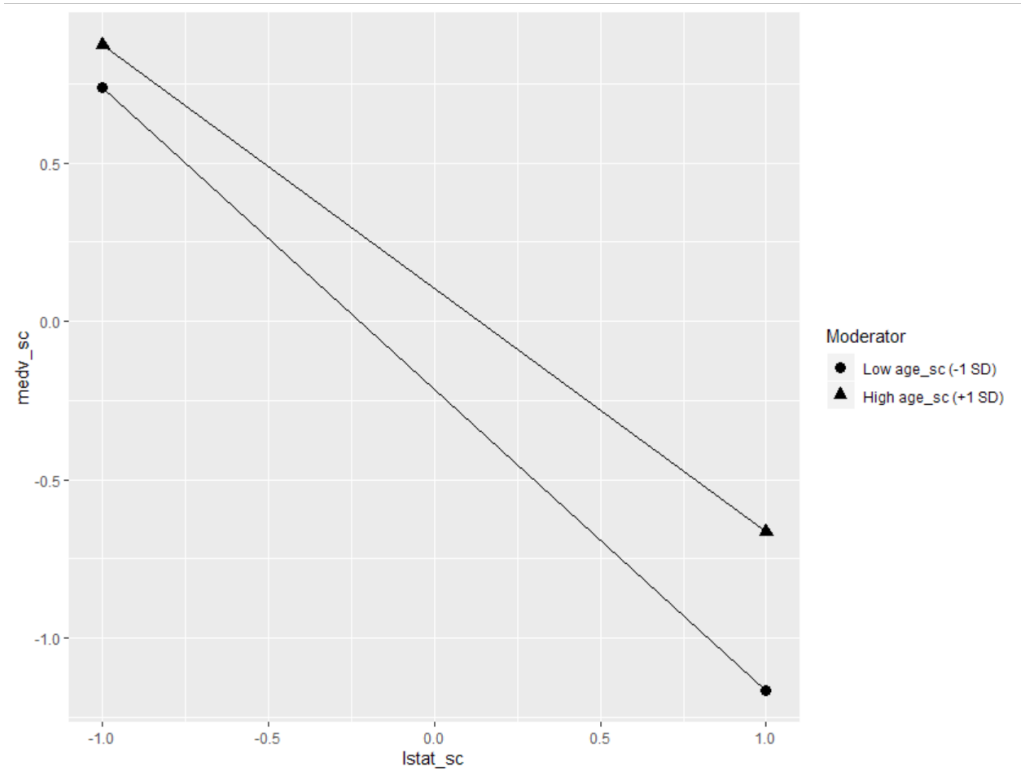
```
Simple Slope:
                    simple slope standard error   t-value p.value
Low age_sc (-1 SD)    -0.9504548     0.07617810 -12.47675       0
High age_sc (+1 SD)   -0.7687891     0.03999213 -19.22351       0
```

lstat has a significant negative effect on *medv* for high and low values of *age*, but the effect is stronger for lower values of age (i.e., newer neighborhoods)

Visualize Simple slope analysis

```
PlotSlope(sslope)
```

The output of simpleSlope can be given as input to PlotSlope, which draws the simple slope plot



Mediation in R

Load packages and read data

Function **mediate** in package psych implements mediation (there are many other and more sophisticated alternatives, but this one is easy).

The formula syntax is similar to regression, but you need to specify **mediators in parentheses.**

```
fit <- mediate(medv ~ lstat + (age), data = Boston)
summary(fit)
```

Output

```
Direct effect estimates (traditional regression)    (c') X + M on Y
          medv   se      t  df      Prob
Intercept 33.22 0.73  45.46 503 2.94e-180
lstat     -1.03 0.05 -21.42 503  8.42e-73
age        0.03 0.01   2.83 503  4.91e-03

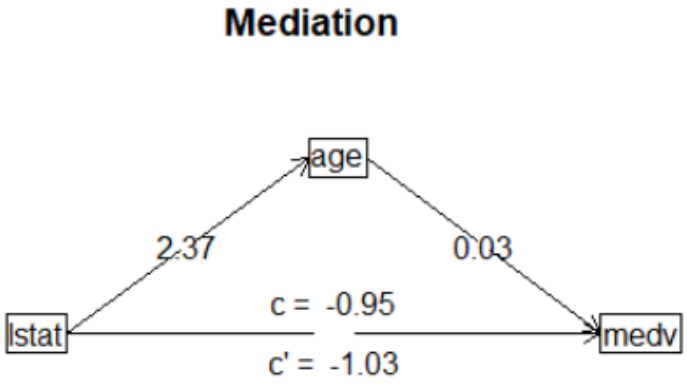
R = 0.74 R2 = 0.55   F = 308.97 on 2 and 503 DF   p-value:  2.98e-88

 Total effect estimates (c) (X on Y)
          medv   se      t  df      Prob
Intercept 34.55 0.56  61.42 504 3.74e-236
lstat     -0.95 0.04 -24.53 504  5.08e-88

 'a'  effect estimates (X on M)
           age   se     t  df     Prob
Intercept 38.53 2.04 18.93 504 9.48e-61
lstat      2.37 0.14 16.94 504 2.78e-51

 'b'  effect estimates (M on Y controlling for X)
    medv   se    t  df    Prob
age 0.03 0.01 2.83 503 0.00491

 'ab'  effect estimates (through mediators)
      medv boot   sd lower upper
lstat 0.08 0.08 0.04  0.01  0.16
```



Values of bootstrspped confidence intervals can be slightly different in R and Jamovi, because they involve random number generation.
The lower and upper limits of the confidence interval for the mediated effect are obtained with boostrap

And if you want beta coefficients?

```
fit <- mediate(medv ~ lstat + (age),
data = Boston, std = TRUE)
```

```
summary(fit)
```

You can also get parallel mediators, by simply adding additional mediators to the formula

MAN(C)OVA in R

Albeit there is a manova() function in R, it only implements Type I sum of squares, which are sequential (your p-values depend on the order in which you entered the predictors!)

A better alternative is function **Manova()** in package **car**

Data are expected in wide format.

Manova in car

1. Fit a multivariate linear model using lm. This includes many DVs.
   ```
   fit <- lm(cbind(DV1, DV2, …) ~ IVs, data = yourdata)
   ```
2. Use function Manova around the fitted object, specifying the desired Type of SS
   ```
   out <- Manova(fit)
   ```
3. Ask for a summary of the fitted object. Option univariate = TRUE gives you also the ANOVAs for each DV
   ```
   summary(out, univariate = TRUE)
   ```

Unlike ez, covariates are treated as continuos predictors.

Exercise

1. Load the car, dplyr, and reard packages
2. Import BPD data
3. Run a multivariate lm regressing the three SCL scales on Sex
4. Get a summary of results

Solution

```
if(!require("pacman")) install.packages("pacman")
p_load("readr", "dplyr", "car")

BPD <- read_csv("data/BPD.csv")

fit <- lm(cbind(SCL_SOM, SCL_PSY, SCL_PHOB) ~ Sex,
data = BPD)
out <- Manova(fit)
summary(out, univariate = TRUE)
```

Adding covariates

Add BPDCL, a continuous predictor, by simply including it as a predictor in the regression

Solution

```
fit <- lm(cbind(SCL_SOM, SCL_PSY, SCL_PHOB) ~ Sex +
BPDCL, data = BPD)
out <- Manova(fit)
summary(out, univariate = TRUE)
```

Using the ez approach to covariates

You can partial out covariates from continuous DVs, by regressing the DVs on the covariates and taking the residuals. The residuals can then be used as DVs in the MANOVA

```
fit1 <- lm(cbind(SCL_SOM, SCL_PSY, SCL_PHOB) ~ BPDCL, data = BPD)

fit2 <- lm(residuals(fit1) ~ Sex, data = BPD )
out <- Manova(fit2)
summary(out, univariate = TRUE)
```

Now you

Perform a MANOVA inspecting whether Sex and BPDCL have an effect on personality (variables Quest_extraversion, Quest_conscientiousness, Quest_openness)

Solution

```
fit <- lm(cbind(Quest_extraversion,
Quest_conscientiousness, Quest_openness) ~
Sex + BPDCL, data = BPD)

out <- Manova(fit)

summary(out, univariate = TRUE)
```

Mixed-model analyses

Experiments with more than one random factor (part I)

Marine Rougier,
Ghent University, Belgium
Marine.Rougier@ugent.be

Jamie Cummins,
Ghent University, Belgium
Jamie.Cummins@ugent.be

Slides based on great mixed-model courses:
Charles M. Judd (University of Colorado, Boulder, USA) Dominique Muller (University Grenoble Alpes, Grenoble, France)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

Important papers

Baayen, R. H. (2008). *Analyzing linguistic data. APractical introduction to statistics using R*. Cambridge University Press.

Baayen, R.H., Davidson, D.J., & Bates, D.M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* 59, 390-412.

Bates, D.M. (in progress, 2010). lme4: Mixed-effects modeling with R

Brauer, M., & Curtin, J.J.(2018). Linear mixed-effects models and the analysis of nonindependent data:A unified framework to analyze categorical and continuous independent variables that vary within- subjects and/or within-items. *Psychological Methods*, 23, 389–411.

Judd, C. M.,Westfall, J., & Kenny, D.A. (2012).Treating stimuli as a random factor in social psychology:A new and comprehensive solution to a pervasive but largely ignored problem. *Journal of Personality and Social Psychology*, *103*, 54-69.

Judd, C. M.,Westfall, J., & Kenny, D.A. (2016). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology, 68,* 601-625.

Wonnacott,T.H. & Wonnacott, R.J. (1990). *Introductory statistics for business and economics* (4th Edition).

Plan

- Introduction: Why we need mixed-models
- Two random factors, one fixed effect
- Two random factors, two fixed effects
- Nested random factors: The school classes example
- Power estimations
- Good practices considerations

Plan

- Introduction: Why we need mixed-models
- Two random factors, one fixed effect
- Two random factors, two fixed effects
- Nested random factors: The school classes example
- Power estimations
- Good practices considerations

Typical experiment in social psychology
(Judd et al., 2012)

- A researcher shows 15 faces of White people and 15 faces of Black people
- For each face, each participant emit a judgment (e.g., how trustworthy is this person?) on a scale from -20 to +20

	Stim 1	Stim 2	Stim 3	Stim 4	Stim 5	Stim 6	Stim 7	Stim 8	Stim 9	Stim 10	Stim 11	Stim 12	…
Pp 1	5	4	6	7	3	8	8	7	9	5	6	5	
Pp 2	4	4	7	8	4	6	9	6	7	4	5	6	
Pp 3	5	3	6	7	4	5	7	5	8	3	4	5	

…

- We typically consider participants as the unit of observation (multiple responses under different conditions)
- Classic analysis of data: ANOVA

‘By-participants’ analysis

- **‘By-participants’ analysis** (standard ANOVA), that is, using participants as the unit of analysis:

  We will test whether the Cond difference (Black vs. White faces) within Participants is significant.

  → *participants = ‘random’ factor/effect*

	Stim 1	Stim 2	Stim 3	Stim 4	Stim 5	Stim 6	Stim 7	Stim 8	Stim 9	Stim 10	Stim 11	Stim 12	…
Pp 1	5	4	6	7	3	8	8	7	9	5	6	5	
Pp 2	4	4	7	8	4	6	9	6	7	4	5	6	
Pp 3	5	3	6	7	4	5	7	5	8	3	4	5	

…

‘By-participant’ analysis

- We recondition the long format data frame (each row = one response) by computing the two mean values per condition and per participant.

```
DF_pp <- cast(DF, pp ~ Cond, value = "y", mean, fill = NA)
```

- Then, we calculate a score (called here ‘W1’) that codes for the difference between the two conditions (to control for the non independence of the residuals)

```
DF_pp$W1 <- DF_pp$White - DF_pp$Black
```

	pp	stim	Cond	y
1	1	1	Black	-2.30
2	1	2	Black	-0.56
3	1	3	Black	-4.47
4	1	4	Black	-2.11
5	1	5	Black	2.55
6	1	6	Black	-1.08
7	1	7	Black	-3.65
8	1	8	Black	2.15
9	1	9	Black	-3.17
10	1	10	Black	-2.90
11	1	11	Black	-5.43
12	1	12	Black	-0.32
13	1	13	Black	-9.15
14	1	14	Black	-4.51
15	1	15	Black	-0.46
16	1	16	White	9.27
17	1	17	White	8.77
18	1	18	White	0.77

pp	Black	White
1	-2.3606667	1.75200000
2	-4.0513333	-0.05733333
3	-2.6113333	0.53600000
4	-3.6866667	2.63466667
5	-0.5306667	2.39400000
6	1.8766667	5.60600000
7	-0.3260000	4.68533333
8	3.5906667	2.44666667
9	-0.1346667	1.86600000
10	-0.8106667	0.00200000
11	-6.1480000	-7.34200000
12	-1.2440000	0.31533333
13	-3.1366667	1.89333333
14	1.6660000	1.12466667
15	-1.8773333	0.24866667
16	-4.4960000	1.26666667
17	-3.7793333	0.79200000
18	-3.7120000	-0.41600000

pp	Black	White	W1
1	-2.3606667	1.75200000	4.11266667
2	-4.0513333	-0.05733333	3.99400000
3	-2.6113333	0.53600000	3.14733333
4	-3.6866667	2.63466667	6.32133333
5	-0.5306667	2.39400000	2.92466667
6	1.8766667	5.60600000	3.72933333
7	-0.3260000	4.68533333	5.01133333
8	3.5906667	2.44666667	-1.14400000
9	-0.1346667	1.86600000	2.00066667
10	-0.8106667	0.00200000	0.81266667
11	-6.1480000	-7.34200000	-1.19400000
12	-1.2440000	0.31533333	1.55933333
13	-3.1366667	1.89333333	5.03000000
14	1.6660000	1.12466667	-0.54133333
15	-1.8773333	0.24866667	2.12600000
16	-4.4960000	1.26666667	5.76266667
17	-3.7793333	0.79200000	4.57133333
18	-3.7120000	-0.41600000	3.29600000

‘By-participant’ analysis

- Finally, we can perform the regression corresponding to the following model: $W_{1i} = \alpha_0 + \varepsilon_i$ (corresponds to a within-participants ANOVA).
- To perform this regression in R, we use:

```
fit.lm <- lm(W1~1, DF_pp)
```
=> 1 indicates that we need to estimate the intercept

```
summary(fit.lm)
```

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.5211     0.4567   5.521 5.98e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

‘By-participant’ analysis

- We managed to solve the stat issue of the non independence of residuals (with W1)
- The mean difference is likely to generalize on other participants
- However, by computing the average value of each cond per participant we lost a critical information : the variance between targets

	Stim 1	Stim 2	Stim 3	Stim 4	Stim 5	Stim 6	Stim 7	Stim 8	Stim 9	Stim 10	Stim 11	Stim 12
Pp 1	5	4	6	7	3	8	8	7	9	5	6	5
Pp 2	4	4	7	8	4	6	9	6	7	4	5	6
Pp 3	5	3	6	7	4	5	7	5	8	3	4	5

- If we ONLY aim at generalizing the results to other individuals, ANOVA is completely adequate. But this is rarely what we do —> we implicitly infer that the results generalize on other stimuli as well

Introduction: Why we need mixed-models

- The variation between faces, words, etc. can be substantial (e.g., some faces producing very high ratings and other a very low rating)
- Ignoring this variation and concluding that results can **be generalized to other people and other faces** can in turn result in a **substantial bias**.
- Responsible (at least in part) for failures to replicate effects
- To generalize to other stimuli (and thus increase replicability), we have to take into account stimuli's variability.
- Alternative: By-stimuli analysis?

‘By-stimuli’ analysis

- **‘By-stimuli’ analysis** (alternative ANOVA), that is, using stimuli as the unit of analysis:

  We will test an ANOVA with Cond (Black vs. White) as a between-stimulus factor.

→ *Stimuli = ‘random’ factor/effect*

	Stim 1	Stim 2	Stim 3	Stim 4	Stim 5	Stim 6	Stim 7	Stim 8	Stim 9	Stim 10	Stim 11	Stim 12	…
Pp 1	5	4	6	7	3	8	8	7	9	5	6	5	
Pp 2	4	4	7	8	4	6	9	6	7	4	5	6	
Pp 3	5	3	6	7	4	5	7	5	8	3	4	5	

…

‘By-stimuli’ analysis

- We recondition the data frame, we compute one mean value **per stimulus**.

stim	Cond	y	Condc
1	Black	-1.70133333	-0.5
2	Black	-3.99066667	-0.5
3	Black	-4.12733333	-0.5
4	Black	1.13566667	-0.5
5	Black	-0.79933333	-0.5
6	Black	0.20033333	-0.5
7	Black	0.91466667	-0.5
8	Black	-1.22633333	-0.5
9	Black	-3.66266667	-0.5
10	Black	-0.68400000	-0.5
11	Black	-3.10000000	-0.5
12	Black	-0.31833333	-0.5
13	Black	-4.35133333	-0.5
14	Black	1.10266667	-0.5
15	Black	-1.00633333	-0.5
16	White	-1.35466667	0.5
17	White	6.13700000	0.5
18	White	-1.34800000	0.5
19	White	1.15000000	0.5

- Here, no issue with non independence of the residuals for stimuli (given that stimuli are only in one condition of the Cond variable)

- Finally, we can perform the regression corresponding to the following model: $y = \alpha_0 + \alpha_1 \text{Condc} + \varepsilon_i$ (corresponds to a between-stimuli ANOVA).

‘By-stimuli’ analysis

$$y = \alpha_0 + \alpha_1 \text{Condc} + \varepsilon_i$$

- To perform this regression in R, we use:

`fit.lm <- lm(y ~ 1 + Condc, DF_stim)` => 1 is not necessary (it is automatically implied when indicating a slope). Here we estimate an intercept and a slope.

`summary(fit.lm)`

```
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.1804     0.3737  -0.483  0.63297
Condc         2.5211     0.7473   3.373  0.00219 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

‘By-stimuli’ analysis

- We managed to take into account the variance between stimuli
- However, by computing the average value for each stimulus we lost a critical information. We ignored two sources of variance from the participants:
  - Participants’ mean (to what extent participants have, on average, higher/lower ratings than others)
  - Participants’ condition difference (to what extent participants differ in judging differently White and Black faces)

	Stim 1	Stim 2	Stim 3	Stim 4	Stim 5	Stim 6	Stim 7	Stim 8	Stim 9	Stim 10	Stim 11	Stim 12	Mean
Pp 1	5	4	6	7	3	8	8	7	9	5	6	5	6,08
Pp 2	4	4	7	8	4	6	9	6	7	4	5	6	5,83
Pp 3	5	3	6	7	4	5	7	5	8	3	4	5	5,16

- Thus, we still have the generalization issue: we cannot make inferences on whether we should obtain the same results on other participants

Solution?

- The former permits generalization to different participants but same targets
- The latter permits generalization to different targets but same participants
- Do both ‘by-participants’ and ‘by-stimuli’ analyses? Their **conjunction** does not permit generalization to future studies with **different participants *and* different targets**
- Mixed-models allowing multiple random factors and thus modeling multiple sources of ‘error’ variance at the same time

Introduction: Why we need mixed-models

- To generalize the observed results on other participants and other stimuli at the same time
- To control for Type I error (and thus increase replicability)
- It allows a great flexibility in the analyses (e.g., when using continuous within-participants IVs)
- To keep a maximum of information
- To test for the variability of the effects

Introduction: Why we need mixed-models

- What is a random/fixed factor?

Random factor/effect = almost infinite number of levels, requiring a random sampling —> we aim at making inference on other levels of the factor (e.g., participants)

Fixed factor/effect (independent variable) = finite number of levels, not requiring a random sampling —> we do not make inferences on other levels of the factor (e.g., cognitive load)

- What is a crossed/nested random factor?

Crossed factor (= within): when the levels of the random factor (e.g., each participant) appear for each level of the fixed factor

Nested factor (= between): when the levels of the random factor appear only for some levels of the fixed factor

Plan

- Introduction: Why we need mixed-models
- Two random factors, one fixed effect
- Two random factors, two fixed effects
- Nested random factors: The school classes example
- Power estimations
- Good practices considerations

Typical experiment in social psychology
(Judd et al., 2012)

- A researcher shows 15 faces of White people and 15 faces of Black people
- For each face, each participant emit a judgment on a scale from -20 to +20
- IV: targets' race (White vs. Black) → fixed factor

—> participants are fully crossed with targets' race (i.e., within-participants variable) and targets are nested in the race condition (a given target is either Black or White).

- DV: judgment

	Stim 1	Stim 2	Stim 3	Stim 4	Stim 5	Stim 6	Stim 7	Stim 8	Stim 9	Stim 10	Stim 11	Stim 12	…
Pp 1	5	4	6	7	3	8	8	7	9	5	6	5	
Pp 2	4	4	7	8	4	6	9	6	7	4	5	6	
Pp 3	5	3	6	7	4	5	7	5	8	3	4	5	

…

Two random factor, one fixed effect

- Basic concepts
- Model declaration
- Fixed effects
- Random effects

Two random factor, one fixed effect

- Basic concepts
- Model declaration
- Fixed effects
- Random effects

Two random factor, one fixed effect

- Instead of « reconditionning » the data frame, we keep data in a ‘long’ format (i.e., with all the information we have)
- Instead of working on a score of difference (as for the by-participants analysis) we will model the sources of the non independence

pp	stim	Cond	y	Condc
1	1	Black	-2.30	-0.5
1	2	Black	-0.56	-0.5
1	3	Black	-4.47	-0.5
1	4	Black	-2.11	-0.5
1	5	Black	2.55	-0.5
1	6	Black	-1.08	-0.5
1	7	Black	-3.65	-0.5
1	8	Black	2.15	-0.5
1	9	Black	-3.17	-0.5
1	10	Black	-2.90	-0.5
1	11	Black	-5.43	-0.5
1	12	Black	-0.32	-0.5
1	13	Black	-9.15	-0.5
1	14	Black	-4.51	-0.5
1	15	Black	-0.46	-0.5
1	16	White	9.27	0.5
1	17	White	8.77	0.5
1	18	White	0.77	0.5
1	19	White	-0.33	0.5

Two random factor, one fixed effect

$$Y_{ij} = \beta_{0ij} + \beta_{1i}C_{ij} + \varepsilon_{ij}$$

i = Participants

j = Stimuli

Y = DV

C = Condc, the condition variable that is contrast-coded

ε_{ij} = the residual error for a given participant i and a given stimulus j

- Modeling the sources of non-independence (for participants) and variances (for both participants and stimuli) implies to decompose ε_{ij}.

Two random factor, one fixed effect

$$Y_{ij} = \beta_{0ij} + \beta_{1i} C_{ij} + \varepsilon_{ij}$$

In this case, ε_{ij} can refer to 4 sources of variation:

- The **intercept** can vary from one **participant** to another: Some participants have judgments more positive or more negative compared to the average value of judgment
- The **intercept** can vary from one **stimulus** to another: Some stimuli have judgments more positive or more negative compared to the average value of judgment
- The **slope** can vary from one **participant** to another: Some participants are more or less sensitive to conditions, so there will be variations around the slope
- What remains in addition to these variations (i.e., the error)

Why the slope cannot vary from one stimulus to another?

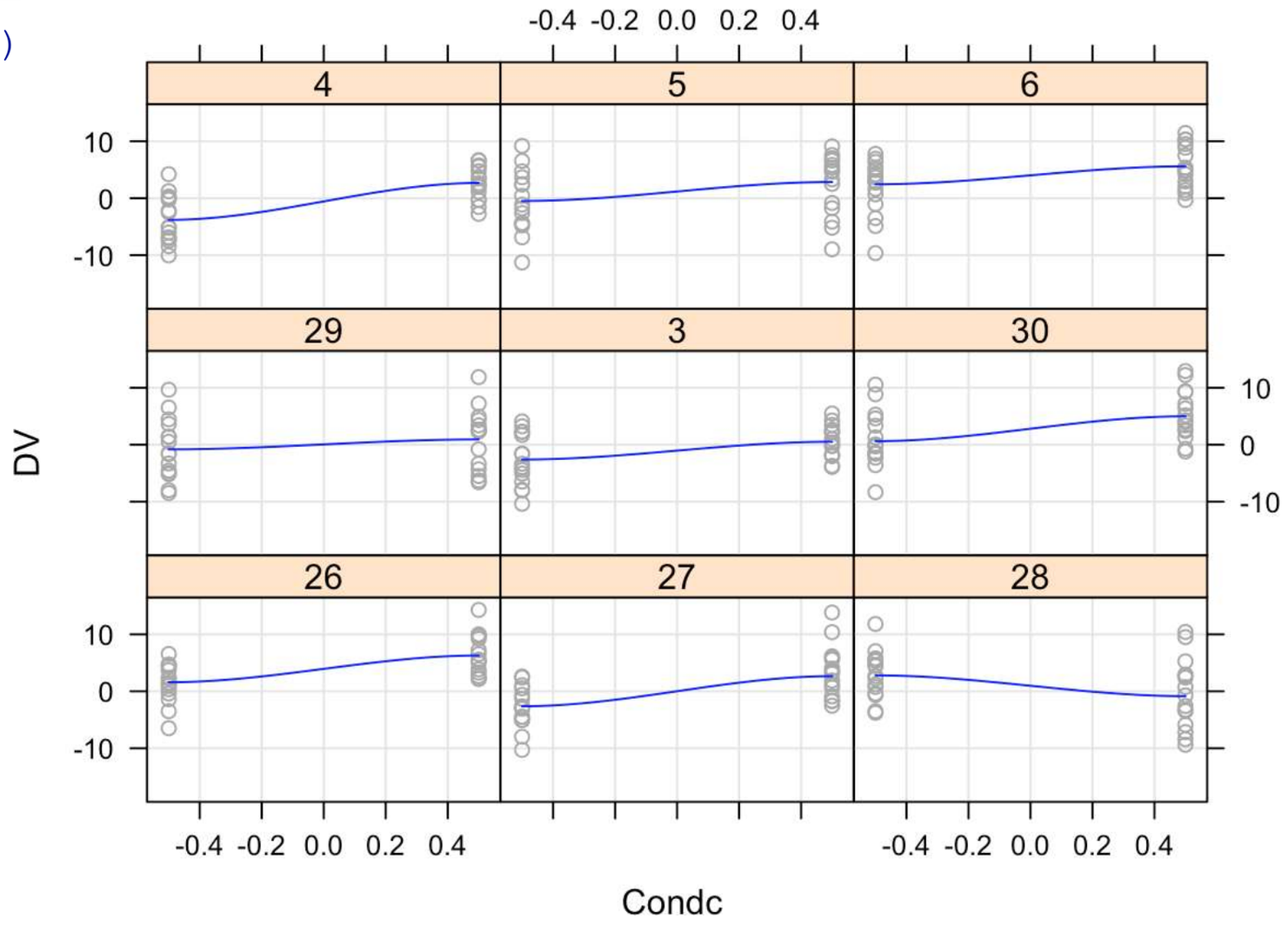
Two random factor, one fixed effect

Why the slope cannot vary from one stimulus to another?

```
xylowess.fnc(y ~ Condc | pp, data =
DF, ylab = « DV",layout = c(3,3,3) )
```

« languageR » function

It makes sense to estimate an intercept and a slope for each participant because they are **crossed with the condition**



Two random factor, one fixed effect

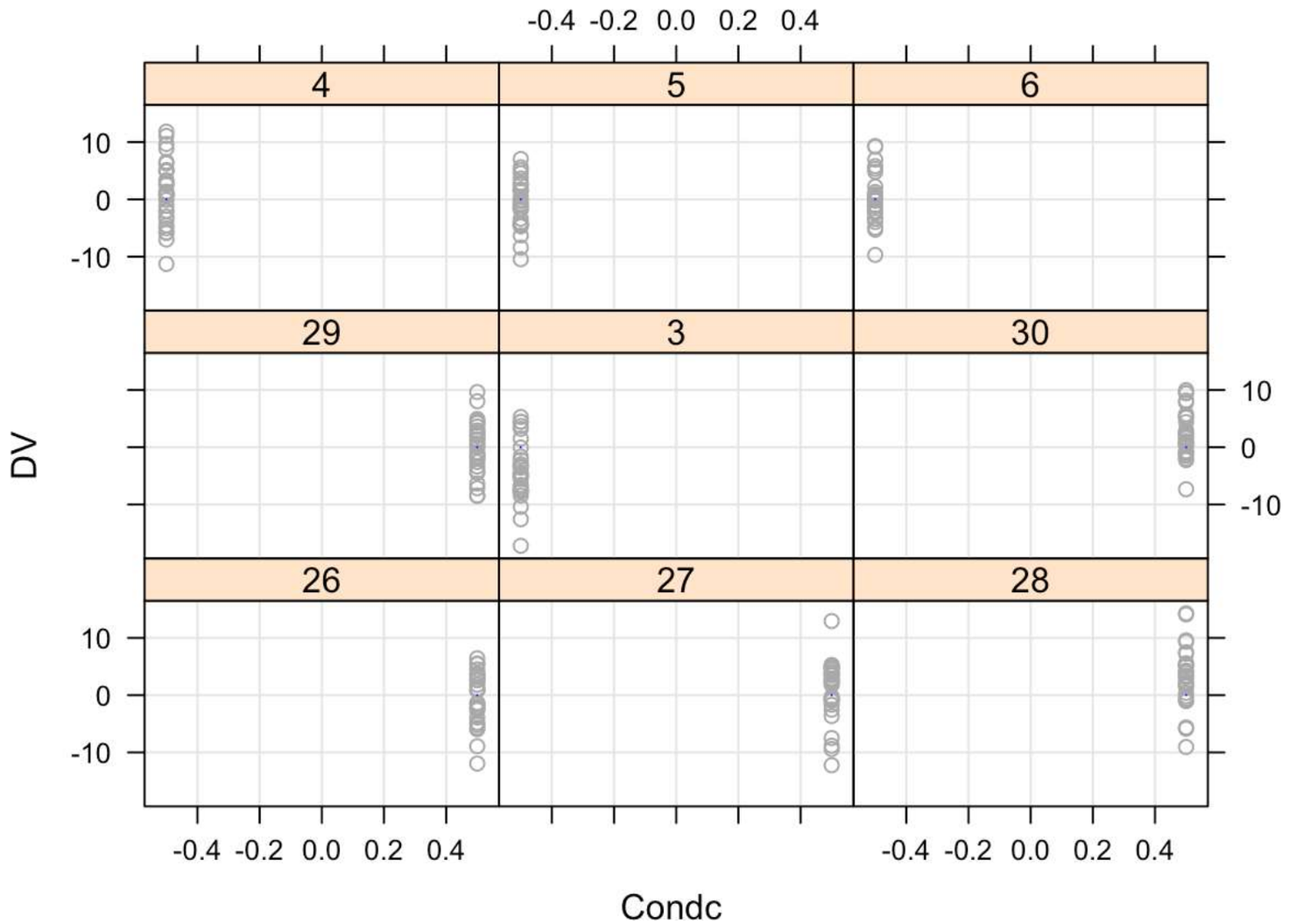
Why the slope cannot vary from one stimulus to another?

```
xylowess.fnc(y ~ Condc | stim, data =
DF, ylab = "DV",layout = c(3,3,3))
```

« languageR » function

It makes sense to estimate an intercept for each stimulus.

However it is not correct to estimate a slope for stimuli because they are **nested within the condition**



Two random factor, one fixed effect

General model

$$Y_{ij} = \beta_{0ij} + \beta_{1i} C_{ij} + \varepsilon_{ij}$$

$$\alpha_0 + \mu_{0i} + \mu_{0j} \qquad\qquad \alpha_1 + \mu_{1i}$$

$$Y_{ij} = (\alpha_0 + \mu_{0i} + \mu_{0j}) + (\alpha_1 + \mu_{1i}) C_{ij} + \varepsilon_{ij}$$

$$Y_{ij} = \underbrace{\alpha_0 + \alpha_1 C_{ij}}_{\text{Fixed effects}} + \underbrace{\mu_{0i} + \mu_{0j} + \mu_{1i} C_{ij}}_{\text{Random effects}} + \varepsilon_{ij}$$

Two random factor, one fixed effect

Mixed-model

$$Y_{ij} = \alpha_0 + \alpha_1 C_{ij} + \mu_{0i} + \mu_{0j} + \mu_{1i} C_{ij} + \varepsilon_{ij}$$

Fixed effects *Random effects*

- Fixed effect: effect of one variable for which the levels can be reproduced on other samples
- Random effects: how the effects vary in the selected sample

Two random factor, one fixed effect

- Basic concepts
- Model declaration
- Fixed effects
- Random effects

Two random factor, one fixed effect

Model estimation

$$Y_{ij} = \alpha_0 + \alpha_1 C_{ij} + \mu_{0i} + \mu_{0j} + \mu_{1i} C_{ij} + \varepsilon_{ij}$$

- Most popular package for mixed-models: lme4
- With the DF in a long format we will use the lmer function:

```
fit.lmer <- lmer(y ~ 1 + Condc + (1 + Condc | pp) + (1 | stim),
                 data=DF)
```

- To have the results:

```
summary(fit.lmer)
```

Note: when we declare a slope, the "1" becomes useless.
Thus "1 + Condc" is equivalent to "Condc"

Two random factor, one fixed effect

Results (lme4)

```
Linear mixed model fit by REML ['lmerMod']
Formula: y ~ 1 + Condc + (1 + Condc | pp) + (1 | stim)
   Data: DF

REML criterion at convergence: 5179

Scaled residuals:
     Min       1Q   Median       3Q      Max
-3.11189 -0.66409 -0.01626  0.62175  2.80907

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pp       (Intercept)  4.294   2.072
          Condc        4.182   2.045    0.27
 stim     (Intercept)  3.670   1.916
 Residual             15.557   3.944
Number of obs: 900, groups:  pp, 30; stim, 30

Fixed effects:
            Estimate Std. Error t value
(Intercept)  -0.1804     0.5317  -0.339
Condc         2.5211     0.8354   3.018

Correlation of Fixed Effects:
      (Intr)
Condc 0.086
```

Random effects

Fixed effects

Two random factor, one fixed effect

- Classic approach: within-participants ANOVA
- Mixed-model: model the errors
  - Basic concepts
  - Model declaration
  - Fixed effects
  - Random effects

Two random factor, one fixed effect

Fixed effects: Interpretation

```
Fixed effects:
            Estimate Std. Error t value
(Intercept)  -0.1804     0.5317  -0.339
Condc         2.5211     0.8354   3.018
```

The estimated parameter for the fixed effects can be interpreted as usual:

- Intercept a0 = -0.18 —> it is the prediction for a value of 0 on Condc. Given that Condc is contrast coded, it is the average prediction
- Slope a1 = 2.52 —> it is how much the prediction changes when we increase of 1 unit on Condc. Given that we have 1 unit of difference between the two conditions (-0.5 for Black and +0.5 for White), it is the mean difference. In this example, faces of White people are judged as 2.52 higher than faces of Black people.

Two random factor, one fixed effect

Fixed effects: Interpretation

```
Fixed effects:
            Estimate Std. Error t value
(Intercept)  -0.1804     0.5317  -0.339
Condc         2.5211     0.8354   3.018
```

- lmer (from lme4 package) does not give any *p*-values
- One way to have this information is to use the lmerTest package (Satterthwhaite method for this estimation)

Two random factor, one fixed effect

Fixed effects: Interpretation

- We have to activate the lmerTest package
- (Re)create our model:

```
fit.lmer <- lmer(y ~ 1 + Condc + (1 + Condc | pp)+ (1 | stim), data=DF)
```

- Use the summary function:

```
summary(fitA.lmer)
```

```
Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  -0.1804     0.5317 50.4747  -0.339   0.7358
Condc         2.5211     0.8354 38.5167   3.018   0.0045 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

One random factor, one fixed effect

- Basic concepts
- Model declaration
- Fixed effects
- Random effects

Two random factor, one fixed effect

Random effects

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pp       (Intercept)  4.294   2.072
          Condc        4.182   2.045    0.27
 stim     (Intercept)  3.670   1.916
 Residual             15.557   3.944
```

The table of the random effects gives the variance (and SD) of the various random effects:

- The estimated variance of the random participant effect for the intercept ($\sigma^2 \mu_{0i}$) is 4.294
- The estimated variance of the random stimulus effect for the intercept ($\sigma^2 \mu_{0j}$) is 3.670
- The estimated variance of the random participant effect for the slope ($\sigma^2 \mu_{1i}$) is 4.182
- The estimated variance for the residual error ($\sigma^2 \varepsilon_{ij}$) is 15.557

Two random factor, one fixed effect

Random effects

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pp       (Intercept)  4.294   2.072
          Condc        4.182   2.045    0.27
 stim     (Intercept)  3.670   1.916
 Residual             15.557   3.944
```

- These are the variances of the non adjustments values that are applied to the observations (i.e., difference between the average coefficient and those we truly observe for each random factor)
- We can obtain these adjustments values with the ranef function (for random effects)

Two random factor, one fixed effect

Adjustment values: Participants

```
ranef(fitA.lmer)
$pp
   (Intercept)      Condc
1  -0.07708646  1.0295864
10 -0.23395560 -1.1346206
11 -5.90160350 -2.9692881
12 -0.27177811 -0.6517655
13 -0.33972272  1.6026877
14  1.33497204 -1.8705777
15 -0.57064065 -0.3104588
```

- For participant 1, the adjustment for the intercept (μ_{0i}) is -0.08, meaning that we apply an adjustment of this value for this observation to attain the observed coefficients
- For this participant, the slope adjustment (μ_{1i}) is 1.03, meaning that we apply an adjustment of this value for this observation to attain the observed coefficients

$$Y_{ij} = \alpha_0 + \alpha_1 C_{ij} + \mu_{0i} + \mu_{0j} + \mu_{1i} C_{ij} + \varepsilon_{ij}$$

$$Y_{ij} = -0.18 + 2.52 C_{ij} + \mu_{0i} + \mu_{0j} + \mu_{1i} C_{ij} + \varepsilon_{ij}$$

$$Y_{ij} = (-0.18 + \mu_{0i}) + (2.52 C_i + \mu_{1ij}) + \mu_{0j} + \varepsilon_{ij}$$

For participant 1:

$Y_{1j} = (-0.18 - 0.08) + (2.52 + 1.03) C_{1j}$

$Y_{1j} = -0.26 + 3.55 C_{1j}$ => In R we can find these coefficients with the function coef(fitA.lmer)

Two random factor, one fixed effect

Adjustment values : Stimuli

```
ranef(fitA.lmer)
 $stim
    (Intercept)
 1  -0.22814340
 10  0.66324558
 11 -1.45365721
 12  0.98364326
 13 -2.55007734
 14  2.22872558
 15  0.38081664
 16 -2.13335591
```

$$Y_{ij} = \alpha_0 + \alpha_1 C_{ij} + \mu_{0i} + \mu_{0j} + \mu_{1i} C_{ij} + \varepsilon_{ij}$$

$$Y_{ij} = -0.18 + 2.52 C_{ij} + \mu_{0i} + \mu_{0j} + \mu_{1i} C_{ij} + \varepsilon_{ij}$$

$$Y_{ij} = (-0.18 + \mu_{0j}) + (2.52 C_i + \mu_{1ij}) + \mu_{0i} + \varepsilon_{ij}$$

For stimulus 1:

Y1j = (-0.18 - 0.23) + 2.52C1j

Y1j = 0.41 + 2.52C1j

- For stimulus 1, the adjustment for the intercept ($\mu 0i$) is -0.23, meaning that we apply an adjustment of this value for this observation to attain the observed coefficient

Two random factor, one fixed effect

Correlation of random effects

$$Y_{ij} = \alpha_0 + \alpha_1 C_{ij} + \mu_{0i} + \mu_{0j} + \mu_{1i} C_{ij} + \varepsilon_{ij}$$

```
fit.lmer <- lmer(y ~ 1 + Condc + (1 + Condc | pp)+ (1 | stim), data=DF)
```

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pp       (Intercept)  4.294   2.072
          Condc        4.182   2.045   0.27
 stim     (Intercept)  3.670   1.916
 Residual             15.557   3.944
```

→ Correlation between the random intercept and the random slope

- In the model declaration, the fact to have the intercept and the slope in the same random term indicates implicitly that we authorize their correlation:

```
(1 + Condc | pp)
```

- To remove the estimation of this correlation, we need 2 separate terms:

```
fit.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim),
data=DF)
```

- Or alternatively the '||' sign:

```
fit.lmer <- lmer(y ~ 1 + Condc + (1 + Condc || pp)+ (1 | stim), data=DF)
```

Two random factor, one fixed effect

Correlation of random effects

$$Y_{ij} = \alpha_0 + \alpha_1 C_{ij} + \mu_{0i} + \mu_{0j} + \mu_{1i} C_{ij} + \varepsilon_{ij}$$

```
fitA.lmer <- lmer(y ~ 1 + Condc + (1 + Condc | pp)+ (1 | stim), data=DF)

fitB.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim), data=DF)
```

- We have two models (A and B) with model A having something more than the other (i.e., the correlation between the random intercept and the random slope)
- To test whether the correlation adds something to the model, we can use the anova function:

```
anova(fitA.lmer,fitB.lmer)
```

```
Models:
fitB.lmer: y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim)
fitA.lmer: y ~ 1 + Condc + (1 + Condc | pp) + (1 | stim)
           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
fitB.lmer  6 5194.4 5223.2 -2591.2   5182.4
fitA.lmer  7 5195.1 5228.7 -2590.5   5181.1 1.3022      1     0.2538
```

- Note: we have 1 df of difference between the two model.
- The A model does not perform better than the B model, meaning that the correlation is not significant (some would say ‘useless’)

Two random factor, one fixed effect

Random effects: tests

$$Y_{ij} = \alpha_0 + \alpha_1 C_{ij} + \mu_{0i} + \mu_{0j} + \mu_{1i} C_{ij} + \varepsilon_{ij}$$

- This is the model B without the correlation:

```
fitB.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim), data=DF)
```

- We will compare the model B with other models in which we removed one target random parameter:

```
fitC.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp)+ (1 | stim), data=DF)

fitD.lmer <- lmer(y ~ 1 + Condc + (1 | pp) + (1 | stim), data=DF)

fitE.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp), data=DF)
```

- Model C is for testing the importance of the random intercept for participants, model D for the random slope for participants, and model E is for the random intercept for stimuli

Exercise

- Data ‘dataJuddetal’
- By-participant analysis
- By-stimuli analysis
- Mixed-model analysis
  - Equivalent to the by-participants analysis
  - Equivalent to the by-stimuli analysis
  - Full mixed-model: interpretation of the fixed effects and test of the random effects

Exercise

Each participant row of data – paired sample t-test on participant means for each race

```
> fit.lm<-lm(W1~1, DF_pp)
> summary(fit.lm)

Call:
lm(formula = W1 ~ 1, data = DF_pp)

Residuals:
    Min      1Q  Median      3Q     Max
-5.8464 -1.5217  0.5149  1.6999  3.8003

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.5211     0.4567   5.521 6.98e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.501 on 29 degrees of freedom
```

By participant analysis equivalent to omitting random intercept due to targets
Long form data

```
> fitA.lmer <- lmer(y ~ 1 + Condc + (Condc | pp), data=DF)
> summary(fitA.lmer) # we have the same results
Linear mixed model fit by REML. t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula: y ~ 1 + Condc + (Condc | pp)
   Data: DF

REML criterion at convergence: 5298.5

Scaled residuals:
    Min      1Q  Median      3Q     Max
-3.3658 -0.6958  0.0038  0.6779  2.7859

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pp       (Intercept)  4.171   2.042
          Condc        3.693   1.922    0.29
 Residual             19.228   4.385
Number of obs: 900, groups:  pp, 30

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  -0.1804     0.4005 29.0009  -0.450    0.656
Condc         2.5211     0.4567 28.9997   5.521 6.98e-06 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
      (Intr)
Condc 0.209
```

Exercise

Each stimulus row of data – paired sample t-test on stimuli means for each race

```
> fit.lm<-lm(y~Condc, DF_stim)
> summary(fit.lm)

Call:
lm(formula = y ~ Condc, data = DF_stim)

Residuals:
    Min      1Q  Median      3Q     Max
-2.9104 -1.6013 -0.0952  1.2548  5.0569

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  -0.1804     0.3737  -0.483  0.63297
Condc         2.5211     0.7473   3.373  0.00219 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.047 on 28 degrees of freedom
Multiple R-squared:  0.289,     Adjusted R-squared:  0.2636
F-statistic: 11.38 on 1 and 28 DF,  p-value: 0.002188
```

By participant analysis equivalent to omitting random intercept due to targets
Long form data

```
> fitA.lmer <- lmer(y ~ 1 + Condc + (1 | stim), data=DF)
> summary(fitA.lmer) # we have the same results as in the by-stimuli analysis
Linear mixed model fit by REML. t-tests use Satterthwaite's method [
lmerModLmerTest]
Formula: y ~ 1 + Condc + (1 | stim)
   Data: DF

REML criterion at convergence: 5340.4

Scaled residuals:
    Min      1Q  Median      3Q     Max
-1.6654 -0.66776  0.01705  0.68137  2.89705

Random effects:
 Groups   Name        Variance Std.Dev.
 stim     (Intercept)  3.492   1.869
 Residual             20.897   4.571
Number of obs: 900, groups:  stim, 30

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  -0.1804     0.3737 28.0000  -0.483  0.63297
Condc         2.5211     0.7473 28.0000   3.373  0.00219 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
      (Intr)
Condc 0.000
```

Exercise

Both participant and target as random factors
Target contributes random intercepts only
Participant contributes random intercepts and slopes

```
> fitA.lmer <- lmer(y ~ 1 + Condc + (1 + Condc | pp) + (1 | stim), data=DF)
> summary(fitA.lmer)
Linear mixed model fit by REML. t-tests use Satterthwaite's
  method [lmerModLmerTest]
Formula: y ~ 1 + Condc + (1 + Condc | pp) + (1 | stim)
   Data: DF

REML criterion at convergence: 5179

Scaled residuals:
     Min       1Q   Median       3Q      Max
-3.11189 -0.66409 -0.01626  0.62175  2.80907

Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pp       (Intercept)  4.294   2.072
          Condc        4.182   2.045    0.27
 stim     (Intercept)  3.670   1.916
 Residual             15.557   3.944
Number of obs: 900, groups:  pp, 30; stim, 30

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  -0.1804     0.5317 50.4747  -0.339   0.7358
Condc         2.5211     0.8354 38.5167   3.018   0.0045 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Correlation of Fixed Effects:
      (Intr)
Condc 0.086
```

Exercise

Testing for the correlation (random slope - random intercept pp)

```
> # Testing for the correlation (random slope - random intercept pp)
> fitA.lmer <- lmer(y ~ 1 + Condc + (1 + Condc | pp)+ (1 | stim), data=DF)
> fitB.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim), data
=DF)
>
> anova(fitA.lmer, fitB.lmer)
refitting model(s) with ML (instead of REML)
Data: DF
Models:
fitB.lmer: y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim)
fitA.lmer: y ~ 1 + Condc + (1 + Condc | pp) + (1 | stim)
           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
fitB.lmer  6 5194.4 5223.2 -2591.2   5182.4
fitA.lmer  7 5195.1 5228.7 -2590.5   5181.1 1.3022      1     0.2538
```

Exercise

Testing for the random intercept pp

```
> # Testing for the intercept pp
> fitA.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim), data
=DF)
> fitB.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp)  + (1 | stim), data=DF)
> anova(fitA.lmer, fitB.lmer)
refitting model(s) with ML (instead of REML)
Data: DF
Models:
fitB.lmer: y ~ 1 + Condc + (0 + Condc | pp) + (1 | stim)
fitA.lmer: y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim)
          Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
fitB.lmer  5 5338.2 5362.2 -2664.1   5328.2
fitA.lmer  6 5194.4 5223.2 -2591.2   5182.4 145.86      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Exercise

Testing for the random slope pp

```
> # Testing for the slope pp
> fitA.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim), data
=DF)
> fitB.lmer <- lmer(y ~ 1 + Condc + (1 | pp) + (1 | stim), data=DF)
>
> anova(fitA.lmer, fitB.lmer)
refitting model(s) with ML (instead of REML)
Data: DF
Models:
fitB.lmer: y ~ 1 + Condc + (1 | pp) + (1 | stim)
fitA.lmer: y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim)
           Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
fitB.lmer  5 5216.6 5240.6 -2603.3   5206.6
fitA.lmer  6 5194.4 5223.2 -2591.2   5182.4 24.274      1  8.355e-07 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Exercise

Testing for the random intercept stim

```
> # Testing for the intercept stim
> fitA.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim), data
=DF)
> fitB.lmer <- lmer(y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp), data=DF)
> anova(fitA.lmer, fitB.lmer)
refitting model(s) with ML (instead of REML)
Data: DF
Models:
fitB.lmer: y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp)
fitA.lmer: y ~ 1 + Condc + (0 + Condc | pp) + (1 | pp) + (1 | stim)
          Df    AIC    BIC  logLik deviance  Chisq Chi Df Pr(>Chisq)
fitB.lmer  5 5310.0 5334.0 -2650.0   5300.0
fitA.lmer  6 5194.4 5223.2 -2591.2   5182.4 117.66      1  < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Mixed-model analyses

Experiments with more than one random factor (part II)

Marine Rougier,
Ghent University, Belgium
Marine.Rougier@ugent.be

Jamie Cummins,
Ghent University, Belgium
Jamie.Cummins@ugent.be

Slides based on great mixed-model courses:
Charles M. Judd (University of Colorado, Boulder, USA) Dominique Muller (University Grenoble Alpes, Grenoble, France)

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

Https://bit.ly/2XAtNY9

Experiments with more than one random factor

West University of Timisoara, September 2021

Marine Rougier
Marine.Rougier@ugent.be

Jamie Cummins
Jamie.Cummins@ugent.be

Slides relying on great mixed-model courses:
Charles M. Judd (University of Colorado, Boulder, USA)
Dominique Muller (University Grenoble Alpes, Grenoble, France)

Plan

- Introduction: Why we need this
- Two random factors, one fixed effect
- Two random factors, two fixed effects
- Nested random factors: The school classes example
- Power estimations
- Good practices considerations

Two random factors, two fixed effects

- Correll et al. (2012) example

Journal of Personality and Social Psychology
2002, Vol. 83, No. 6, 1314–1329

0022-3514/02/$5.00 DOI: 10.1037//0022-3514.83.6.1314

The Police Officer's Dilemma: Using Ethnicity to Disambiguate Potentially Threatening Individuals

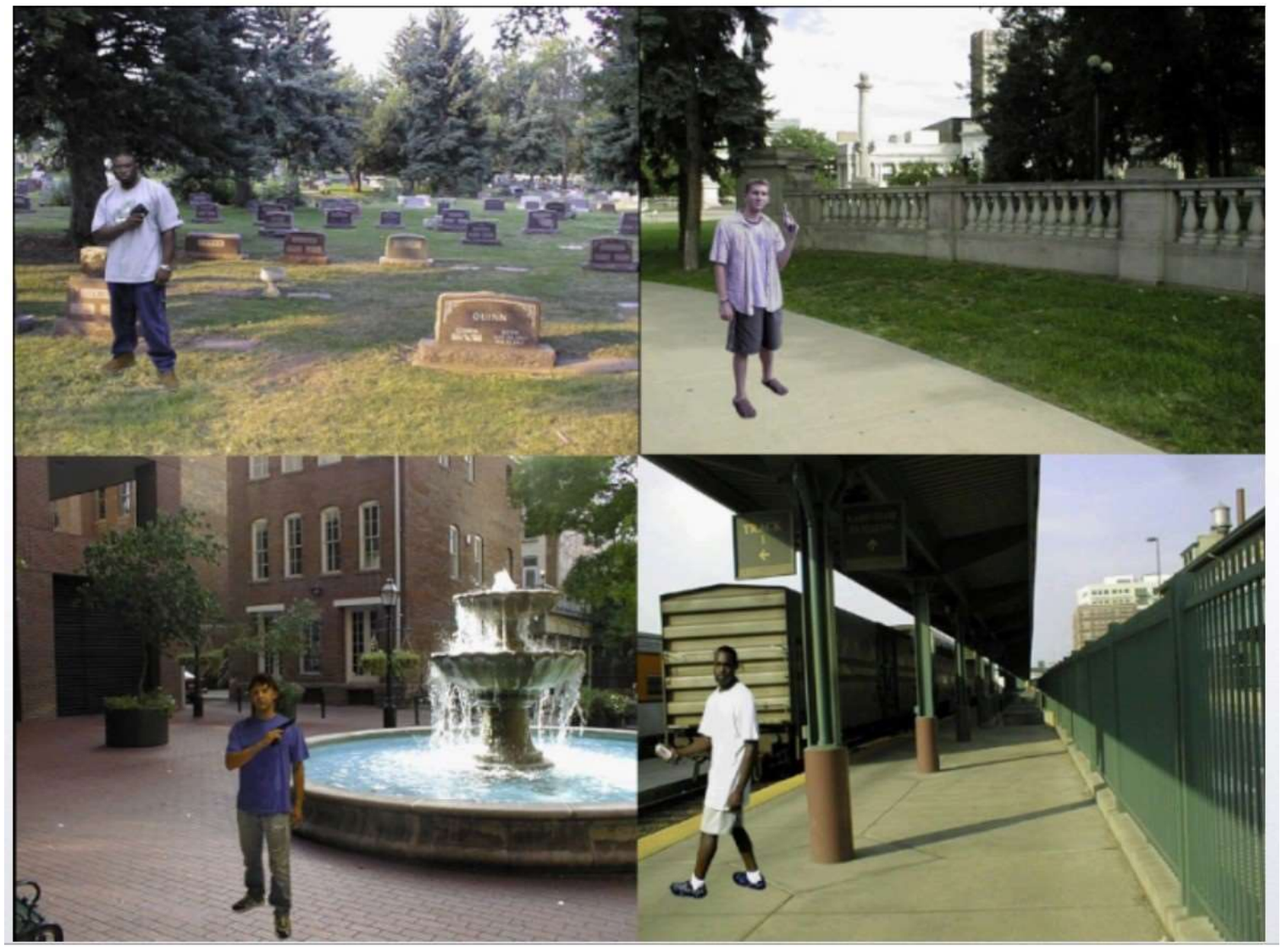
Joshua Correll, Bernadette Park, and Charles M. Judd
University of Colorado at Boulder

Bernd Wittenbrink
University of Chicago

Using a simple videogame, the effect of ethnicity on shoot/don't shoot decisions was examined. African American or White targets, holding guns or other objects, appeared in complex backgrounds. Participants were told to "shoot" armed targets and to "not shoot" unarmed targets. In Study 1, White participants made the correct decision to shoot an armed target more quickly if the target was African American than if he was White, but decided to "not shoot" an unarmed target more quickly if he was White. Study 2 used a shorter time window, forcing this effect into error rates. Study 3 replicated Study 1's effects and showed that the magnitude of bias varied with perceptions of the cultural stereotype and with levels of contact, but not with personal racial prejudice. Study 4 revealed equivalent levels of bias among both African American and White participants in a community sample. Implications and potential underlying mechanisms are discussed.

Two random factors, two fixed effects

- Correll et al. (2012) example



Two random factors, two fixed effects

- Correll et al. showed 25 White targets and 25 Black targets
- Each target stimulus appeared once with a gun and once without a gun (e.g., cellphone)
    - Participants saw all the 4 cases of the design (race X object), that is, they are fully crossed with both race and object
    - Target stimuli are either Black or White (nested with race) and they appeared both with a gun and without (crossed with object)
- At each trial, participants have to indicate as fast as possible if they need to shoot or not

Two random factors, two fixed effects

- Definition of the model
- Study of random effects
- Study of fixed effects
- Theoretical interest of random effects
- Exercise

Two random factors, two fixed effects

- Definition of the model
- Study of random effects
- Study of fixed effects
- Theoretical interest of random effects
- Exercise

Two random factors, two fixed effects

Fixed effects

- IV1: race with people being either Black (-0.5) or White (+0.5)

```
DF$racec <- -0.5*(DF$race=="black") + 0.5*(DF$race=="white")
```

- IV2: object that are either a gun (-0.5) or not a gun (+0.5)

```
DF$objectc <- -0.5*(DF$object=="gun") + 0.5*(DF$object=="nogun")
```

- The authors predict an interaction:
- When the target has a gun, participants will be faster to respond ‘shoot’ if the target is Black rather than White
- When the target does not have a gun, participants will be faster to respond ‘no shoot’ if the target is White rather than Black

Two random factors, two fixed effects

Random effects

- Each participant see the 4 conditions of the design, so we have 4 random effects: the intercept, one slope for the race effect, one slope for the object effect, and one slope for the race*object interaction
- Each target appears once with a gun and once without a gun, but each target is either Black OR White, so we have 2 random effects: the intercept, and one slope for the object effect.

Two random factors, two fixed effects

Model

$$Yij = \alpha 0 + \alpha 1 Raij + \alpha 2 Obij + \alpha 1 Raij * Obij + \mu 0i + \mu 1 iRaij + \mu 2 iObij + \mu 3 iRaij * Obij + \mu 0j + \mu 1 jObij + \varepsilon ij$$

Fixed effects | **Random effects**

```
m0 <- lmer(Time ~ 1 + racec + objectc + racec:objectc + (1 + racec +
    objectc + racec:objectc|pp) + (1 + objectc|person), data=DF2)
```

Equivalent to:

```
m0 <- lmer(Time ~ racec*objectc + (racec*objectc|pp) +
            (objectc|person), data=DF2)
```

Two random factors, two fixed effects

- Definition of the model
- Study of random effects
- Study of fixed effects
- Theoretical interest of random effects
- Exercise

Two random factors, two fixed effects

Study of random effects

- When having a complex model like this one, it is possible that the model will not converge (because it is over-specified).
- Procedure of Bates, Kliegl, Vasishth, and Baayen (2015) to remove useless random terms:

1. We run the total model (i.e., with all the random terms)
2. We run the model without the correlations (i.e., with the double bar ‘||’), we identify the small variances and we test them
3. We put back the correlations and we test them
4. At each stage, we make sure that the model is not over-specified

Two random factors, two fixed effects

Study of random effects

We will need the ‘RePsychLing’ package that is not available on CRAN. To download it, we have to use the following lines:

```
install.packages("devtools") #for package development

library(devtools)

install_github("dmbates/RePsychLing") #download the package (in
development) from Github

library(RePsychLing)
```

Two random factors, two fixed effects

Study of random effects

1. We run the total model (i.e., with all the random terms)

```
m0 <- lmer(Time ~ 1 + racec + objectc + racec:objectc + (1 + racec +
objectc + racec:objectc|pp) + (1 + objectc|person), data=DF2)
```

```
> summary(rePCA(m0))
$person
Importance of components:
                          [,1]   [,2]
Standard deviation     0.9112 0.4228
Proportion of Variance 0.8229 0.1771
Cumulative Proportion  0.8229 1.0000

$pp
Importance of components:
                          [,1]   [,2]      [,3] [,4]
Standard deviation     0.3909 0.2744 0.0006749    0
Proportion of Variance 0.6700 0.3300 0.0000000    0
Cumulative Proportion  0.6700 1.0000 1.0000000    1
```

→ Principal component analysis

→ Says that the matrix is ‘singular’ and the model is over-specified

Two random factors, two fixed effects

Study of random effects

2. We run the model without the correlations (i.e., with the double bar ‘||’), we identify the small variances and we test them

```
m1 <- lmer(Time ~ racec*objectc + (racec*objectc||pp) + (objectc||person),
data=DF2)
```

```
> VarCorr(m1)
 Groups   Name           Std.Dev.
 person   objectc        64.0566021
 person.1 (Intercept)    31.1232349
 pp       racec:objectc   0.0000000
 pp.1     objectc        17.7056636
 pp.2     racec           0.0031924
 pp.3     (Intercept)    27.6961418
 Residual                70.9389250
```

→ Extracts the variances (here SD) of the summary

→ Theoretical importance of this zero variance (I come back to this later)

Two random factors, two fixed effects

Study of random effects

2. We run the model without the correlations (i.e., with the double bar '||'), we identify the small variances and we test them

```
m1 <- lmer(Time ~ racec*objectc + (racec*objectc||pp) + (objectc||person),
data=DF2)
```

We set a model without the random effects of racec and the interaction for pp and we test whether the difference is significant:

```
m2 <-lmer(Time ~ racec*objectc + (1+objectc||pp) + (1+objectc||person),
data=DF2)

anova(m1, m2)
```

```
Data: DF2
Models:
m2: Time ~ racec * objectc + (1 + objectc || pp) + (1 + objectc ||
m2:     person)
m1: Time ~ racec * objectc + (racec * objectc || pp) + (1 + objectc ||
m1:     person)
   Df   AIC   BIC logLik deviance Chisq Chi Df Pr(>Chisq)
m2  9 40201 40257 -20092    40183
m1 11 40205 40273 -20092    40183     0      2          1
```

→ The difference is not significant so we can keep m2

Two random factors, two fixed effects

Study of random effects

```
m2 <-lmer(Time ~ racec*objectc + (1+objectc||pp) + (1+objectc||person),
data=DF2)
```

```
> summary(rePCA(m2))
$person
Importance of components:
                        [,1]   [,2]
Standard deviation     0.903 0.4387
Proportion of Variance 0.809 0.1910
Cumulative Proportion  0.809 1.0000

$pp
Importance of components:
                        [,1]   [,2]
Standard deviation     0.3904 0.2496
Proportion of Variance 0.7098 0.2902
Cumulative Proportion  0.7098 1.0000
```

No small variances anymore

Two random factors, two fixed effects

Study of random effects

Finally, we can test for the random effects as before:

```
m2 <-lmer(Time ~ racec*objectc + (1+objectc||pp) + (1+objectc||person), data=DF2)


# tests for the intercept pp
m5 <-lmer(Time ~ racec*objectc + (0+objectc||pp) + (1+objectc||person), data=DF2)
# tests for the slope pp
m6 <-lmer(Time ~ racec*objectc + (1||pp) + (1+objectc||person), data=DF2)

# tests for the intercept stim
m7 <-lmer(Time ~ racec*objectc + (1+objectc||pp) + (0+objectc||person), data=DF2)
# tests for the slope stim
m8 <-lmer(Time ~ racec*objectc + (1+objectc||pp) + (1||person), data=DF2)
```

Two random factors, two fixed effects

- Definition of the model
- Study of random effects
- Study of fixed effects
- Exercise
- Theoretical interest of random effects

Two random factors, two fixed effects

Study of fixed effects

```
m4 <-lmer(Time ~ racec*objectc + (1+objectc||pp) + (1+objectc||person), data=DF2)
summary(m4)
```

```
Fixed effects:
             Estimate Std. Error      df t value Pr(>|t|)
(Intercept)   601.320      6.490  74.315  92.658  < 2e-16 ***
racec          -2.102      9.124  47.852  -0.230   0.8187
objectc        61.396      9.825  55.839   6.249 6.03e-08 ***
racec:objectc -37.719     18.743  47.900  -2.012   0.0498 *
```

$$Y_{ij} = 601.32 - 2.10Ra_{ij} + 61.40Ob_{ij} - 37.72Ra_{ij} * Ob_{ij}$$

$$Y_{ij} = (601.32 - 2.10)Ra_{ij} + (61.40 - 37.72Ra_{ij})Ob_{ij}$$

- a1 = -2.10: when we increase of one unit on race (i.e., from black to white), RT decrease of 2.10 milliseconds
- a2 = 61.40: when we increase of one unit on object (i.e., from Gun to No Gun), RT increases of 61.40 milliseconds. Participants are slower when the object is not a gun.
- a3 = -37.72: The difference observed between gun vs. nogun decreases when we increase of one unit on race (i.e., from black to white). So it is less true for white people.

Two random factors, two fixed effects

- Definition of the model
- Study of random effects
- Study of fixed effects
- Exercise
- Theoretical interest of random effects

Exercise

- Data ‘shooterLong’ https://osf.io/9bw82/?view_only=5eb2e58145ed4ac5abe8639397a64742
- Study of the random effects (Bates method)
- Study of the fixed effects

Two random factors, two fixed effects

- Definition of the model
- Study of random effects
- Study of fixed effects
- Exercise
- Theoretical interest of random effects

Two random factors, two fixed effects

Theoretical interest of random effects

```
m1 <- lmer(Time ~ racec*objectc + (racec*objectc||pp) + (objectc||person),
data=DF2)
```

```
> VarCorr(m1)
 Groups   Name             Std.Dev.
 person   objectc          64.0566021
 person.1 (Intercept)      31.1232349
 pp       racec:objectc     0.0000000
 pp.1     objectc          17.7056636
 pp.2     racec             0.0031924
 pp.3     (Intercept)      27.6961418
 Residual                  70.9389250
```

Theoretical importance of this zero variance

Two random factors, two fixed effects

Theoretical interest of random effects

- If the shooter bias depends on a psychological construct situated at the individual level (e.g., prejudice), we would have expected it to vary between individuals
- One possible explanation: Cultural influence (e.g., city-level or country-level construct)
- Importantly, aiming at moderating an effect that does not vary between individuals with an individual-level variable (e.g., prejudice) will be an issue

Plan

- Introduction: Why we need this
- Two random factors, one fixed effect
- Two random factors, two fixed effects
- Nested random factors: The school classes example
- Power estimations
- Good practices considerations

Nested random factors: The school example

- At times, random effects may be nested within one another
- Example: comparing two interventions for reading abilities
- Students are assigned to complete one of two interventions (SMART or dual n-back); students are collected from multiple different schools. After 8 weeks fluency on reading 5 different texts is assessed
- Participant random effect as before; nested within intervention
- Can also consider school: crossed with intervention
- But these two random factors are *not independent* from each other; participants are nested within school!
- Need to model this non-independence

Nested random factors: The school example

- So “(1 | school/participant)” implies modelling (i) the nested random effect of participant within schools, and (ii) the random effect of school
- AKA: (1 | school:students) + (1 | school)
- BUT school is crossed with intervention, not nested; we need to model (1 + intervention | school).
- Can’t do this as “(1 + intervention | school/participant)” since participant is not crossed with intervention
- Therefore, replace (1 | school) in “(1 | school:students) + (1 | school)” with (1 + intervention | school)
- So…

Nested random factors: The school example

Model

```
m0 <- lmer(reading_score ~ intervention + (1 + intervention | school) +
                (1 | school:participant), data=df)
```

Fixed effects

Random effects

Plan

- Introduction: Why we need this
- Two random factors, one fixed effect
- Two random factors, two fixed effects
- Nested random factors: The school classes example
- Power estimations
- Good practices considerations

Power estimations

- For power estimations, we first need an estimation of the effect size
- Typical effect size
  - Cohen’s d: Expresses the difference between two means relative to their standard deviation
  - Eta-Squared: How much of the total variance is the data is explained by the difference between the means
- In mixed-model, the variance depends on both participants and stimuli.
- How to compute the effect size in mixed-models?

Power estimations

- Let’s go back to our first illustration (15 white faces and 15 black faces judged by all participants)

```
fit.lmer <- lmer(y ~ 1 + Condc + (1 + Condc | pp)+ (1 | stim),
                                        data=DF)
```

  - Westfall et al. (2014; see also Judd et al., 2017 for more complex designs)

$$d = \frac{\textit{difference between the means}}{\sqrt{\textit{varintercept}_{\textit{part}} + \textit{varintercept}_{\textit{item}} + \textit{varslope}_{\textit{part}} + \textit{var}_{\textit{residual}}}}$$

```
Random effects:
 Groups   Name        Variance Std.Dev. Corr
 pp       (Intercept)  4.294   2.072
          Condc        4.182   2.045    0.27
 stim     (Intercept)  3.670   1.916
 Residual             15.557   3.944
Number of obs: 900, groups:  pp, 30; stim, 30

Fixed effects:
            Estimate Std. Error      df t value Pr(>|t|)
(Intercept)  -0.1804     0.5317 50.4747  -0.339   0.7358
Condc         2.5211     0.8354 38.5167   3.018   0.0045 **
---
```

```
d <- 2.5211/(sqrt(4.294+4.182+3.670+15.557)) # 0.4789902
```

Power estimations

- The ‘r2glmm’ package (Jaeger et al., 2017; several method possible)

```
r2nsj = r2beta(fit.lmer, method = 'nsj', partial = TRUE) # pR = .061
```

```
> r2nsj
  Effect   Rsq upper.CL lower.CL
1  Model 0.061    0.094    0.034
2  Condc 0.061    0.094    0.034
```

Power estimations

- As for the effect size computation, power depends on **both** the numbers of **participants** and the numbers of **targets**
- One way to have power estimations: https://jakewestfall.shinyapps.io/two_factor_power/

Power Analysis with Random Targets and Participants

Judd, C. M., Westfall, J., & Kenny, D. A. (2016). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology.*

Article | Supplemental Appendix: Additional topics | Code for this app (using package 'shiny' in R) | Back to JakeWestfall.org

Note: when sharing the link to this app, please use the stable redirecting page at jakewestfall.org/two_factor_power/, as the app's current URL is possibly subject to change.

Choose a design:

CCC

Standardized or Unstandardized input:

Standardized

Note: with Standardized input, all of the Variance Partitioning Coefficients (VPCs) must sum to 1.

Enter the design parameters below.

To compute power estimates, enter an X for the variable you wish to solve for, then click the 'Solve

Design schematic

(The interpretation of this design schematic is explained in the accompanying paper.)

	Target1	Target2	Target3	Target4	Target5	Target6
Participant1	AB	AB	AB	AB	AB	AB
Participant2	AB	AB	AB	AB	AB	AB
Participant3	AB	AB	AB	AB	AB	AB
Participant4	AB	AB	AB	AB	AB	AB
Participant5	AB	AB	AB	AB	AB	AB
Participant6	AB	AB	AB	AB	AB	AB

Solution from power analysis

Power estimations

- The relative importance of participants and stimuli **depends on the design and the magnitude of their variance components**
    - In general, the larger the variance component(s) of a random factor the more increasing the n of that factor will help
    - But this depends on the design, in general increasing the n of the factor that is nested under condition will help more than increasing the n of the factor that is crossed with condition
- In many experimental designs, common assumptions about sufficient number of targets (and participants) are seriously mistaken

Plan

- Introduction: Why we need this
- Two random factors, one fixed effect
- Two random factors, two fixed effects
- Nested random factors: The school classes example
- Power estimations
- Good practices considerations

How to write your results section

"We used R (R Core Team, 2012) and lme4 (Bates, Maechler & Bolker, 2012; version 1.1-21) to perform a linear mixed effects analysis of the relationship between pitch and politeness. As fixed effects, we entered politeness, gender, and their interaction term into the model. As random effects, we had intercepts for subjects and items, as well as by-subject and by-item random slopes for the effect of politeness. P-values were obtained by Satterthwhaite approximation with the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2015; version 3.1-0). Effect sizes were estimated with the 'r2glmm' package (Jaeger et al., 2017; 'nsj' method; version 0.1.2). The analysis revealed..."

- Indicate the fixed and random effects and what exactly is estimated for the random effects (i.e., which slopes)
- Specify whether you followed the Bates method for over-specified models
- Indicate the package version
- sjPlot

Basics of power analysis
(part I)

Marco Perugini,

Università degli Studi di Milano-Bicocca

marco.perugini@unimib.it

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

Outline

- **Basic statistical concepts**
- **Errors of inference**
- **Power analysis**
  -Two independent groups and repeated measures
  -Contrasts
  -Regression
  -Moderation
  -ANOVA (Between, Within, Mixed)
  -Main problems in power analysis

Back to basics

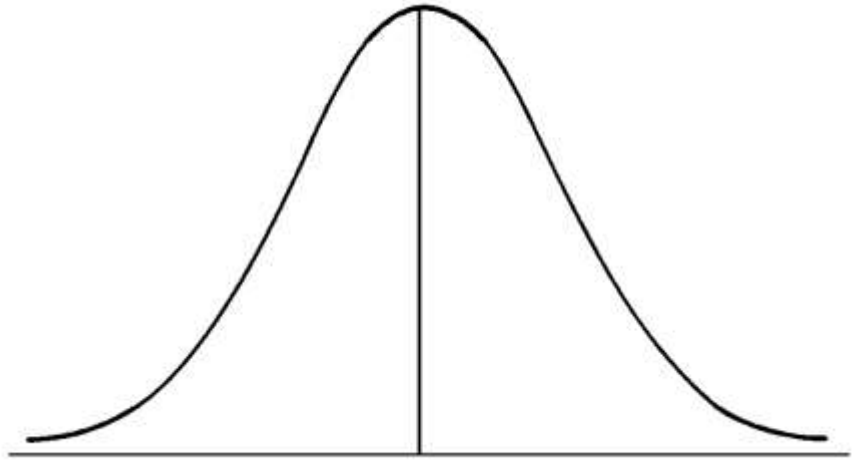
- Power analysis is an important tool in planning studies testing hypotheses

- We need first to go back to a few basic concepts

Mean

- A single value that reflects the central point of a distribution
- If the distribution is normal, it is also the best simple way to summarize it

$$\overline{X} = \frac{\sum X_i}{N}$$

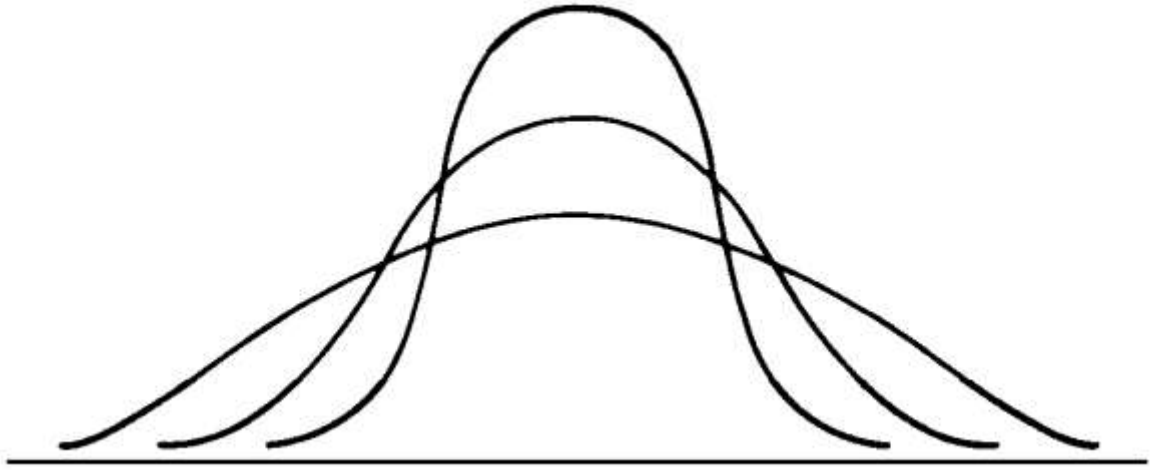


Variance and standard deviation

- Reflects the dispersion (variability) around the mean

$$s^2 = \frac{\sum (X_i - \overline{X})^2}{N} = \frac{\sum X^2}{N} - \overline{X}^2$$

$$s = \sqrt{s^2}$$



Standard error

- When we measure something, more data means less measurement error

- Exit polls are more accurate (less error) the more the sampled voters or polling stations

- We have a sample but would like to say something about the underlying population (or anyway something that generalizes beyond that sample)

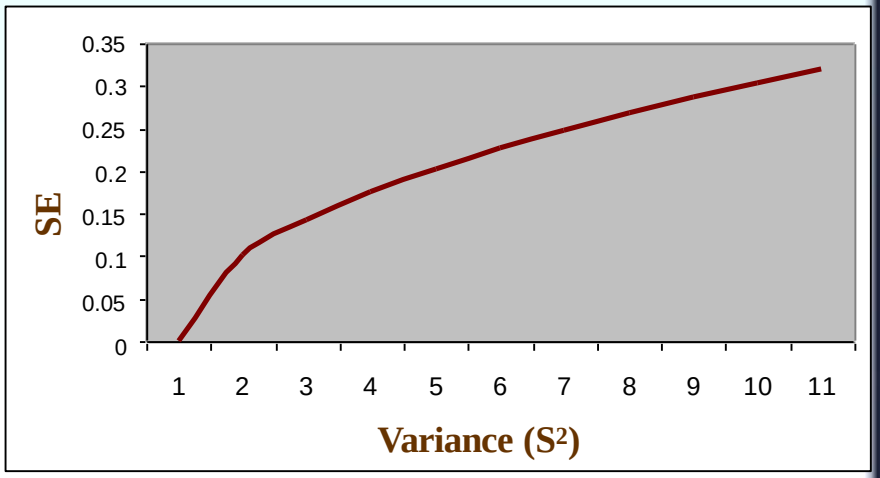
Standard error

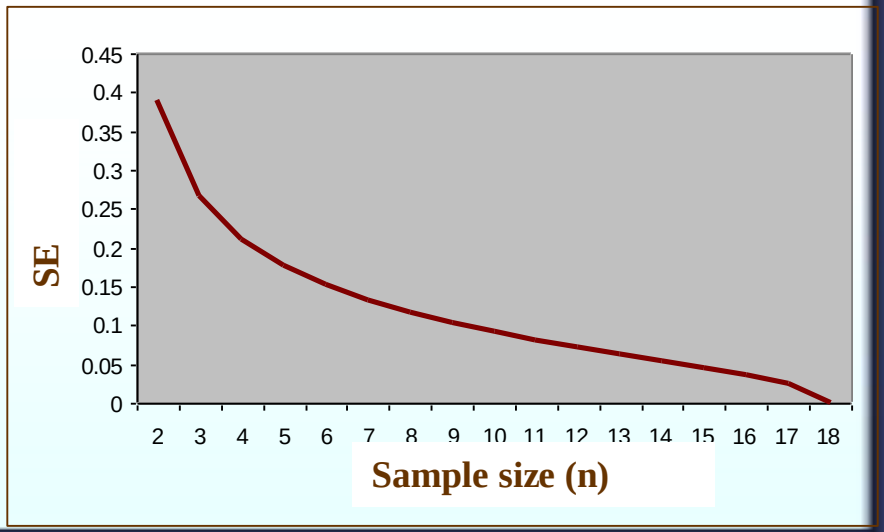
- Error in estimating a population parameter (e.g., mean) from a sample

Goes up with increasing variance

$$SE =$$

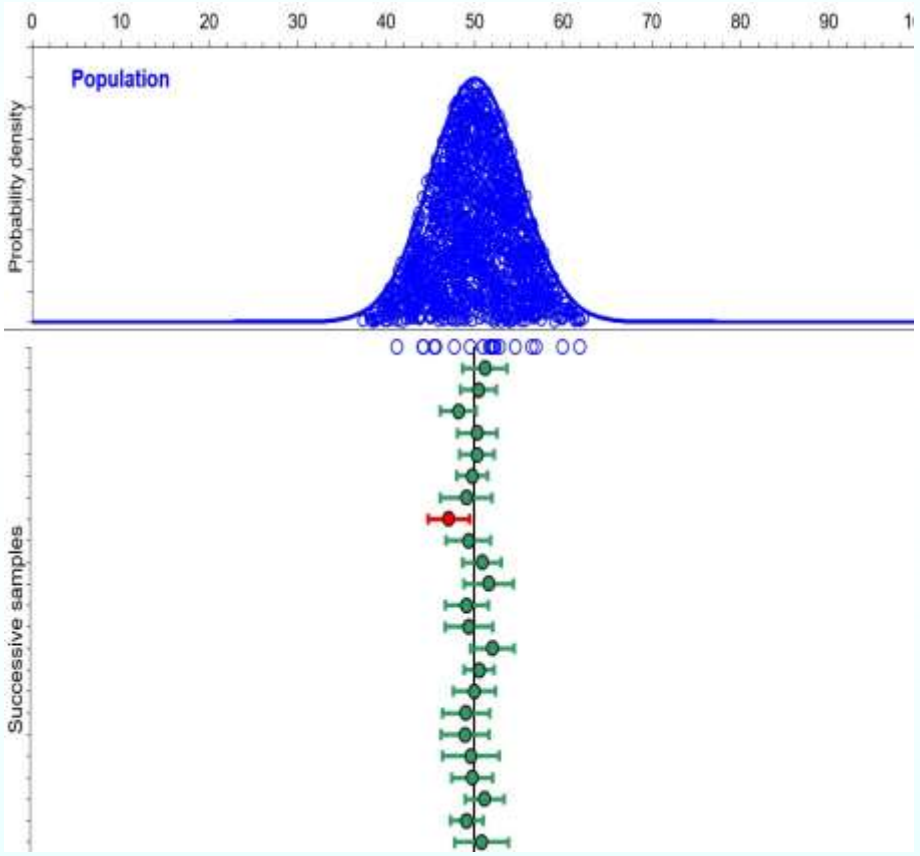
Goes down with increasing sample size



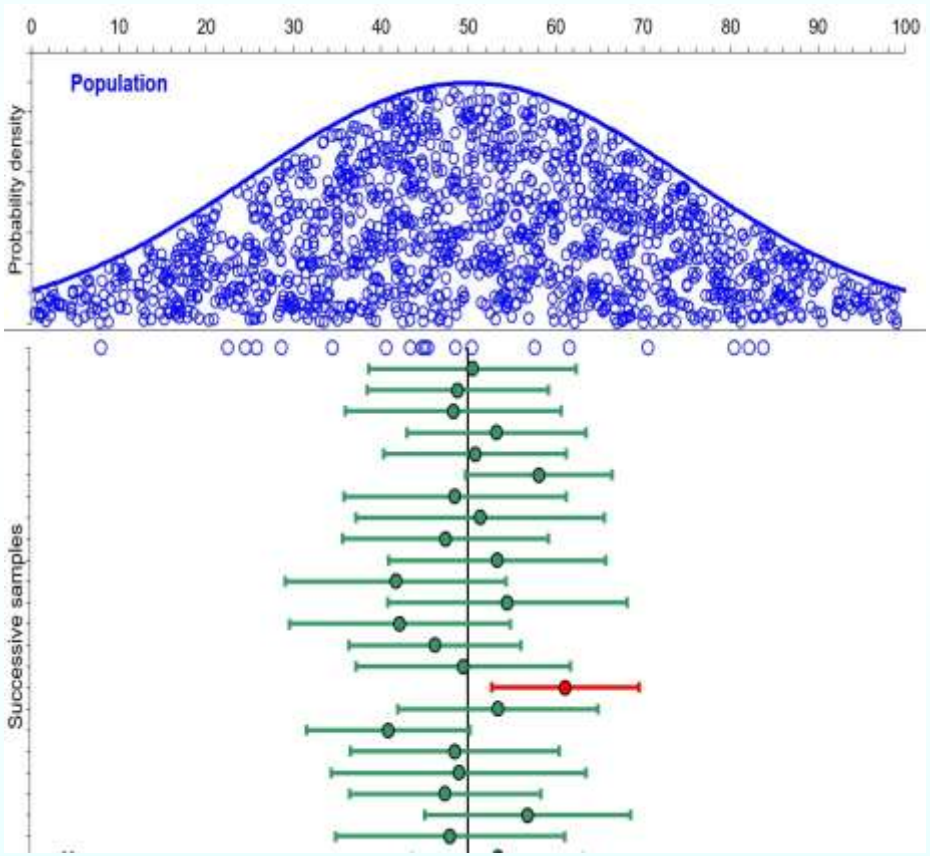


Parameter estimation: Error and variability

N=20

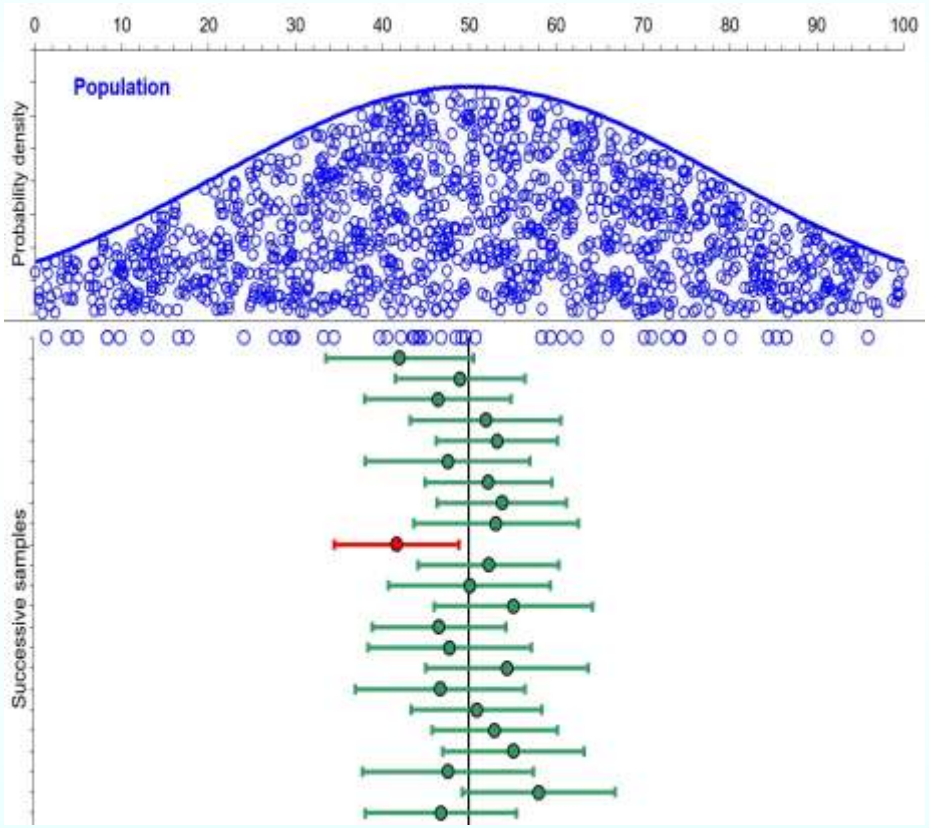


N=20

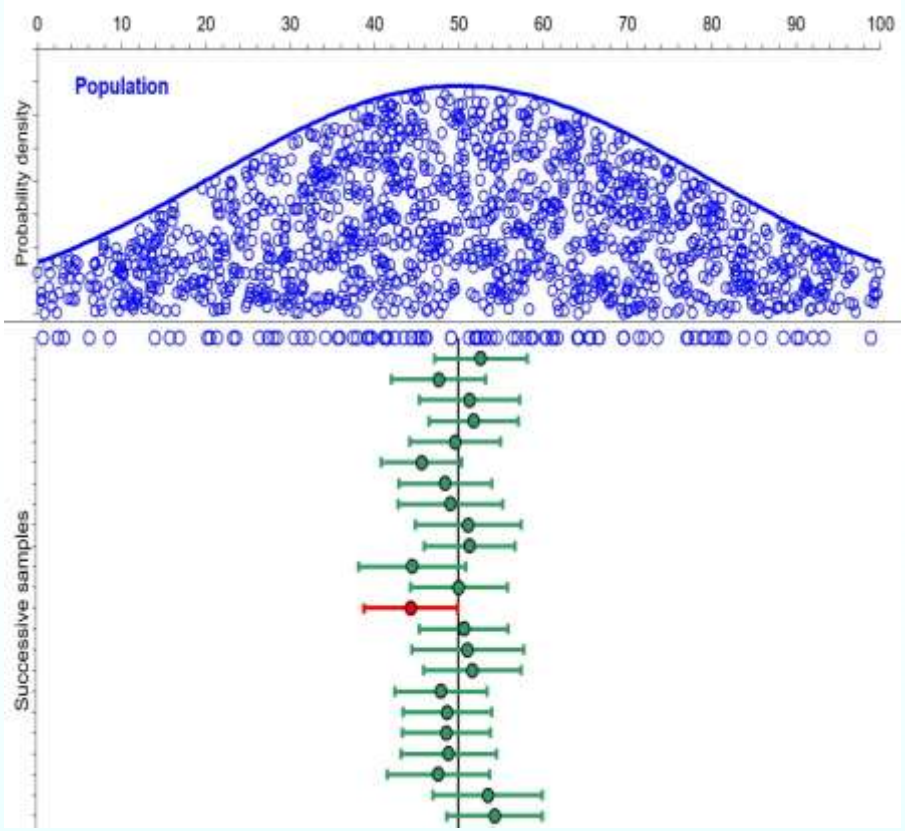


Error and variability

N=50

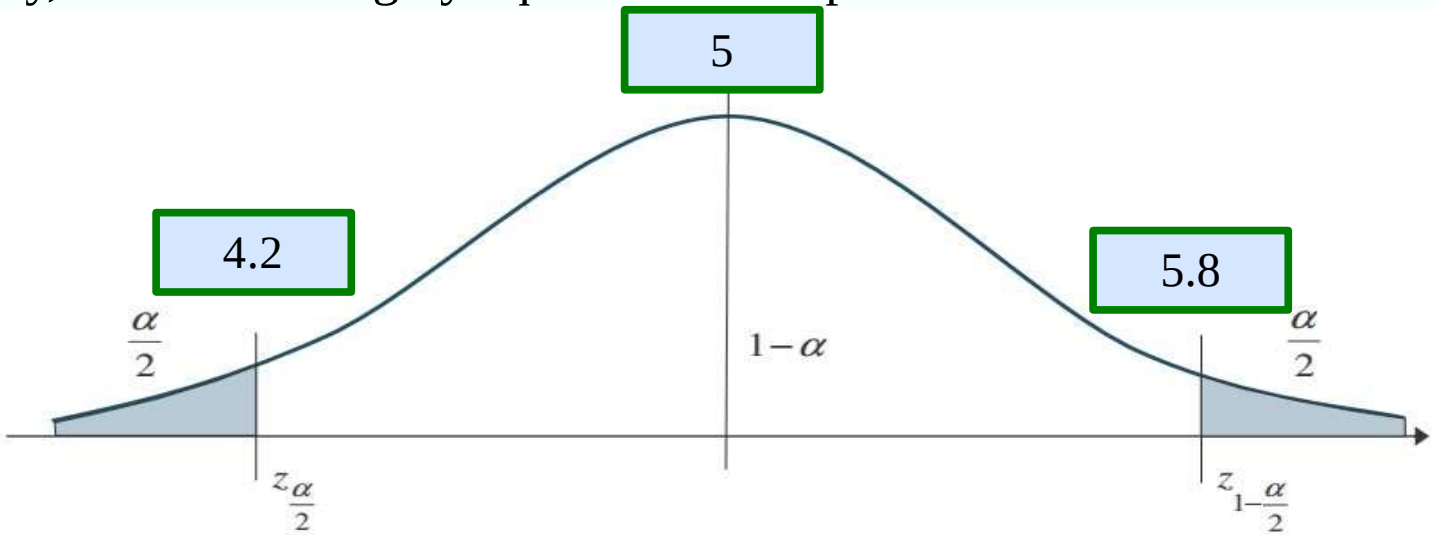


N=100



From SE to Confidence Interval (CI)

The sample estimate does not correspond to the population value.
Confidence Interval provides a range of values that contain the population value with a certain likelihood (e.g., 95%), should the study be repeated many times
To simplify, CI 95% is roughly equal to the sample mean +/- 2 SE



For example: M = 5; DS = 4 N=100

SE = $\sqrt{\frac{4^2}{100}}$ o $\frac{4}{\sqrt{100}}$ =0.4
Range: 2 x SE = 0.8
95% CI = [4.2, 5.8]

$$\mu \in \left(\overline{X}_n \pm t^{(n-1)}_{1-\frac{\alpha}{2}} \sqrt{\frac{\overline{s}_n^2}{n}} \right)$$

Standard Error

The Confidence Interval (CI)

The CI reflects the concept of **accuracy** in estimating a parameter

Imagine this research scenario. We want to understand the efficacy of 2 ads for a product (e.g., snack). N=100
We computed the mean evaluation of the two ads

A) M = +3.10; DS = 15, $p<.05$
B) M = +2.50; DS = 10, $p<.05$

Which is the best ad? It is not obvious that it is A

A) 95% CI= [0.16, 6.04]
B) 95% CI= [0.54, 4.46]

A can be 6.04, but it can also be 0.16.

B is more accurate, so its possible values are less spread: it is very unlikely that its mean is lower than 0.54

Effect size

There are several ways to measure effect size

Most common: Cohen's *d*

$$d = \frac{M_1 - M_2}{pooled\ SD}\ ,\ \text{with } pooled\ SD = \sqrt{\frac{(n_1-1)SD_1^{\,2}+(n_2-1)SD_2^{\,2}}{n_1+n_2-2}},$$

If the two groups have the same sample size,

$$pooled\ SD = \sqrt{\frac{SD_1^{\,2}+SD_2^{\,2}}{2}}$$

But also r (correlation coefficient)
From do to r and viceversa

$$d = \frac{2r}{\sqrt{1-r^2}}$$

Example of Cohen's d

$$pooled\ SD = \sqrt{\frac{(n_1-1)SD_1^2+(n_2-1)SD_2^2}{n_1+n_2-2}} = \sqrt{\frac{(246-1)1.107^2+(219-1)1.197^2}{246+219-2}} = 1.150$$

$$\boldsymbol{d = \frac{M_1-M_2}{pooled\ SD} = \frac{7.79-7.57}{1.150} = 0.19}$$

Statistiche di gruppo

	forma Forma.email	N	Media	Deviazione std.	Errore std. Media
persuasività Giudizio. Persuasività	1 Tu	246	7,57	1,107	,071
	2 Lei	219	7,79	1,197	,081

Conventional values:

d = 0.2 small → *r* = .10
d = 0.5 medium → *r* = .24
d = 0.8 large → *r* = .37

Other effect size indexes (from Ellis, 2010)

Table 1.1 *Common effect size indexes*

Measures of group differences (the *d* family)		Measures of association (the *r* family)	
(a) Groups compared on dichotomous outcomes		**(a) Correlation indexes**	
RD	The risk difference in probabilities: the difference between the probability of an event or outcome occurring in two groups	r	The Pearson product moment correlation coefficient: used when both variables are measured on an interval or ratio (metric) scale
RR	The risk or rate ratio or relative risk: compares the probability of an event or outcome occurring in one group with the probability of it occurring in another	ρ (or r_s)	Spearman's rho or the rank correlation coefficient: used when both variables are measured on an ordinal or ranked (non-metric) scale
OR	The odds ratio: compares the odds of an event or outcome occurring in one group with the odds of it occurring in another	τ	Kendall's tau: like rho, used when both variables are measured on an ordinal or ranked scale; tau-b is used for square-shaped tables; tau-c is used for rectangular tables
(b) Groups compared on continuous outcomes		r_{pb}	The point-biserial correlation coefficient: used when one variable (the predictor) is measured on a binary scale and the other variable is continuous
d	Cohen's d: the uncorrected standardized mean difference between two groups based on the pooled standard deviation	φ	The phi coefficient: used when variables and effects can be arranged in a 2×2 contingency table
Δ	Glass's delta (or d): the uncorrected standardized mean difference between two groups based on the standard deviation of the control group	C	Pearson's contingency coefficient: used when variables and effects can be arranged in a contingency table of any size
g	Hedges' g: the corrected standardized mean difference between two groups based on the pooled, weighted standard deviation	V	Cramér's V: like C, V is an adjusted version of phi that can be used for tables of any size
PS	Probability of superiority: the probability that a random value from one group will be greater than a random value drawn from another	λ	Goodman and Kruskal's lambda: used when both variables are measured on nominal (or categorical) scales

(cont.)

Table 1.1 (*cont.*)

Measures of group differences (the *d* family)	Measures of association (the *r* family)	
	(b) Proportion of variance indexes	
	r^2	The coefficient of determination: used in bivariate regression analysis
	R^2	R squared, or the (uncorrected) coefficient of multiple determination: commonly used in multiple regression analysis
	$_{adj}R^2$	Adjusted R squared, or the coefficient of multiple determination adjusted for sample size and the number of predictor variables
	f	Cohen's f: quantifies the dispersion of means in three or more groups; commonly used in ANOVA
	f^2	Cohen's f squared: an alternative to R^2 in multiple regression analysis and ΔR^2 in hierarchical regression analysis
	η^2	Eta squared or the (uncorrected) correlation ratio: commonly used in ANOVA
	ε^2	Epsilon squared: an unbiased alternative to η^2
	ω^2	Omega squared: an unbiased alternative to η^2
	R^2_C	The squared canonical correlation coefficient: used for canonical correlation analysis

General logic behind ES

$$\eta^2 = \frac{SS_{Effect}}{SS_T},$$

$$\eta_p^2 = \frac{SS_{Effect}}{SS_{Effect} + SS_{s/Cells}},$$

$$\hat{\omega}_p^2 = \frac{SS_{Effect} - df_{Effect} MS_{s/Cells}}{SS_{Effect} + (N - df_{Effect})\, MS_{s/Cells}}$$

$$d = \frac{M_1 - M_2}{pooled\ SD}$$

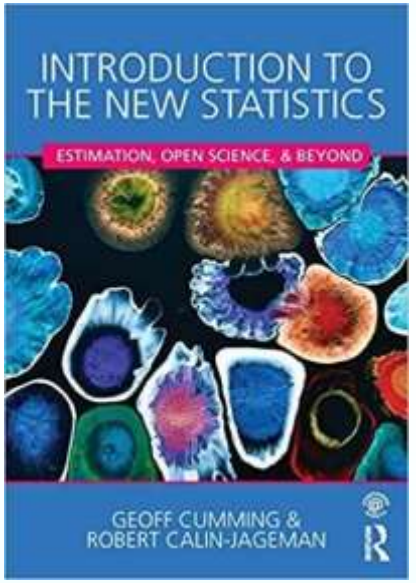
$$r = \frac{Covariance\ (x,y)}{S.D.(x)S.D.(y)}$$

$$r = \frac{n\Sigma xy - (\Sigma x)(\Sigma y)}{\sqrt{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]}}$$

- Effect sizes go up when "**signal**" (*numerator*) increases relative to "**noise**" (*denominator*)

Errors of inference

- Frequentist approach
- There are three types of errors
- NHST*: Type I error (False positives)
  Type II error (False negatives)
- CI (aka "The New Statistics":
  Estimate error (imprecision)



NHST= Null Hypothesis Significance Testing
(what you have been taught as a student)
H0 vs. H1

Errors of inference in NHST

Real World (**POPULATION**)

Conclusion of the significance test (**SAMPLE**)	Null is true (H0 is correct)	Null is false (H1 is correct)
Null is true	Correct decision (1-α)	Type II error (β)
Null is false	Type I error (α)	Correct decision (1-β)

Errors of inference in NHST

- **Type I error**: ***Erroneously rejecting the null hypothesis (<u>False positive</u>)***.
  The result in the sample is significant ($p < .05$), so the null hypothesis is rejected, but the null hypothesis is actually true in the population.

- **Type II error**: ***Erroneously accepting the null hypothesis (<u>False negative</u>)***. The result in the sample is not significant ($p > .05$), so the null hypothesis is not rejected, but it is actually false in the population.

How to control Type I errors?

- The Type I error rate (***False positive***) is controlled by the researcher.

- It is called the **alpha rate** and corresponds to the probability cut-off that one uses in a significance test (*p value threshold*).

- Conventionally, researchers use an alpha rate (α) of .05. This means that the null hypothesis is rejected when a value such as the one found is likely to occur 5% of the time or less when the null hypothesis is true.

- The test can be two-tailed (more common) or one-tailed (directional)

One-tailed and two-tailed test

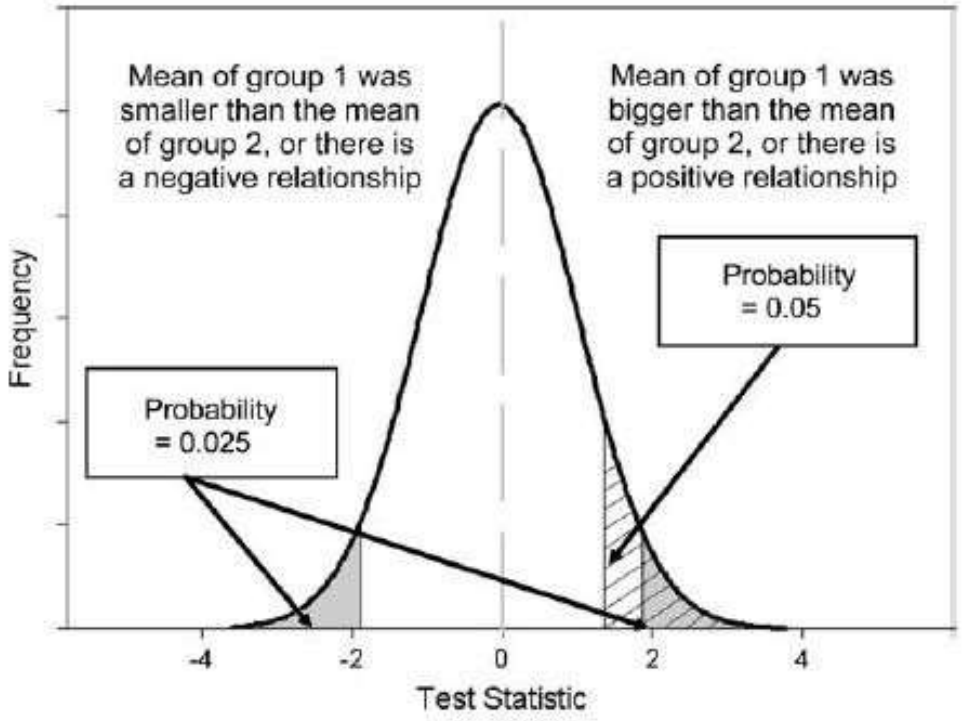


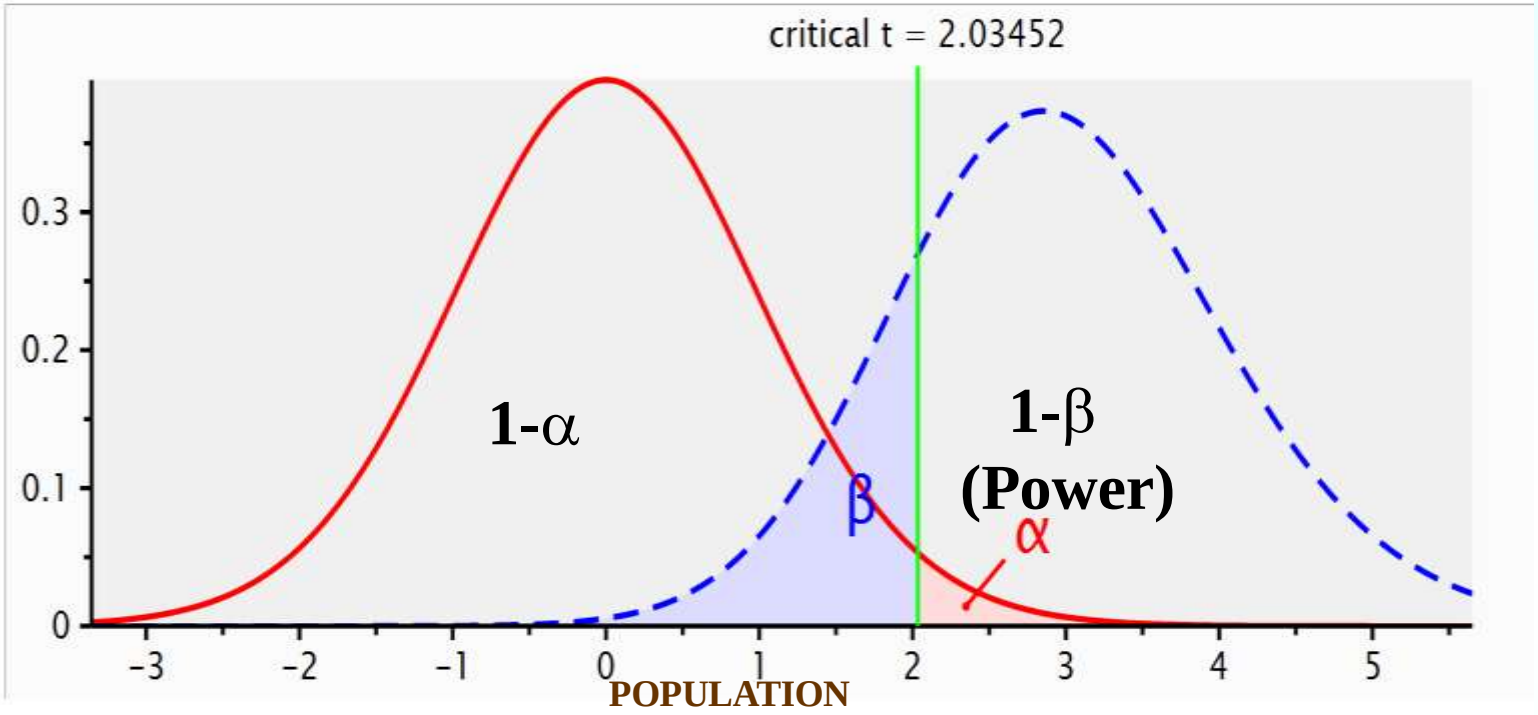
FIGURE 2.10
Diagram to show the difference between one- and two-tailed tests

How to control Type II errors?

- The Type II error (***False negative***) can also be controlled by the experimenter.
- The Type II error rate is called **beta** (β) as a complement to alpha.
- How can the beta rate be controlled? The easiest way to control Type II errors is by increase the **statistical power** of a test.
- **Statistical power**= probability of finding an effect, if it exists
- **Power** = **1** – β
- Conventionally a power of at least .80 (β=.20) is considered as acceptable

Power Analysis

What is power?



SAMPLE	Null is true (H0 is correct)	Null is false (H1 is correct)
Null is true	Correct decision (1-α)	Type II error (β)
Null is false	Type I error (α)	Correct decision (1-β)

The key determinants of power

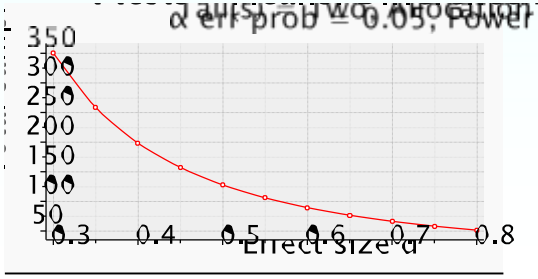
- Power is determined by four elements

1) Decision criterion (α)
2) Sample size (n)
3) Effect size (δ)
4) Desired power ($1-\beta$)

- Fixing one of the elements one can derive the others

A simple example

- Fix α=.05 and (1- β)=.80
- Plot sample size and effect size for a two sample t-test



What affects power?

- Power goes up with larger effect sizes and sample sizes, given a certain decision criterion (e.g., α=.05)

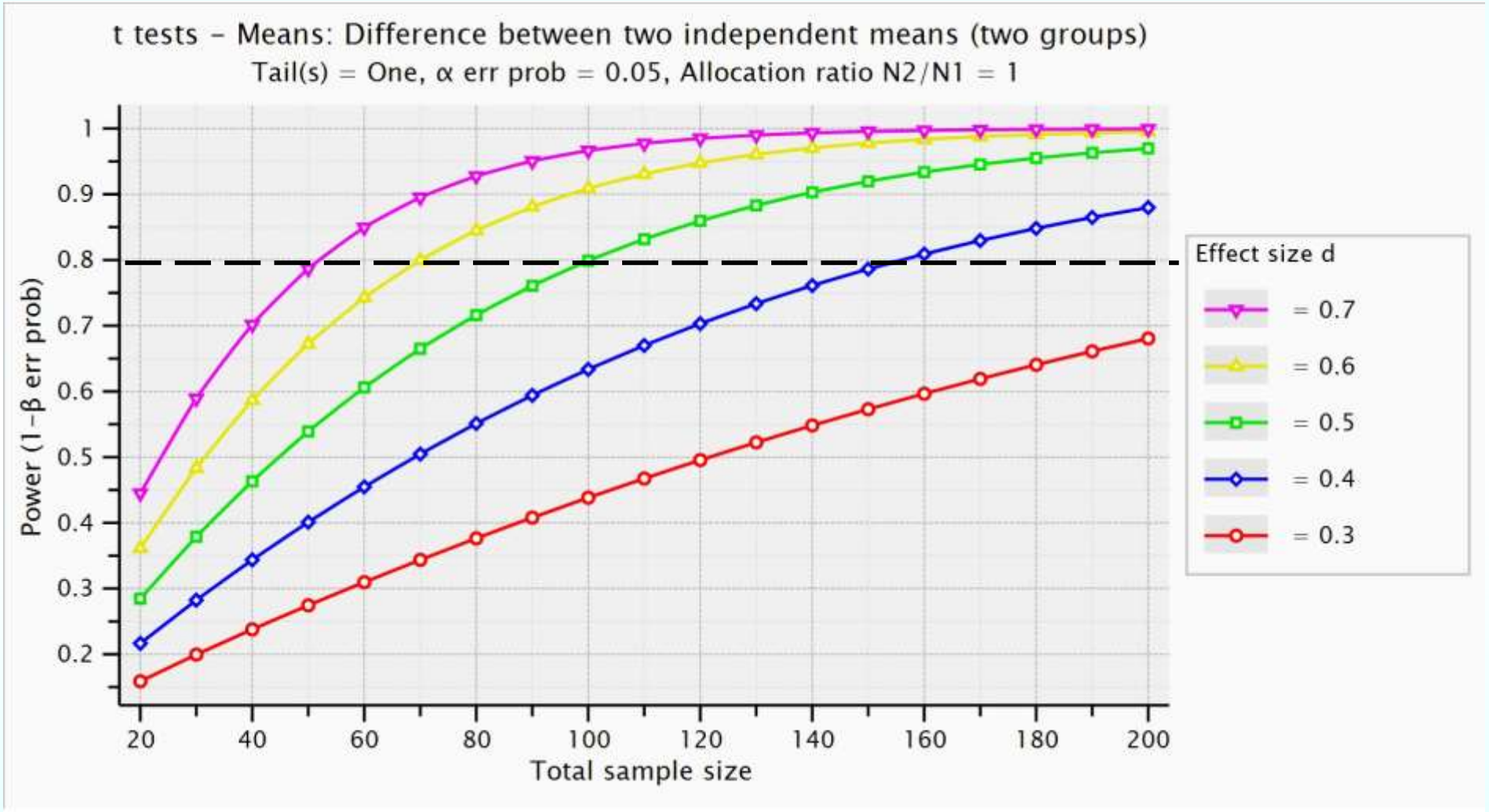
- When effect sizes become larger? When the portion of variability (difference) ascribed to the effect of interest grows more than the general (non specific) variability

$$d = \frac{M_1 - M_2}{pooled\ SD}$$

$$\hat{\eta}^2 = \frac{SS_{Effect}}{SS_T}$$

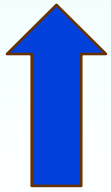
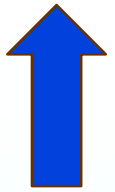
$$r(v,x) = \frac{\mathrm{cov}(v,x)}{sd(v) * sd(x)}$$

Power as a function of ES and N



How to increase power?

Power is affected by

- ***Sample size*** 
- ***Construct-related (i.e., SIGNAL) variance*** 
- ***Construct-unrelated (i.e., NOISE) variance*** 

What is affected by power?

Higher power means

- *Less False Negatives*
- *Lower overall errors of inference (crucial error rates)*

Lower power means

- *with multiple outcomes and HARKing: body of conflicting evidence in the literature*
- *with publication bias: presence of many false-positives in the literature*

Why power analysis to plan studies?

- Without logistical constraints (infinite resources and no costs), only accuracy in estimating parameters should matter (e.g., AIPE, Maxwell, 2008 Ann Rew Psych)
- In an accuracy (precision) approach, one thing matters a lot: sample size, the bigger, the better (*ceteris paribus*)
- The point is not whether some effect exists (or not) but how precise is our estimate of it
- All effects exist given an infinite sample size (Cohen)
- Increased accuracy means less inference errors (both Type I and Type II)
- ***If you want to get it right, increase sample size***

Precision vs. Power

- They have different aims

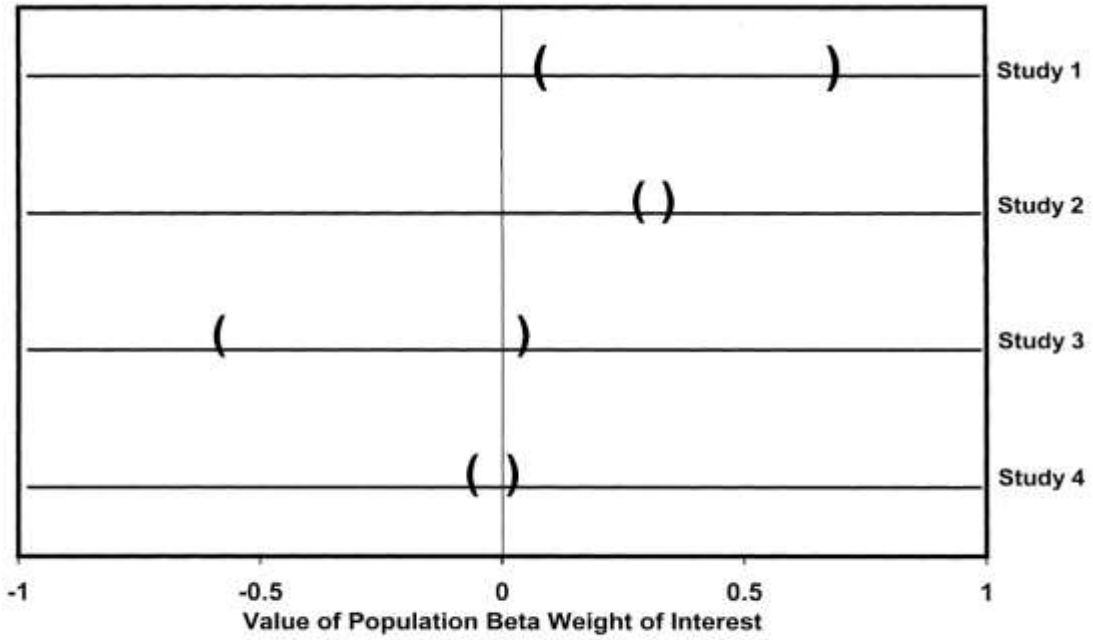
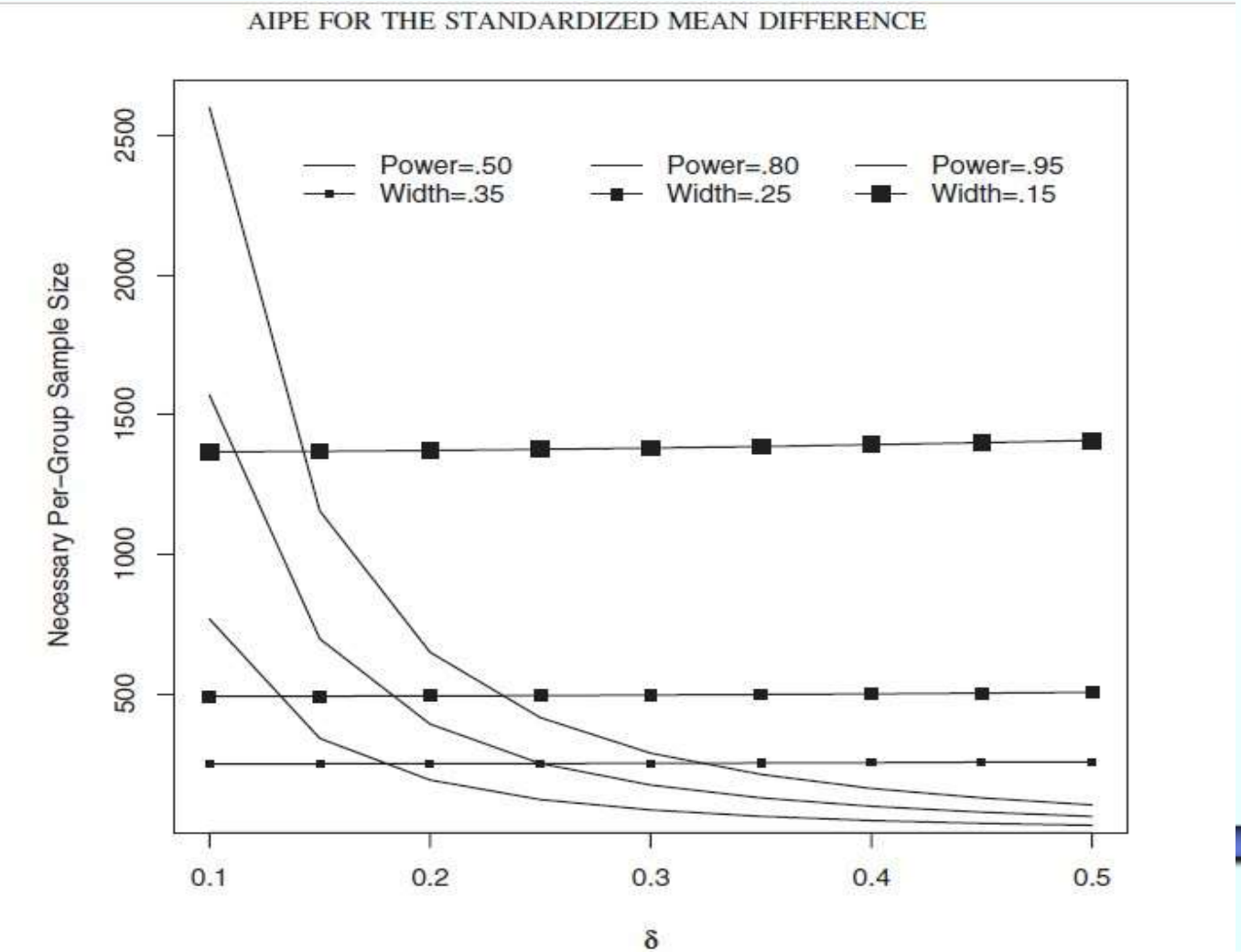


Figure 1. Illustration of possible scenarios in which planned sample size was considered a "success" or "failure" according to the accuracy in parameter estimation and the power analysis frameworks. Parentheses are used to indicate the width of the confidence interval.

- Precision is valuable no matter everything else

a MINOR practical problem…

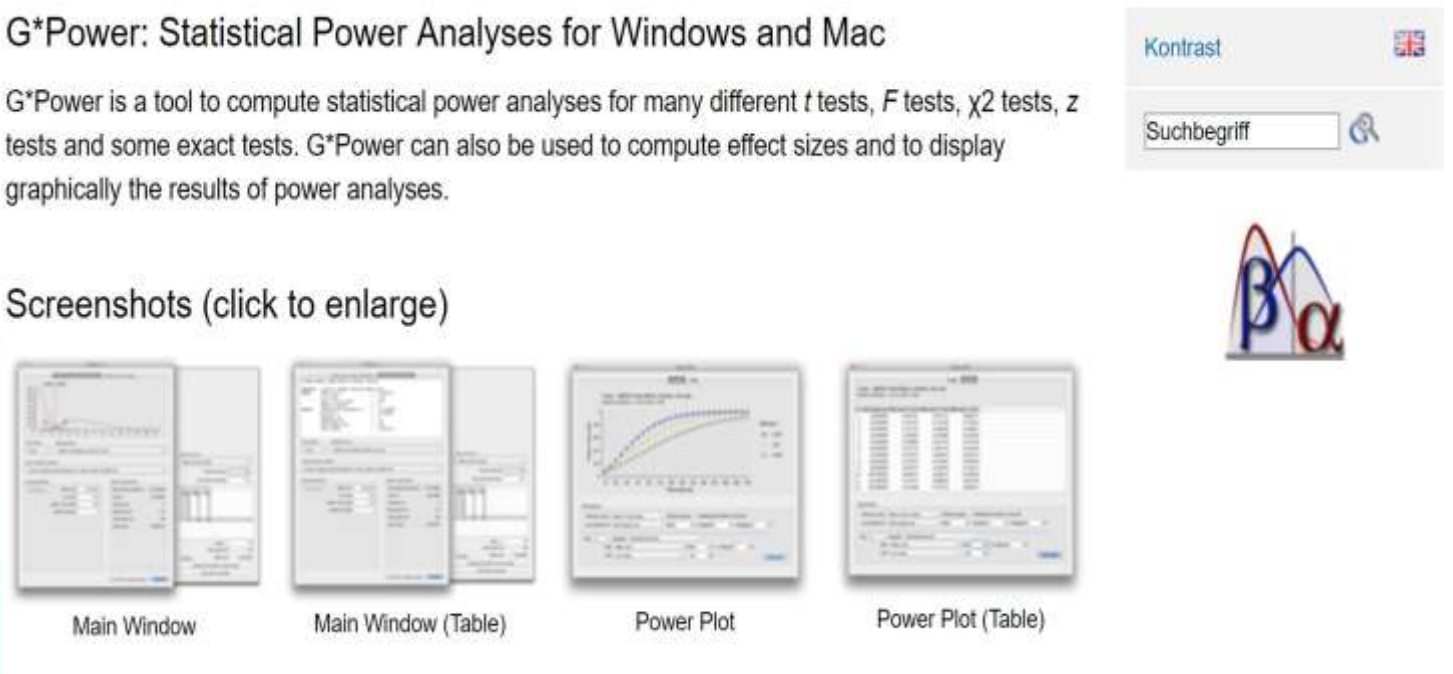
- **Big** sample sizes are needed for precise estimates no matter the effect size



How to calculate power

- Different software and routines (e.g., in R)
- A free comprehensive package is G*Power

http://www.gpower.hhu.de/



Power analysis calculations

- Examples of calculation of power analysis for some simple designs
- Also based on

Perugini, M., et al. (2018). A Practical Primer To Power Analysis for Simple Experimental Designs. *International Review of Social Psychology*, 31(1): 20, 1–23, DOI: https://doi.org/10.5334/irsp.181

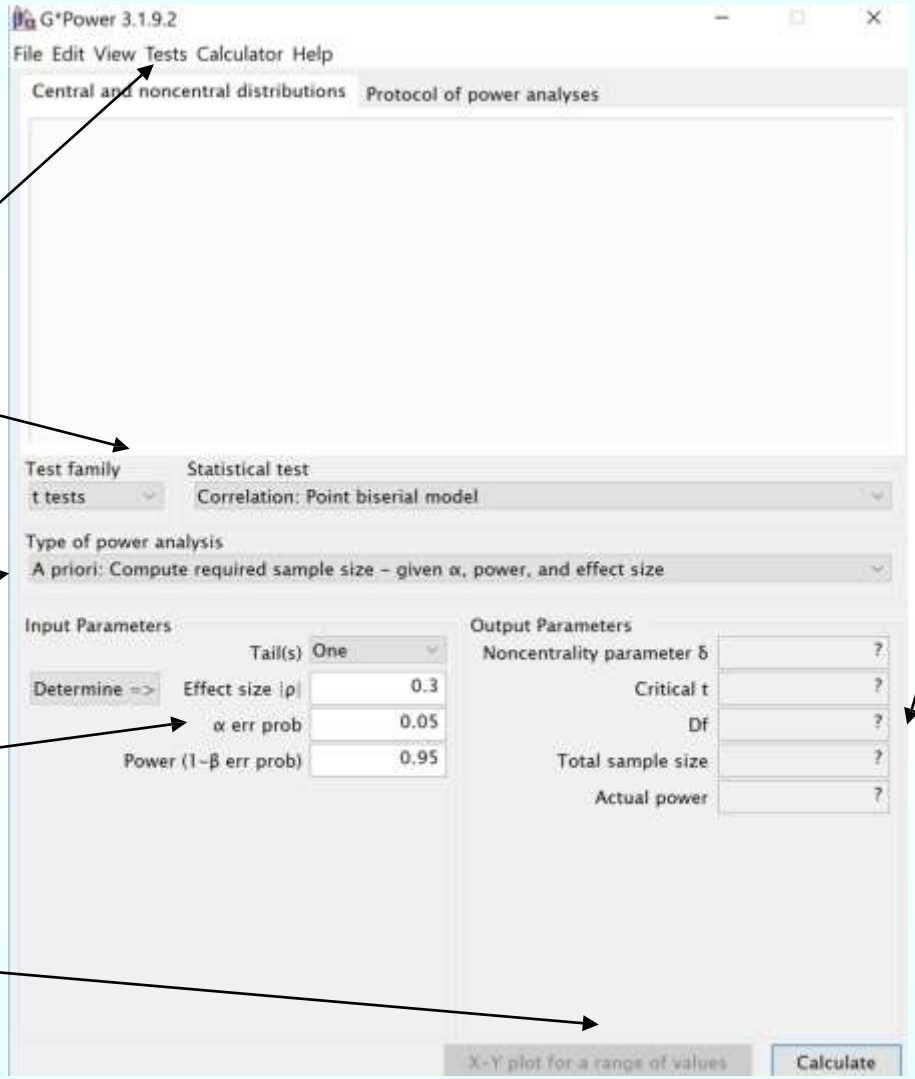
RESEARCH ARTICLE

A Practical Primer To Power Analysis for Simple Experimental Designs

Marco Perugini, Marcello Gallucci and Giulio Costantini

SM and routines available at https://github.com/mcfanda/primerPowerIRSP

G*Power



Analysis

Type of power analysis

Inputs

Action buttons

Output

Example: Two independent groups

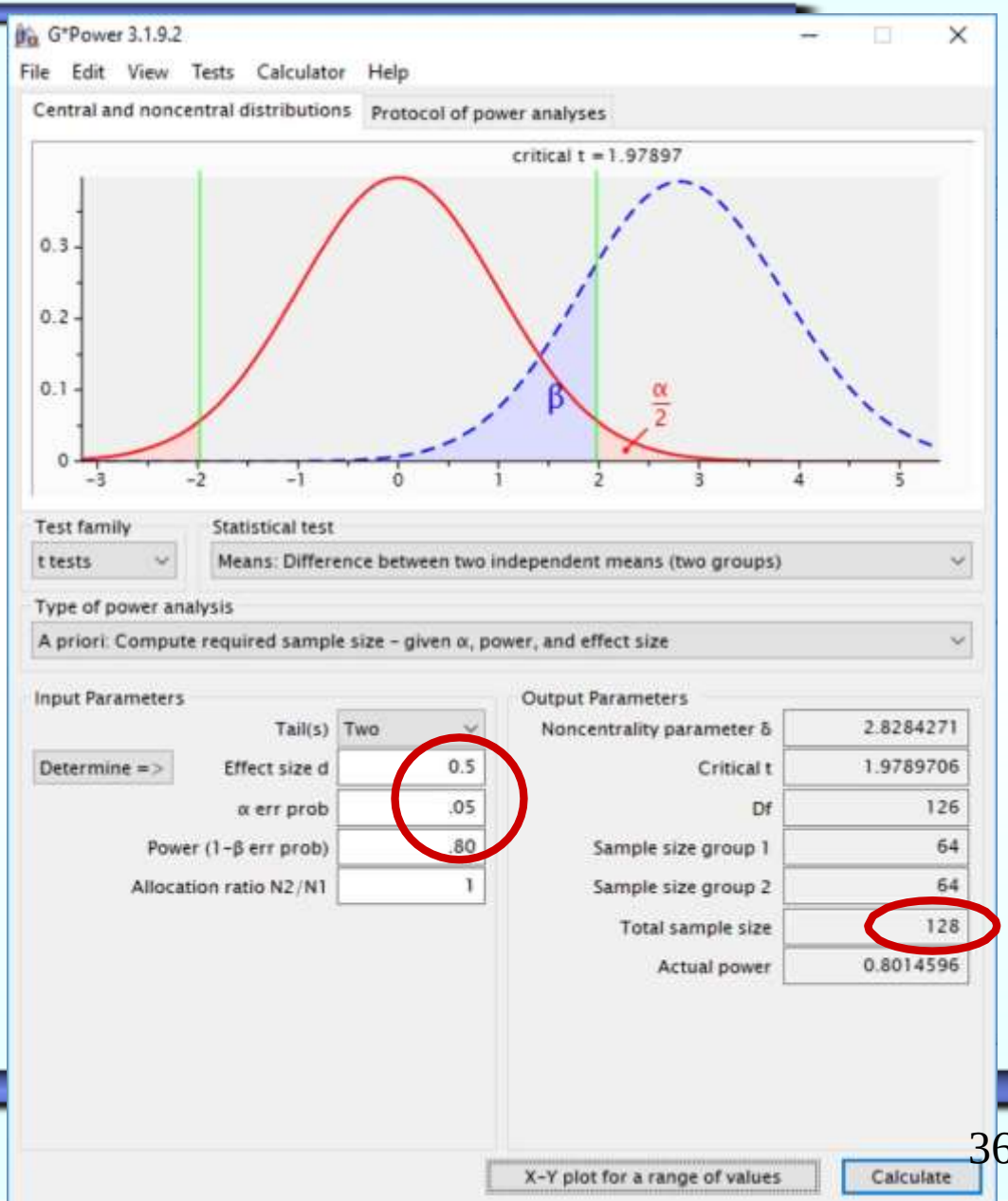
Standard pre-study planning approach

Fix ES, α, 1-β

Calculate needed N

Good practice

We consider also sensitivity analysis



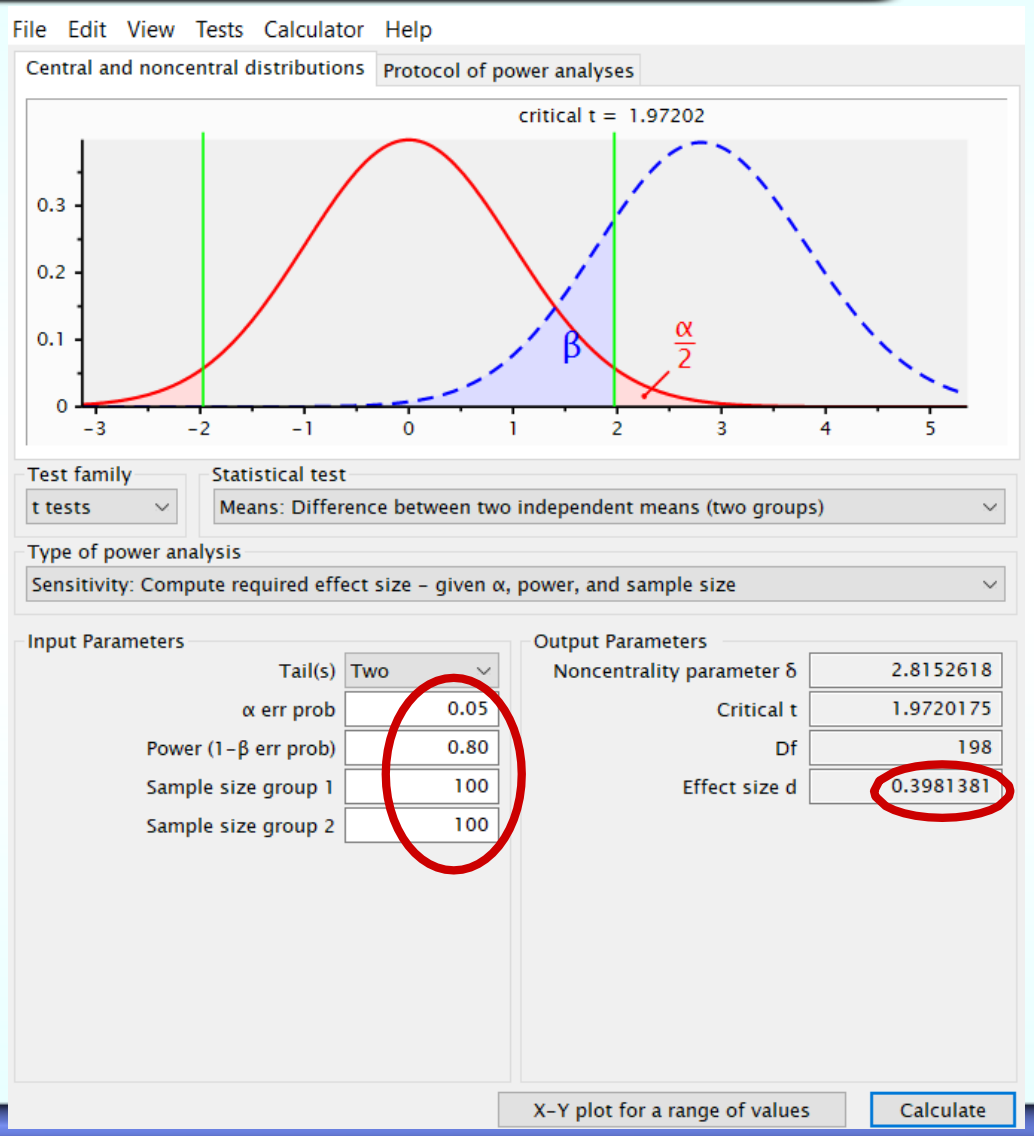
Sensitivity analysis: Starting from N

"Sometimes"
resources are fixed

You know that you can collect a certain N

The question becomes what ES can be found with sufficient power

Sensitivity analysis



Sensitivity plot: N by ES

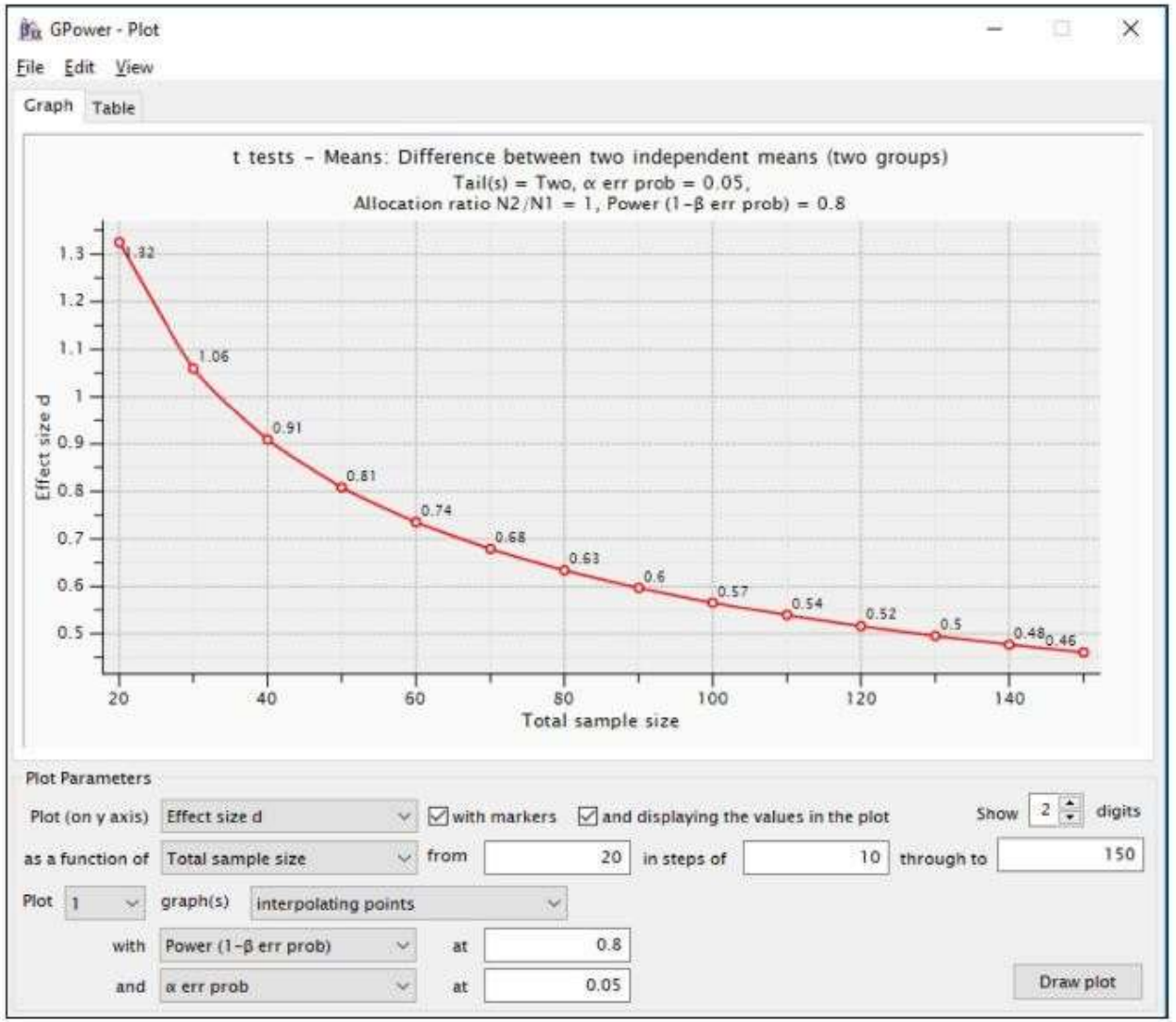


Figure 2: Sensitivity Plot of G*Power calculating the power of a two independent samples t-test: Lowest detectable effect size as a function of required *N*.

Sensitivity plot: N by Power

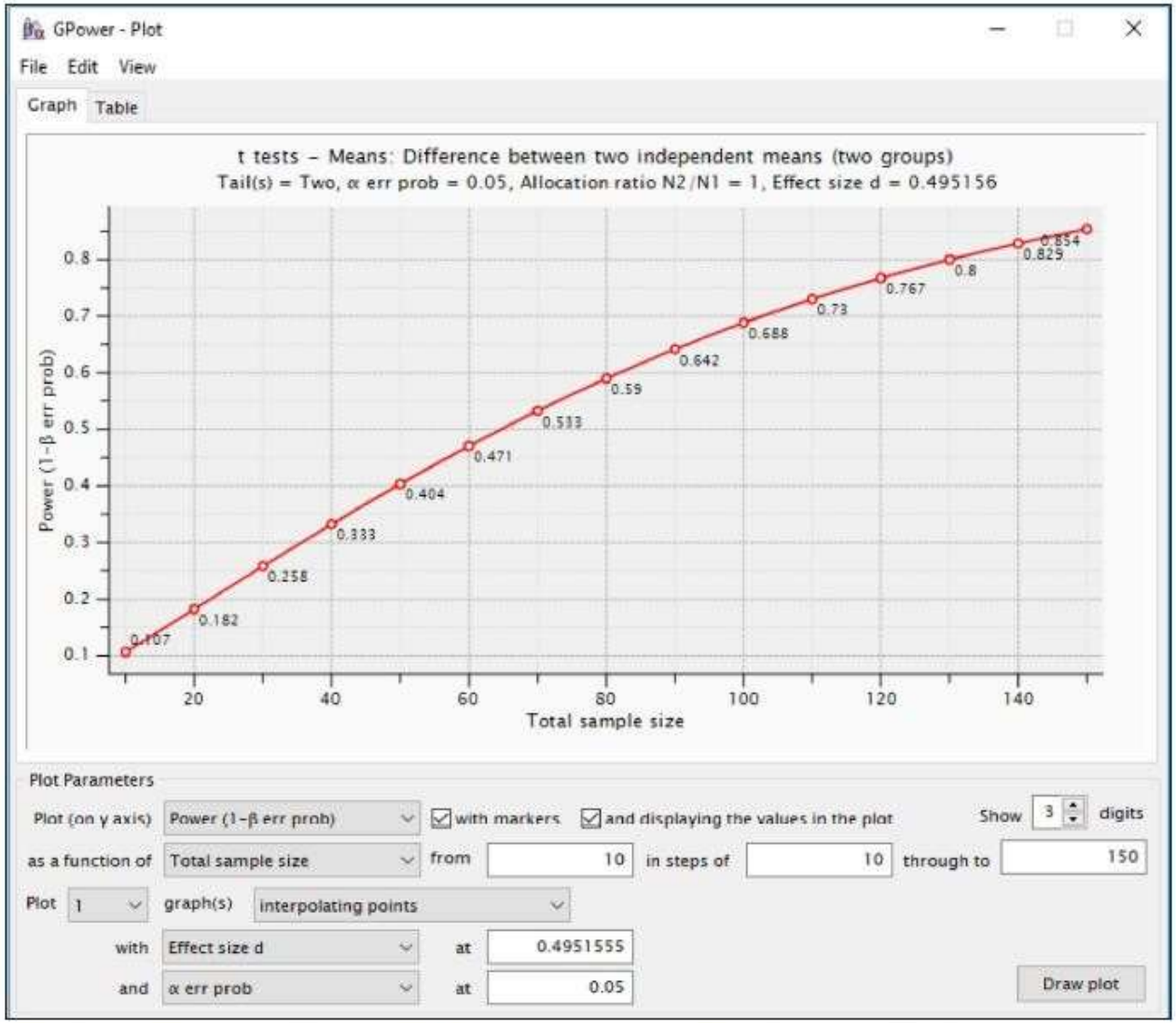
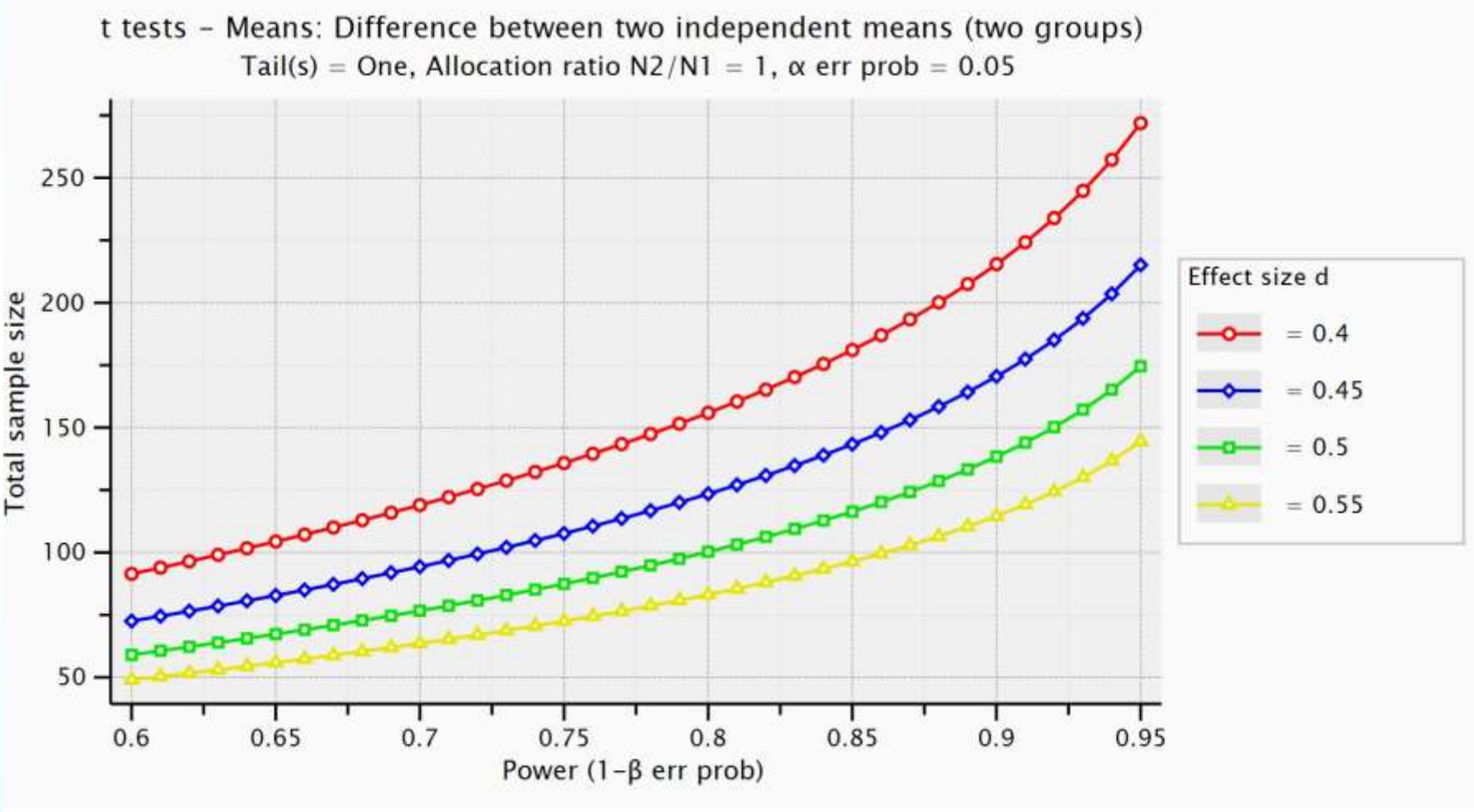
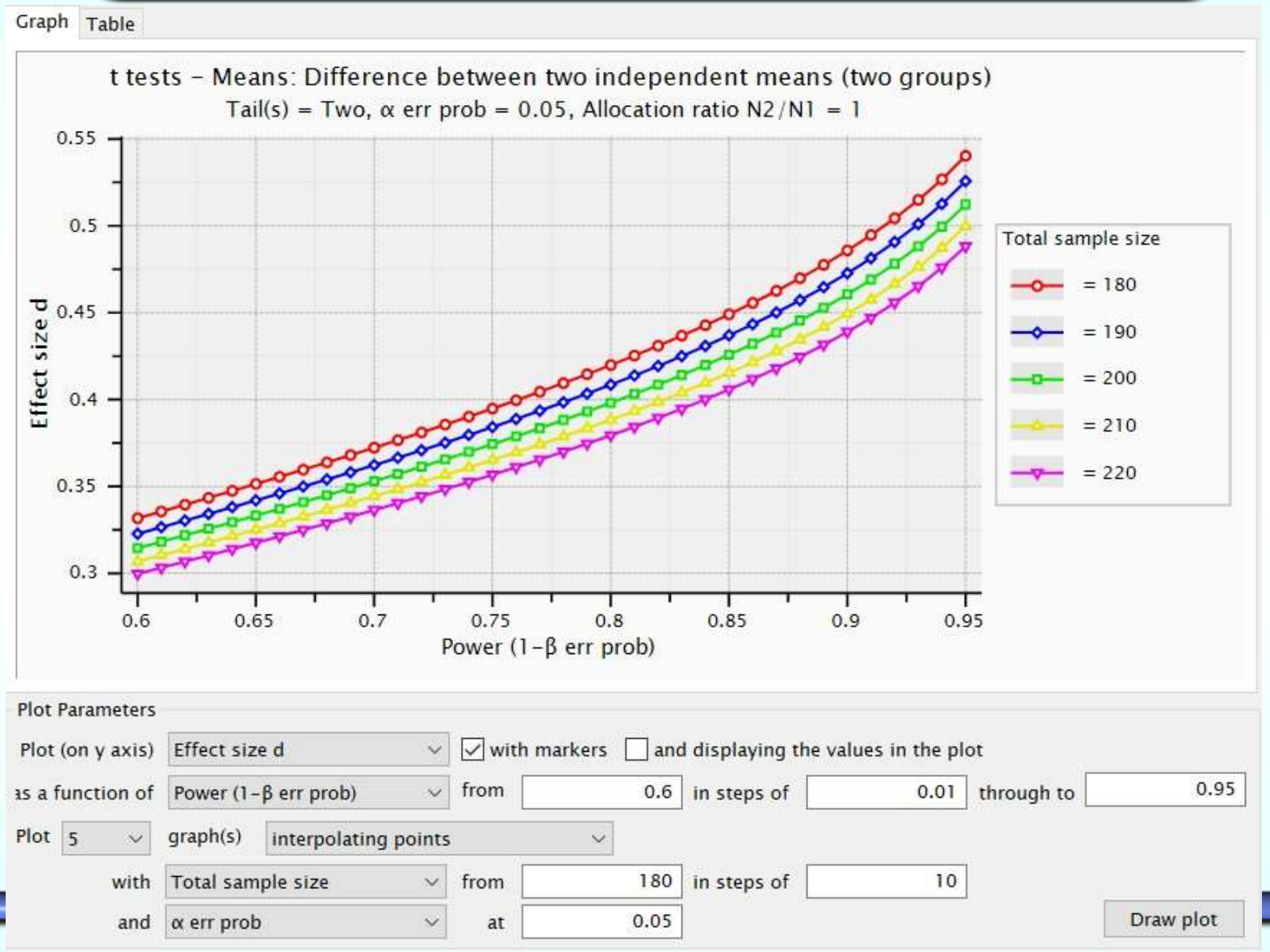


Figure 3: Sensitivity Plot of G*Power calculating the power of a two independent samples t-test: Power as a function of required N for fixed effect size.

Inspecting scenarios around effect sizes



Inspecting scenarios around N

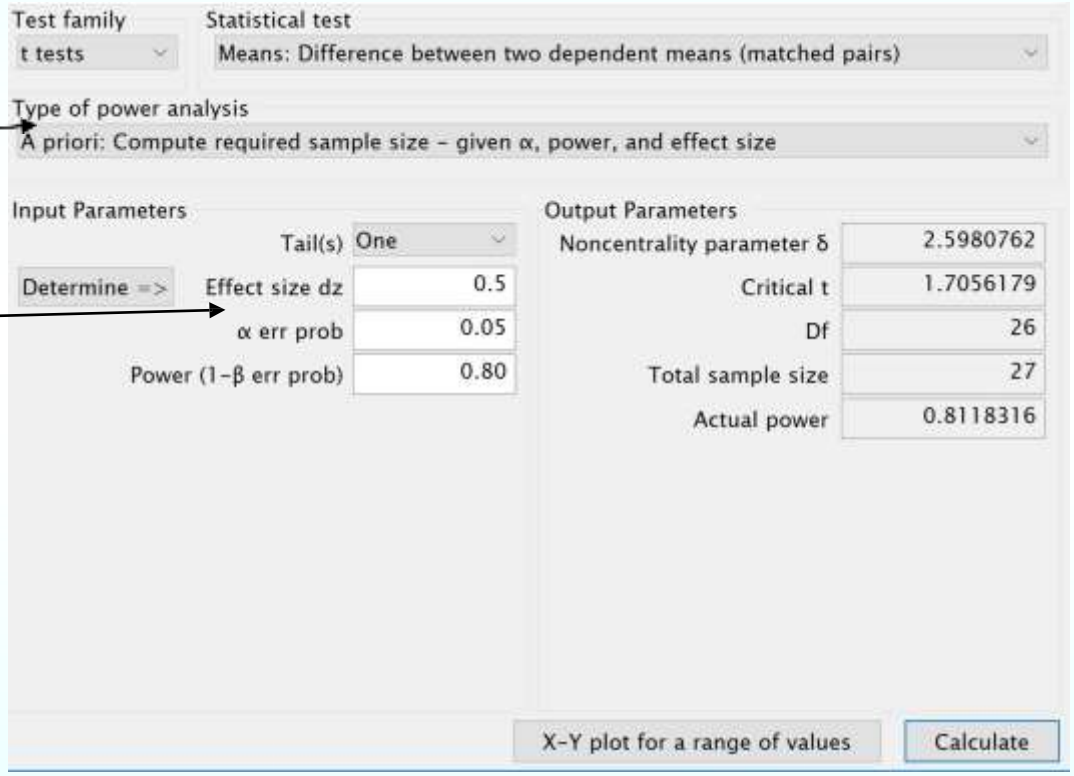


Two repeated measures

Analysis

Type of power analysis

Inputs



Output

Two repeated measures

The ES for a paired means design is $d_z = \Delta / sd.$

This is not the same as Cohen's d, but calculated with the difference score divided by its SD. To double check with available previous results $d_z = \frac{t}{\sqrt{N}}$

If no previous results, could guess ES as if the two are independent groups and how much the measures are correlated (r/ρ)

$\boldsymbol{d_z} = \frac{\boldsymbol{d}}{\sqrt{\boldsymbol{2(1-r)}}}$, e.g., with d=0.5 and r=.55, $d_z = \frac{0.5}{\sqrt{2(1-0.55)}} = 0.527$

and the other way round

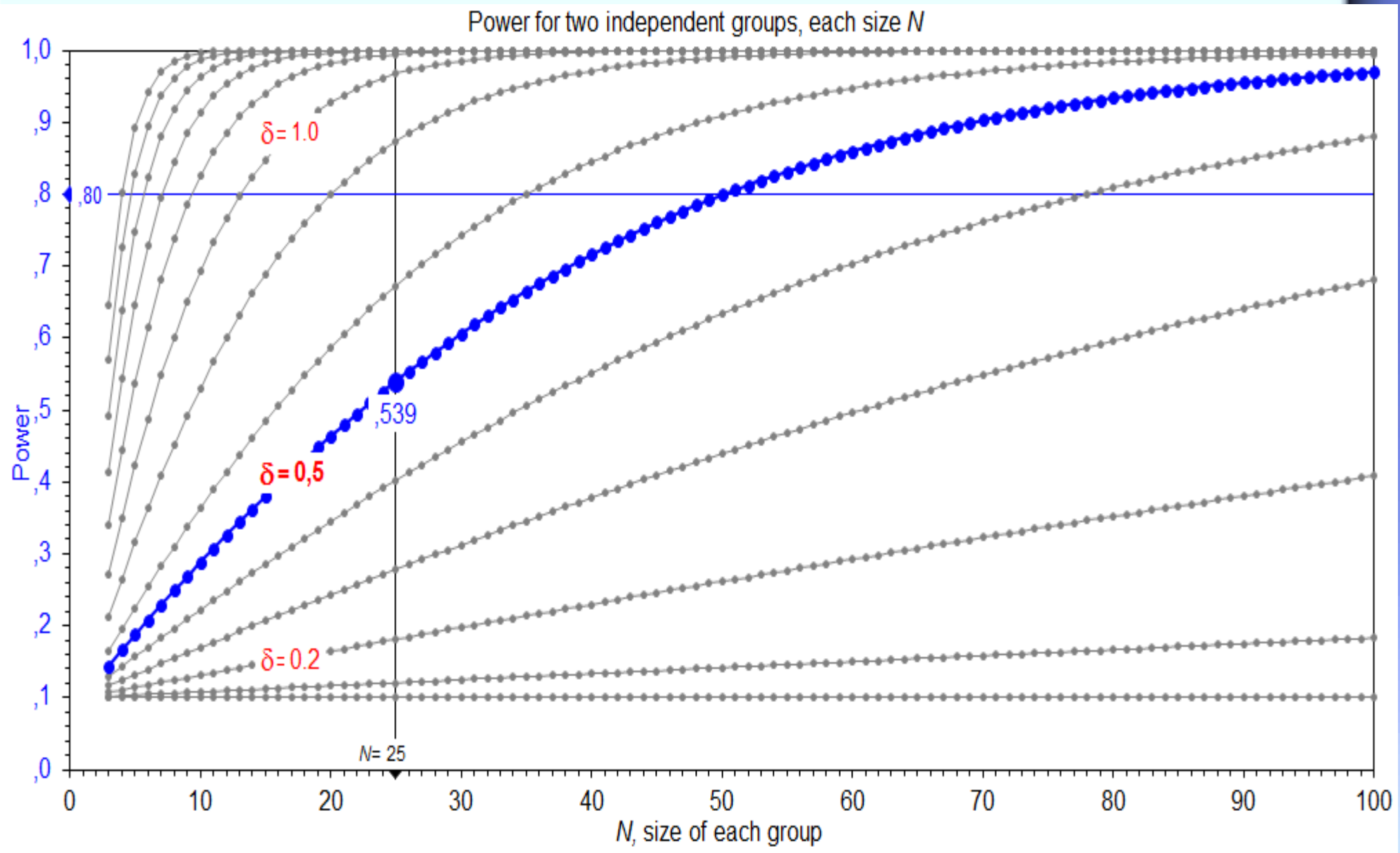
$\mathbf{d} = \boldsymbol{d_z} * \sqrt{\boldsymbol{2(1-\rho)}}$, e.g. with d_z=0.527 and r=.55,

$$d = 0.527 * \sqrt{2(1-.55)} = 0.527 * 0.95 = 0.50$$

Between vs. Within

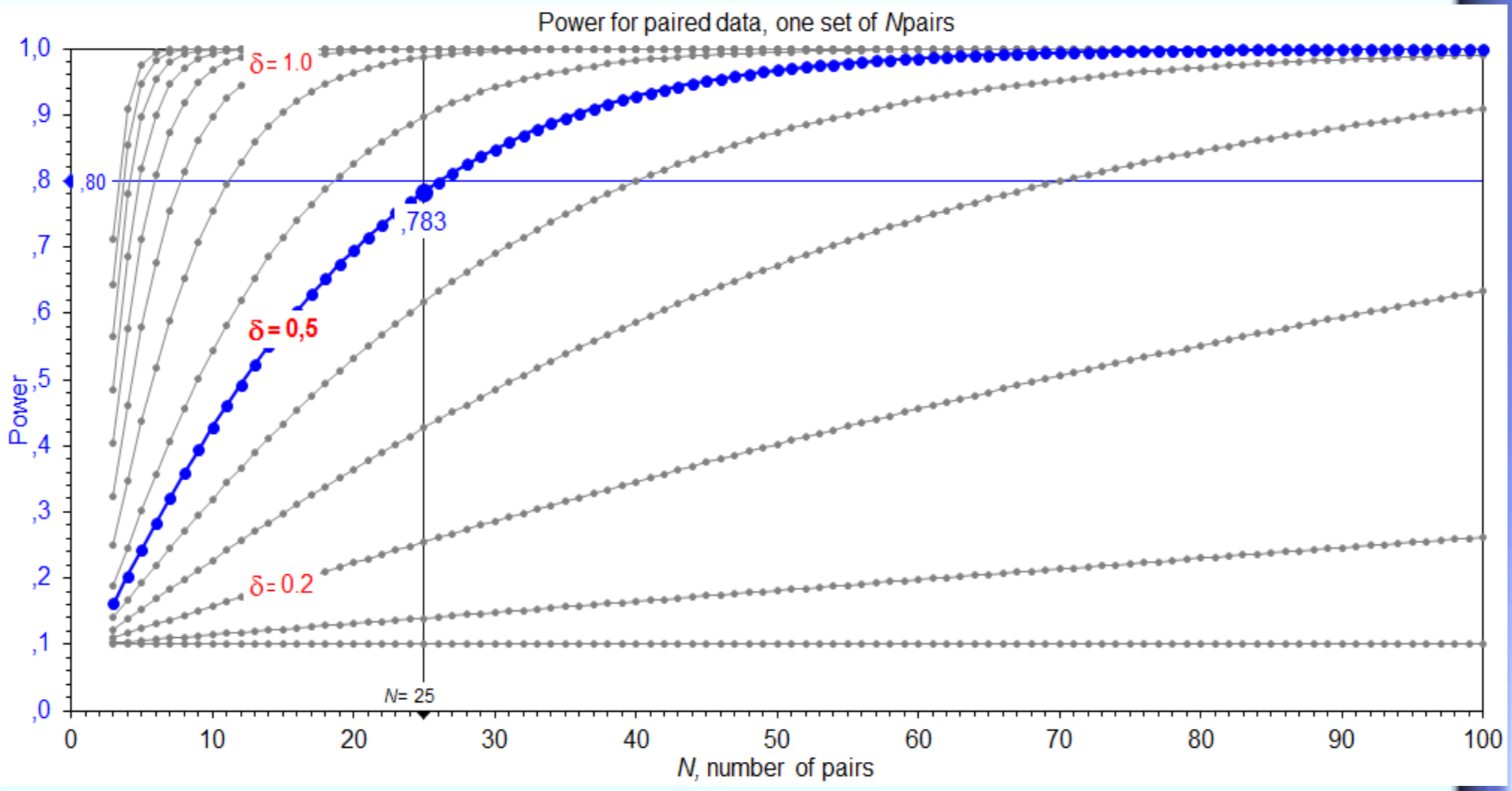
- Everything else being equal, within studies are more powerful than between studies
- Example with a simple two groups/two measures design
- Later on more articulated examples

Power Between Ss

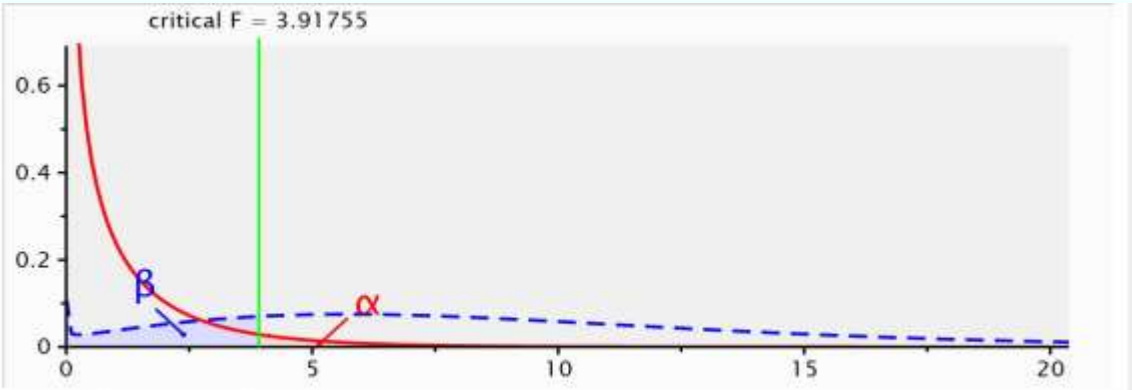


Power Within Ss

- Power for within Ss studies is greater (*ceteris paribus*) but depends also on r (e.g., r = .50) between DVs



ANOVA 2 x 2



Analysis

Test family: F tests

Statistical test: ANOVA: Fixed effects, special, main effects and interactions

Type of power analysis: A priori: Compute required sample size – given α, power, and effect size

Inputs

Input Parameters

Determine => Effect size f	0.25
α err prob	0.05
Power (1–β err prob)	0.80
Numerator df	1
Number of groups	4

Output Parameters

Noncentrality parameter λ	8.0000000
Critical F	3.9175498
Denominator df	124
Total sample size	128
Actual power	0.8013621

Output

Correspondence between some ES

Effect Size	d	0.20	Effect Size	d	0.50	Effect Size	d	0.80
d	0.2		d	0.5		d	0.8	
r	0.0995		r	0.2425		r	0.3714	
η^2	0.0099		η^2	0.0588		η^2	0.1379	
f	0.1		f	0.25		f	0.4	
Odds Ratio	1.4373		Odds Ratio	2.4766		Odds Ratio	4.2675	

https://www.psychometrica.de/effect_size.html#transform

Contrast approach (from Means)

More complex designs can be sometimes simplified with a focused contrast approach

The key point is to be explicit about the focal hypothesis and the pattern of expected means and to use contrast weights that reflect the focal hypothesis

CONTRAST WEIGHTS

	A1	A2
B1	1	-1
B2	1	-1

MAIN EFFECT (A) {1, -1, 1, -1}

CONTRAST WEIGHTS

	A1	A2
B1	1	1
B2	-1	-1

MAIN EFFECT (B) {1, 1, -1, -1}

CONTRAST WEIGHTS

	A1	A2
B1	1	-1
B2	-1	1

INTERACTION (AxB) {1, -1, -1, 1}

Contrast approach

Suppose you expect this pattern of means

Table 2: Example of 2 × 2 design.

	A1	A2
B1	10	0
B2	0	0

Note: Pooled standard deviation is equal to 5.

Calculate ***f*** for main effects and interactions, after coding

$$f = \frac{\left|\sum c_i \cdot \mu_i\right|}{\sqrt{k \cdot \sum c_i^2 \cdot \sigma^2}}$$

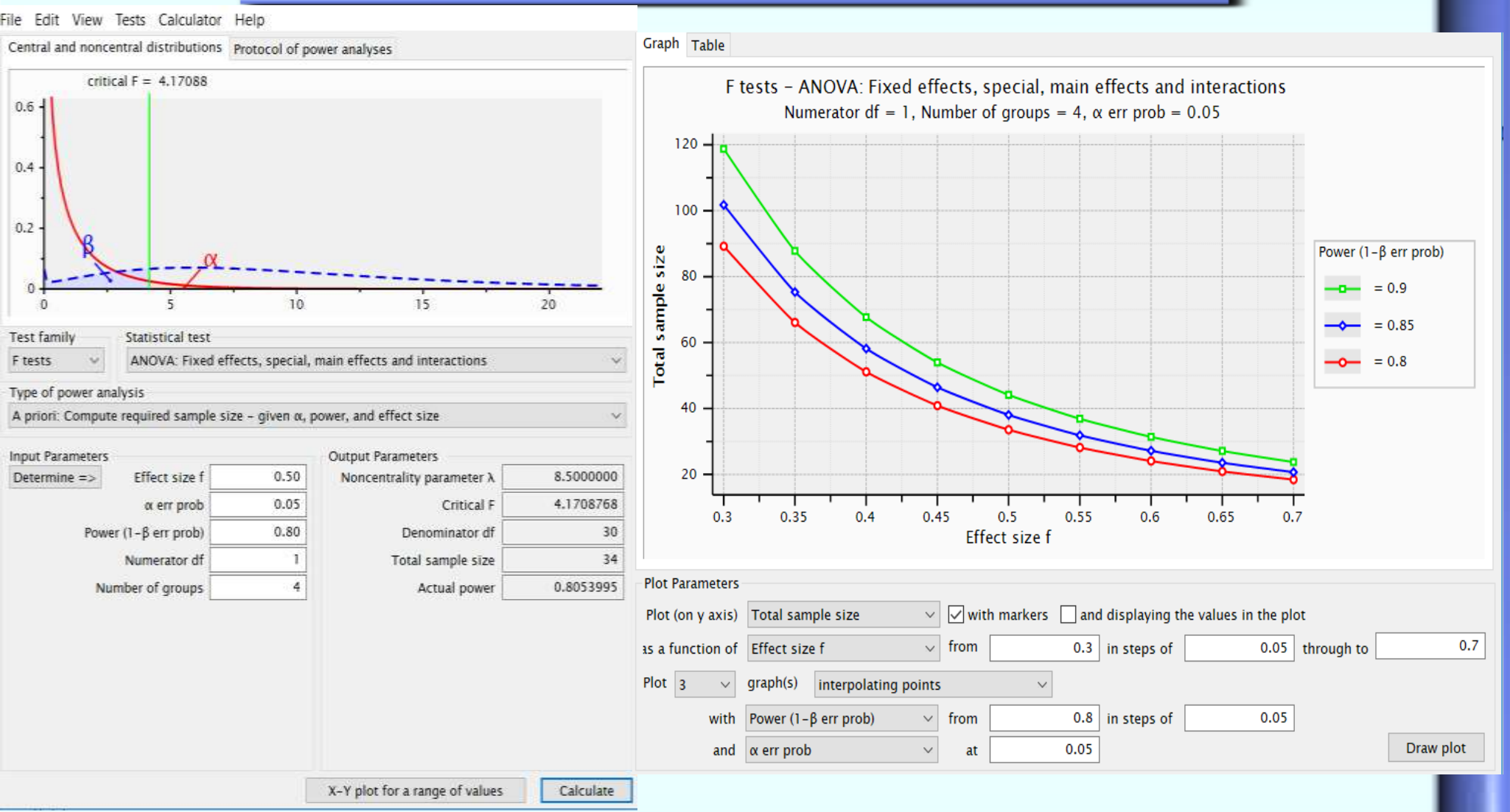
WEIGHTS (c_i)		
	1	1
	-1	-1

$$f = \frac{10}{\sqrt{4 \cdot 4 \cdot 25}} = .50$$

Might be useful to recall that $f = \frac{d}{2}$ or $d = 2f$

Tip: if it helps, you can think with standardized means (e.g., SD=1, transform EMs in standardized expected differences, as for Cohens d)

Plugging it in G*Power



Planning a “moderation” study

A common research scenario

Study 1: Main effect

Study 2: Test of a moderator of the main effect

Table 3: Example of 2 × 2 design expected results.

		Case 1		Case 2	
		A1	A2	A1	A2
B	replicated	5	2	5	2
	moderated	0	0	2	5

Note: Pooled standard deviation is equal to 1.

Key point: Power calculations will change depending on design and type of expected moderation

Using a contrast approach

Case 1: Expect to replicate main effect and moderator suppresses it

Table 3: Example of 2 × 2 design expected results.

		Case 1	
		A1	A2
B	replicated	5	2
	moderated	0	0

Note: Pooled standard deviation is equal to 1.

Power calculation for main effect in original study (2 conditions, A1 vs. A2)

$$f = \frac{\left|\sum c_i \cdot \mu_i\right|}{\sqrt{k \cdot \sum c_i^2 \cdot \sigma^2}}$$

$$C_A = (1) \cdot 5 + (-1) \cdot 2 = 3$$

$$f = \frac{3}{\sqrt{2 \cdot 2 \cdot 1}} = 1.50$$

Using a contrast approach

Case 1: Expect to replicate main effect and moderator suppresses it

Table 3: Example of 2 × 2 design expected results.

		Case 1	
		A1	A2
B	replicated	5	2
	moderated	0	0

Note: Pooled standard deviation is equal to 1.

Power calculation for interaction effect

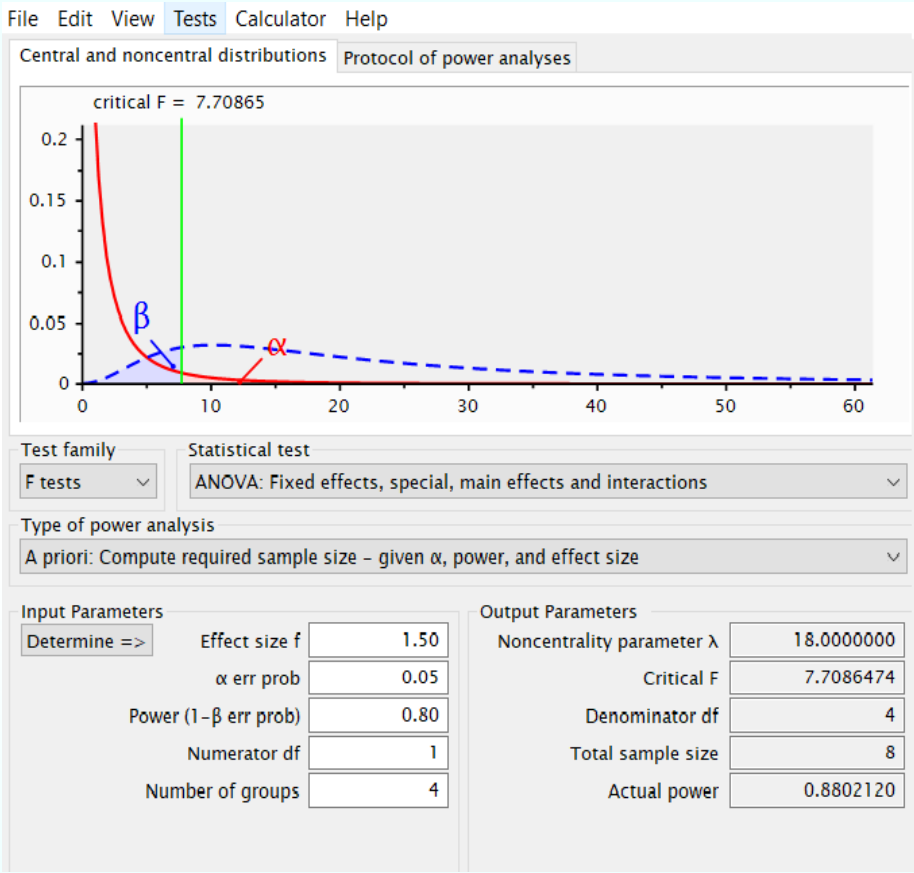
$$f = \frac{\left|\sum c_i \cdot \mu_i\right|}{\sqrt{k \cdot \sum c_i^2 \cdot \sigma^2}}$$

$$C_{AB} = (1) \cdot 5 + (-1) \cdot 2 + (-1) \cdot 0 + (1) \cdot 0 = 3$$

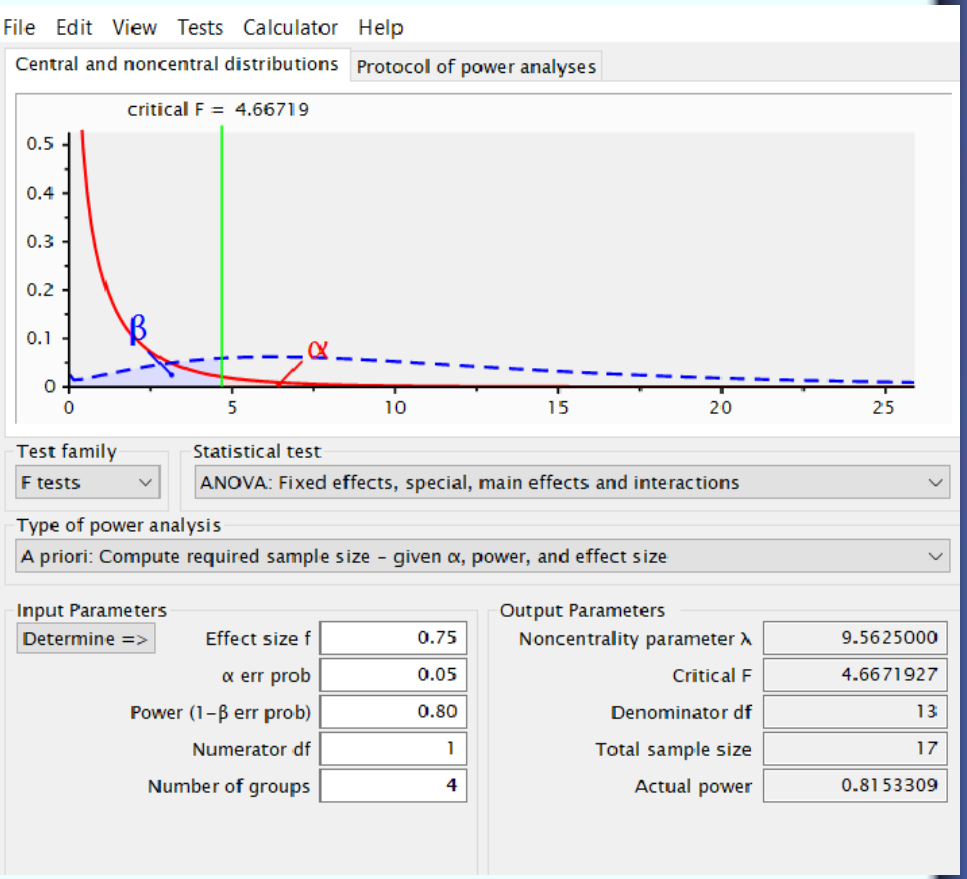
$$f = \frac{3}{\sqrt{4 \cdot 4 \cdot 1}} = 0.75$$

Plugging it in G*Power

Main effect



Moderated (suppression) effect



Needed sample size is about doubled

Using a contrast approach

Case 2: Expect to replicate main effect and moderator reverts it

Table 3: Example of 2 × 2 design expected results.

		Case 2	
		A1	A2
B	replicated	5	2
	moderated	2	5

Note: Pooled standard deviation is equal to 1.

Power calculation for interaction effect

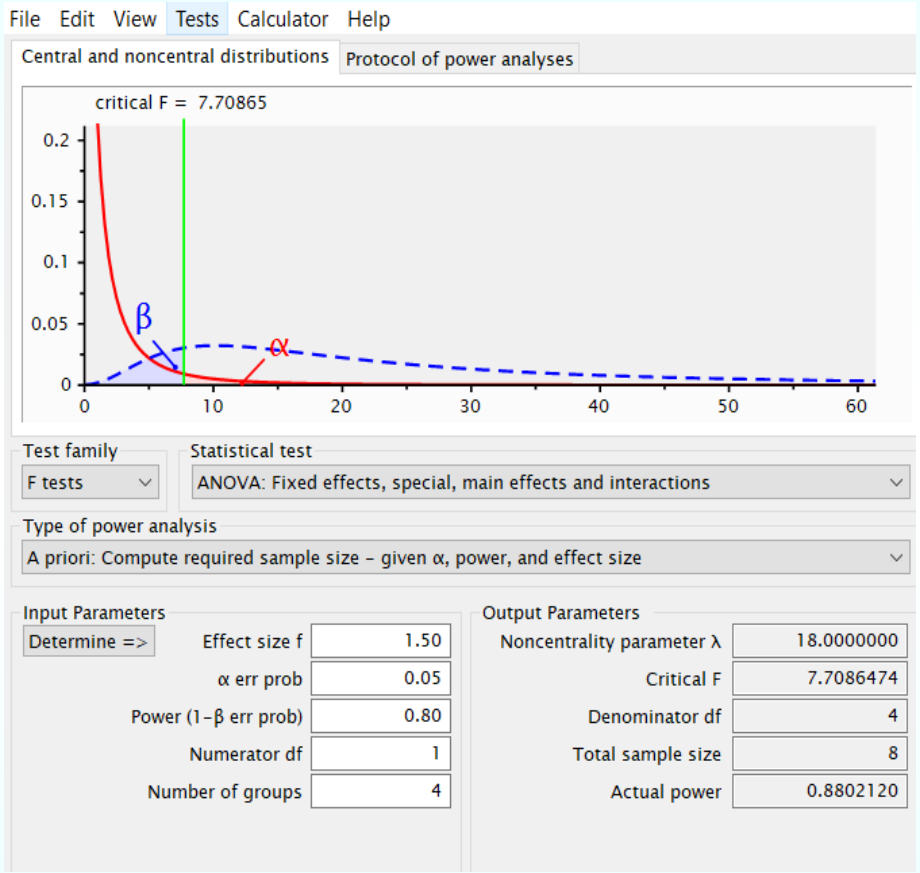
$$f = \frac{\left|\sum c_i \cdot \mu_i\right|}{\sqrt{k \cdot \sum c_i^2 \cdot \sigma^2}}$$

$$C_{AB} = (1) \cdot 5 + (-1) \cdot 2 + (-1) \cdot 2 + (1) \cdot 5 = 6$$

$$f = \frac{6}{\sqrt{4 \cdot 4 \cdot 1}} = 1.50$$

Plugging it in G*Power

Main effect or Moderated (reverted) effect



Needed sample size is the same !

General rule

One can think in terms of **percentage (expected change) of moderation effect**

$$f_n = \frac{p_m}{100} \cdot f_o \cdot \sqrt{\frac{k_o}{k_n \cdot l}}$$

f_n = expected ES of moderator for planned research

f_o = observed ES of original effect

k_o = number of cells in original study

k_n = number of cells in planned research

l = number of levels of moderator

p_m : 0%= no moderation (replicated effect)
100%= moderation as suppression
200%= moderation as reverted effect

Applying the rule

$$f_n = \frac{p_m}{100} \cdot f_o \cdot \sqrt{\frac{k_o}{k_n \cdot l}}$$

CASE 1 (suppressed moderation)

$p_m = 100\%$

$f_o = 0.50$

$k_o = 2;\ k_n = 4;\ l = 2$

$$f_n = 1.50 * \frac{1}{2} = 0.75$$

CASE 2 (reverted moderation)

$p_m = 200\%$

$$f_n = 2 * 1.50 * \frac{1}{2} = 1.50$$

CASE 3 (weak suppressed moderation)

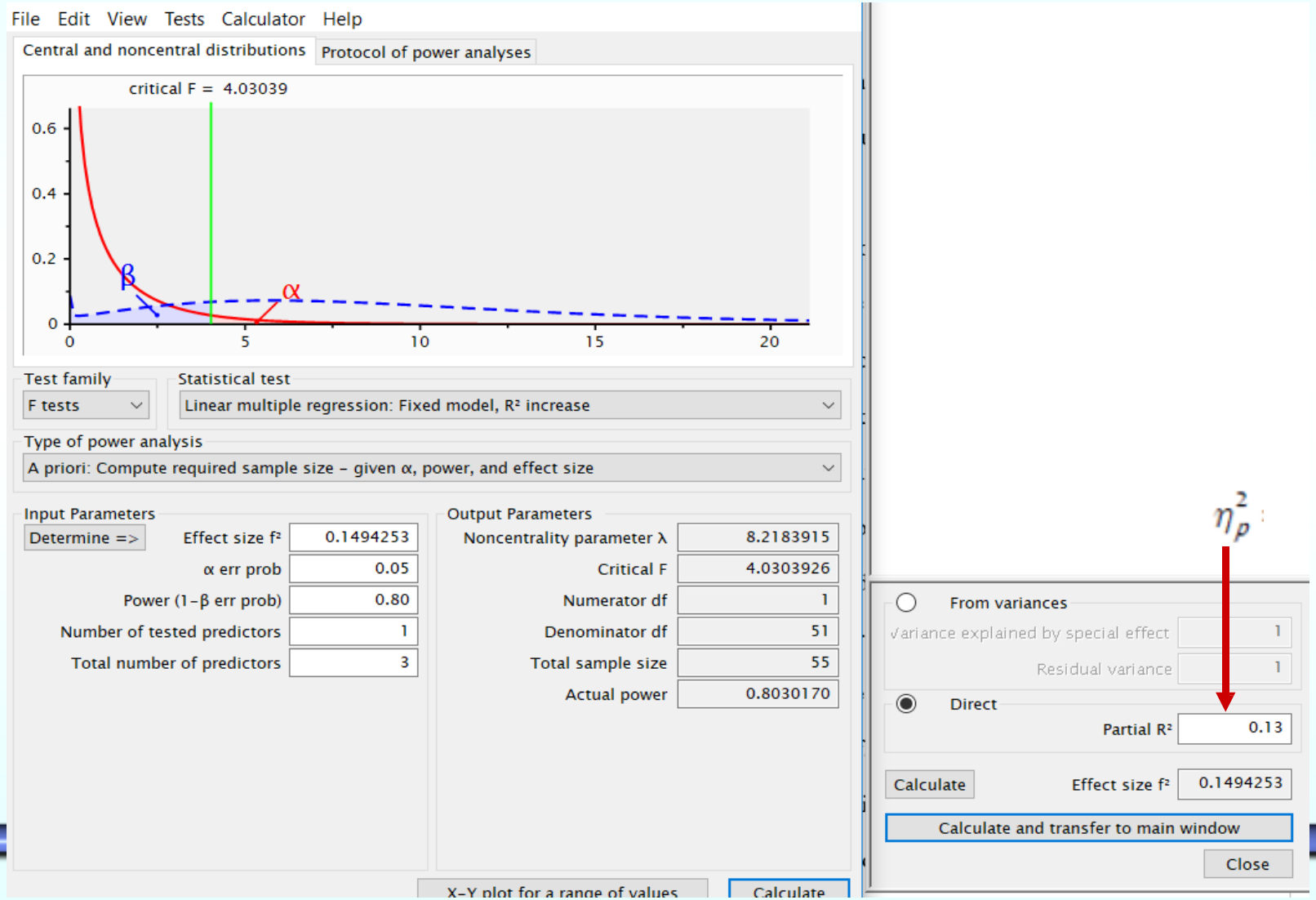
$p_m = 50\%$

$$f_n = 0.50 * 1.50 * \frac{1}{2} = 0.375$$

No need to guess Means and SD !

Regression analysis

Power calculation for a term is straightforward



Moderators in regression analysis

But ES are sometimes not easy to guess

Interaction effects (moderator) can explain not much variance (also for technical reasons) yet be theoretically key

One likely easier way to guess ES is to think in terms of difference between standardized coefficients (β)

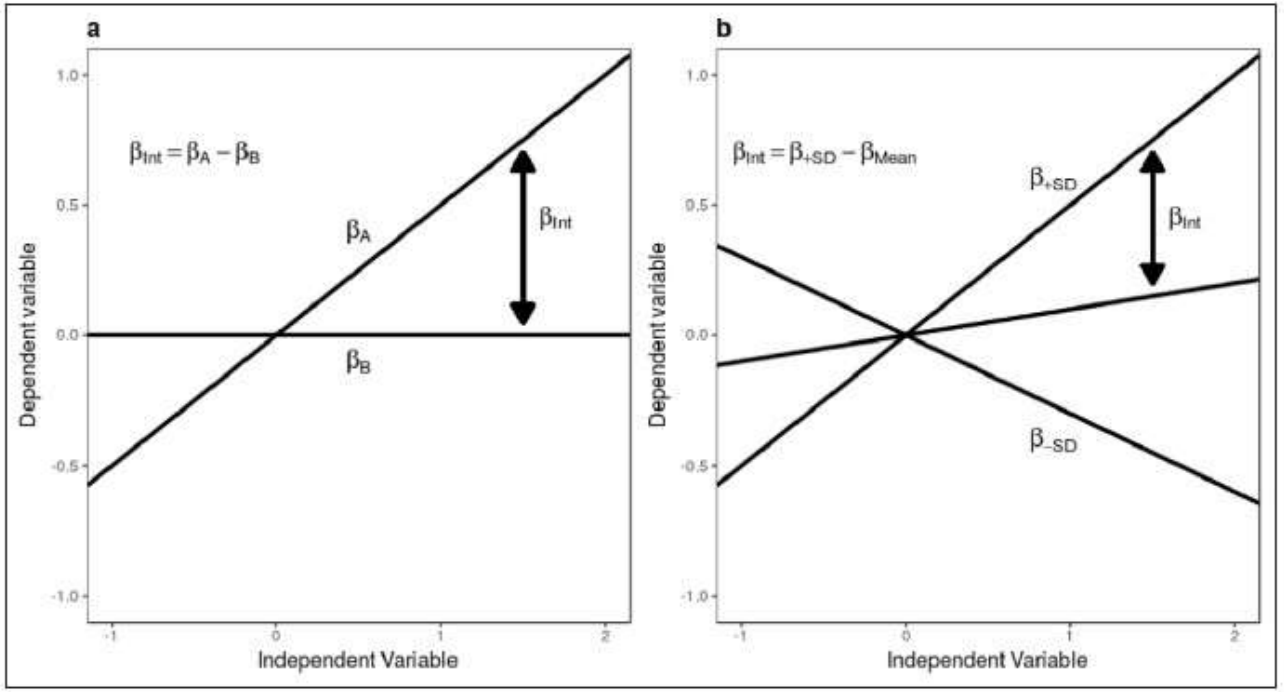


Figure 7: Geometrical interpretation of the interaction beta coefficient, with a dichotomous moderator (a) and a continuous moderator (b).

MLR with dichotomous moderator

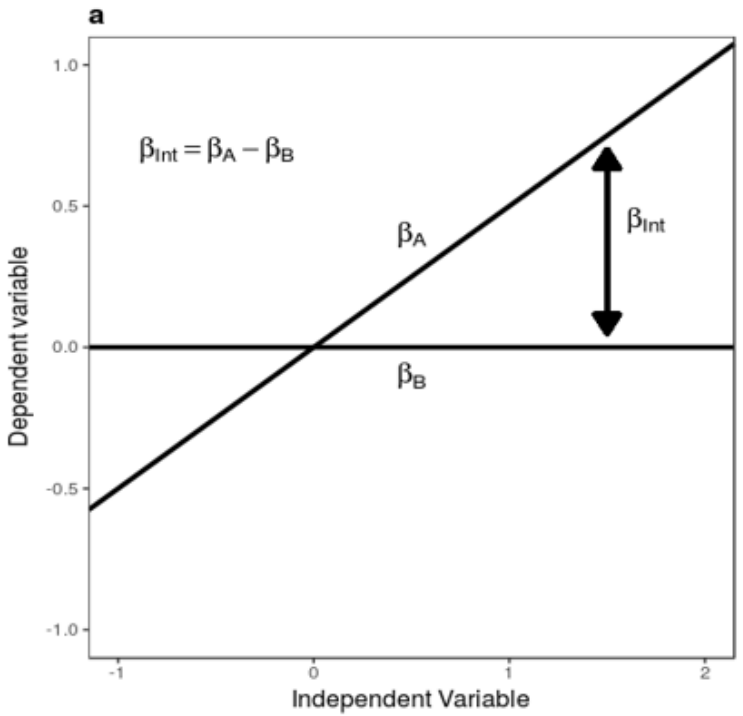
The moderation effect can be approximated as

$$f^2 \approx \frac{\beta_{int}^2}{2 \cdot (2 - r_a^2 - r_b^2)}$$

EFFECT SIZE OF INTERACTION BETWEEN A CONTINUOUS AND A DICHOTOMOUS VARIABLE

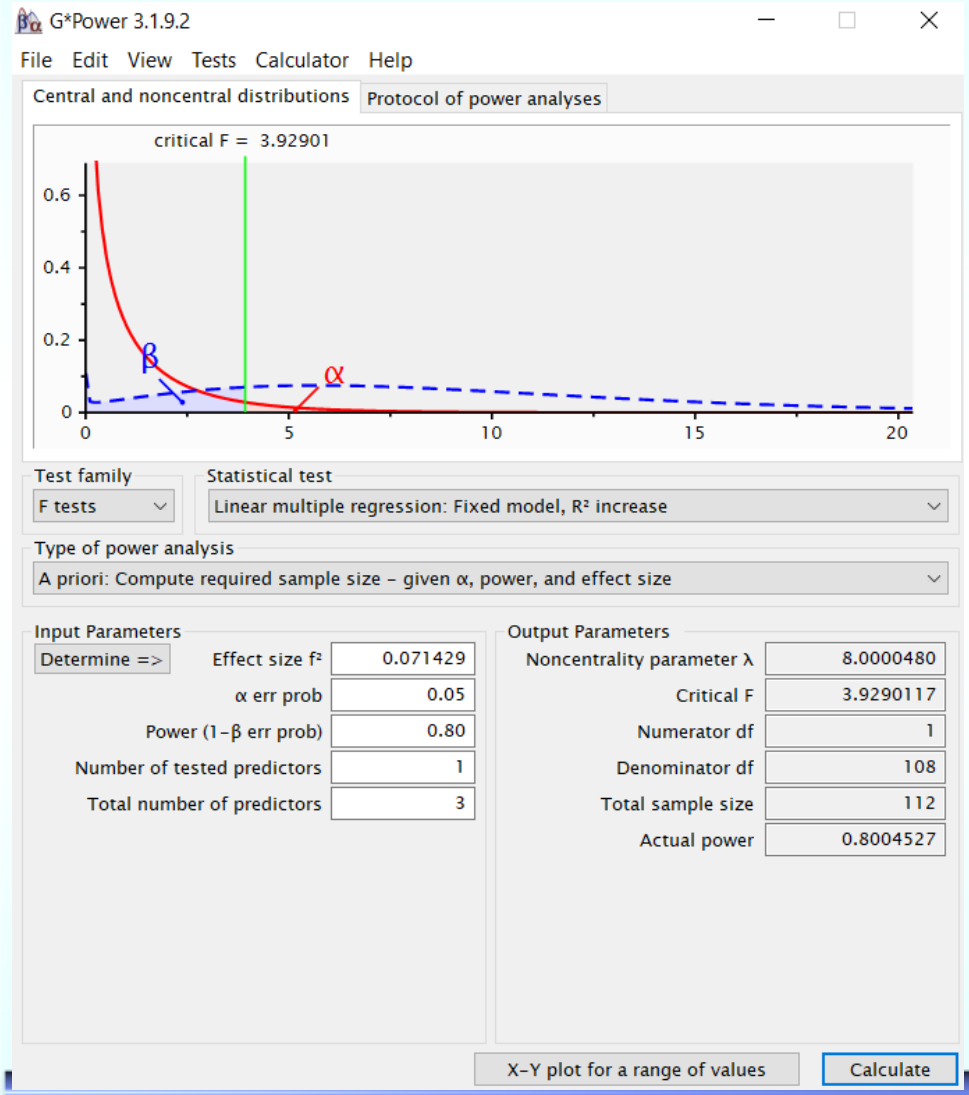
Input the expected correlations between the continuous independent variable X and the dependent variable Y at the two levels A and B of the moderator (referred to as β_a and β_b, or equivalently as r_a and or r_b). In the yellow cells you can see the implied interaction term (β_{int}) and the interaction effect size (f_n)

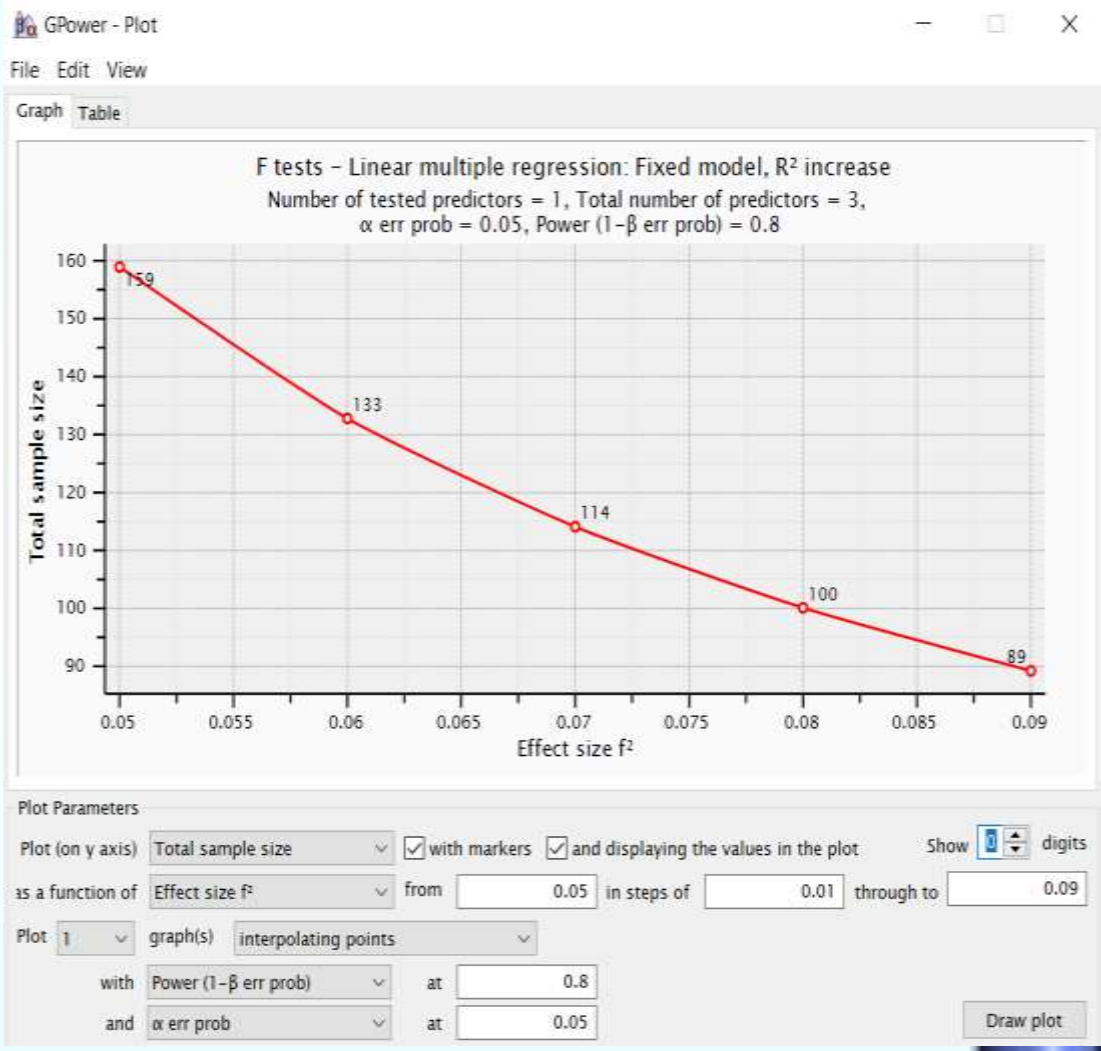
β_a or r_a	0,5
βa or r_b	0
β_{int}	0,5
f_n	0,071429



Routine Excel at https://github.com/mcfanda/primerPowerIRSP

Plugging it in G*Power





MLR with continuous moderator

The moderation effect can be approximated as $$f^2 \approx \frac{\beta_{int}^2}{1 - r_{yx}^2 - r_{ym}^2}$$

INTERACTION BETWEEN TWO CONTINUOUS VARIABLES

Option 1: Input the values of correlations between the dependent variable Y and the predictor X when the moderator M is equal to its mean (r_{yx}), the correlation between the dependent variable Y and the moderator M (r_{ym}), and the expected percentage of moderation (p_m)

r_{yx}	0,35	
r_{ym}	0,25	
p_m	50	%
β_{int}	0,175	
f_n	0,037577	

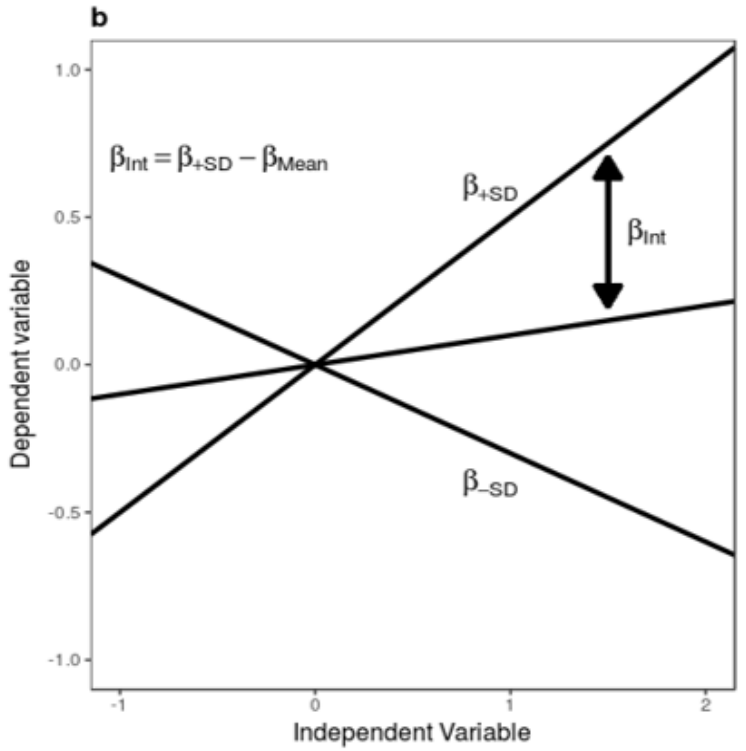
Routine Excel at https://github.com/mcfanda/primerPowerIRSP

MLR with continuous moderator

The moderation effect can be approximated as $f^2 \approx \frac{\beta_{int}^2}{1 - r_{yx}^2 - r_{ym}^2}$

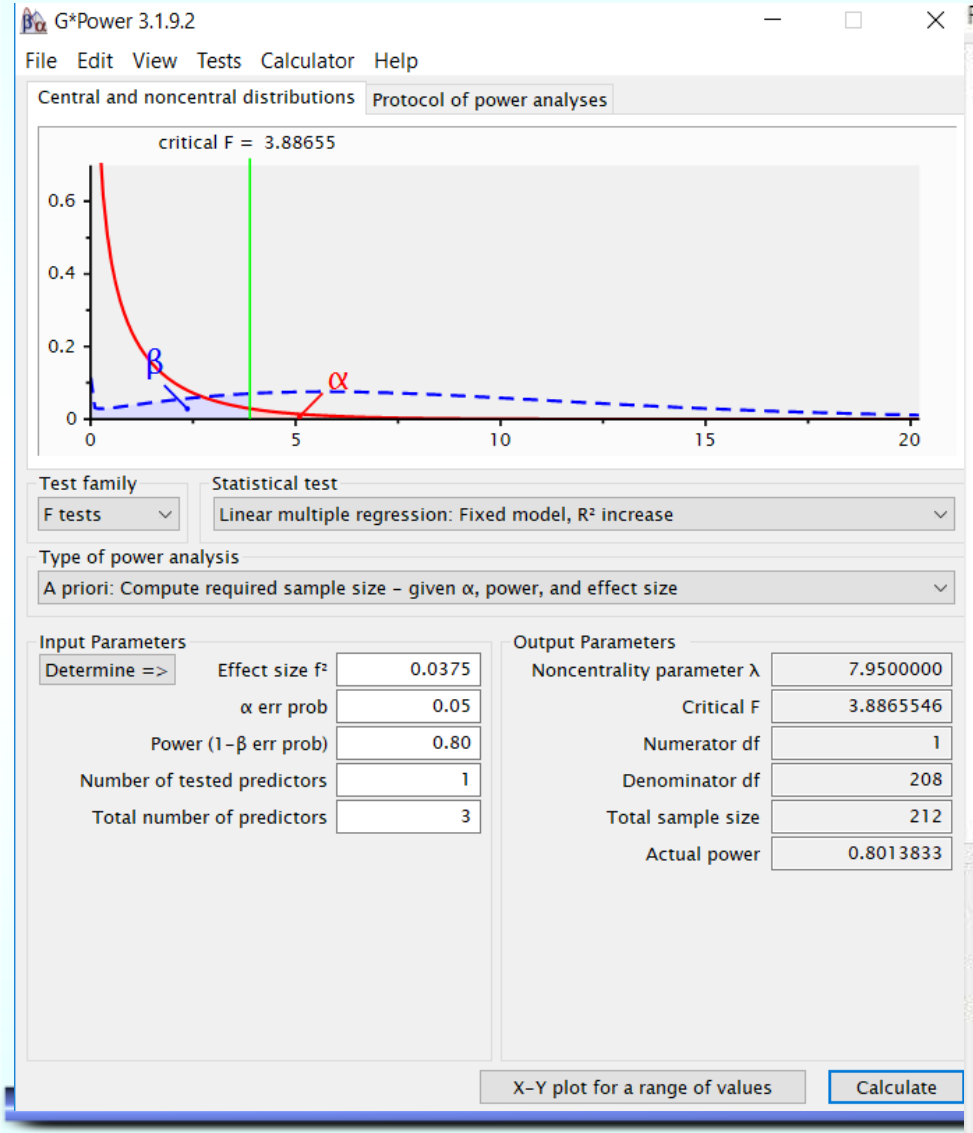
Option 2 Input the expected value of correlations between the dependent variable Y and the predictor X when the moderator M is equal to its mean (referred to as r_{yx} or equivalently as β_{Mean}), the correlation between the dependent variable and the moderator r_{ym}, and the expected correlation between the dependent variable Y and the predictor X when the moderator is 1SD above (or below) its mean, referred to as r_{yx} ±SD or β_{Mean} ±SD

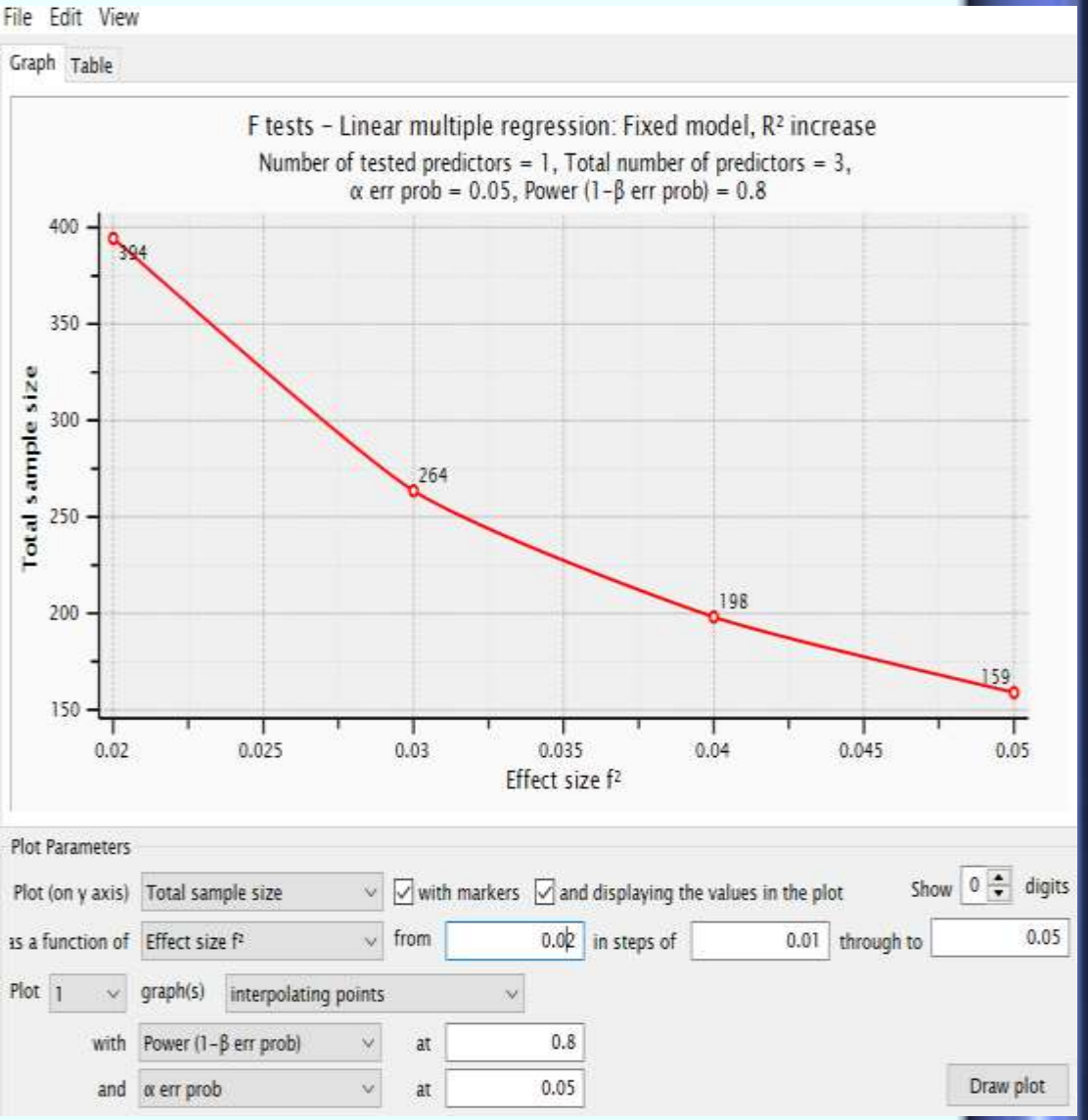
r_{yx} or β_{Mean}	0,35
r_{ym}	0,25
r_{yx} ±SD or β_{Mean} ±SD	0,525
β_{int}	0,175
f_n	0,037577



Routine Excel at https://github.com/mcfanda/primerPowerIRSP

Plugging it in G*Power





Mediation in regression analysis

Not all statistical models have available easy or generally valid analytic solutions for power

For mediation analysis, G*Power is not of help

Simulation approach can be a solution (**see Giulio after!**)

R-packages are available

powerMediation (Qiu, 2017)

Bmem (Zhang, 2014)

Also Shiny app based on R

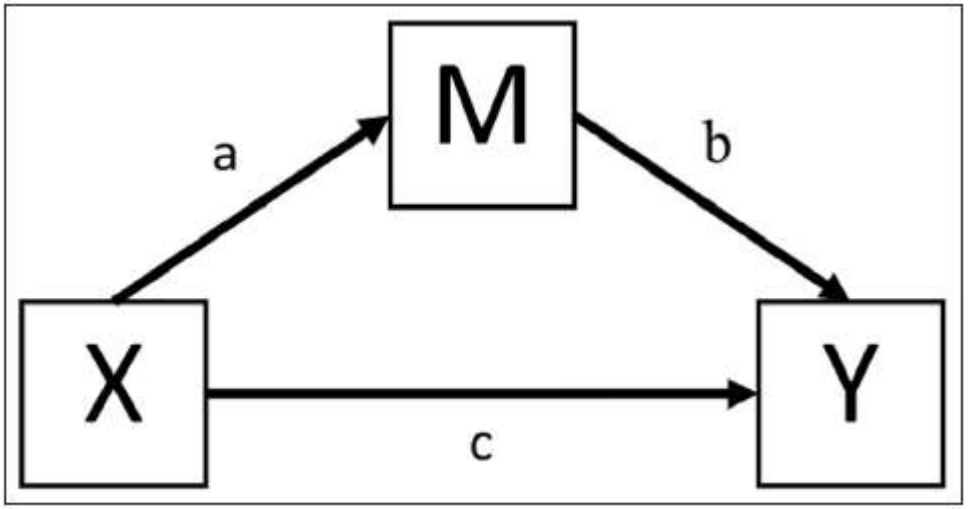
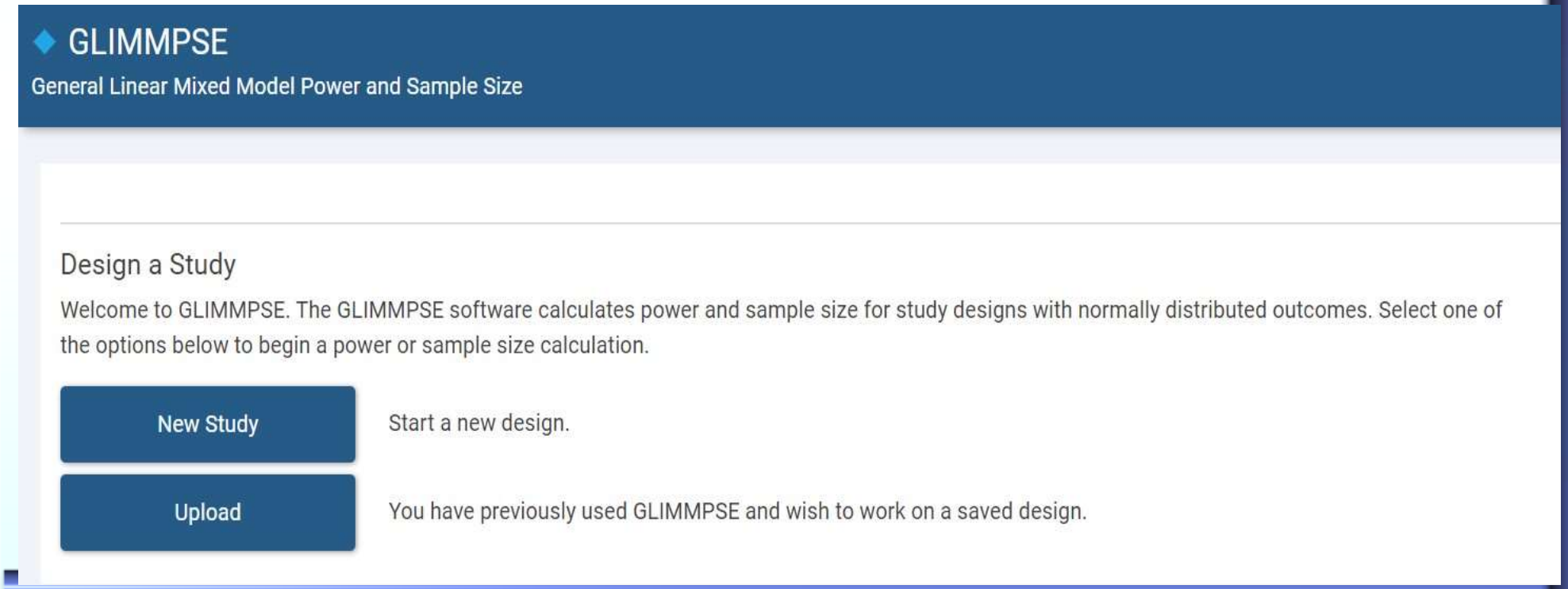


Figure 9: Mediation model.

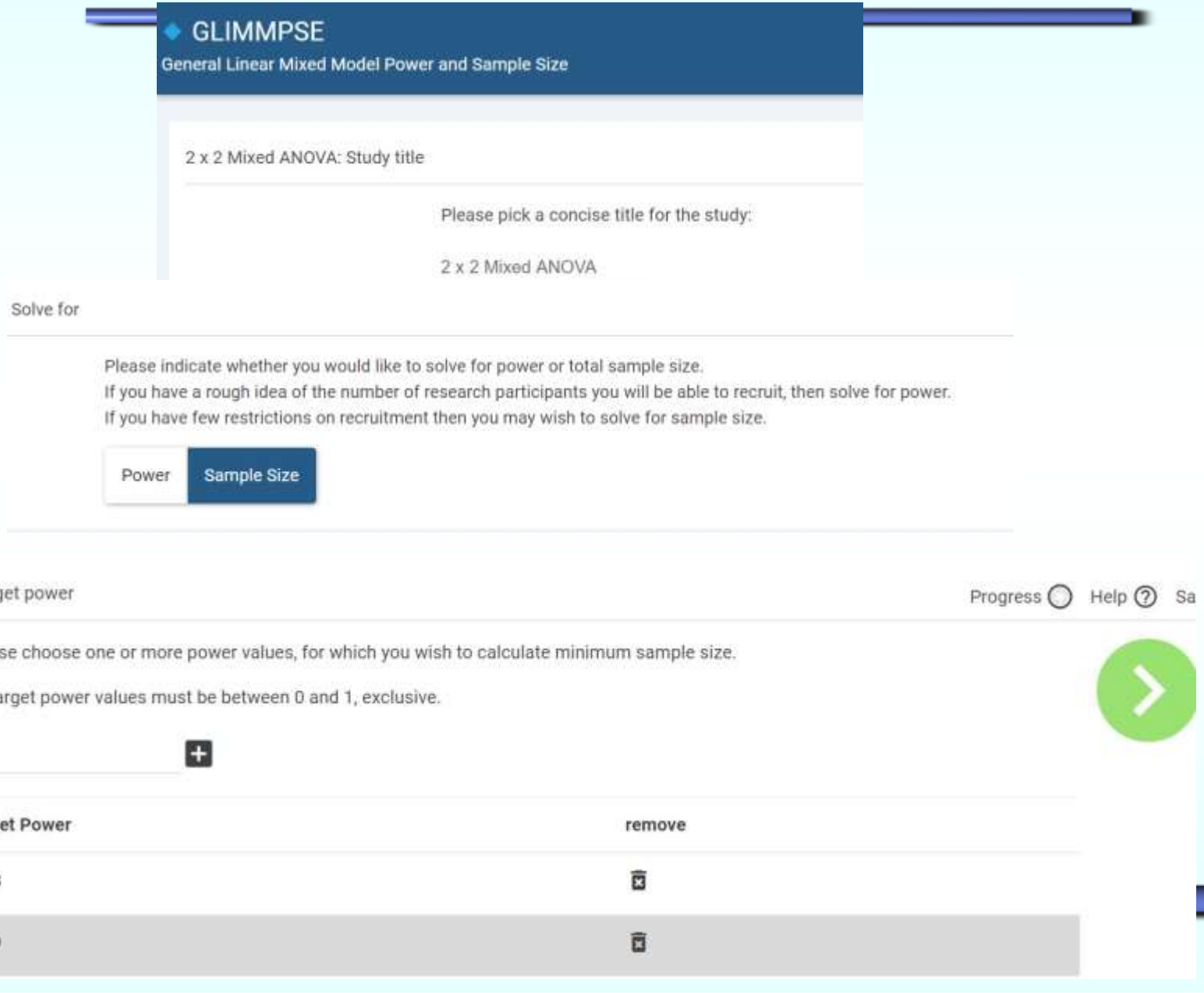
ANOVA Within and Mixed

Web app: GLIMMPSE (https://glimmpse.samplesizeshop.org) but check also https://samplesizeshop.org/

2 x 2 Mixed ANOVA



Solve for sample size



Define test and alpha

es : Statistical tests

Progress Help Sa

Please choose one or more statistical tests. If you are unsure which to pick, we recommend the Hotelling Lawley Trace test due to its equivalence to a mixed model test.

- [x] Hotelling Lawley Trace
- [ ] Pillai-Bartlett Trace
- [ ] Wilks Likelihood Ratio
- [ ] Box Corrected
- [ ] Geisser-Greenhouse Corrected
- [ ] Huynh-Feldt Corrected
- [ ] Uncorrected

: Type I error rates

Progress Help Sa

A Type I error occurs when a scientist declares a difference when none is present in the population. The Type I error rate is the probability of that kind of error, a false positive, and is often referred to as α (alpha). A Type I error rate can range from 0 to 1. Although the most commonly used value is 0.05, we recommend 0.01.

Type I Error Rate	remove
0.05	

Define outcomes and Within factor

: Outcomes

Progress Help Sav

Enter the name of each outcome variable one at a time in the underlined space below. For example, in a study investigating cholesterol-lowering medication, the outcome variables could be HDL, LDL, and total cholesterol.

Note that repeated measurement information will be addressed on the next screen.

Please name the one or more outcomes.

Outcome	remove
Performance	

: Repeated measures

Progress Help Sa

GLIMMPSE allows you to define within-participant factors, specified as repeated measures. An independent sampling unit provides one or more observations such that observations from one unit are statistically independent from any other distinct unit while observations from the same unit may be correlated. Repeated measures are present when a response variable is measured on each independent sampling unit on two or more occasions or under two or more conditions. The values of the repeated measures (that is, the levels of the within-participant factors) distinguish the occasions or conditions.

If the study includes repeated measures, click "Add Repeated Measure" and follow the prompts.

You may specify up to 5 repeated measures. Each repeated measure you add will apply to each outcome you specified on the previous page.

Define Repeated Measure

Define Within factor

What is the name of the dimension you will be measuring?

The text entered in the "Dimension" text box indicates the dimension over which measures were taken (e.g. time, days, locations, etc.). The choice of "Type" indicates whether the repeated measures are numeric (e.g. time), or categorical (e.g. arm, leg, hand)

Dimension: Repeated measures

time

Cancel | Next: Type

What type of data is time?

Categorical | Numeric

Cancel | Back | Next: No. Measurements

Number of measurements of time?

2

You must have between 2 and 10 repeats (inclusive)

Cancel | Back | Next: Spacing

Spacing

If the repeated measures are numeric, the spacing values must be unique nonnegative integers, in ascending order.

Set values myself | Select values by series

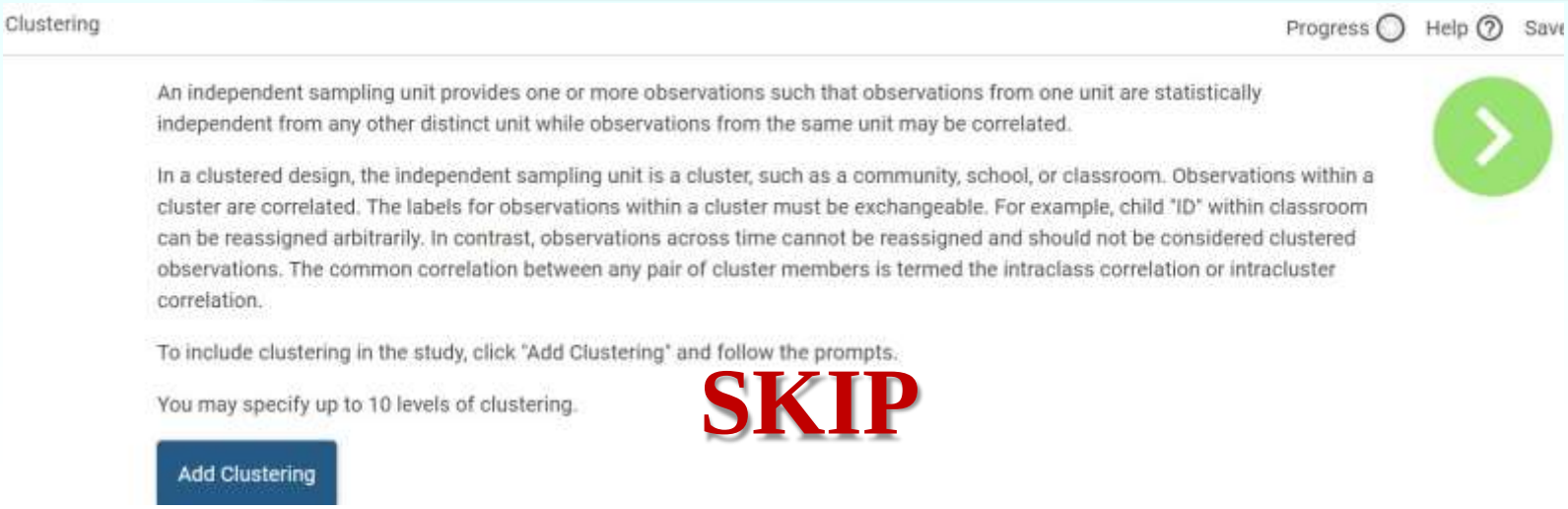
Measurement #1 at 1

Measurement #2 at 2

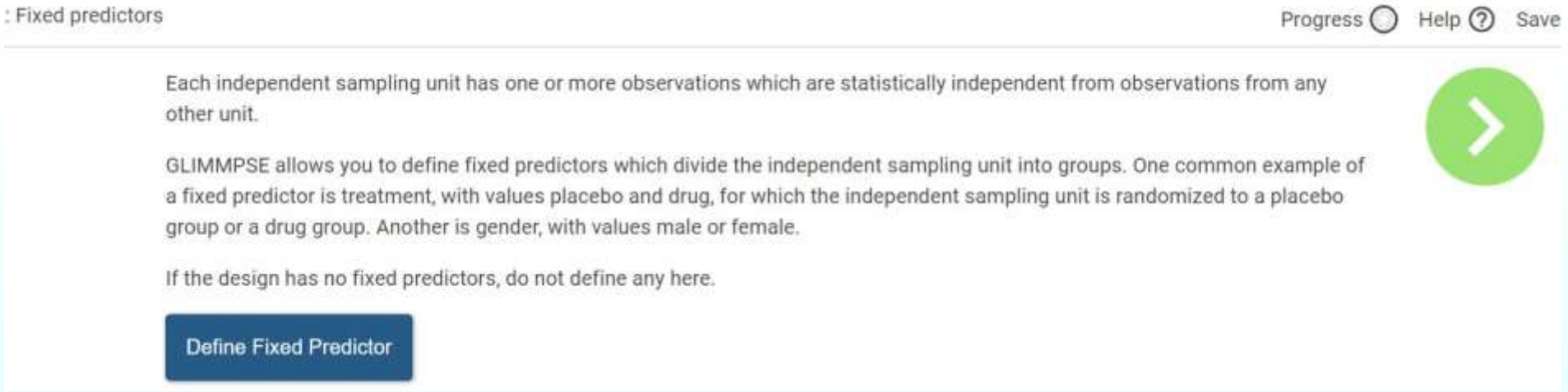
Cancel | Back | Update repeated measure

Repeated Measure Dimension	Type	Measurements	Edit	Remove
time	Numeric	["1", "2"]		

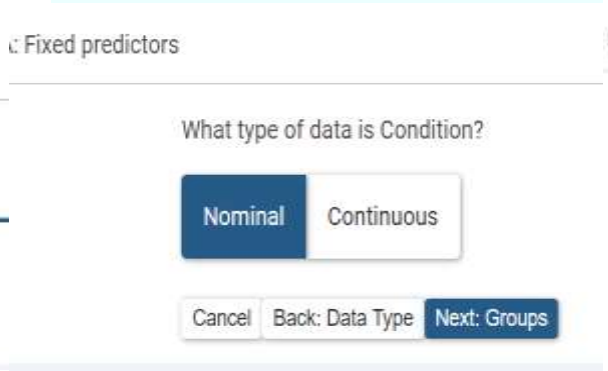
Define Between factor

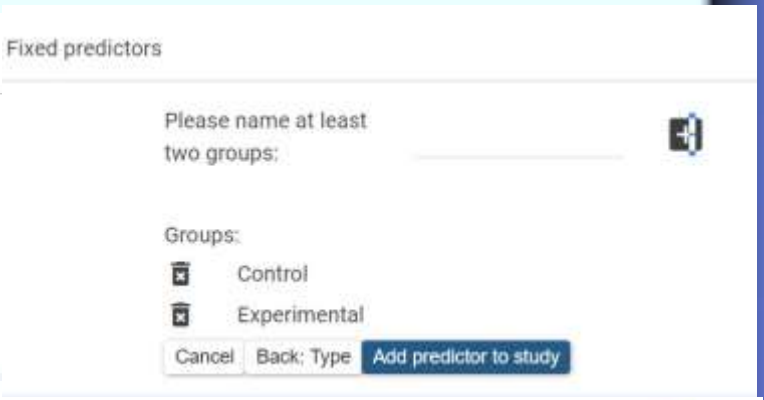


SKIP



Define Between factor





Each independent sampling unit has one or more observations which are statistically independent from observations from any other unit.

GLIMMPSE allows you to define fixed predictors which divide the independent sampling unit into groups. One common example of a fixed predictor is treatment, with values placebo and drug, for which the independent sampling unit is randomized to a placebo group or a drug group. Another is gender, with values male or female.

If the design has no fixed predictors, do not define any here.

Define Fixed Predictor

Fixed Predictors

Name	Type	Units	Groups	Remove	Edit
Condition	NOMINAL		["Control", "Experimental"]		

Select key hypothesis for power analysis

Hypothesis choice

Progress Help Save

Each power or sample size calculation is based on selecting a specific study hypothesis. The options below show the hypotheses which are available for the current study design. Specify the hypothesis that represents your scientific question.

GLIMMPSE chooses sensible contrast matrices based on cell means coding. Should you wish to define your own contrast matrices, pick the highest order interaction and choose from the advanced options in the hypothesis components.

Select a hypothesis from the list.

	Effects Available for Consideration	Nature of Variation
◉	Condition x time: Interaction	Between x Within
○	time: Main Effect	Within
○	Condition: Main Effect	Between
○	Grand Mean	Between

Test hypothesis

: Hypothesis

Progress

What type of contrast do you wish among the means defined by your groups and repeated measures?

All mean differences zero

A parameter is a characteristic of a population.The parameters of interest are differences between groups at individual repeated measures.

The null hypothesis is that all pairwise differences between groups are the same among all pairs of repeated measures.

Show Advanced Options

Theta 0

Progress Help Sa

A hypothesis compares parameters to a constant, the contrast comparison constant, Θ_0. This is almost always zero. If you choose a value other than zero, be sure that you understand that the hypothesis you define is scientifically meaningful. Also note that the description and interpretation of your hypothesis given when choosing your contrasts will be affected.

0

Group size ratios

Progress

For equal group sizes, input a "1" in the block next to each group. This is the default study design.

For unequal group sizes, specify the ratio of the group sizes. For example, consider a design with an active drug group and a placebo group. If twice as many study participants receive the placebo, a value of "2" would be selected for the placebo group, and a value of "1" would be selected for the active drug group.

Group size ratios

Condition	Control	1
	Experimental	1

Expected means under key hypothesis

Marginal means

Progress (

The table below shows the mean values for outcome **Performance** within each group in the study. Each group is represented by a row in the table, and each repeated measure dimension is represented by a column.

Enter the mean values you expect to observe for outcome **Performance** within each group. The table should contain at least one value that is non-zero. Also, at least two groups should have means which differ by a scientifically meaningful amount.

Expected mean values, per group, for *Performance*

		time	
		1	2
Condition	Control	5	5
	Experimental	5	6

Set blank values to value

Scale factors (different scenarios) and SD

Scale factor for the marginal means

Progress

In power analysis, it is not possible to know the exact values of means before the experiment is observed. Scale factors allow you to consider alternative values for the means by scaling the values entered on the previous screen.
For example, entering the scale factors 0.5, 1, and 2 would compute power for the mean values divided by 2, the mean values as entered, and the mean values multiplied by 2.

Enter a scale factor:

number > 0 +

Scale Factor	remove
1	

Variability across outcomes

Enter the standard deviation you expect to observe for each outcome.

Outcome	Standard Deviation
Performance	1

: Repeated measure standard deviation ratios

Progress

Define the ratios of standard deviations for time. One of your values should be 1 and the others should represent the ratio of that value to that value:

For example, if you believe that the standard deviation doubles at each time, enter the values 1, 2, 4, 8... etc.

time	Standard Deviation Ratio
1	1
2	1

Repeated measures correlations and scale factors

Repeated measure correlation

Progress

For a given research participant, responses vary across outcomes and across repeated measurements. The amount of variability can dramatically impact power and sample size.

Define the **time** correlation matrix, by entering correlations you expect to observe among the chosen spacing values of **time**:

Unstructured | LEAR

time

	1	2
1	1	0,5
2	0,5	1

(each off-diagonal correlation must be between -1 and 1, exclusive)

Scale factor variance

Progress

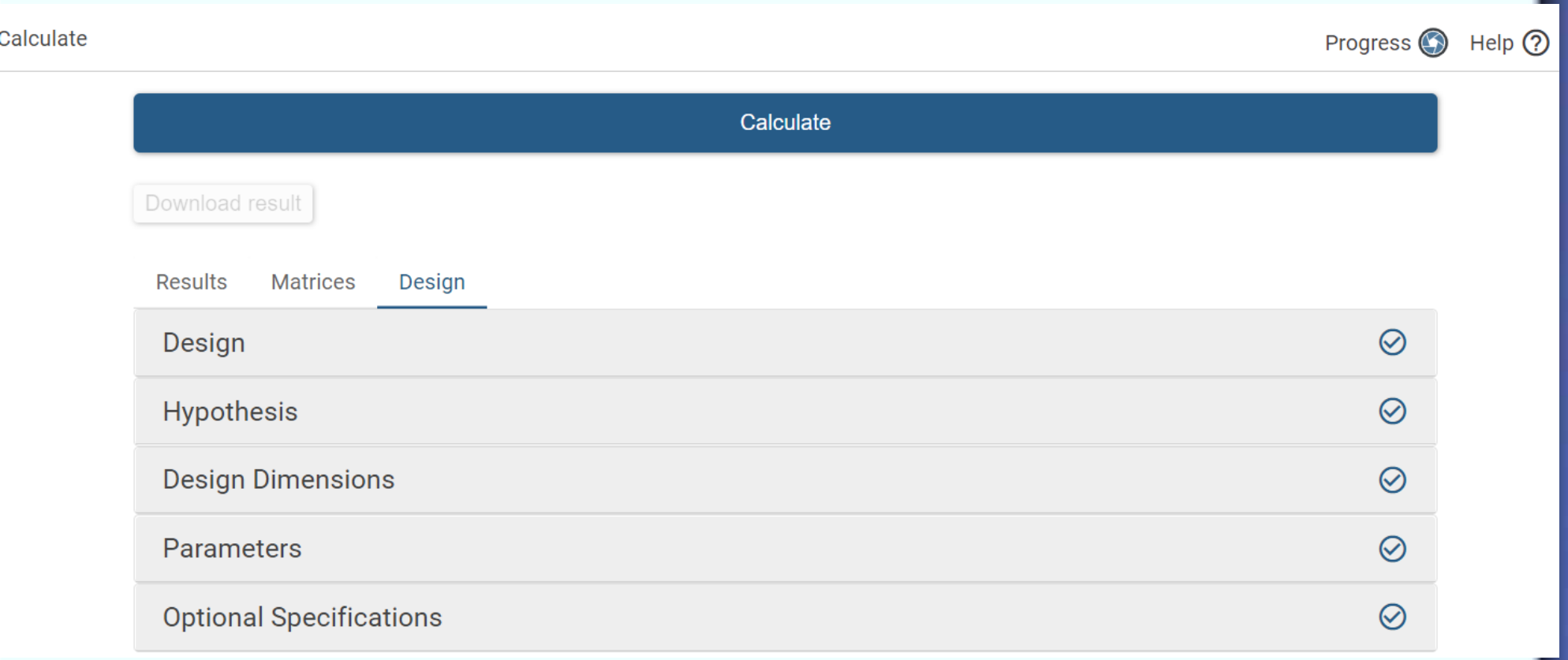
Changes in variability can dramatically affect power and sample size results. It is not possible to know the variability until the experiment is observed. Scale factors allow you to consider alternative values for variability by scaling the calculated covariance matrix. For example, entering the scale factors 0.5, 1, and 2 would compute power for the covariance matrix divided by 2, the covariance matrix as entered, and the covariance matrix multiplied by 2.

You may add up to 10 scale factors.

Choose a number greater than zero +

Scale Factor	remove
1	

Finally, the calculation…



…and the results!

Progress Help Save

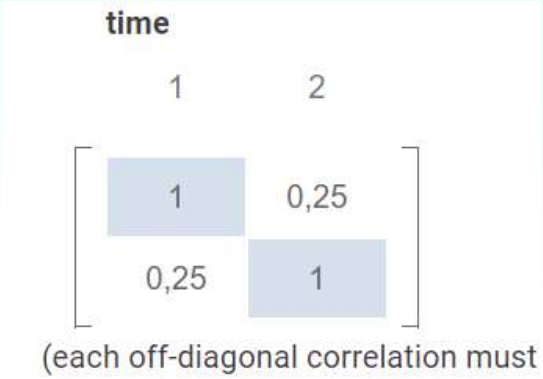
Calculate

Download result

Results | Matrices | Design

Power	Total Sample Size	Target Power	Means Scale Factor	Variability Scale Factor	Test	Power Method	Type I Error Rate
0.807	34	0.8	1	1	Hotelling Lawley Trace	conditional	0.05
0.912	46	0.9	1	1	Hotelling Lawley Trace	conditional	0.05

Suppose expected correlation is lower



Calculate

Download result

Results | Matrices | Design

Power	Total Sample Size	Target Power	Means Scale Factor	Variability Scale Factor	Test	Power Method	Type I Error Rate
0.807	50	0.8	1	1	Hotelling Lawley Trace	conditional	0.05
0.904	66	0.9	1	1	Hotelling Lawley Trace	conditional	0.05

Suppose no correlation

time

	1	2
1	1	0
2	0	1

(each off-diagonal correlation p

Calculate

Download result

Results | Matrices | Design

Power	Total Sample Size	Target Power	Means Scale Factor	Variability Scale Factor	Test	Power Method	Type I Error Rate
0.808	66	0.8	1	1	Hotelling Lawley Trace	conditional	0.05
0.906	88	0.9	1	1	Hotelling Lawley Trace	conditional	0.05

Recap Examples Mixed ANOVA

1) The design was a 2 x 2 Mixed ANOVA
2) We varied the expected correlations

Required N for power at .80

- **r=.00, N= 66**
- **r=.25, N= 50**
- **r=.50, N= 34**

Required N goes down as the correlation between DVs of the Within factor goes up

Suppose instead a 2 x 2 Between Ss

The design you've described, means that every level of **Group** occurs at every level of **Condition**. This concept applies to every combination of fixed predictors.

Define Fixed Predictor

Fixed Predictors

Name	Type	Units	Groups	Remove	Edit
Condition	NOMINAL		["Control", "Experimental"]		
Group	NOMINAL		["No previous experience", "Previous experience"]		

Effects Available for Consideration	Nature of Variation
Condition x Group: Interaction	Between x Between

Expected mean values, per group, for *Performance*

Condition, Group	
Control, No previous experience	5
Control, Previous experience	5
Experimental, No previous experience	5
Experimental, Previous experience	6

Calculate

Download result

Results | Matrices | Design

Power	Total Sample Size	Target Power	Means Scale Factor	Variability Scale Factor	Test	Power Method	Type I Error Rate
0.801	128	0.8	1	1	Hotelling Lawley Trace	conditional	0.05
0.903	172	0.9	1	1	Hotelling Lawley Trace	conditional	0.05

Suppose instead a 2 x 2 Within Ss (r=.25)

The design you have described, means that every level of **Variable B** is measured at every level of **Variable A**. This concept applies to every combination of repeated measures.

Define Repeated Measure

Repeated Measure Dimension	Type	Measurements	Edit	Remove
Variable A	Categorical	["1", "2"]		
Variable B	Categorical	["1", "2"]		

Effects Available for Consideration	Nature of Variation
Variable A x Variable B: Interaction	Within x Within

Expected mean values, per group, for *Performance*

Variable A, Variable B

1,1	1,2	2,1	2,2
5	5	5	6

Variable A

1	0,25
0,25	1

Variable B

1	0,25
0,25	1

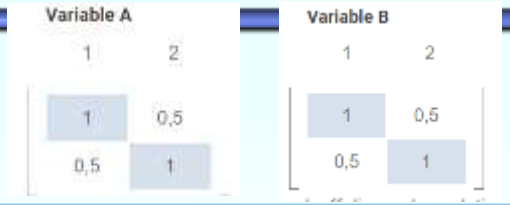
Calculate

Download result

Results | Matrices | Design

Power	Total Sample Size	Target Power	Means Scale Factor	Variability Scale Factor	Test	Power Method	Type I Error Rate
0.807	20	0.8	1	1	Hotelling Lawley Trace	conditional	0.05
0.904	26	0.9	1	1	Hotelling Lawley Trace	conditional	0.05

2 x 2 Within Ss with r=.50 and r=.0

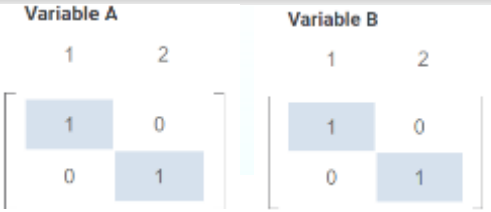


Calculate

Download result

Results | Matrices | Design

Power	Total Sample Size	Target Power	Means Scale Factor	Variability Scale Factor	Test	Power Method	Type I Error Rate
0.803	10	0.8	1	1	Hotelling Lawley Trace	conditional	0.05
0.911	13	0.9	1	1	Hotelling Lawley Trace	conditional	0.05



Results | Matrices | Design

Power	Total Sample Size	Target Power	Means Scale Factor	Variability Scale Factor	Test	Power Method	Type I Error Rate
0.808	34	0.8	1	1	Hotelling Lawley Trace	conditional	0.05
0.900	44	0.9	1	1	Hotelling Lawley Trace	conditional	0.05

Power Comparison

- Three 2 x 2 ANOVA designs (Mixed, Between, Within)
- In each design the same pattern of expected means

	A1	A2
B1	5	5
B2	5	6

- Always SD=1
- Always powered for interaction effect
- Required N for power at .80

  -Between = 128
  -Mixed (r=.00) = 64
  -Mixed (r=.25) = 50
  -Mixed (r=.50) = 34
  -Within (r=.00) = 34
  -Within (r=.25) = 20
  -Within (r=.50) = 10

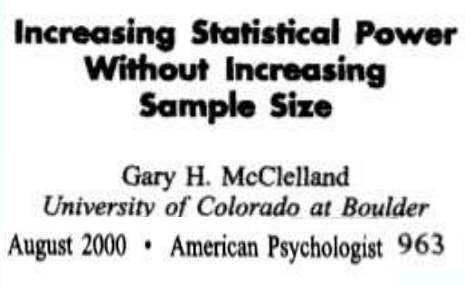
- You can draw your own conclusion…

How to increase power?

- **<u>Increase sample size</u> (also multi-lab collaborations)**
- Use blocking or repeated measures (**<u>within Ss design</u>**) BUT sometimes can be inappropriate
- Administer stronger treatments (e.g., experimental manipulation) BUT be wary of possible reduced ecological validity
- Avoid restrictions of range for dependent variables
- Standardize experimental procedures
- Increase reliability of measures
- Use more homogenous subject samples BUT increased risks to generalizability of results
- Meta-analytic mindset

Increasing power without increasing sample size I

$$SE = \sqrt{\frac{S^2}{n}}$$


Increasing Statistical Power Without Increasing Sample Size

Gary H. McClelland
University of Colorado at Boulder
August 2000 • American Psychologist 963


Increasing the Power of Your Study by Increasing the Effect Size

TOM MEYVIS
STIJN M. J. VAN OSSELAER
Journal of Consumer Research Feb2018, Vol. 44 Issue 5, p1157-1173

- Standard errors depend on N and SD (smaller SD means smaller SE)
- SD can be reduced with more reliable measures, more precise experimental designs, less Ss variability (e.g., also within Ss designs)
- Plan your design as simple and as clean as possible

Increasing power without increasing sample size II

Power, Dominance, and Constraint: A Note on the Appeal of Different Design Traditions

Jeffrey N. Rouder[1,2] and Julia M. Haaf[2]
[1]Department of Cognitive Sciences, University of California, Irvine, and [2]Department of Psychological Sciences, University of Missouri

Advances in Methods and Practices in Psychological Science
2018, Vol. 1(1) 19–26

DOI: 10.1177/2515245917745058
www.psychologicalscience.org/AMPPS
SAGE

Psychological Methods

Power Contours: Optimising Sample Size and Precision in Experimental Psychology and Human Neuroscience

Daniel H. Baker, University of York
Greta Vilidaite, University of Southampton
Freya A. Lygo and Anika K. Smith, University of York
Tessa R. Flack, University of Lincoln
André D. Gouws and Timothy J. Andrews, University of York

- The SD (σ) depends on both N (variance between Ss) and K (trials, variance across trials within Ss)
- This means that N can be (to some extent) traded for K

$$\lambda = \sqrt{\mu^2 \times \frac{IK}{K\sigma_\beta^2 + 2\sigma^2}}.$$

$$\sigma_s = \sqrt{\sigma_b^2 + \frac{\sigma_w^2}{k}}$$

$$\sigma_b = \sqrt{\sigma_s^2 - \frac{\sigma_w^2}{k}},$$

Increasing power without increasing sample size II

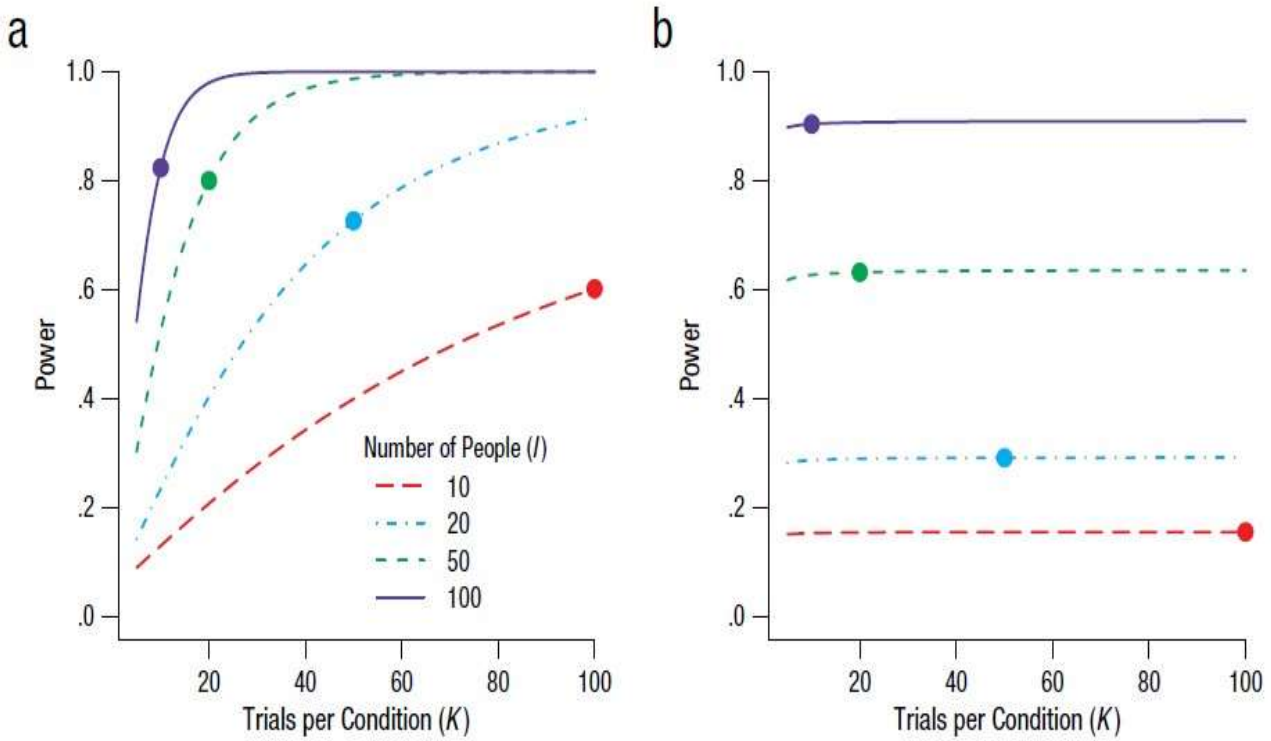
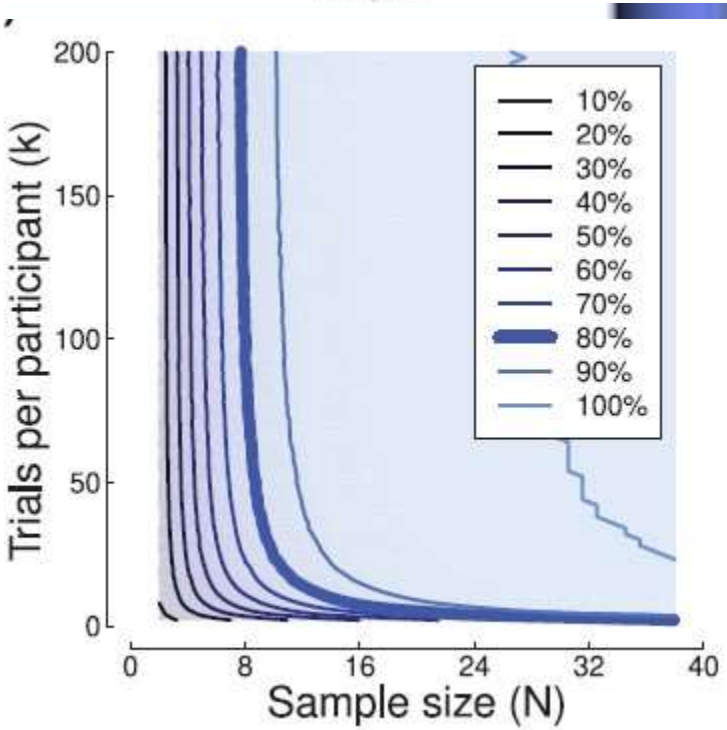


Fig. 1. Illustration of how power changes as a function of *I*, the number of participants, and *K*, the number of observations per condition per participant. The graph in (a) shows the trade-off between these two design parameters when population variability (σ_β = 28 ms) is small relative to trial variability (σ = 300 ms). The graph in (b) shows the trade-off when population variability (σ_β = 3) is large relative to trial variability (σ = 1). The plotted point on each curve indicates the values of *I* and *K* that provide 1,000 observations.

Psychological Methods

Power Contours: Optimising Sample Size and Precision in Experimental Psychology and Human Neuroscience

Daniel H. Baker, University of York
Greta Vilidaite, University of Southampton
Freya A. Lygo and Anika K. Smith, University of York
Tessa R. Flack, University of Lincoln
André D. Gouws and Timothy J. Andrews, University of York



Power, Dominance, and Constraint: A Note on the Appeal of Different Design Traditions

Advances in Methods and Practices in Psychological Science
2018, Vol. 1(1) 19–26
© The Author(s) 2018
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/2515245917745058
www.psychologicalscience.org/AMPPS
SAGE

Jeffrey N. Rouder[1,2] and Julia M. Haaf[2]
[1]Department of Cognitive Sciences, University of California, Irvine, and [2]Department of Psychological Sciences, University of Missouri

The signal and the noise

$$SE = \sqrt{\frac{S^2}{n}}$$

Distinguish conceptually between unnecessary ("added noise") and necessary ("natural") variance

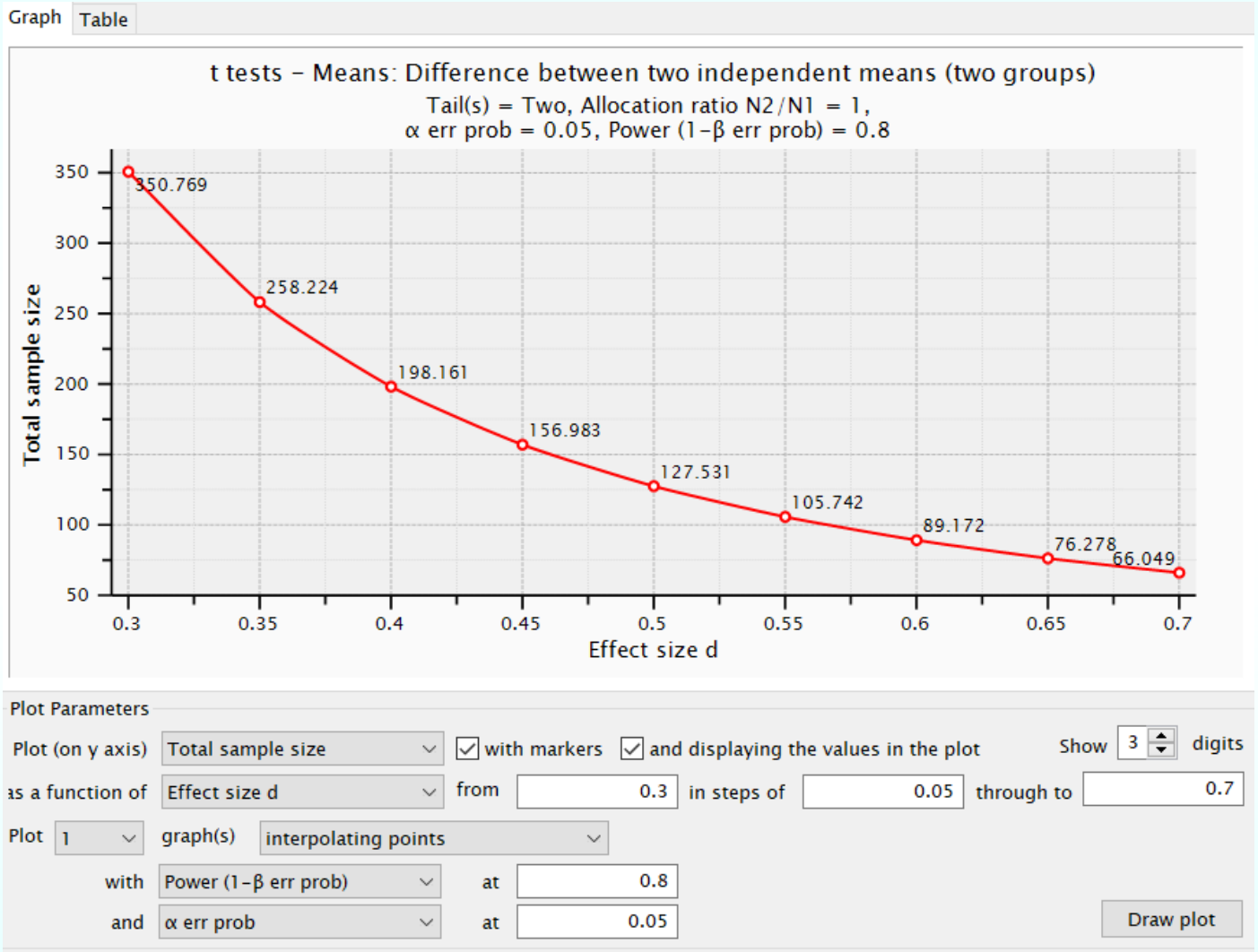
Improve your design. Optimize it. Think carefully about it. Few extra hours spent on this can be worth hundreds of extra participants (and avoid frustrations…)

Reduce the **noise**! Increase the **signal**!

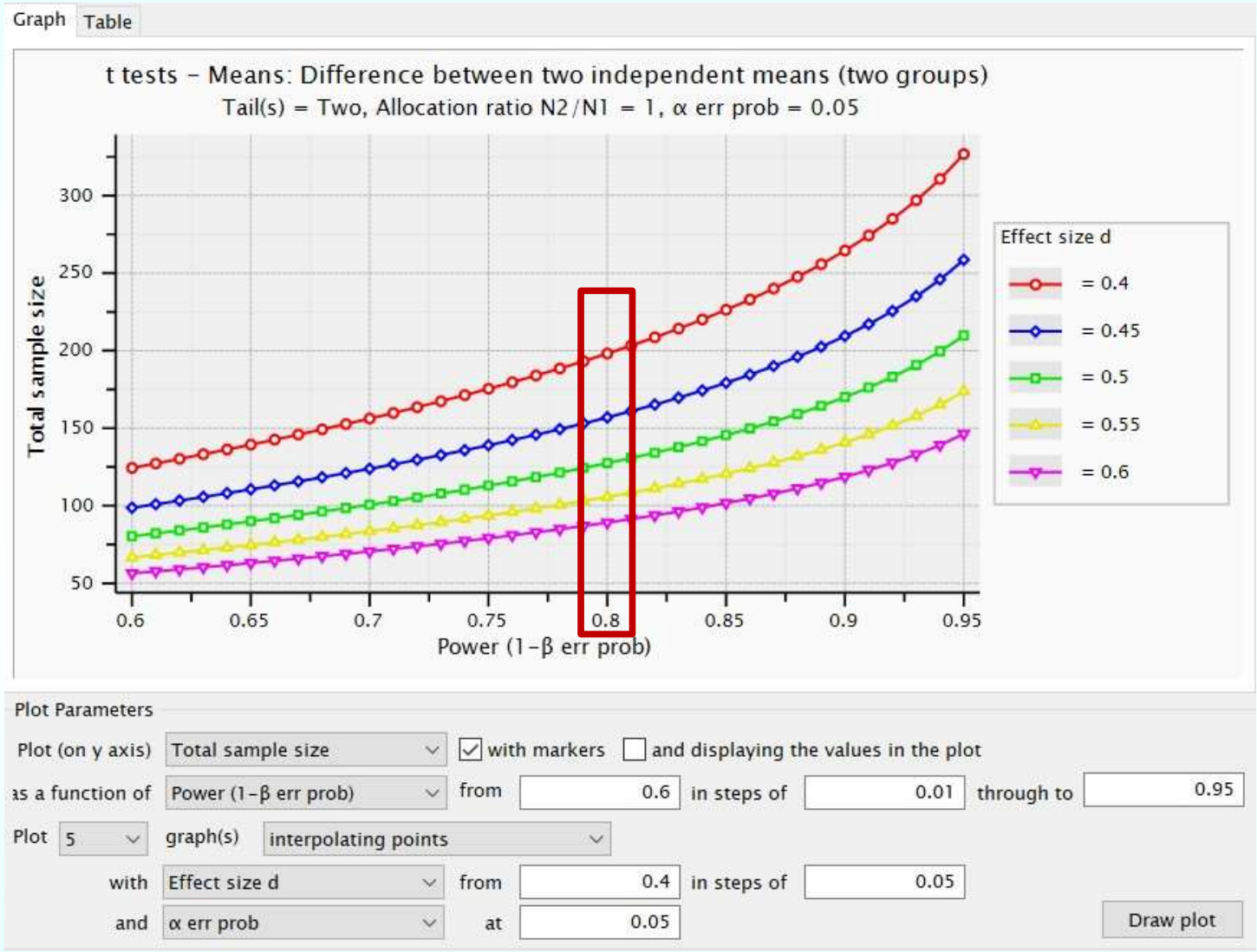
Problems in power analysis

- **Main error**: ***post-hoc power (calculated after the results) is trivial and misleading***. **Sensitivity analysis** is much better
- **Main problem**:
  -One key element of power analysis for planning studies is the Effect Size (ES)
  -We cannot know the ES. If we knew it, we wouldn't need to run the study…
  -At best we can guess/estimate ES from a meta-analysis or from previous studies, often based on a hunch. **Uncertainty** of the estimate.
  - What happens if the ES estimate is incorrect?

Uncertainty of ES



Uncertainty of ES



Uncertainty of ES

Graph | Table

t tests - Means: Difference between two independent means (two groups)
Tail(s) = Two, Allocation ratio N2/N1 = 1, α err prob = 0.05

		Effect size d = 0.4	Effect size d = 0.45	Effect size d = 0.5	Effect size d = 0.55	Effect size d = 0.6
#	Power (1-β err prob)	Total sample size	Total sample size	Total sample size	Total sample size	Total sample size
16	0.750000	175.449	139.039	112.997	93.7315	79.0806
17	0.760000	179.664	142.369	115.695	95.9607	80.9535
18	0.770000	184.029	145.818	118.488	98.2689	82.8927
19	0.780000	188.556	149.395	121.385	100.663	84.9042
20	0.790000	193.261	153.112	124.396	103.151	86.9946
21	0.800000	198.161	156.983	127.531	105.742	89.1716
22	0.810000	203.275	161.024	130.804	108.446	91.4438
23	0.820000	208.626	165.252	134.228	111.276	93.8216
24	0.830000	214.241	169.689	137.822	114.246	96.3167
25	0.840000	220.153	174.359	141.605	117.372	98.9433
26	0.850000	226.397	179.293	145.601	120.675	101.718

Plot Parameters

Plot (on y axis) [Total sample size] ☑ with markers ☐ and displaying the values in the plot

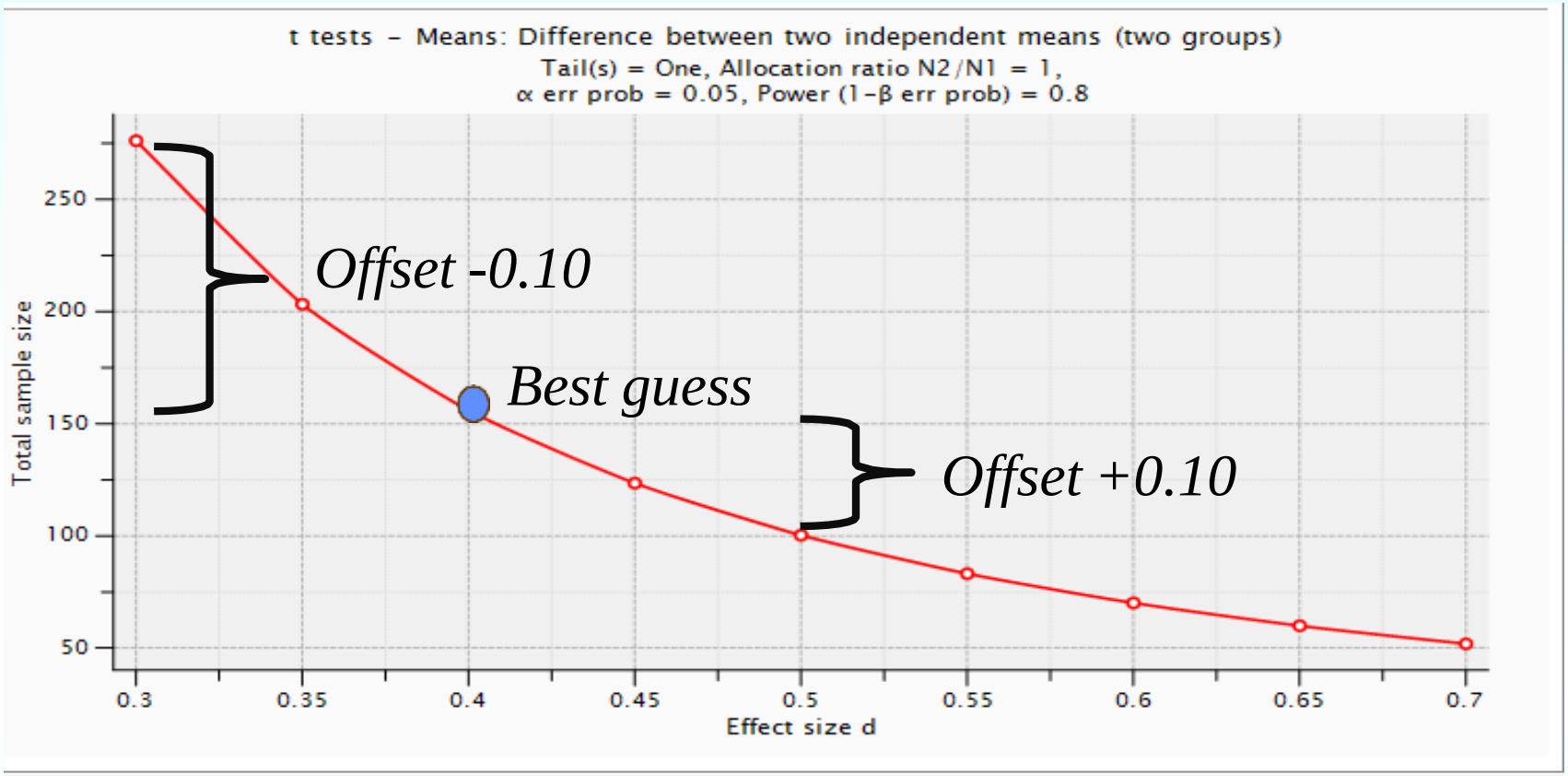
as a function of [Power (1-β err prob)] from [0.6] in steps of [0.01] through to [0.95]

Plot [5] graph(s) [interpolating points]

with [Effect size d] from [0.4] in steps of [0.05]

and [α err prob] at [0.05] Draw plot

Asymmetry of ES errors



Plot Parameters

Plot (on y axis) Total sample size ☑ with markers ☐ and displaying the values in the plot

s a function of Effect size d from 0.3 in steps of 0.05 through to 0.7

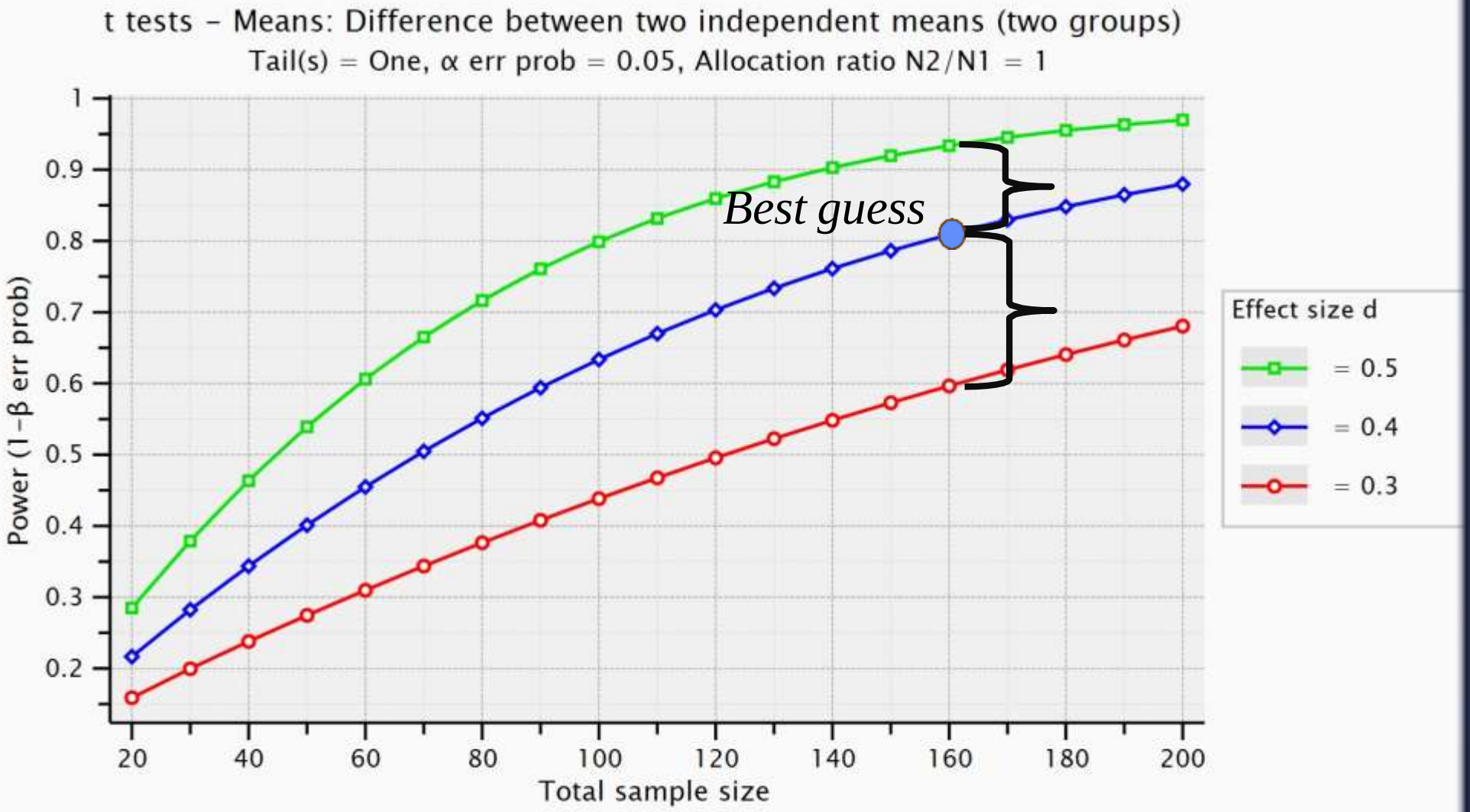
lot 1 graph(s) interpolating points

with Power (1−β err prob) at 0.80

and α err prob at 0.05

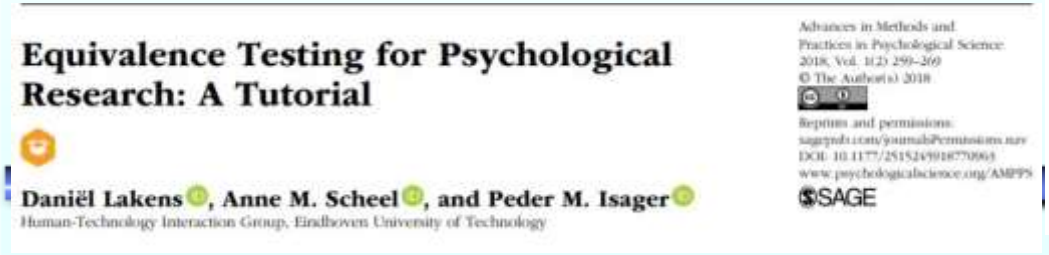
Draw plot

Asymmetry of ES errors



What to do then?

- Power depends on estimated ES (we don't know the "true" ES)
- ES over-estimation is more common (*optimistic bias*) and more influential than under-estimation (*asymmetric effect*)
- Should consider different scenarios rather than a single value
- Could consider minimum effect of interest (SESOI, Lakens, 2014)
- Could consider sensitivity analysis
- Could consider safeguarding yourself against "optimistic" ES estimates

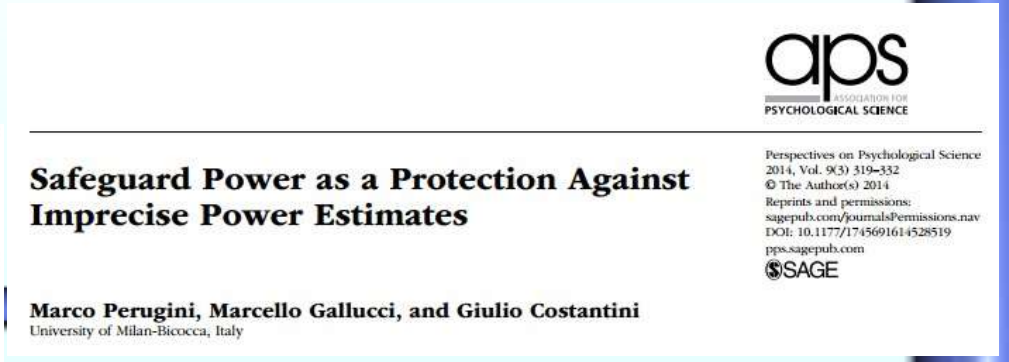

Equivalence Testing for Psychological Research: A Tutorial

Daniël Lakens, Anne M. Scheel, and Peder M. Isager

Human-Technology Interaction Group, Eindhoven University of Technology

Advances in Methods and Practices in Psychological Science
2018, Vol. 1(2) 259–269
© The Author(s) 2018

SAGE


aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Safeguard Power as a Protection Against Imprecise Power Estimates

Marco Perugini, Marcello Gallucci, and Giulio Costantini

University of Milan-Bicocca, Italy

Perspectives on Psychological Science
2014, Vol. 9(3) 319–332
© The Author(s) 2014

SAGE

Summing up Power Analysis I

- Power analysis is one important way to efficiently plan a study
- Try to power your study adequately
- A main problem is to best guess a predicted ES
- Beware of the uncertainty of ES estimates and the asymmetric impact of ES estimate errors
- Wise to consider uncertainty in the ES estimate (e.g., by running different scenarios)
- Think in terms of range of values rather than a specific value

Power as fuel in the tank

- Have enough fuel to find what you are looking for (hoping that it is there) in a place at a distance that you hope have guessed reasonably well

Summing up Power Analysis II

- Increasing power means to decrease inference errors in general (direct effect on false negatives and indirect effect on false positives)
- More complex designs (e.g., multi-level) are challenging
- Sometimes no algorithmic solutions are available
- A simulation approach can be a valid solution (**see Giulio's presentation**)

Some readings for some advanced issues

Contrast, regression, moderation, and mediation effects

- Perugini, M., Gallucci, M., & Costantini, G. (2018). A Practical Primer To Power Analysis for Simple Experimental Designs. *International Review of Social Psychology, 31(1).*

Within and Mixed ANOVA

- Guo, Y., Logan, H. L., Glueck, D. H., & Muller, K. E. (2013). Selecting a sample size for studies with repeated measures. *BMC medical research methodology*, *13*(1), 100
- Rouder, J. N., & Haaf, J. M. (2018). Power, dominance, and constraint: A note on the appeal of different design traditions. *Advances in Methods and Practices in Psychological Science*, *1*, 19–26.
- Web app: GLIMMPSE (https://glimmpse.samplesizeshop.org)

Mixed/Multilevel Models

- Judd, C. M., Westfall, J., & Kenny, D. A. (2016). Experiments with more than one random factor: Designs, analytic models, and statistical power. *Annual Review of Psychology.* Web app: https://jakewestfall.shinyapps.io/two_factor_power/ See also Brysbaert, M., & Stevens, M. (2018). Power analysis and effect size in mixed effects models: a tutorial. *Journal of Cognition*, *1*(1).
- Kelcey, B., Xie, Y., Spybrook, J., & Dong, N. (2020). Power and sample size determination for multilevel mediation in three-level cluster-randomized trials. *Multivariate Behavioral Research* https://www.causalevaluation.org/power-analysis.html

Simulation based power analysis

- Gelman, A., Hill, J. (2006) *Data analysis using regression and multilevel/hierarchical models*. Cambridge: Cambridge University Press.

Advanced models and exemplary R code

- Liu, X. S. (2014). *Statistical Power Analysis for the Social and Behavioral Sciences: Basic and Advanced Techniques*. New York: Routledge.

and so…


KEEP
CALM
YOU'VE
GOT THE
POWER


I'VE GOT
THE POWER


KEEP
CALM
BECAUSE
WE'VE GOT
THE POWER

Power analysis
(part II)

Giulio Costantini,

Università degli Studi di Milano-Bicocca

giulio.costantini@unimib.it

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

Analytic power analysis

In simple cases, such as t-test or ANOVA, power analysis can be performed analytically.

Given three among **power**, **alpha level**, **effect size**, and **sample size**, you can determine the fourth analytically.

G*Power is a free software that is more practical than R for most simple situations.

https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

Power analysis for t-test

```
power.t.test(n = 100, # sample size in each group
             delta = .4, # true difference in means
             sd = 1, # SD
             sig.level = .05, # alpha
             power = NULL, # power
             type = "two.sample",
             alternative = "one.sided")
```

The element equal to NULL is the one estimated.

Instead of asking Cohen's d, the function asks for delta and sd. If sd = 1 (default), that's Cohen's d.

Output

```
     Two-sample t test power calculation

              n = 100
          delta = 0.4
             sd = 1
      sig.level = 0.05
          power = 0.8797802
    alternative = one.sided

NOTE: n is number in *each* group
```

Now you

Try exploring different scenarios:

- What if you change sample size?
- What if you change effect size?
- What if you run a paired-samples t-test instead?

Power and Simulation

The idea behind using simulation for power analysis is quite straightforward.

If the power of a statistical test is **the probability of successfully rejecting H0 if H1 is true**, one can determine power by

1. defining the expected values of the population parameters under H1
2. generating a sample of size N from the population parameters
3. testing the significance of the target effect using the preferred statistical method
4. replicate steps 2 and 3 a large number of times
5. estimating power as the proportion of simulated samples in which H0 is rejected

The necessary ingredients

- A set of values of population parameters under H1 (~effect size),
- A method for simulating data,
- A statistical test for deciding whether to reject H0.

Needless to say, R is great in dealing with this kind of situation!

Power and mediation analysis

In mediation analysis, we are interested in the indirect effect that a predictor X has on a response Y through a mediator M.

Within Baron & Kenny's (1986) framework, the indirect effect can be estimated as the product of two path coefficients, $a*b$. We will assume that all variables are standardized with *M* = 0 and *SD* = 1.

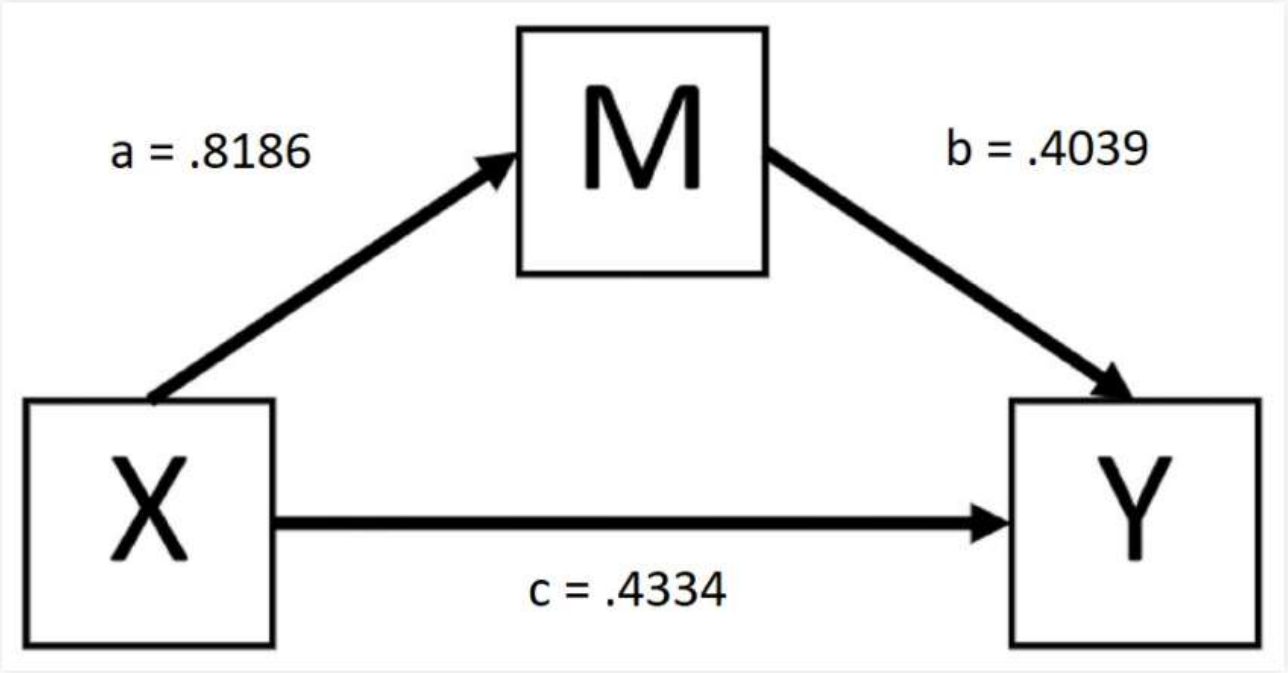
Refresher on mediation analysis (1)

In mediation analysis, we are interested in the indirect effect that a predictor X has on a response Y through a mediator M. Within Baron & Kenny's (1986) framework, the indirect effect can be estimated as the product of two path coefficients, $a * b$. We will assume that all variables are standardized with *M* = 0 and *SD* = 1.

Coefficient a is the slope of a simple linear regression

$$M = aX$$

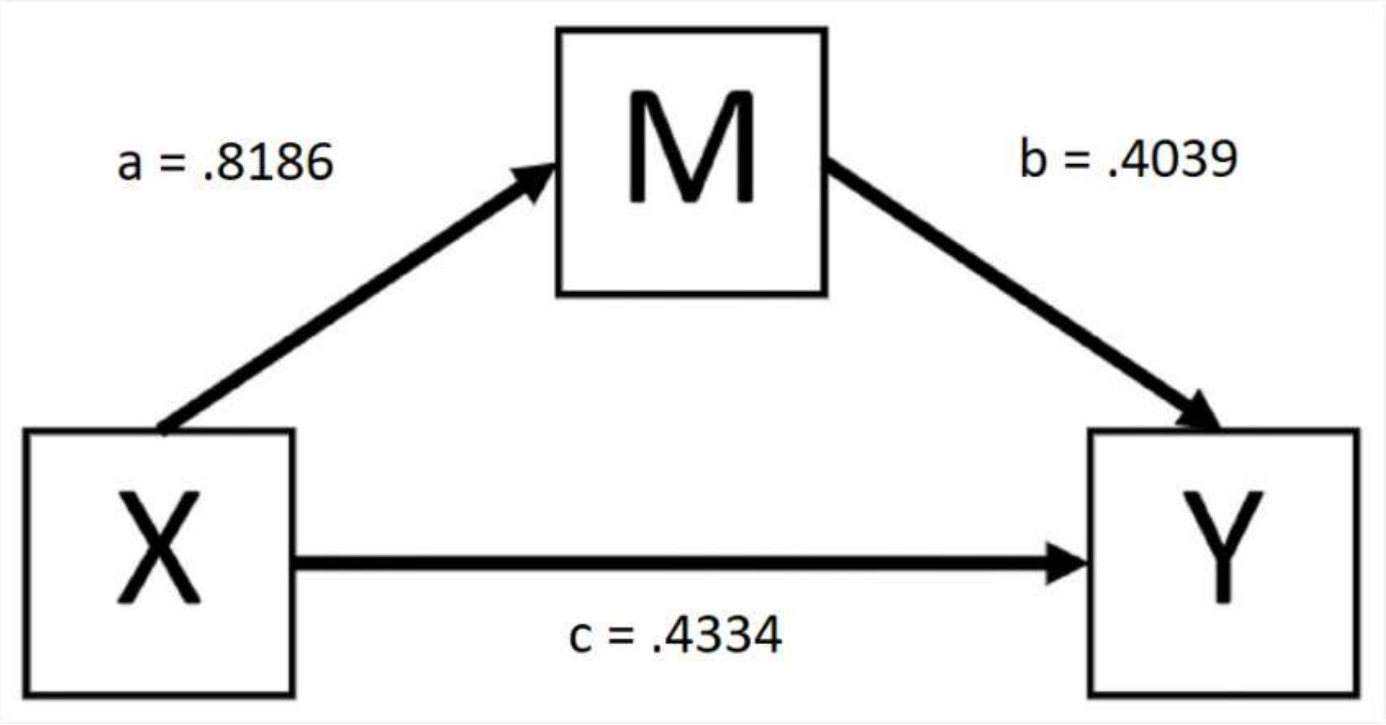
Coefficients b and c are the slopes of a multiple regression 0 $Y = bM + cX$



Refresher on mediation analysis (2)

In the example below, the indirect effect is $.8186 * .4039 = .3306$. This means that, when X increases of 1*SD*, Y increases through M of .3306 *SDs* and it increases of additional .4334 *SDs* independent of M.

Under the assumption that $a * b$ is normally distributed, the indirect effect can be tested through Sobel's (1982) test. Power for Sobel test can be determined analytically, for example with package **powerMediation** (Qiu, 2017).



But…

The assumption that the sampling distribution of the indirect effect a*b is normal is untenable. For this reason, the indirect effect is often tested using bootstrap (e.g., Preacher & Hayes, 2004).

If bootstrap is used, analytic formulas for power are not available.

As we have learned, in cases like this one, simulation can help. Fortunately, we do not need to set up a simulation from scratch!

Schoemann's approach

A very smart solution to do simulations quickly, by Schoemann (2017)

- Instead of bootstrap CIs, they use the so-called **Monte Carlo CIs**, which assume that a and b are normally distributed, without making assumptions on the distribution of their product (for details, see Preacher & Selig, 2012).
- Second, they use a **varying sample size approach**: Instead of fixing N for all simulated samples, a value of N is picked at random in the range of specified values. Each simulated sample has thus a different N.
- Third, the fact that a significant (vs. nonsignificant) result is obtained is predicted from N using a **logistic regression approach**.
- **Power** (i.e., the probability of obtaining a significant result) can be then **predicted from the logistic regression** equation given any sample size N within the range of interest.

Also, no programming required!

They developed a Shiny app, i.e., a web app running R in the background

https://schoemanna.shinyapps.io/mc_power_med/

Monte Carlo Power Analysis for Indirect Effects

Written by Alexander M. Schoemann (Contact), Aaron J. Boulton, & Stephen D. Short

Model: One Mediator

Objective: Set N, Find Power

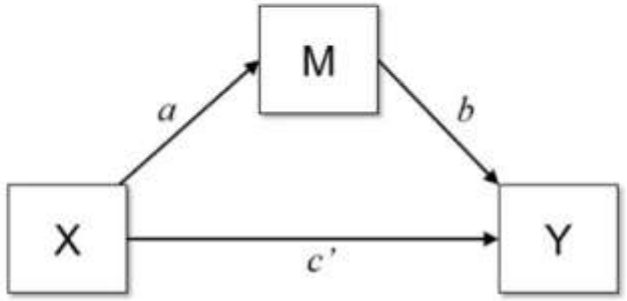
Sample Size (N): 100

# of Replications: 1000

Monte Carlo Draws per Rep: 200(

Random Seed: 1234

Confidence Level (%): 95



Instructions

To use this app, follow these steps:

1. Select Model. The user should first select the mediation model containing the indirect effect(s) of interest. Models may be selected in the drop-down menu in the left-most column of the app. Note that when a different mediation model is selected, the model graphic and input-value sections in the middle column will be altered.

Input Method: Standardized Coefficients

a	0.8186
b	0.4039
c'	0.4334

	X	M	Y
Std. Deviation	1.00	1.00	1.00

Calculate Power

Parameter	N	Power
ab	100	0.96

Power for mixed models

- We will use package **simr** (Green et al., 2016), which is in turn based on the the most widespread package for GLMM, **lme4** (Bates, 2015).
- We will focus on power for detecting fixed effects, which I assume to be by far the most widespread issue (see Green et al., 2016 or type ?powerSim for additional possibilities).

Mixed models and p-values?

A crucial ingredient of power analysis is a statistical test to reject H0.

- However, lme4 does not return p-values, and for a good reason: Degrees of freedom cannot be computed in mixed models (see this famous post by Bates, https://stat.ethz.ch/pipermail/r-help/2006-May/094765.html)
- Package **lmerTest** is a useful companion to lme4, as it includes p-values back in lme4 using the Satterthwaite degrees of freedom approximation (Kuznetsova et al., 2017).
- Since Satterthwaite approximation is quite widespread has been shown to perform satisfactorily (Luke, 2017), I will show how to implement power considering lmerTest's p-values.

Parameters of mixed models?

Another important ingrendient of power analysis is a **specification of model parameters under H1**.

This can be very complex in mixed models: The easiest way to use package simr is by having available a dataset (e.g., a previous study, a pretest) and by varying some parameters of interest (see Green et al., 2016 for how to specify a model from scratch).

In our examples, we will use the simon dataset.

Load packages and data

```
if(!require("pacman")) install.packages("pacman")
require(pacman)
p_load("dplyr", "simr", "lme4", "lmerTest")

load("data/simon.RData")
```

A simple mixed model

Let's fit a mixed model, in which RTs are predicted by congruency and target color. Only a random intercept by ID is included

```
model1 <- lmer(RT ~ congruency*target_color + (1|subject),
               data = simon)
```

A simple mixed model

Simon effect turns out to be significant: A facilitation of about 30ms is observed for congruent trials ($p < .001$), irrespective of target color.

```
Random effects:
 Groups   Name        Variance Std.Dev.
 subject  (Intercept)  5206     72.15
 Residual             65012    254.98
Number of obs: 26624, groups:  subject, 16

Fixed effects:
                            Estimate Std. Error        df t value Pr(>|t|)
(Intercept)                  399.164     18.307    15.678  21.804 3.78e-13 ***
congruency1                  -29.843      4.420 26605.000  -6.752 1.49e-11 ***
target_colorred               -5.834      4.420 26605.000  -1.320    0.187
congruency1:target_colorred    7.651      6.251 26605.000   1.224    0.221
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Power in simr

We want to know the power that we had to detect a Simon effect (i.e., main effect of congruency), with the current setup (number of individuals and trials).

In package **simr**, this is simply achieved with the function **powerSim**.

Power in simr

```
ps1 <- powerSim(model1, # fitted lme4 object

                # specification of the test of interest, i.e., a fixed
                # effect of congruency1 with t-test using Satterthwaite's
                # approximation
                test = fixed("congruency1", method = "t"),

                # number of simulated resamples
                nsim = 100)
```

This takes a while to compute. Lower nsim to get (imprecise) results quickly

Output

R reminds us that we should not do observed power analysis!

```
Power for predictor 'congruency1', (95% confidence interval):
      100.0% (96.38, 100.0)

Test: t-test with Satterthwaite degrees of freedom (package lmerTest)
      Effect size for congruency1 is -30.

Based on 100 simulations, (0 warnings, 0 errors)
alpha = 0.05, nrow = 26624

Time elapsed: 0 h 1 m 13 s

nb: result might be an observed power calculation
```

Modifying parameters

In simr, you start from some data and then you can easily change parameters and sample sizes.

e.g., let's decrease the effect of congruency to -10ms

```
model2 <- model1
fixef(model2)["congruency1"] <- -10
ps2 <- powerSim(model2,
                test = fixed("congruency1", method = "t"),
                nsim = 100)
```

Output

```
Power for predictor 'congruency1', (95% confidence interval):
      62.00% (51.75, 71.52)

Test: t-test with Satterthwaite degrees of freedom (package lmerTest)
      Effect size for congruency1 is -10.

Based on 100 simulations, (0 warnings, 0 errors)
alpha = 0.05, nrow = 26624
```

Let's change N and trials

Extend the sample size to N = 30

```
model3 <- extend(model2, along = "subject", n = 30)

ps3 <- powerSim(model3,
                test = fixed("congruency1", method = "t"),
                nsim = 100)
ps3
```

Let's reduce the #of obs within each subject (i.e, trials) to 50

```
model4 <- extend(model3, within = "subject", n = 50)
ps4 <- powerSim(model4,
                test = fixed("congruency1", method = "t"),
                nsim = 100)
ps4
```

Output

```
Power for predictor 'congruency1', (95% confidence interval):
      93.00% (86.11, 97.14)

Test: t-test with Satterthwaite degrees of freedom (package lmerTest)
      Effect size for congruency1 is -10.

Based on 100 simulations, (0 warnings, 0 errors)
alpha = 0.05, nrow = 49920
```

```
Power for predictor 'congruency1', (95% confidence interval):
       8.00% ( 3.52, 15.16)

Test: t-test with Satterthwaite degrees of freedom (package lmerTest)
      Effect size for congruency1 is -10.

Based on 100 simulations, (0 warnings, 0 errors)
alpha = 0.05, nrow = 1500
```

To sum up

These are only few examples of what you can achieve with simulations for power analysis.

Much more in simr as well: Check the paper

Green, P., & MacLeod, C. J. (2016). SIMR: an R package for power analysis of generalized linear mixed models by simulation. Methods in Ecology and Evolution, 7(4), 493-498. https://doi.org/10.1111/2041-210X.12504

Evaluative Conditioning and Evaluative Learning

Jan De Houwer,
Ghent University, Belgium
jan.dehouwer@ugent.be

Slides available on:
https://osf.io/dyjbq/

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

De Houwer (2007)

Lip Lab

"evaluative conditioning can be defined as an observed change in liking that is due to the pairing of stimuli"

<u>Defining elements:</u>

- observed change in liking : DV
- due to pairing of stimuli : IV

= effect = function: Liking = f(Pairings)

= hypothesis / claim

=> Not easy to establish EC but given experimental control, strong claims can be made

Useful definition?

1. It is verifiable (see previous slide)

2. It has adequate scope

2a. It excludes on the basis of

- Nature of response (e.g., not fear)
- Cause of change in response (e.g., not other regularities, maturation, priming, generalization, …)

Why?: Goal of EC research is to study one type of cause of changes in liking: pairings

Other possible causes of changes in liking:

- Other spatio-temporal regularities (e.g., repeated presence of one stimulus, operant contingency)
- Presence of other stimuli (e.g., priming)

prime: target = like
CS-US : CS = liked

no prime: target = neutral *no CS-US : CS = neutral*

- Similarity with other stimuli (e.g., generalization)

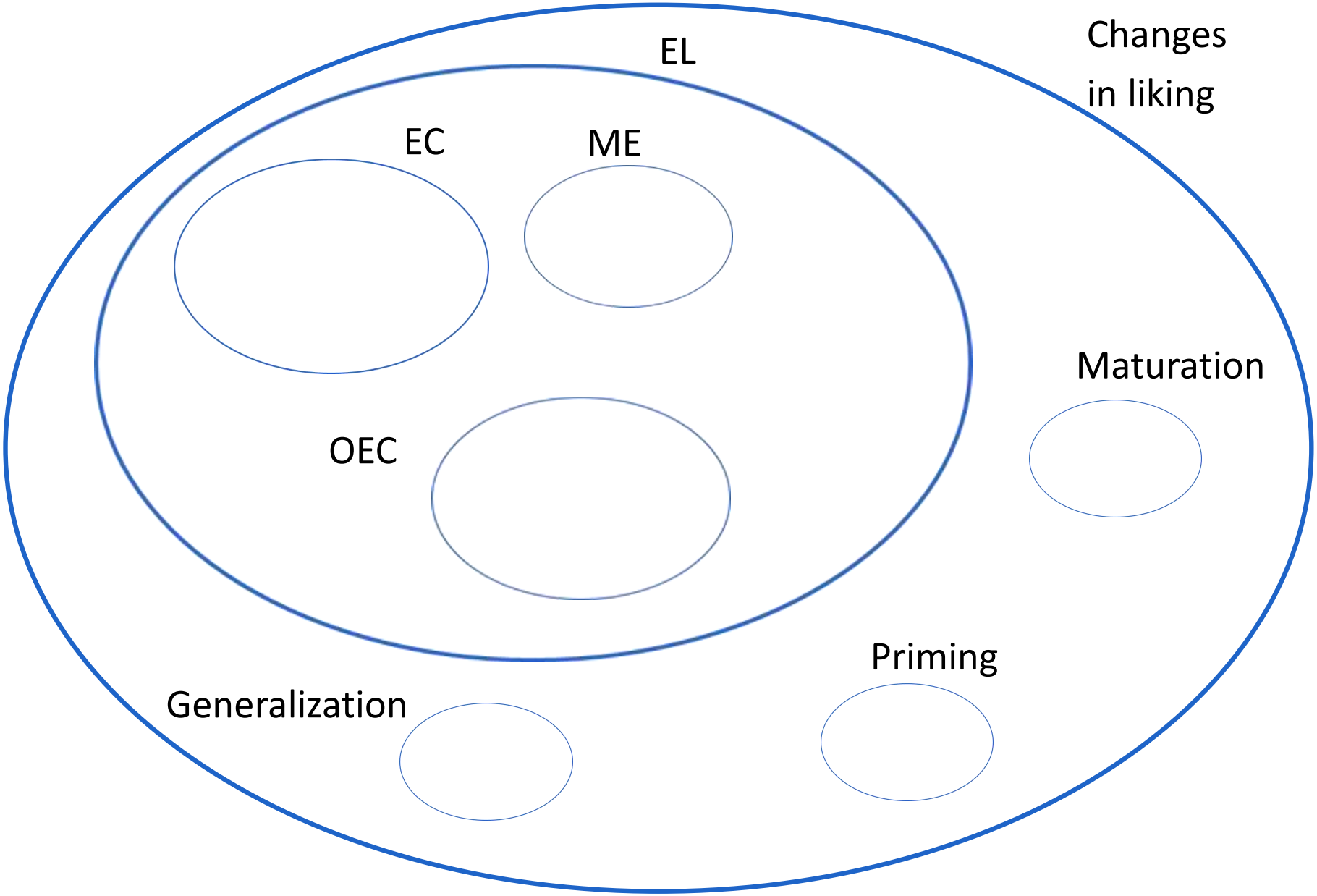
100Hz = liked => 90Hz = liked

10Hz = liked => 90Hz = neutral

=> generalization <> learning

=> “generalization” due to spatio-temporal similarity = instance of learning

=> Also see shared features effect (transfer, transformation) and effects of IR


Changes
in liking
EL
EC
ME
OEC
Maturation
Priming
Generalization

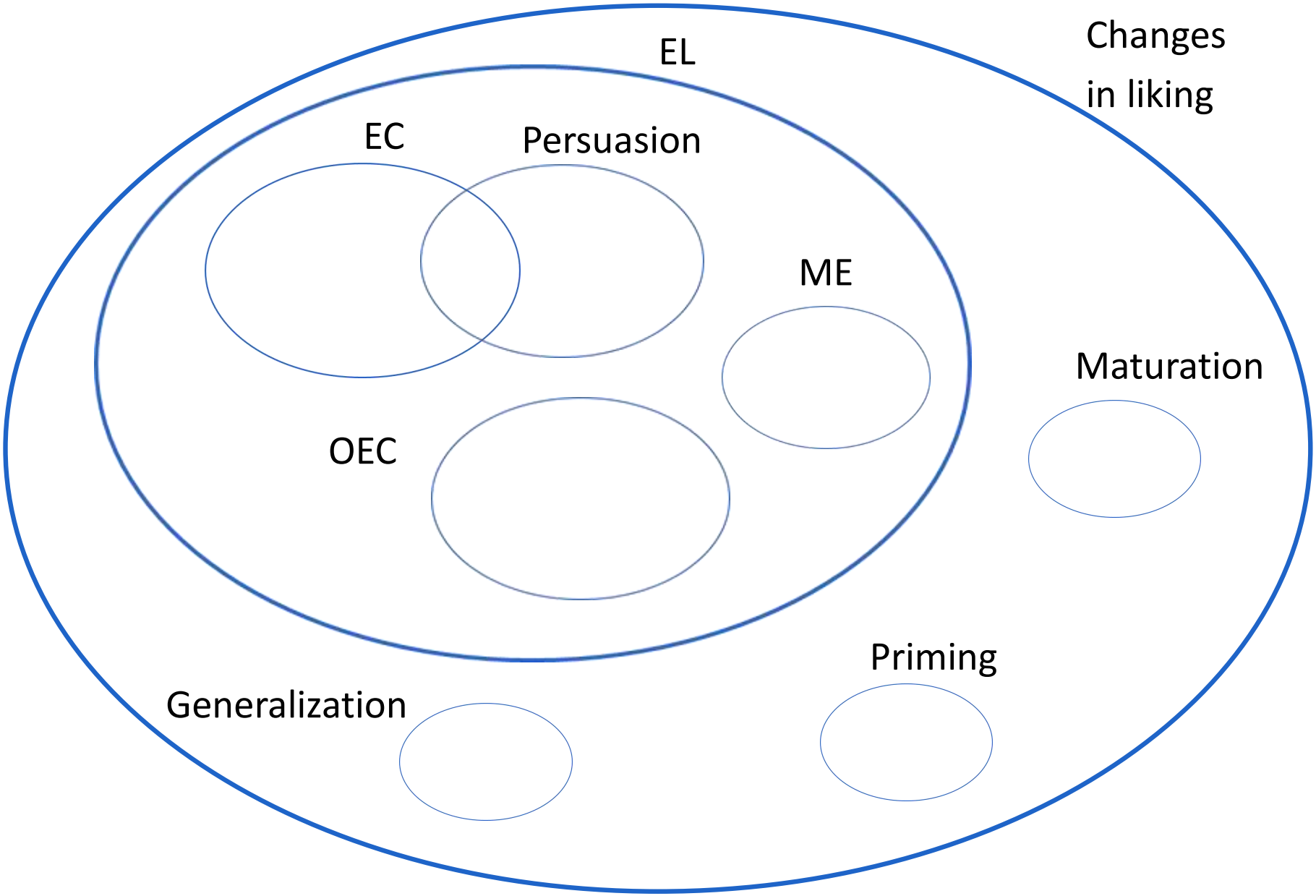
2a. It does not exclude on the basis of

- Nature of change (assimilation vs. contrast)
- Nature of US (biological vs. non-biological)
- Nature of "pairing" (co-occurrence vs. contingency): pairing = any regularity in spatio-temporal presence of two stimuli
- Nature of mental mechanism (association vs. proposition)
- Function of the pairing (simple cause vs. symbol)

=> pairing "stands for" (also see Morse code)

=> pairing = CS "is similar to" US

= (low validity/diagnisticity/how sure can you be) persuasive argument


Changes
in liking
EL
EC
Persuasion
ME
Maturation
OEC
Priming
Generalization

Recap: Is the definition useful?

1. It is verifiable

2. It has adequate scope

3. It helps us understand the mental mechanisms via which pairings influence liking

NOT: EC = proxy of association formation (i.e., presence & strength of EC = presence & strength of associations)

YES: Moderators of EC constrain theories about mental mechanisms that mediate EC (e.g., formation of associations or formation of proposition)

Types of moderators (De Houwer, 2011; DH & H, 2020):

- Nature of stimuli (e.g., CS, US, CS-US combination)
- Nature of (evaluative) behaviors (e.g., ratings, performance on implicit measures, approach/avoid)
- Nature of organism (e.g., “personality”, brain damage)
- Nature of broader context (e.g., secondary task, contextual cues)
- Nature of pairing (e.g., contiguity, contingency, conditional contingency/blocking eff, changes in contingency)

Task of cognitive models of EC is to explain existing and predict new knowledge about moderators of EC

=> Good model is not necessarily “true” but has large heuristic and predictive value

Recap: Is the definition useful?

1. It is verifiable

2. It has adequate scope

3. It helps us understand the mental mechanisms via which pairings influence liking

4. It helps us predict-and-influence behavior in real life?

Typical introduction of EC paper:

=> Evaluations (likes & dislikes, attitudes) are crucial for understanding behavior (psychopathology, consumer behavior, politics, …)

=> Hence vital to understand how evaluations are acquired

= aim of evaluative learning research

But: What has 45 years of EC research changed in the real world (outside of the lab)?: Sean, Jamie, NTBN

- Advertisements: are there but are they different/more effective because of EC research?
- Politics: Did EC research actually change (effectiveness of) political campaigns?
- Psychopathology: Did EC research actually help reduce psychological suffering?

=> No problem if ultimate aim is to understand mental mechanism but if ultimate aim is to predict-and-influence behavior, then we need to do reality checks (De Houwer et al., 2017; De Houwer, 2021: Pragmatic cognitive approach)

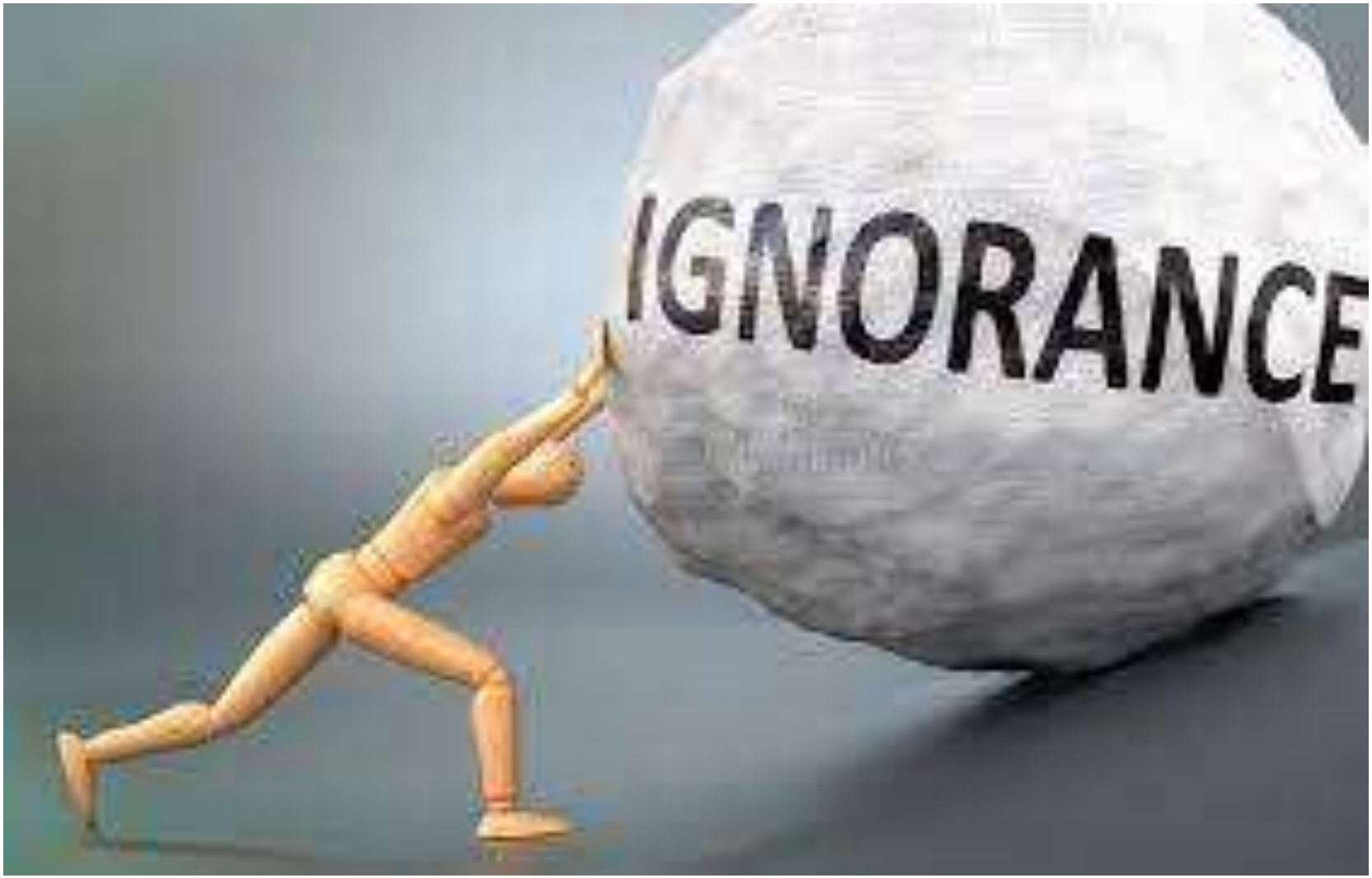
Possible reasons for limited real life impact?:

- Conceptual confusion: What is EC? What is "liking/evaluation"? (e.g., subjective feeling, approach/avoid "tendency")

- Loosing sight of distal goals: trapped into proximal goal of understanding mechanisms and moderators (asso vs. prop vs. dual) because of current academic incentives

- Urge to act while studying real life behavior is difficult

…


IGNORANCE


ALASKA
MONSTERS

Progress has been made:

- Conceptually
- Moderators of EC
- Mental mechanisms underlying EC
- Applied implications of more sophisticated view on (evaluative) conditioning and habits (e.g., in addiction research; e.g., Hogarth, 2020)

But maybe we can still do better:

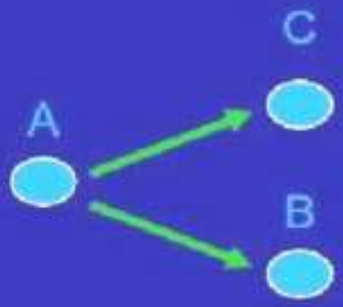
- Shift away from (measuring & changing) evaluation? Maybe "attitude" is not the most important concept in (social) psychology => both mentally & behaviorally at best one step in a chain of steps: focus on one step is useful only if measuring & influencing this step and downstream consequences are uncontroversial.
- Shift toward focus on changing real life behavior instead (also see Baumeister) or lab-based functional equivalent (which requires detailed functional analysis of real life behavior)
- As a community, push for clarity on goals and **reality checks** (be willing to kill rather than wait for death)
- Studying real life behavior is difficult

Causal models

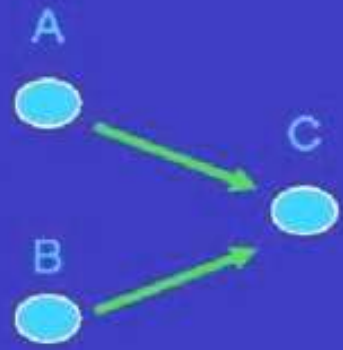
simple

chain

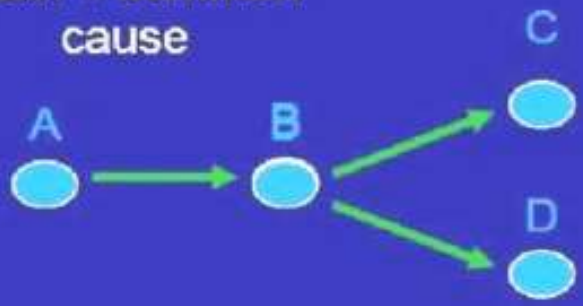
common cause



common effect



chain + common cause



long chain

Additional readings

- De Houwer, J. (2011). Evaluative conditioning: A review of functional knowledge and mental process theories. In T. R. Schachtman & S. Reilly (Eds). *Applications of learning and conditioning* (pp. 399-416). Oxford, UK: Oxford University Press.
- De Houwer, J., & Hughes, S. (2020). Learning to like or dislike: Revealing similarities and differences between evaluative learning effects. *Current Directions in Psychological Science, 29*, 487-491.
- De Houwer, J., & Hughes, S. (2020). *The Psychology of Learning: An Introduction From a Functional-Cognitive Perspective.* The MIT press.
- De Houwer, J. (2021). On the challenges of cognitive psychopathology research and possible ways forward: Arguments for a pragmatic cognitive approach. *Current Opinion in Psychology, 41*, 96-99.

Neuroticism and Emotional Vulnerability

Emanuele Preti,
University of Milano-Bicocca
emanuele.preti@unimib.it

Rossella Di Pierro,
University of Milano-Bicocca
rossella.dipierro@unimib.it

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

DIFFERENT BACKGROUND

-

COMMON LANGUAGE?

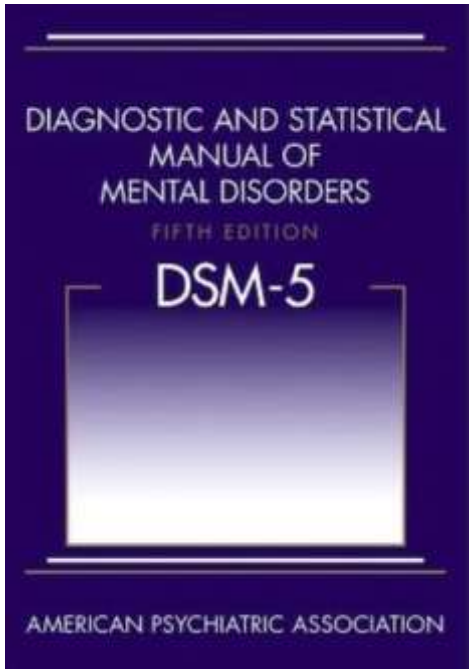
PDlab

www.pdlab.it

Overview

- **Dimensional models of psychopathology**
  AMPD (e.g., BPD), HiTop, Rdocs
- **Introduction to neuroticism**
- **Neuroticism and psychopathology**
  Models
  Example on narcissism

DSM & ICD: A categorical approach to psychopathology



- Categorical diagnosis
- Polythetic system
- Nomothetic approach
- A-theoretical system

Major Depressive Disorder

Diagnostic Criteria

A. Five (or more) of the following symptoms have been present during the same 2-week period and represent a change from previous functioning; at least one of the symptoms is either (1) depressed mood or (2) loss of interest or pleasure.

Note: Do not include symptoms that are clearly attributable to another medical condition.

1. Depressed mood most of the day, nearly every day, as indicated by either subjective report (e.g., feels sad, empty, hopeless) or observation made by others (e.g., appears tearful). (**Note:** In children and adolescents, can be irritable mood.)
2. Markedly diminished interest or pleasure in all, or almost all, activities most of the day, nearly every day (as indicated by either subjective account or observation).
3. Significant weight loss when not dieting or weight gain (e.g., a change of more than 5% of body weight in a month), or decrease or increase in appetite nearly every day. (**Note:** In children, consider failure to make expected weight gain.)
4. Insomnia or hypersomnia nearly every day.
5. Psychomotor agitation or retardation nearly every day (observable by others, not merely subjective feelings of restlessness or being slowed down).
6. Fatigue or loss of energy nearly every day.
7. Feelings of worthlessness or excessive or inappropriate guilt (which may be delusional) nearly every day (not merely self-reproach or guilt about being sick).
8. Diminished ability to think or concentrate, or indecisiveness, nearly every day (either by subjective account or as observed by others).
9. Recurrent thoughts of death (not just fear of dying), recurrent suicidal ideation without a specific plan, or a suicide attempt or a specific plan for committing suicide.

B. The symptoms cause clinically significant distress or impairment in social, occupational, or other important areas of functioning.

C. The episode is not attributable to the physiological effects of a substance or to another medical condition.

Borderline Personality Disorder

Diagnostic Criteria **301.83 (F60.3)**

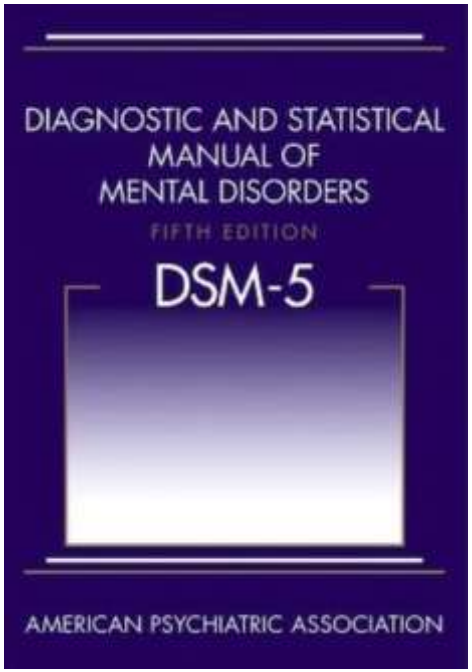
A pervasive pattern of instability of interpersonal relationships, self-image, and affects, and marked impulsivity, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:

1. Frantic efforts to avoid real or imagined abandonment. (**Note:** Do not include suicidal or self-mutilating behavior covered in Criterion 5.)
2. A pattern of unstable and intense interpersonal relationships characterized by alternating between extremes of idealization and devaluation.
3. Identity disturbance: markedly and persistently unstable self-image or sense of self.
4. Impulsivity in at least two areas that are potentially self-damaging (e.g., spending, sex, substance abuse, reckless driving, binge eating). (**Note:** Do not include suicidal or self-mutilating behavior covered in Criterion 5.)
5. Recurrent suicidal behavior, gestures, or threats, or self-mutilating behavior.
6. Affective instability due to a marked reactivity of mood (e.g., intense episodic dysphoria, irritability, or anxiety usually lasting a few hours and only rarely more than a few days).
7. Chronic feelings of emptiness.
8. Inappropriate, intense anger or difficulty controlling anger (e.g., frequent displays of temper, constant anger, recurrent physical fights).
9. Transient, stress-related paranoid ideation or severe dissociative symptoms.

Flaws in the Categorical Approach

- **Polythetic** (e.g., any 5 of 9 criteria for BPD) approach leads to heterogeneity within one personality disorder diagnosis
- Massive "**comorbidity**" of the personality diagnoses
- Factor analytic studies of DSM-IV PD at the criterion level **do not replicate** the categorical structure (Wright & Zimmermann, 2015)
- Across studies BPD criteria **interrelated** with criteria from every other PD
- **Poor clinical utility**: Cliniclians often do not use criteria (Westen & Arkowitz-Westen, 1998)
- Absence of **severity** assessment

DSM-5 Alternative Model of Personality Disorders (AMPD)



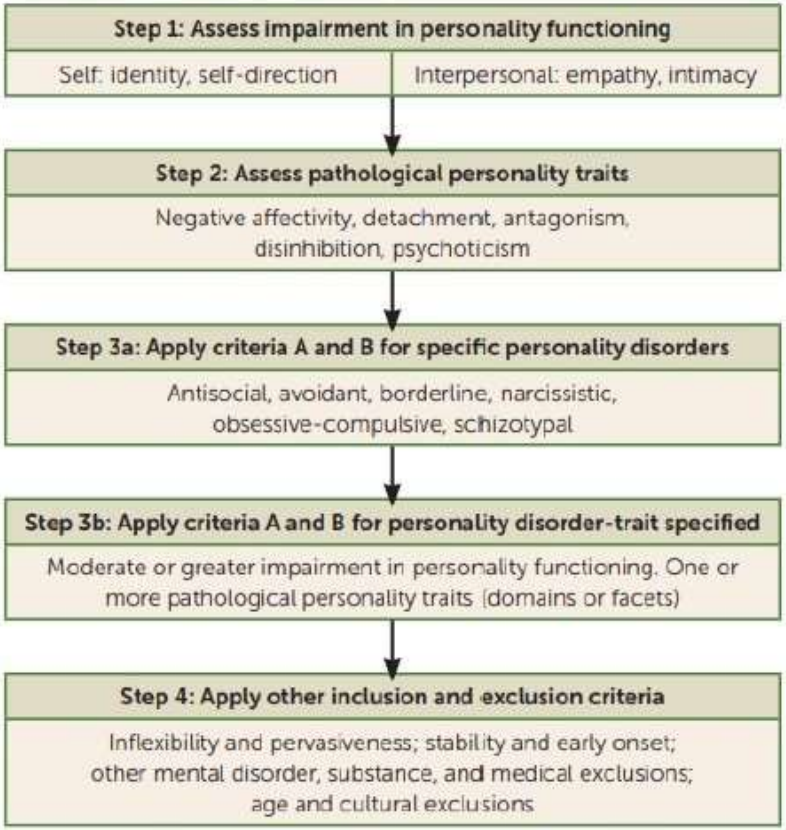
DSM-5 Section III

Hybrid model

PDs as combinations of maladaptive traits

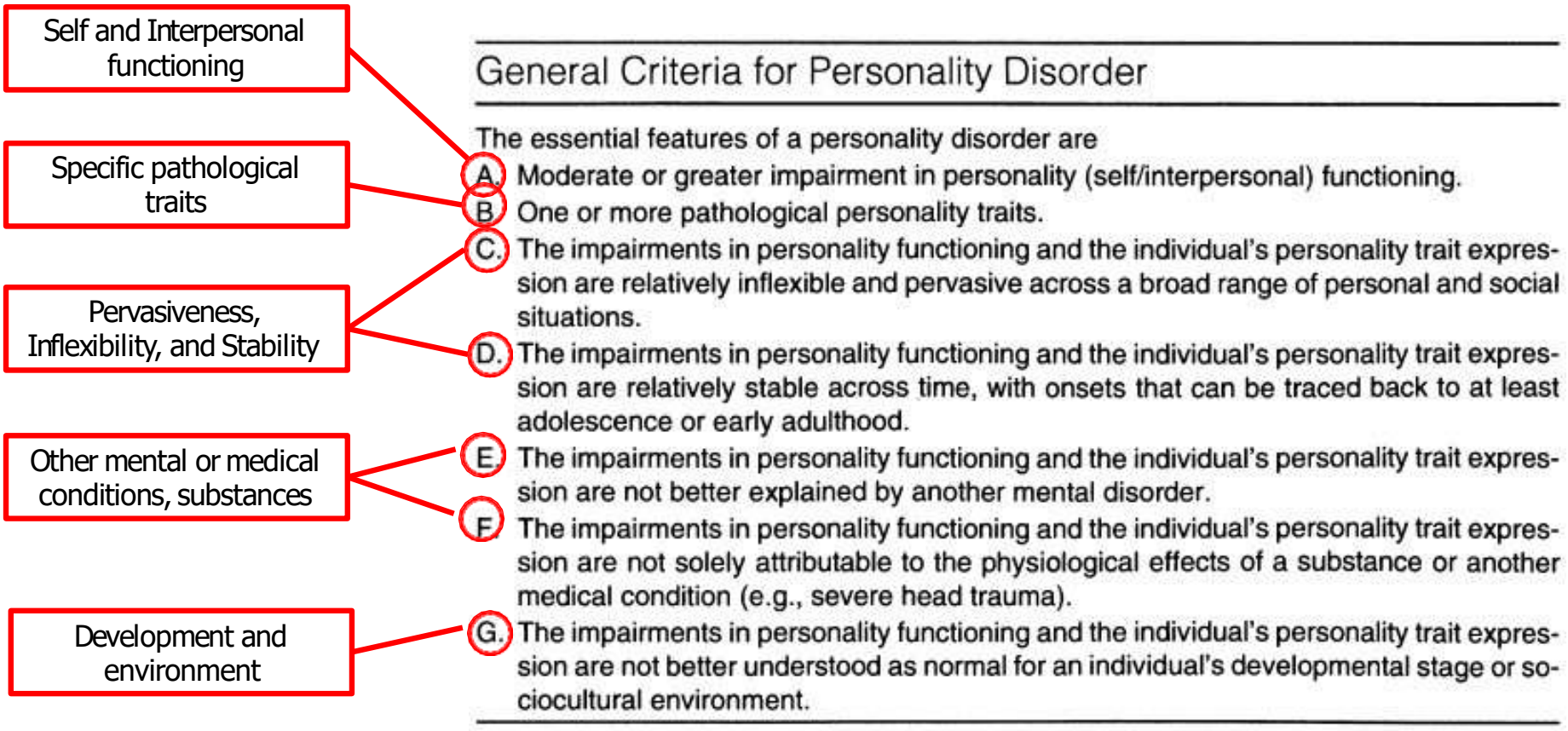
Level of Personality Functioning

FIGURE 1. Stepwise Approach to Assessment According to the General Criteria of the Alternative DSM-5 Model for Personality Disorders



Skodol, A.E. et al. (2015). The alternative DSM-5 model for personality disorders: a clinical application. *American Journal of Psychiatry, 172*(7), 606-613.

DSM-5 AMPD: General criteria for personality disorder



General Criteria for Personality Disorder

The essential features of a personality disorder are

A. Moderate or greater impairment in personality (self/interpersonal) functioning.

B. One or more pathological personality traits.

C. The impairments in personality functioning and the individual's personality trait expression are relatively inflexible and pervasive across a broad range of personal and social situations.

D. The impairments in personality functioning and the individual's personality trait expression are relatively stable across time, with onsets that can be traced back to at least adolescence or early adulthood.

E. The impairments in personality functioning and the individual's personality trait expression are not better explained by another mental disorder.

F. The impairments in personality functioning and the individual's personality trait expression are not solely attributable to the physiological effects of a substance or another medical condition (e.g., severe head trauma).

G. The impairments in personality functioning and the individual's personality trait expression are not better understood as normal for an individual's developmental stage or sociocultural environment.

Criterion A

TABLE 1 Elements of personality functioning

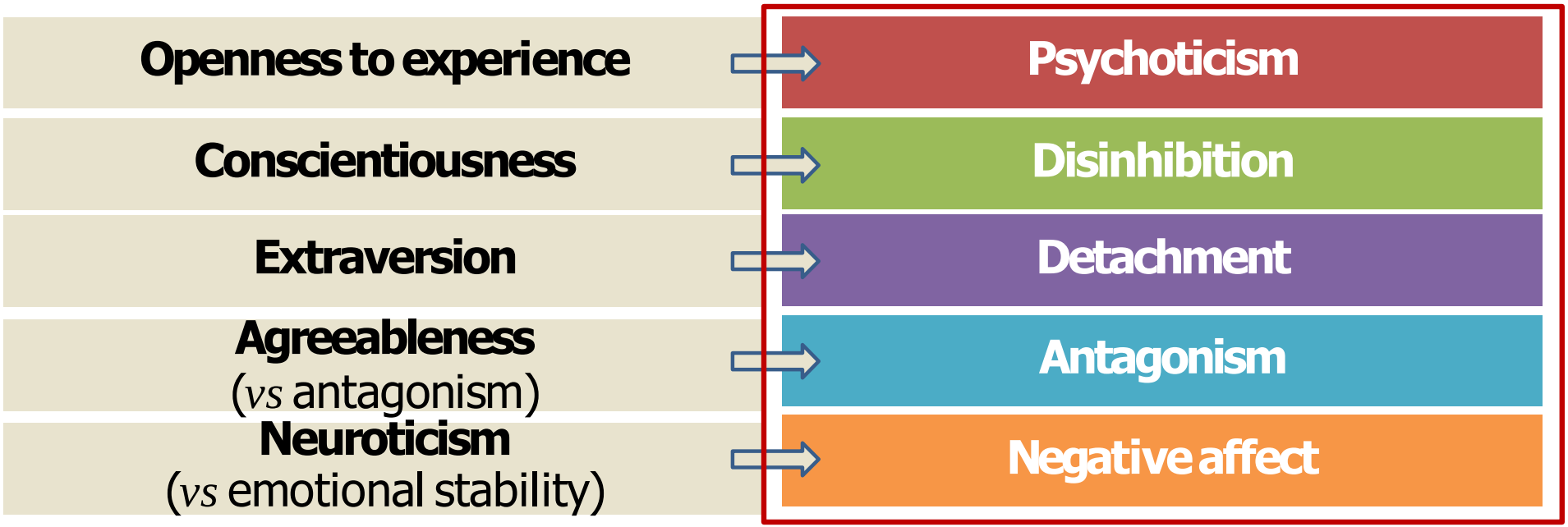
Self:

1. ***Identity:*** Experience of oneself as unique, with clear boundaries between self and others; stability of self-esteem and accuracy of self-appraisal; capacity for, and ability to regulate, a range of emotional experience.
2. ***Self-direction:*** Pursuit of coherent and meaningful short-term and life goals; utilization of constructive and prosocial internal standards of behavior; ability to self-reflect productively.

Interpersonal:

1. ***Empathy:*** Comprehension and appreciation of others' experiences and motivations; tolerance of differing perspectives; understanding the effects of one's own behavior on others.
2. ***Intimacy:*** Depth and duration of connection with others; desire and capacity for closeness; mutuality of regard reflected in interpersonal behavior.

Criterion B



Widiger, T.A. & McCabe, G.A. (2020). The Alternative Model of Personality Disorders (AMPD) from the perspective of the Five-Factor Model. *Psychopathology, 53*, 149-156.

Borderline Personality Disorder

Typical features of borderline personality disorder are instability of self-image, personal goals, interpersonal relationships, and affects, accompanied by impulsivity, risk taking, and/or hostility. Characteristic difficulties are apparent in identity, self-direction, empathy, and/or intimacy, as described below, along with specific maladaptive traits in the domain of Negative Affectivity, and also Antagonism and/or Disinhibition.

Proposed Diagnostic Criteria

A. Moderate or greater impairment in personality functioning, manifested by characteristic difficulties in two or more of the following four areas:

1. ***Identity:*** Markedly impoverished, poorly developed, or unstable self-image, often associated with excessive self-criticism; chronic feelings of emptiness; dissociative states under stress.
2. ***Self-direction:*** Instability in goals, aspirations, values, or career plans.
3. ***Empathy:*** Compromised ability to recognize the feelings and needs of others associated with interpersonal hypersensitivity (i.e., prone to feel slighted or insulted); perceptions of others selectively biased toward negative attributes or vulnerabilities.
4. ***Intimacy:*** Intense, unstable, and conflicted close relationships, marked by mistrust, neediness, and anxious preoccupation with real or imagined abandonment; close relationships often viewed in extremes of idealization and devaluation and alternating between overinvolvement and withdrawal.

B. Four or more of the following seven pathological personality traits, at least one of which must be (5) Impulsivity, (6) Risk taking, or (7) Hostility:

1. ***Emotional lability*** (an aspect of **Negative Affectivity**): Unstable emotional experiences and frequent mood changes; emotions that are easily aroused, intense, and/or out of proportion to events and circumstances.
2. ***Anxiousness*** (an aspect of **Negative Affectivity**): Intense feelings of nervousness, tenseness, or panic, often in reaction to interpersonal stresses; worry about the negative effects of past unpleasant experiences and future negative possibilities; feeling fearful, apprehensive, or threatened by uncertainty; fears of falling apart or losing control.
3. ***Separation insecurity*** (an aspect of **Negative Affectivity**): Fears of rejection by—and/or separation from—significant others, associated with fears of excessive dependency and complete loss of autonomy.
4. ***Depressivity*** (an aspect of **Negative Affectivity**): Frequent feelings of being down, miserable, and/or hopeless; difficulty recovering from such moods; pessimism about the future; pervasive shame; feelings of inferior self-worth; thoughts of suicide and suicidal behavior.
5. ***Impulsivity*** (an aspect of **Disinhibition**): Acting on the spur of the moment in response to immediate stimuli; acting on a momentary basis without a plan or consideration of outcomes; difficulty establishing or following plans; a sense of urgency and self-harming behavior under emotional distress.
6. ***Risk taking*** (an aspect of **Disinhibition**): Engagement in dangerous, risky, and potentially self-damaging activities, unnecessarily and without regard to consequences; lack of concern for one's limitations and denial of the reality of personal danger.
7. ***Hostility*** (an aspect of **Antagonism**): Persistent or frequent angry feelings; anger or irritability in response to minor slights and insults.

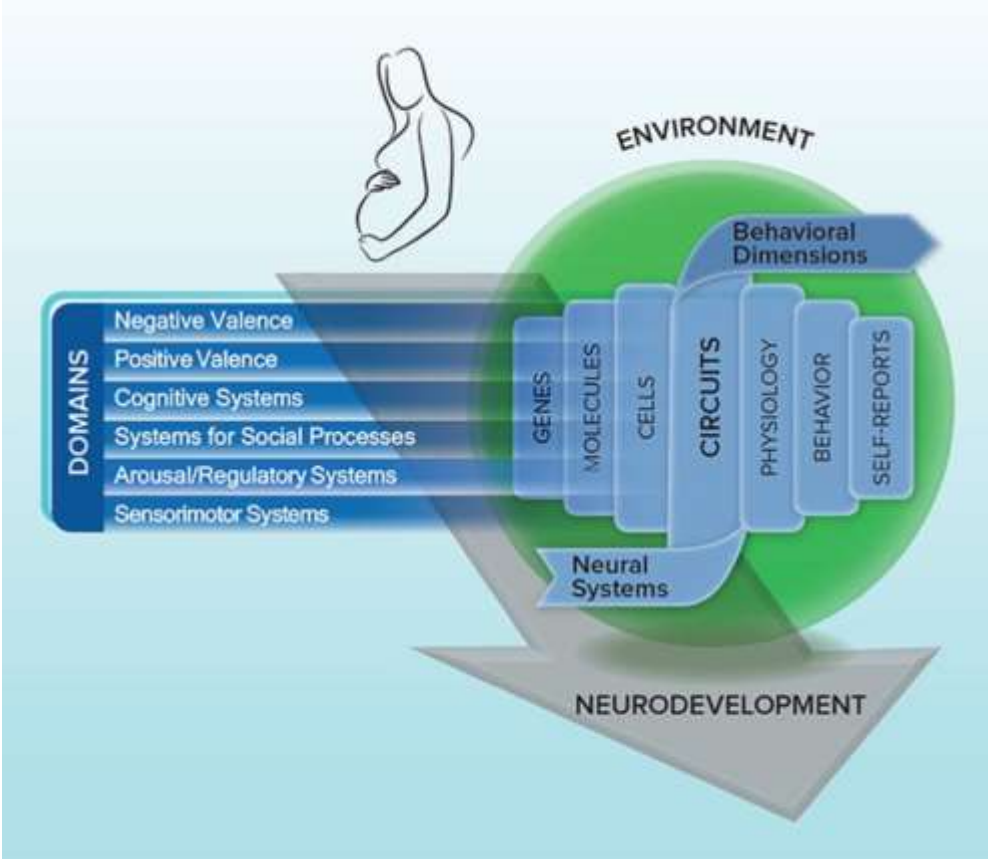
Further dimensional models of psychopathology

- The Research Domain Criteria (RDoC) Framework

NIMH Strategy and Goals Encouraging New Approaches to Diagnosing Mental Disorders to Facilitate Research Is the Foundation for RDoC

Strategy 1.4	Develop, for research purposes, new ways of classifying mental disorders based on dimensions of observable behavior and neurobiological measures.
Specific goals	Initiate a process for bringing together experts in clinical and basic sciences to jointly identify the fundamental behavioral components that may span multiple disorders (e.g., executive functioning, affect regulation, person perception) and that are more amenable to neuroscience approaches. Develop reliable and valid measures of these fundamental components of mental disorders for use in basic studies and in more clinical settings. Determine the full range of variation, from normal to abnormal, among the fundamental components to improve understanding of what is typical versus pathological. Integrate the fundamental genetic, neurobiological, behavioral, environmental, and experiential components that comprise these mental disorders.

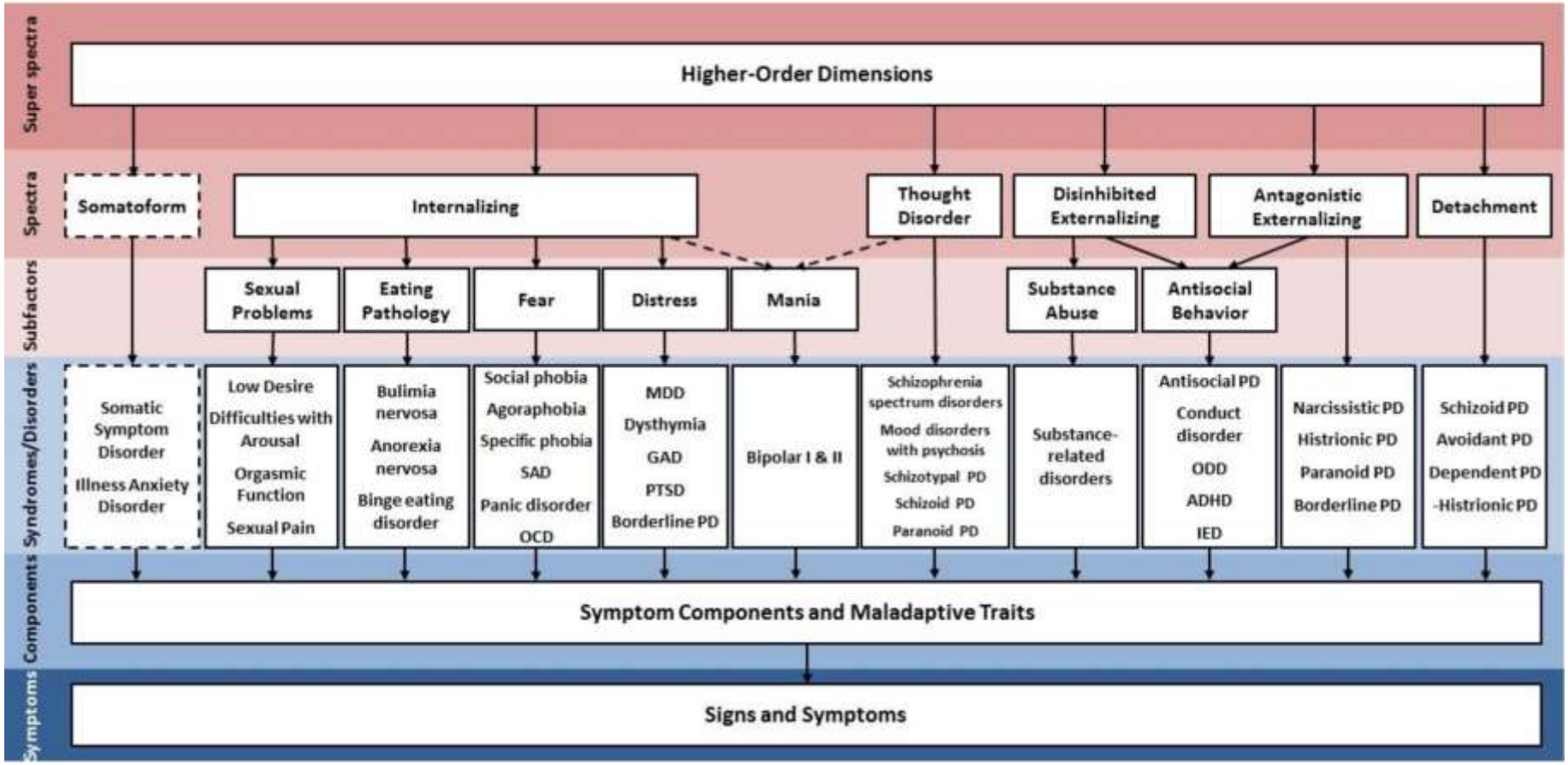
Note. NIMH = National Institute of Mental Health; RDoC = Research Domain Criteria.



Sanislow, et al. (2010). Developing constructs for psychopathology research: research domain criteria. Journal of abnormal psychology, 119(4), 631.

Further dimensional models of psychopathology

- The hierarchical taxonomy of psychopathology (HiTOP)



Kotov et al. (2017). A dimensional alternative to traditional nosologies. Journal of Abnormal Psychology, 126(4), 454–477

Neuroticism

A humble introduction

NEUROTICISM

Several years ago a prominent CEO of a national corporation traveled to our clinic, reporting that she had lost her ability to speak in corporate meetings and that, because this was absolutely essential to her position, her career was on the line. She recounted a meeting in which she had become furious at several participants who seemed to be conspiring against her. She was concerned that she would publicly explode in an unacceptable display of rage, and her anger quickly transformed into a full-blown panic attack, which rendered her unable to speak. After that experience, she stopped attending meetings out of fear of experiencing another panic attack if required to speak, but was running out of excuses. She noted that she typically excelled at running meetings and had been medically screened, so she knew that her inability to perform was a result of anxiety and that there would be no harm in returning to running meetings. In fact, it was the only rational thing to do. But she did not do it, playing out the ageold internal battle between reason and emotion in which, in the case of emotional disorders, emotion always wins.

Barlow, D. H., Curreri, A. J., & Woodard, L. S. (2021). Neuroticism and Disorders of Emotion: A New Synthesis. Current Directions in Psychological Science, 09637214211030253.

Historical journey of emotional disorders

- Mowrer (1950): calls the "**neurotic paradox**" the internal battle between reason and emotion which leads to the engagement of self-defeating and self-perpetuating behavior and deemed it as the "absolutely central problem in neurosis and therapy"

- Eysenck (1947) conceived neuroticism as one of the main biologically based dimensions of personality. Neuroticism is a higher-order temperamental factor related to the experience of frequent and intense negative emotions and it is essential for the development of nonpsychotic mental disorders due the interplay between emotional vulnerability and life stressors.

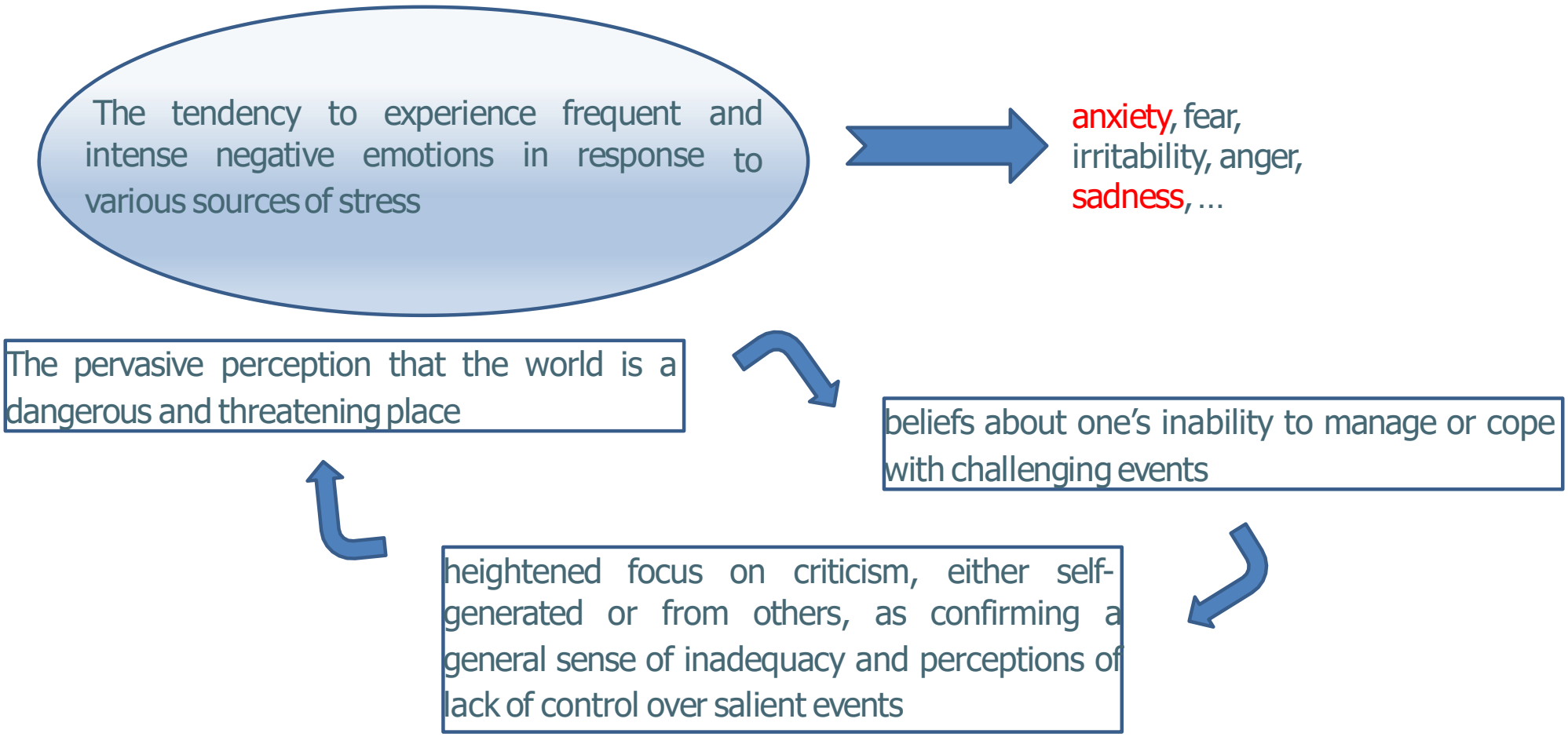
- Nowadays neuroticism occupies a central role in the most well-established models of personality (Clark & Watson, 2008)

NEUROTICISM (N)/(Negative) Emotionality (E)

- N is one of the five higher-order domains in the Five Factor Model (FFM; Costa & McCrae, 1992; Goldberg, 1993), mirroring the proneness toward negative affects (i.e., sadness, anxiety, and anger) and individual responses to threat, frustration, or loss (Widiger, 2009).

- N is included in other trait models of personality, but differences can be found at the facets level (e.g., HEXACO)

NEUROTICISM: definition



NEUROTICISM in the FFM

1. **Anxiety** – Level of Anxiety. How frequent and how easily one feels anxious.

2. **Angry Hostility** – Tendency to feel anger, frustration or bitterness.

3. **Depression** – Tendency to feel guilt, loneliness, depression.

4. **Self-Consciousness** – How easily one experiences Social Anxiety and Shyness.

5. **Impulsiveness** - Tendency to give in to cravings and ability to delay gratification.

6. **Vulnerability** – How one handles stress.

NEUROTICISM in the HEXACO model

Emotionality:

Persons with very high scores on the Emotionality scale experience fear of physical dangers, experience anxiety in response to life's stresses, feel a need for emotional support from others, and feel empathy and sentimental attachments with others.

Conversely, persons with very low scores on this scale are not deterred by the prospect of physical harm, feel little worry even in stressful situations, have little need to share their concerns with others, and feel emotionally detached from others.

EMOTIONALITY facets in the HEXACO model

1. **Fearfulness** - A tendency to experience fear. Low scorers feel little fear of injury and are relatively tough, brave, and insensitive to physical pain, whereas high scorers are strongly inclined to avoid physical harm.

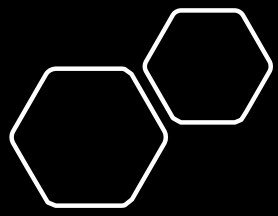
2. **Anxiety** - A tendency to worry in a variety of contexts. Low scorers feel little stress in response to difficulties, whereas high scorers tend to become preoccupied even by relatively minor problems.

3. **Dependence** - One's need for emotional support from others. Low scorers feel self-assured and able to deal with problems without any help or advice, whereas high scorers want to share their difficulties with those who will provide encouragement and comfort.

4. **Sentimentality** - A tendency to feel strong emotional bonds with others. Low scorers feel little emotion when saying good-bye or in reaction to the concerns of others, whereas high scorers feel strong emotional attachments and an empathic sensitivity to the feelings of others.

TABLE 3 Definitions of DSM-5 personality disorder trait domains and facets

DOMAINS (Polar Opposites) and Facets	Definitions
NEGATIVE AFFECTIVITY (vs. Emotional Stability)	Frequent and intense experiences of high levels of a wide range of negative emotions (e.g., anxiety, depression, guilt/ shame, worry, anger) and their behavioral (e.g., self-harm) and interpersonal (e.g., dependency) manifestations.
Emotional lability	Instability of emotional experiences and mood; emotions that are easily aroused, intense, and/or out of proportion to events and circumstances.
Anxiousness	Feelings of nervousness, tenseness, or panic in reaction to diverse situations; frequent worry about the negative effects of past unpleasant experiences and future negative possibilities; feeling fearful and apprehensive about uncertainty; expecting the worst to happen.
Separation insecurity	Fears of being alone due to rejection by—and/or separation from—significant others, based in a lack of confidence in one's ability to care for oneself, both physically and emotionally.
Submissiveness	Adaptation of one's behavior to the actual or perceived interests and desires of others even when doing so is antithetical to one's own interests, needs, or desires.
Hostility	Persistent or frequent angry feelings; anger or irritability in response to minor slights and insults; mean, nasty, or vengeful behavior. *See also* Antagonism.
Perseveration	Persistence at tasks or in a particular way of doing things long after the behavior has ceased to be functional or effective; continuance of the same behavior despite repeated failures or clear reasons for stopping.
Depressivity	*See* Detachment.
Suspiciousness	*See* Detachment.
Restricted affectivity (lack of)	The ***lack of*** this facet characterizes ***low levels*** of Negative Affectivity. *See* Detachment for definition of this facet.



NA in the DSM-5 AMPD

Genetic Influences

- Evidence that N is heritable. Approximately half the variance in N is attributable to additive genetic influences (e.g., Lahey, 2009).
- Only 15% when considering single-nucleotide polymorphisms

Environmental Influences

- **Nonshared** environment accounts for a substantial portion of the variance in trait N (Fullerton, 2006; Lake et al., 2000).
- Parenting and familial factors (e.g., Allen & Lauterbach, 2007), but mostly retrospective (recall) studies.

Neuroticism and Physical Health

- Stronger perceptions of physical health problems: linked to somatic complaints without medical support (e.g., Powers & Oltmanns, 2013)
- Distorted cognition regarding symptoms, which results in greater use of medical services (Goubert, Crombez, & Van Damme, 2004).

Indirect associations (e.g., considering internalizing psychopathology):

- problems with cardiac functioning (Barger & Sydeman, 2005)
- increased mortality (e.g., Robles et al., 2005)
- Disrupted immune functioning (e.g., Pace et al., 2006).

Direct associations:

- Asthma (Huovinen, Kaprio, & Koskenvuo, 2001)
- Atopic eczema (Buske-Kirschbaum, Geiben, & Hellhammer, 2001)
- Cardiovascular disease (Suls & Bunde, 2005)
- Irritable bowel syndrome (Spiller, 2007)

Neuroticism and Quality of Life

"Emotional Stability":

- Higher marital satisfaction (Gattis, Berns, Simpson, & Christensen, 2004)
- Greater occupational success (e.g., Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007)
- Quality of life (e.g., Ozer & Benet-Martínez, 2006).

Neuroticism:

- Social impairment (Mullins-Sweatt & Widiger, 2010)
- Burnout and emotional exhaustion (Armon, Shirom, & Melamed, 2012).

N predicts overall well-being and emotional health more than socioeconomic status (McCann, 2011).

Overlapping genetic influences at least partly account for associations between N and subjective well-being (Weiss, Bates, and Luciano, 2008).

NEUROTICISM AND PSYCHOPATHOLOGY

PSYCHOPATHOLOGY

CLINICAL DISORDERS

- Egodystonic – The patient is aware that something is wrong with him/her
- Psychological distress
- Onset in adulthood
- Moments, situations, period of time

PERSONALITY DISORDERS

- Egosyntonic – The patient is not fully aware of having a disorder
- «*Enduring pattern of inner experience and behavior that is inflexible and pervasive*»
- Early age of onset (adolescence)
- Effects on everyday functioning

INTERNALIZING DISORDERS

- Include a variety of over-inhibited or internally-focused symptoms

- Anxiety symptoms
- Depressive symptoms

EXTERNALIZING DISORDERS

- Include a variety of disinhibited or externally-focused symptoms

- Aggression
- Conduct problems
- Hyperactivity

EMPIRICAL EVIDENCE

- Strong associations with clinical disorders
  - ❑ Anxiety
  - ❑ Depression
  - ❑ Substance use disorders

(e.g., Clark, Watson, & Mineka, 1994; Kotov, Gamez, Schmidt, & Watson, 2010; Lahey, 2009; Malouff, Thorsteinsson, & Schutte, 2005; Ormel, Oldehinkel, & Brilman, 2001; Ormel & Wohlfarth, 1991; Ruiz, Pincus, & Schinka, 2008)

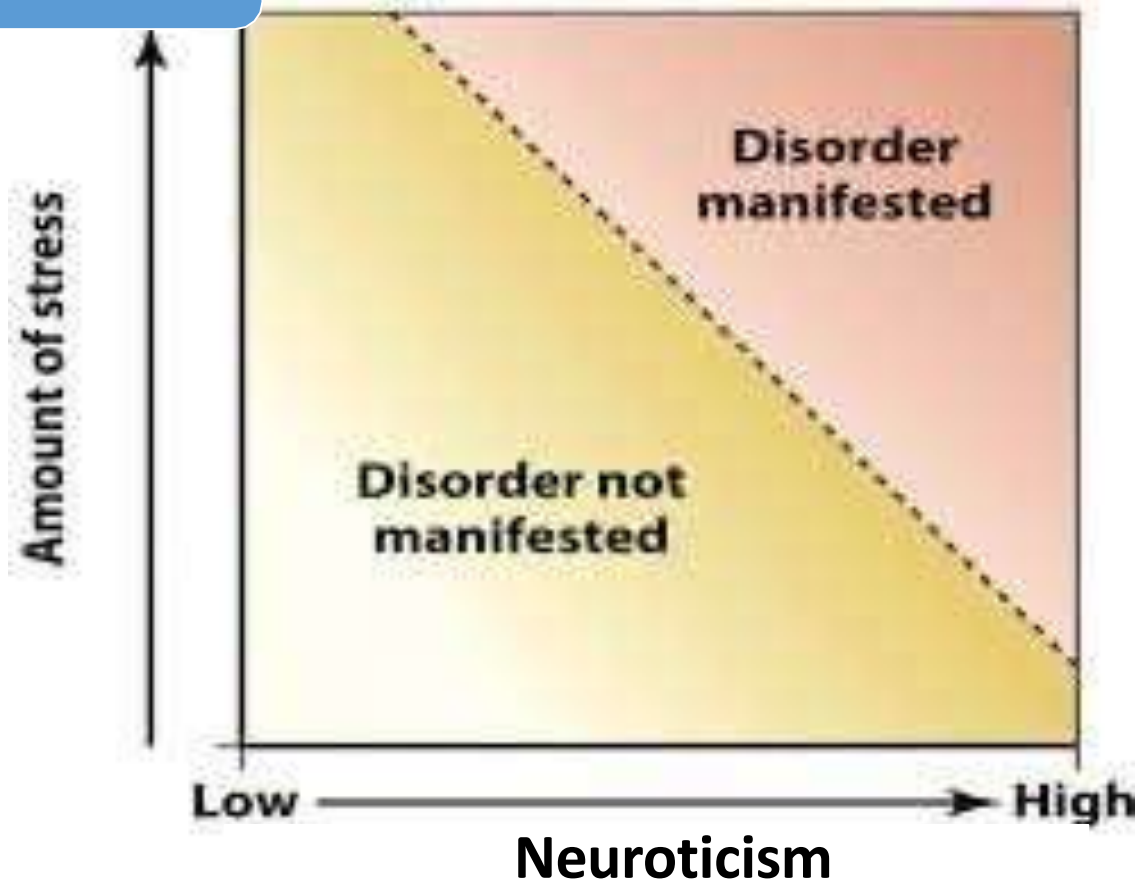
- Associations with Personality Disorders
  - ❑ Borderline
  - ❑ Narcissism

HOW TO EXPLAIN THESE ASSOCIATIONS

- VULNERABILITY MODEL
- SPECTRUM MODEL
- COMMON CAUSE MODEL
- SCAR MODEL
- STATE MODEL

VULNERABILITY MODEL

Neuroticism makes people vulnerable (diathesis) to psychopathology



Empirical findings

N predicts later major depression (Fanous, Neale, Aggen, & Kendler, 2007; Kendler, Neale, Kessler, Heath, & Eaves, 1993) and suicide (Fergusson, Woodward, & Horwood, 2000).

People high in N are at higher risk for internalizing psychopathology following exposure to stressful life events than individuals low in N who are exposed to the same events (Fanous et al., 2002; Hutchinson & Williams, 2007; Jacobs et al., 2006; Kendler et al., 2004; Parslow et al., 2006).

Neuroticism sets in motion processes that lead to psychopathology

- ❑ Negative bias in attention
- ❑ Negative bias in interpretation and recall of information
- ❑ Increased emotional reactivity
- ❑ Ineffective coping

- Some of these processes (e.g., attention bias and repetitive thinking) are psychopathological symptoms (Mathews & MacLeod, 2005; Ehring & Watkins, 2008)

SPECTRUM MODEL

Neuroticism and psychopathology are extremes of a continuum.

- Common determinants account for the association between neuroticism and CMDs (e.g., genetic and environmental factors)
- High neuroticism scores are equivalent to symptoms of clinical disorders (overlap in measurement content)

COMMON CAUSE MODEL

Neuroticism is predictive for psychopathology because the two constructs share genetic and environmental determinants.

(e.g., Carey & DiLalla, 1994; Fanous, Gardner, Prescott, Cancro, & Kendler, 2002; Hettema, Neale, Myers, Prescott, & Kendler, 2006; Mikolajewski, Allan, Hart, Lonigan, & Taylor, 2013; Silberg, Rutter, Neale, & Eaves, 2001; Stein & Stein, 2008; Tackett et al., 2012, 2013).

SCAR MODEL

Neuroticism is shaped by psychopathology

The experience of a major episode of psychopathology has permanent effects on neuroticism (they persist after the episode has remitted)

STATE MODEL

The state model also asserts that neuroticism is shaped by psychopathology but, in contrast with the scar model, argues that the effects of psychopathology on neuroticism are temporary and disappear after the episode has remitted

"It is important to note that the models are not mutually exclusive and that the borders between them are blurry" (Ormel et al., 2013)

Clinical Psychology Review 33 (2013) 686–697

Contents lists available at SciVerse ScienceDirect

Clinical Psychology Review

Neuroticism and common mental disorders: Meaning and utility of a complex relationship

Johan Ormel [a,*], Bertus F. Jeronimus [a,1], Roman Kotov [c,1], Harriëtte Riese [a,b], Elisabeth H. Bos [a], Benjamin Hankin [d], Judith G.M. Rosmalen [a], Albertine J. Oldehinkel [a]

Evaluation of models in light of the evidence on neuroticism (N) and internalizing disorders (INT).

The evidence	State	Scar	Vulnerability	Common cause	Spectrum
Cross-sectional association neuroticism–CMD	++	++	+/−	++	++
Prospective association neuroticism–CMD	-	+/−	++	+/−	-
N enhances life stress effect on psychopathology	-	-	++	+/−	-
Substantial item content overlap	-	-	-	-	++
Partly shared determinants	-	-	-	++	+/−
Partly unique determinants	-	-	++	-	-
Synchrony of change	+/−	+/−	-	+	+/−
N is more stable than INT	+/−	-	+ (1)	+ (2)	-
Post-episode N probably not higher than pre-episode N	+	-	+/−	+/−	+/−
Treatment effect mediated by N	-	+/−	++	+/−	+/−

Note. + (++), (strongly) consistent with model; - (−), (strongly) inconsistent with model.
+/−, lacks a clear implication. 1, if difference is marked. 2, if difference is small.

LACK OF A CONCLUSIVE MODEL

It is important to take into account the specificity of psychopathological conditions

Anxiety disorders: the evidence favors a major role for the common cause model.

Depressive disorders: common cause is the best model.

Substance use disorders: the vulnerability model is empirically supported.

Does neuroticism have any utility in understanding and studying psychopathology?

- ❑WE DO NOT HAVE A CLEAR PICTURE OF THE ROLE OF NEUROTICISM IN PSYCHOPATHOLOGY
- ❑STRONG ASSOCIATIONS WITH SEVERAL PSYCHOPATHOLOGICAL CONDITIONS
- ❑ NEUROTICISM CAN BE CONSIDERED A **TRANSDIAGNOSTIC FEATURE** OF PSYCHOPATHOLOGY (INTERNALIZING CLINICAL DISORDERS)

WHAT ABOUT THE OTHER PERSONALITY TRAITS?

Other personality traits are associated with psychopathology:

Low Conscientiousness
(Dis)inhibition

EXTERNALIZING DISORDERS
(e.g., Substance Use Disorders)

but

- associations are weaker than those with neuroticism
- associations only with specific disorders

Neuroticism and Disorders of Emotion: A New Synthesis

Current Directions in Psychological Science
1–8

DOI: 10.1177/09637214211030253
www.psychologicalscience.org/CDPS
SAGE

David H. Barlow, Andrew J. Curreri, and Lauren S. Woodard
Center for Anxiety and Related Disorders, Boston University

The authors propose a model to explain why and how neuroticism is associated with the development and maintenance of Emotional disorders.

"*At the core of **neuroticism** is the **experience of intense and frequent negative emotionality** accompanied by a **sense of uncontrollability and unpredictability** of stressful or challenging events.*"

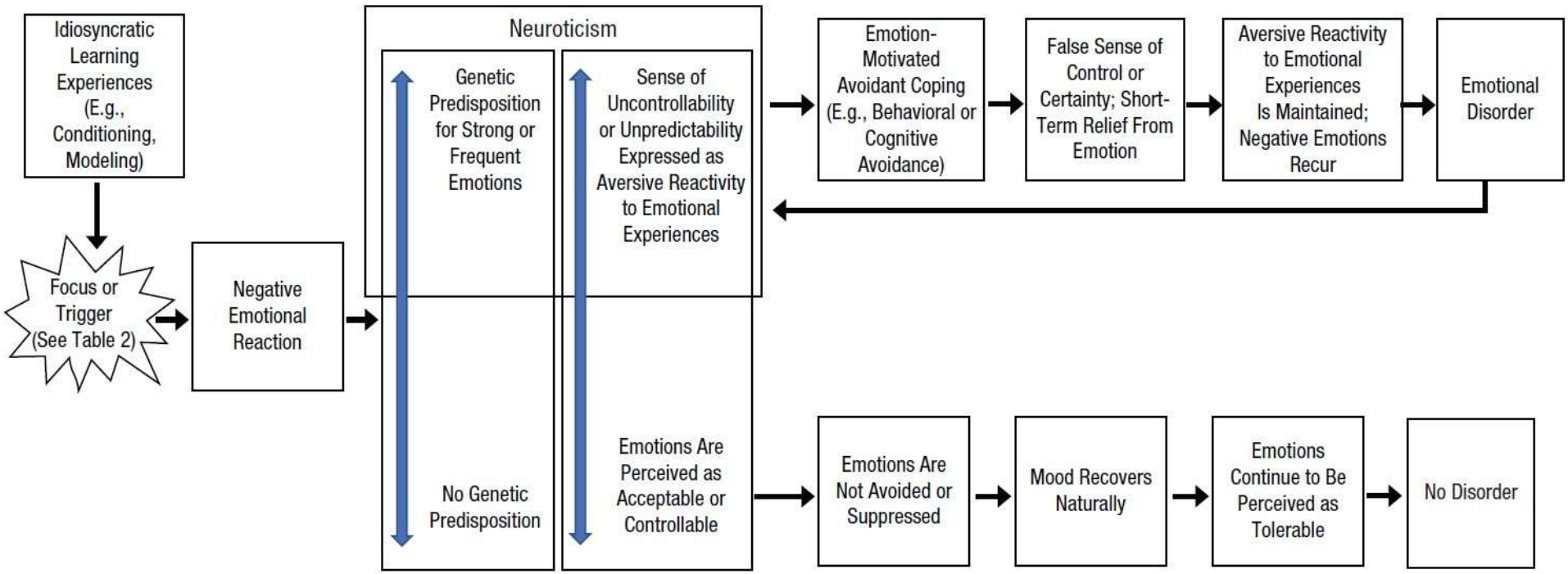


Fig. 1. Model of the functional mechanisms in the development of emotional disorders versus normal emotional experience (adapted from Sauer-Zavala & Barlow, in press). Emotional triggers lead to negative emotional reactions in most people; however, the consequences of these negative emotional reactions differ depending on the individual's level of neuroticism. The top pathway describes how individuals with a more neurotic temperament, who have also developed a sensitivity to certain triggers through learning experiences early in life, tend to engage in a set of processes that culminate in the development and maintenance of emotional disorders. In contrast, the bottom pathway depicts the cascade of adaptive responses to negative emotions that characterizes healthy emotional functioning in individuals lower in neuroticism.

EMOTIONAL-MOTIVATED AVOIDANT COPING

Strategies used to avoid or suppress negative affective states

- **B**ehaviorally
- **C**ognitively

Table 1. Some Examples of Emotion-Motivated Avoidant-Coping Strategies

Category	Examples
Overt behavioral avoidance	Avoiding situations or activities that trigger strong emotions
Subtle behavioral avoidance	Overpreparation Perfectionism Reassurance seeking
Cognitive avoidance	Worry Rumination Distraction
Safety behaviors	"Superstitious" anxiolytic behavior, such as carrying an empty pill bottle

SPECIFICITY OF DISORDERS

Table 2. Precipitating Events (Triggers) for Negative Emotions and Possible Emotional Disorder (Adapted From Sauer-Zavala & Barlow, in press)

Trigger	Disorder
Nonclinical panic attack	Panic disorder
Intrusive ego-dystonic thoughts	Obsessive-compulsive disorder
Uncomfortable somatic sensations	Illness anxiety disorder or panic disorder
Interpersonal conflict	Borderline personality disorder
Insufficient sexual arousal or erectile failure	Sexual dysfunction
Restless, unsatisfying sleep	Insomnia disorder
Loss	Depression or prolonged grief

PERSONALITY PATHOLOGY

The DSM-5 Alternative Model

A. IMPAIRMENT IN PERSONALITY FUNCTIONING
   1) IDENTITY
   2) SELF-DIRECTION
   3) EMPATHY
   4) INTIMACY

B. PATHOLOGICAL PERSONALITY TRAITS
   1) DETACHMENT
   2) NEGATIVE AFFECTIVITY
   3) PSYCHOTICISM
   4) DISINHIBITION
   5) ANTAGONISM

NEUROTICISM AND PERSONALITY DISORDERS

- BORDERLINE PERSONALITY DISORDER
- NARCISSISTIC PATHOLOGY

BORDERLINE PERSONALITY DISORDER

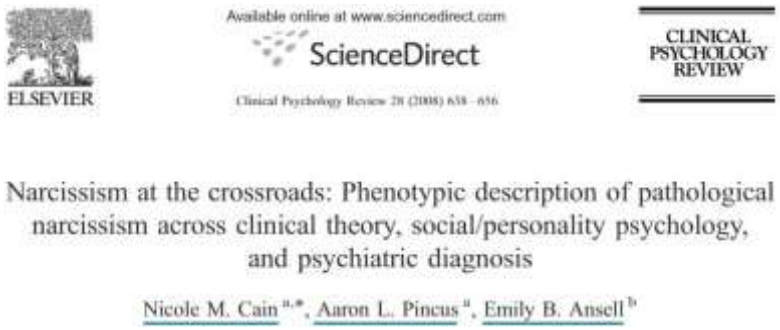
B. Four or more of the following seven pathological personality traits, at least one of which must be (5) Impulsivity, (6) Risk taking, or (7) Hostility:

1. ***Emotional lability*** (an aspect of **Negative Affectivity**): Unstable emotional experiences and frequent mood changes; emotions that are easily aroused, intense, and/or out of proportion to events and circumstances.
2. ***Anxiousness*** (an aspect of **Negative Affectivity**): Intense feelings of nervousness, tenseness, or panic, often in reaction to interpersonal stresses; worry about the negative effects of past unpleasant experiences and future negative possibili-

ties; feeling fearful, apprehensive, or threatened by uncertainty; fears of falling apart or losing control.

3. ***Separation insecurity*** (an aspect of **Negative Affectivity**): Fears of rejection by—and/or separation from—significant others, associated with fears of excessive dependency and complete loss of autonomy.
4. ***Depressivity*** (an aspect of **Negative Affectivity**): Frequent feelings of being down, miserable, and/or hopeless; difficulty recovering from such moods; pessimism about the future; pervasive shame; feelings of inferior self-worth; thoughts of suicide and suicidal behavior.
5. ***Impulsivity*** (an aspect of **Disinhibition**): Acting on the spur of the moment in response to immediate stimuli; acting on a momentary basis without a plan or consideration of outcomes; difficulty establishing or following plans; a sense of urgency and self-harming behavior under emotional distress.
6. ***Risk taking*** (an aspect of **Disinhibition**): Engagement in dangerous, risky, and potentially self-damaging activities, unnecessarily and without regard to consequences; lack of concern for one's limitations and denial of the reality of personal danger.
7. ***Hostility*** (an aspect of **Antagonism**): Persistent or frequent angry feelings; anger or irritability in response to minor slights and insults.

NARCISSISTIC PERSONALITY PATHOLOGY



- Pathological narcissism has a long history in the psychodynamic clinical literature (Kohut, Kernberg..)
- Narcissistic pathology has been understudied in the empirical literature
- Gap between the psychiatric description of NDP and clinical observations of patients with NPD

Narcissistic Personality Disorder

Diagnostic Criteria **301.81 (F60.81)**

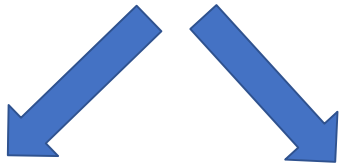
A pervasive pattern of grandiosity (in fantasy or behavior), need for admiration, and lack of empathy, beginning by early adulthood and present in a variety of contexts, as indicated by five (or more) of the following:

1. Has a grandiose sense of self-importance (e.g., exaggerates achievements and talents, expects to be recognized as superior without commensurate achievements).
2. Is preoccupied with fantasies of unlimited success, power, brilliance, beauty, or ideal love.
3. Believes that he or she is "special" and unique and can only be understood by, or should associate with, other special or high-status people (or institutions).
4. Requires excessive admiration.
5. Has a sense of entitlement (i.e., unreasonable expectations of especially favorable treatment or automatic compliance with his or her expectations).
6. Is interpersonally exploitative (i.e., takes advantage of others to achieve his or her own ends).
7. Lacks empathy: is unwilling to recognize or identify with the feelings and needs of others.
8. Is often envious of others or believes that others are envious of him or her.
9. Shows arrogant, haughty behaviors or attitudes.

GRANDIOSE NARCISSISM
Grandiose Fantasies
Entitled attitudes
Disregards for feelings and needs of others
Interpersonal antagonism and dominance

NARCISSISTIC PATHOLOGY IS ESSENTIALLY A PATHOLOGY OF THE SELF
(Caligor, 2013)

INDIVIDUALS WITH NARCISSISTIC PATHOLOGY NEED TO MAINTAIN A POSITIVE SELF-VIEW IN ORDER TO PRESERVE THE INTEGRITY AND STABILITY OF AFFECTIVE AND SELF-EXPERIENCES



Self-promotion
GRANDIOSE NARCISSISM

Self-protection
VULNERABLE NARCISSISM

(Ackerman et al., 2019; Ronningstam &Baskin-Sommers, 2013)

Proposed Diagnostic Criteria

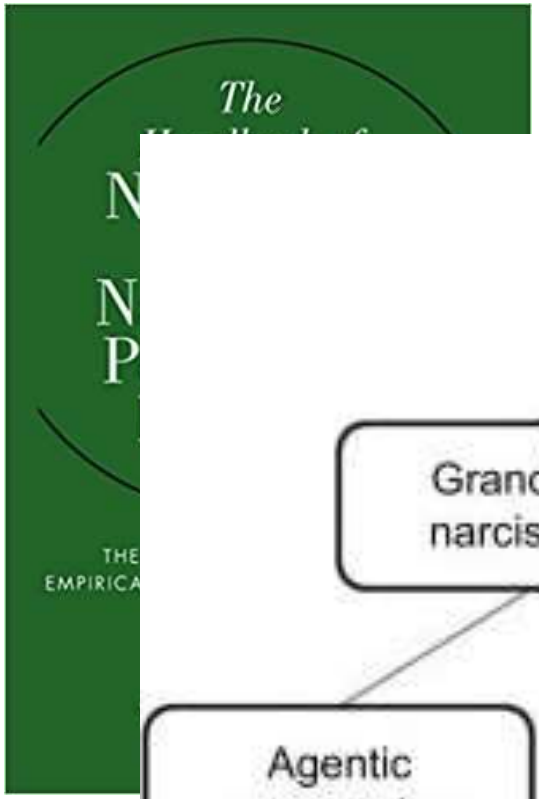
A. Moderate or greater impairment in personality functioning, manifested by characteristic difficulties in two or more of the following four areas:

1. ***Identity:*** Excessive reference to others for self-definition and self-esteem regulation; exaggerated self-appraisal inflated or deflated, or vacillating between extremes; emotional regulation mirrors fluctuations in self-esteem.
2. ***Self-direction:*** Goal setting based on gaining approval from others; personal standards unreasonably high in order to see oneself as exceptional, or too low based on a sense of entitlement; often unaware of own motivations.
3. ***Empathy:*** Impaired ability to recognize or identify with the feelings and needs of others; excessively attuned to reactions of others, but only if perceived as relevant to self; over- or underestimate of own effect on others.
4. ***Intimacy:*** Relationships largely superficial and exist to serve self-esteem regulation; mutuality constrained by little genuine interest in others' experiences and predominance of a need for personal gain.

B. Both of the following pathological personality traits:

1. ***Grandiosity*** (an aspect of **Antagonism**): Feelings of entitlement, either overt or covert; self-centeredness; firmly holding to the belief that one is better than others; condescension toward others.
2. ***Attention seeking*** (an aspect of **Antagonism**): Excessive attempts to attract and be the focus of the attention of others; admiration seeking.

Specifiers. Trait and personality functioning specifiers may be used to record additional personality features that may be present in narcissistic personality disorder but are not required for the diagnosis. For example, other traits of Antagonism (e.g., manipulativeness, deceitfulness, callousness) are not diagnostic criteria for narcissistic personality disorder (see Criterion B) but can be specified when more pervasive antagonistic features (e.g., "malignant narcissism") are present. Other traits of Negative Affectivity (e.g., depressivity, anxiousness) can be specified to record more "vulnerable" presentations. Furthermore, although moderate or greater impairment in personality functioning is required for the diagnosis of narcissistic personality disorder (Criterion A), the level of personality functioning can also be specified.



Grandiose and Vulnerable Narcissism:

email: jmiller@uga.edu
[2]Department of Psychological Sciences, Purdue University, West Lafayette, Indiana 47907

Empirical evidence

Miller et al., 2018
Journal of Personality

- Neuroticism accounted for the greatest amount of variance in the vulnerable narcissism scores (range 56% to 79%).

Table 2 General Dominance Weights for FFM Domains

	Neuroticism	Extraversion	Openness	Agreeableness	Conscientiousness	R^2
S1: Vulnerable narcissism composite	.66**	−.25**	−.04	−.28**	−.16	.52
GD weight	.41 (79%)[a]	.03 (6%)[bc]	.00 (0%)[d]	.07 (13%)[b]	.01 (2%)[c]	
S2: PNI Vulnerable	.64**	−.32**	.02*	−.51**	−.35*	.50
GD weight	.28 (56%)[a]	.04 (8%)[c]	.00 (0%)[d]	.14 (28%)[b]	.04 (8%)[c]	
S3: Vulnerable narcissism composite	.70**	−.30**	.02	−.41**	−.46**	.58
GD weight	.36 (62%)[a]	.04 (7%)[b]	.00 (0%)[c]	.10 (17%)[d]	.08 (14%)[d]	
S4: Vulnerable narcissism composite	.80**	−.43**	−.28**	−.50**	−.33**	.73
GD weight	.46 (63%)[a]	.07 (10%)[b]	.04 (5%)[b]	.12 (16%)[b]	.04 (5%)[b]	

Table 4 Sample 1: Vulnerable Narcissism and Neuroticism and Interpersonal Relations

	Vulnerable Narcissism	Neuroticism	Vulnerable-Partialed
Attachment styles			

Vulnerable Narcissism Is (Mostly) a Disorder of Neuroticism

Journal of Personality 86:2, April 2018

DOI: 10.1111/jopy.12303

Joshua D. Miller,[1] Donald R. Lynam,[2] Colin Vize,[2] Michael Crowe,[1] Chelsea Sleep,[1] Jessica L. Maples-Keller,[1] Lauren R. Few,[1] and W. Keith Campbell[1]

- EMPIRICAL CORRELATES OF VULNERABLE NARCISSISM AND NEUROTICISM WERE NEARLY IDENTICAL

Openness	−.10	.01	−.14
Agreeableness	−.03	.00	−.04
Conscientiousness	.10	.08	.06
Attractiveness	−.07	.02	−.10
Likability	−.11	−.14	−.03
Narcissism	−.01	−.10	.07

OPEN ISSUES

(Related to the LEARNVUL Project)

- **What are we measuring?**

  Trait level vs. Facet level

- **Models of association with CMD**

  Can LEARNVUL project support one or more model?

- **Functional mechanisms**

  Can LEARNVUL project shed light on some of the undelying processes?

Relating personality and learning

Florin Alin Sava,
West University of Timisoara
florin.sava@e-uvt.ro

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

WHY NEUROTICISM SHOULD BE STUDIED FROM AN EVALUATIVE CONDITIONING VIEW

- People scoring high in neuroticism are generally more dissatisfied about their life (EC translation)
- Neuroticism is the main "normal" personality trait seen as a risk factor for internalized psychopathology (e.g. anxiety, and depression)

- **Neuroticism** represents the degree to which a person experiences the world as distressing, threatening, or unsafe. It is not a medical state it is a psychological term.

- GENERAL IDEEA –First showing, and later on, understanding, whether Neuroticism moderates the EC effect in a biased way:

(a) eliciting a stronger change in the negative side (EC effect slope steeper for high N);

(b) assymetrical effect in counter-conditioning (faster change from + to – than from - to +);

(c) more intense transfer of valence ("generalization"/"hallo"/"horn")

(d) AND/OR eliciting more evaluations, both positive and negative (e.g., anxiety)

But not only, EC, maybe OEC, or AC as well.

- STARTING POINT
  - Vogel, Hutter, and Gebauer (2019)
  - Weak-situation context / Inkblot (specific USs in the pairing)

- Is there a moderation effect of N?

Which facets of N?

N

↓

US + CS

Characteristics of US:
(1) Ambivalent USs (competing cues)

(2) Arousal of US

(3) Self vs. non-self relevant

Characteristics of the pairing:
(1) Ambivalent Contingency rate / Inconsistent pairing

(2) Counter-conditioning case

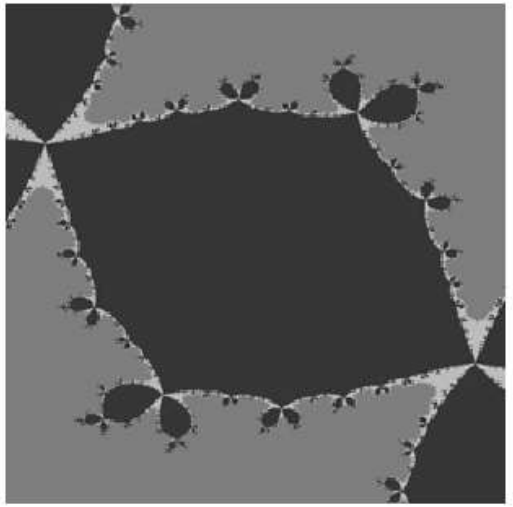
Characteristics of CS:
New CS = CS paired US+ + CS paired US-

- Why does N interacts with EC effect? Memory bias (Contingency Awareness); Attention bias (eye-tracking); Interpretation bias (relational qualifier)

SOME EMPIRICAL RESULTS (SO FAR)

- Casa 1 – The arousal of US (Italian team)
- Case 2 – The ambivalence of US (Romanian team, Bunghez)
- Case 3 – The ambivalence of the pairings (contigency rate) (Romanian team, Bunghez)
- Case 4 – The generalization effect (hallo/horn effects) (Romanian team, Huzoaica)

CASE 2 – AMBIVALENT US (CUE-COMPETING US)



Design: 4 US (US+ vs. US- vs. Us vs. Usamb); Five Categories of N (N++ (+1.5 SD), N+ (.5-1.5 SD), N = (-.5 to +.5) SD … N (NEO PI-R)

Results:
Nsig. for CS paired with pos US or neutral US
Sig results (liner trend)

- $F(1, 1539) = 11.242, p = .001$, ***d* = 0.71** for amb US (in the expected direction)
- $F(1, 1539) = 6.260, p = .01$, ***d* = 0.48** for neg US (in the expected direction)
- 4 facets out six were contributing to the general effect (Anx, Dep, Vul, Shy)

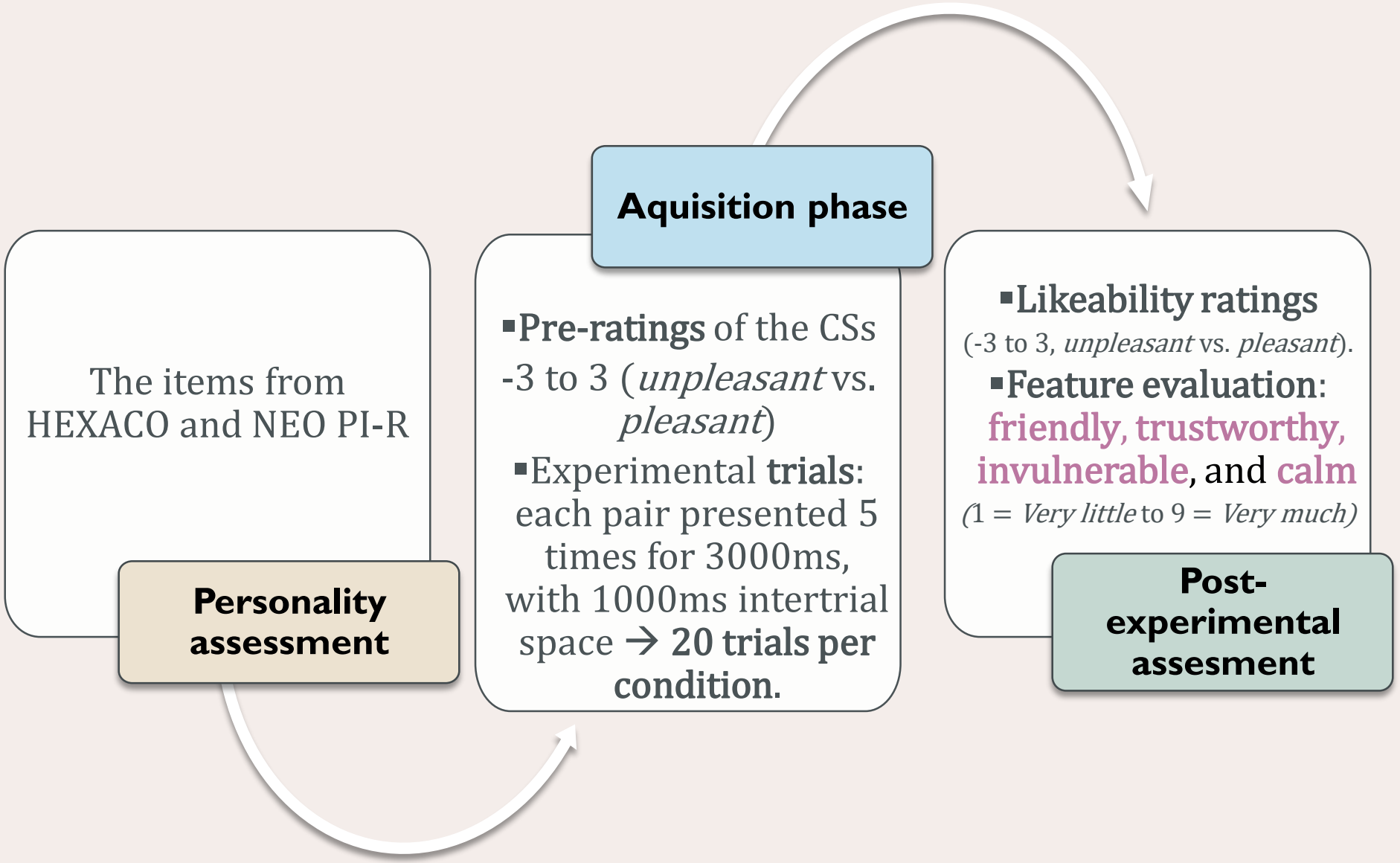
CASE 3 – 50%/50% CONTINGENCY

Within-subject design: 3 types of US-CS pairings (100% pos vs. 50% poz vs. 0%poz)
4 US (US+ vs. US- vs. Us vs. Usamb); Five Categories of N (N++ (+1.5 SD), N+ (.5-1.5 SD), N = (-.5 to +.5) SD …
Several measures of N (NEO PI-R, HEXACO, BFI, BIS/BAS,
But also state measures (N state HEXACO, NA PANAS)

Results:
Nsig. for CS paired with pos US or neutral US
Sig results (liner trend)
for almost all measures of neuroticism, excepting partially on HEXACO emotionality (linear trend on all scales)
Anxiety is the facet that contributes to N effect in all cases.

Not significant for state measures

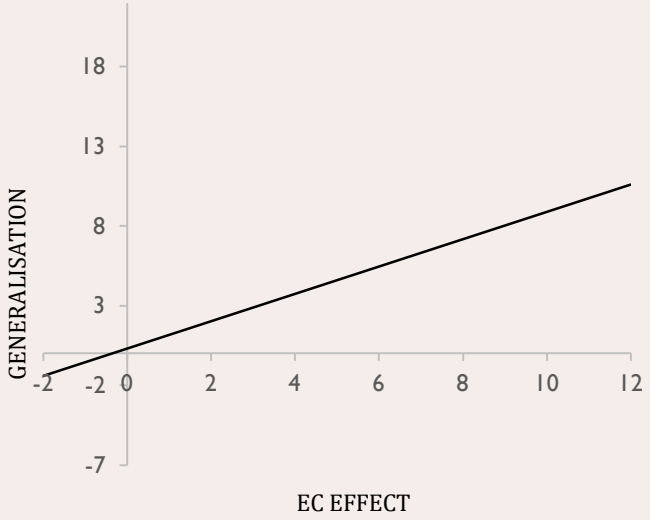
CASE 4 – GENERALIZABILITY (HALLO/HORN EFFECT)

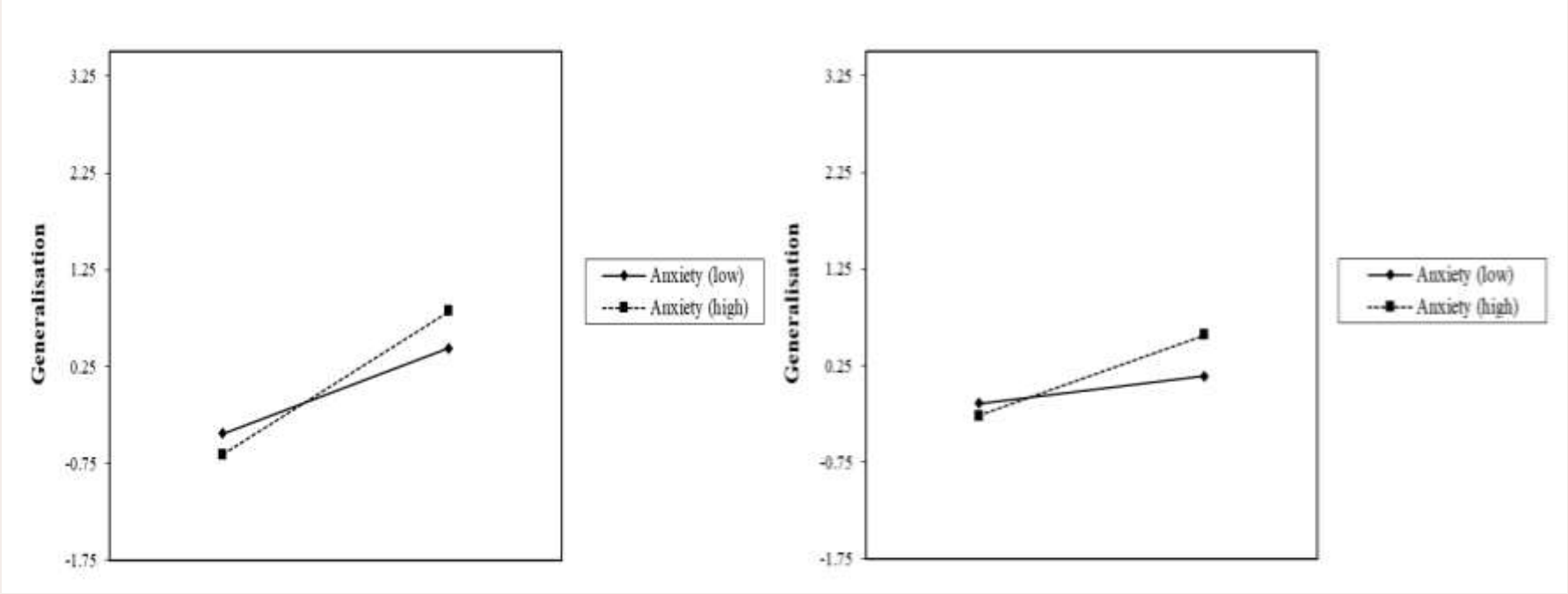


CASE 4 – GENERALIZABILITY (HALLO/HORN EFFECT)

1. Strong generalization (Hallo/Horn) Effect
Generalizability ~ EC Effect (r = .52)

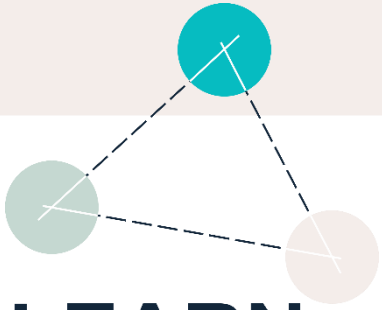
II. Anxiety slightly moderates (steeper the slope for people scoring high on anxiety)





DISCUSSION / CONCLUDING

- It seems Neuroticism moderates EC, particularly in those instances in which pairings or US are congruent with the "weak situation"
- Anxiety is the facet that seems to contribute most to the most to this effect (but it is also advantaged by the structure of the scales)
- Intolerance to uncertainty (weak situation)…the possible story



Principles of data processing

Jamie Cummins,
Ghent University, Belgium
Jamie.Cummins@ugent.be

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

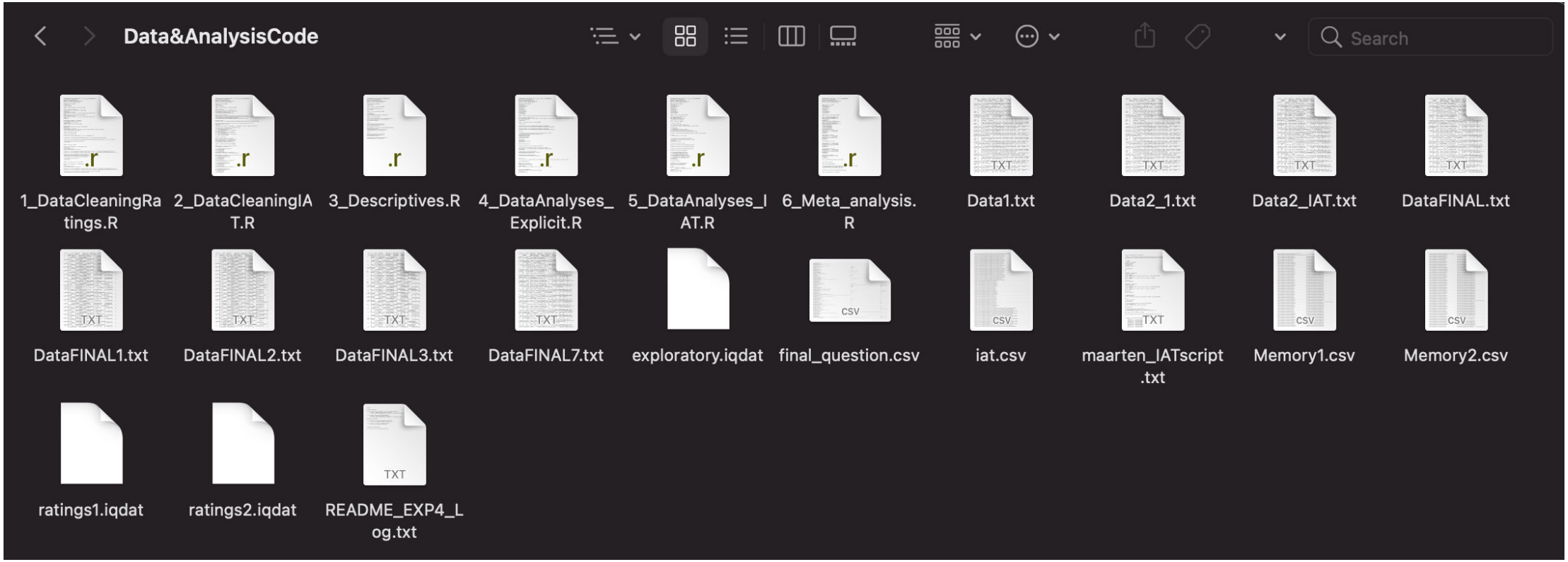
https://osf.io/fqx7a/

Why preprocessing?

- When we get data from a procedure, it can range from relatively tidy (e.g., Inquisit) to a sometimes uninterpretable mess (e.g., in .json files)
- We need reproducible, generalisable workflows to get this data into analysable form
- As psychologists we generally get little formal training in things like:
  - Folder structuring
  - version control (can talk about this if we have time)
  - tidy data
- These are integral to a reproducible workflow

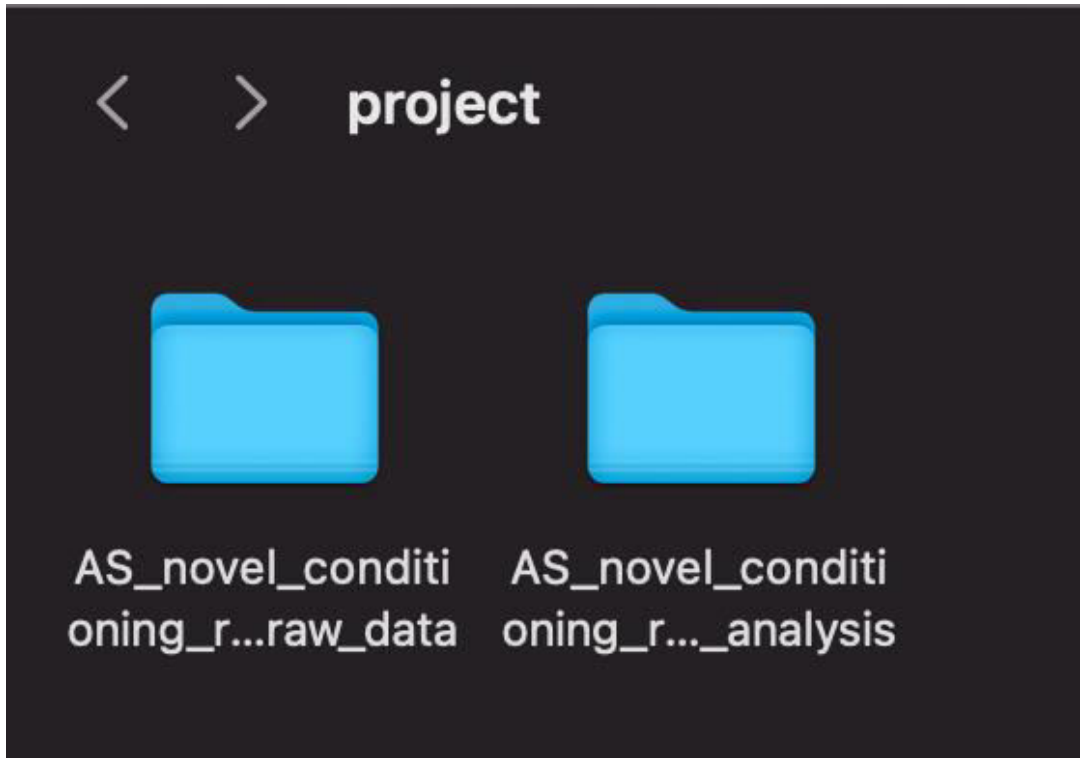
Quiz time!

Round 1: Find the processed, participant-level data…



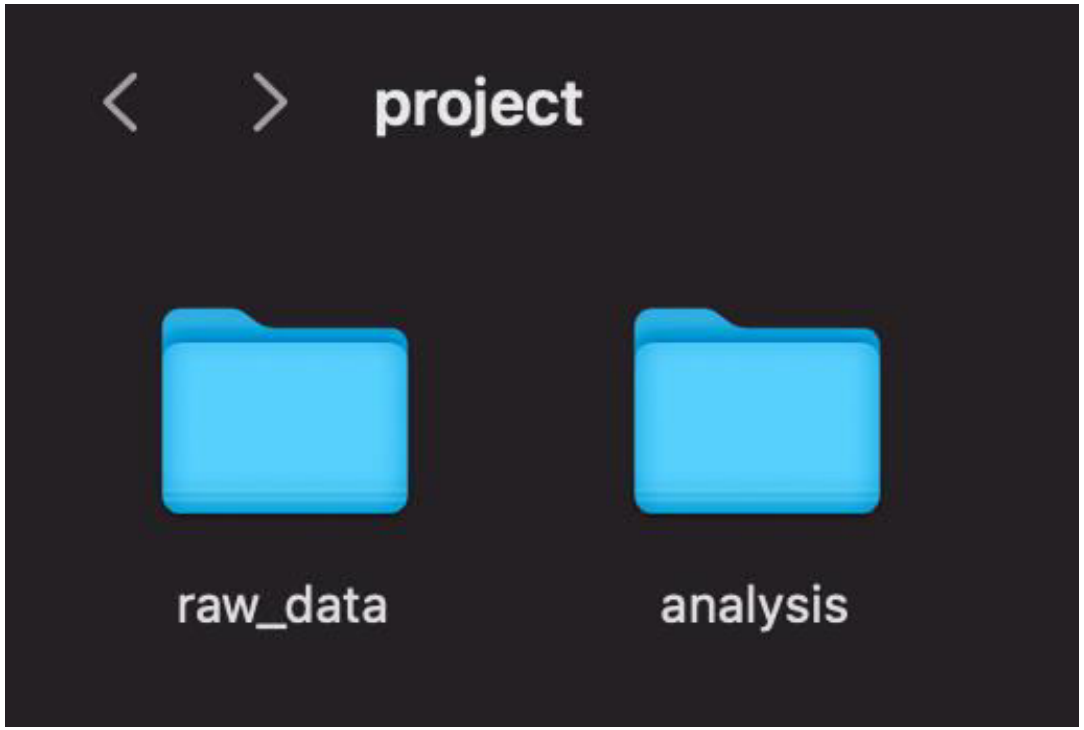
Quiz time!

Round 2: Find the processed, participant-level data…



Quiz time!

Round 2: Find the processed, participant-level data…

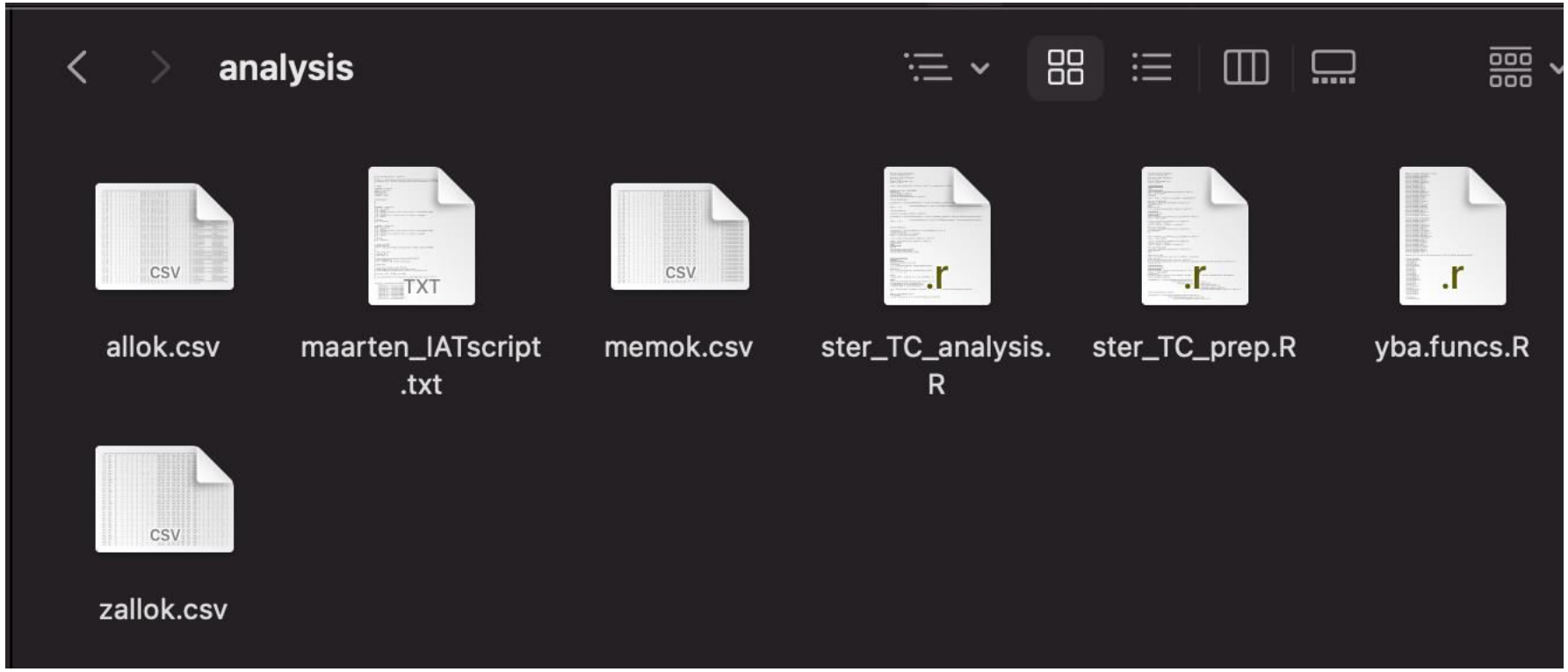


Quiz time!

Round 2: Find the processed, participant-level data…

Quiz time!

Round 2: Find the processed, participant-level data…

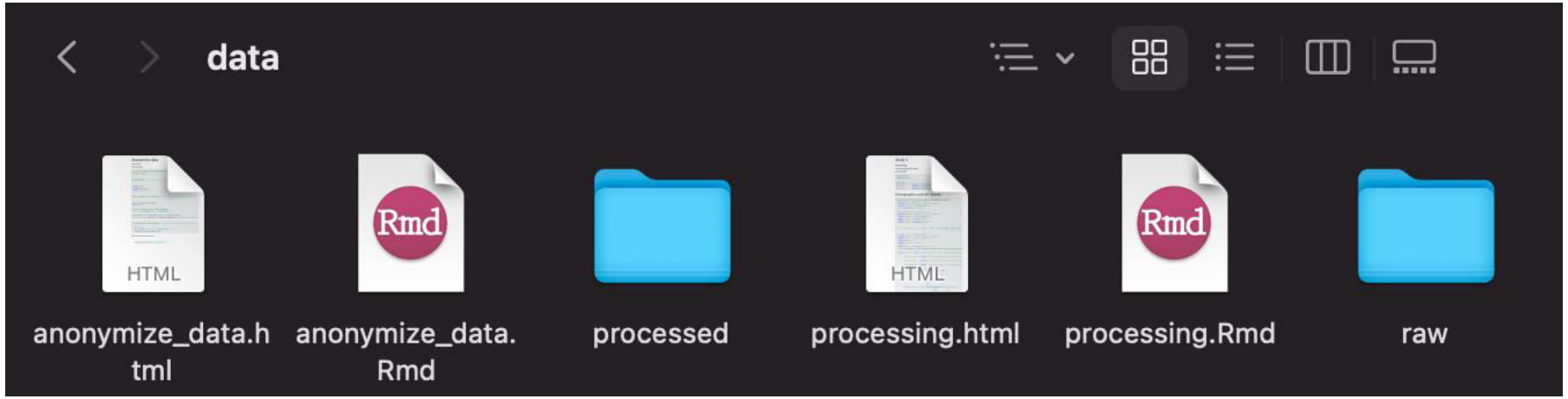


Quiz time!

Round 3: Find the processed, participant-level data…

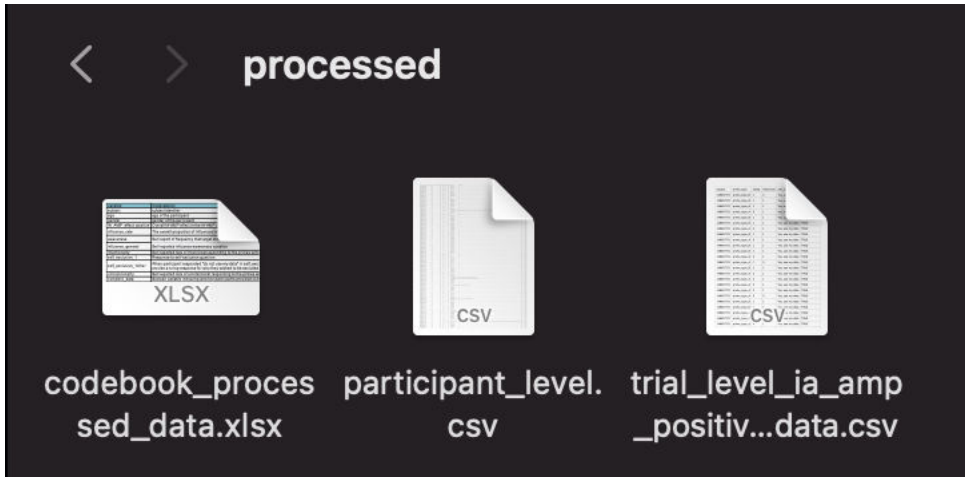
Quiz time!

Round 3: Find the processed, participant-level data…



Quiz time!

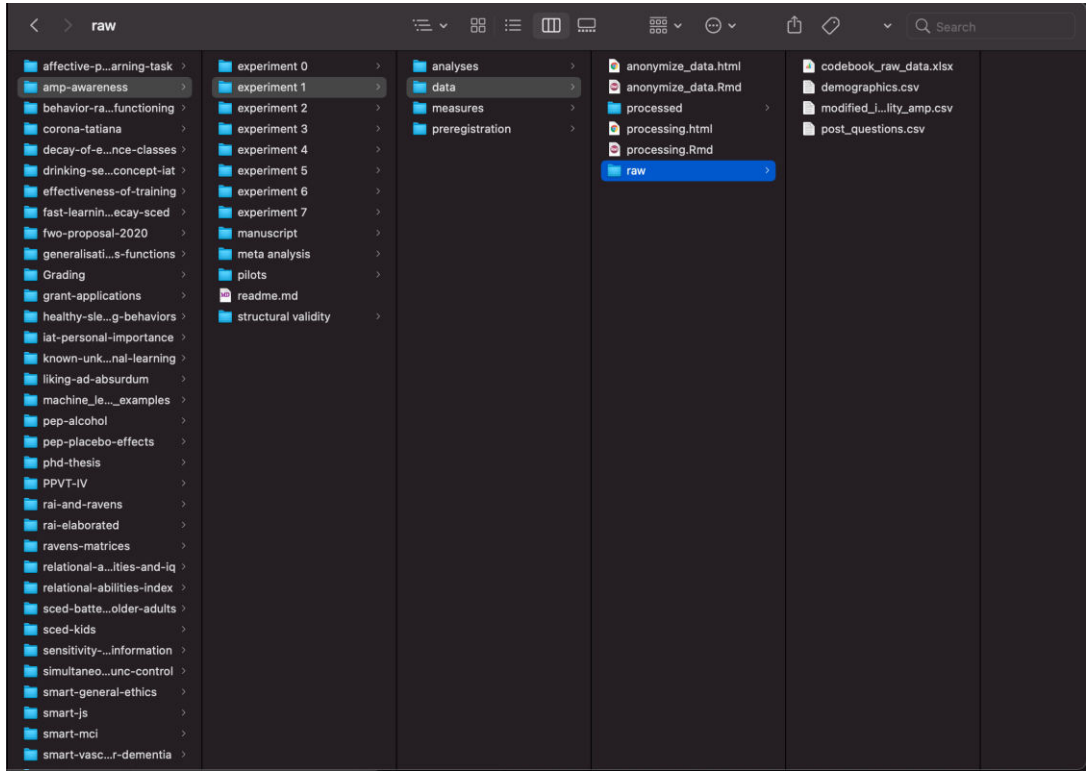
Round 3: Find the processed, participant-level data…

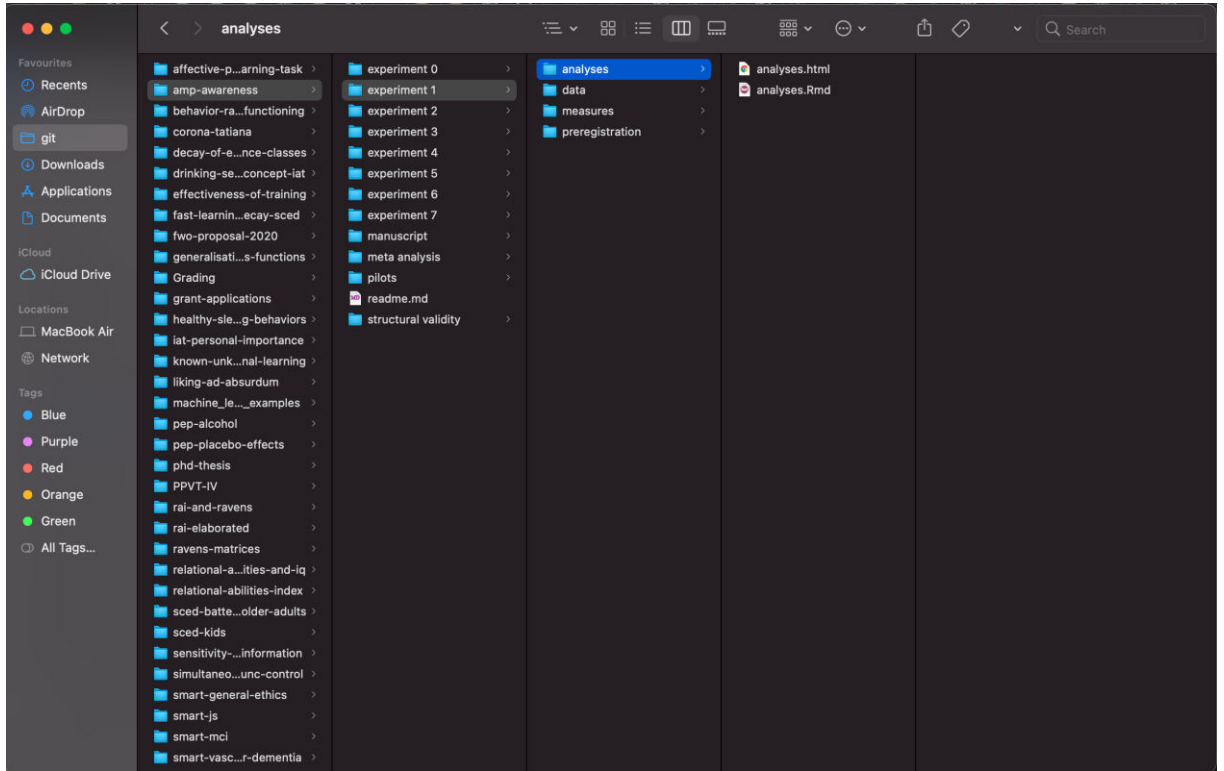


BORING THINGS LIKE FOLDER STRUCTURE ARE SUPER IMPORTANT FOR REPRODUCIBILITY

Folder structuring

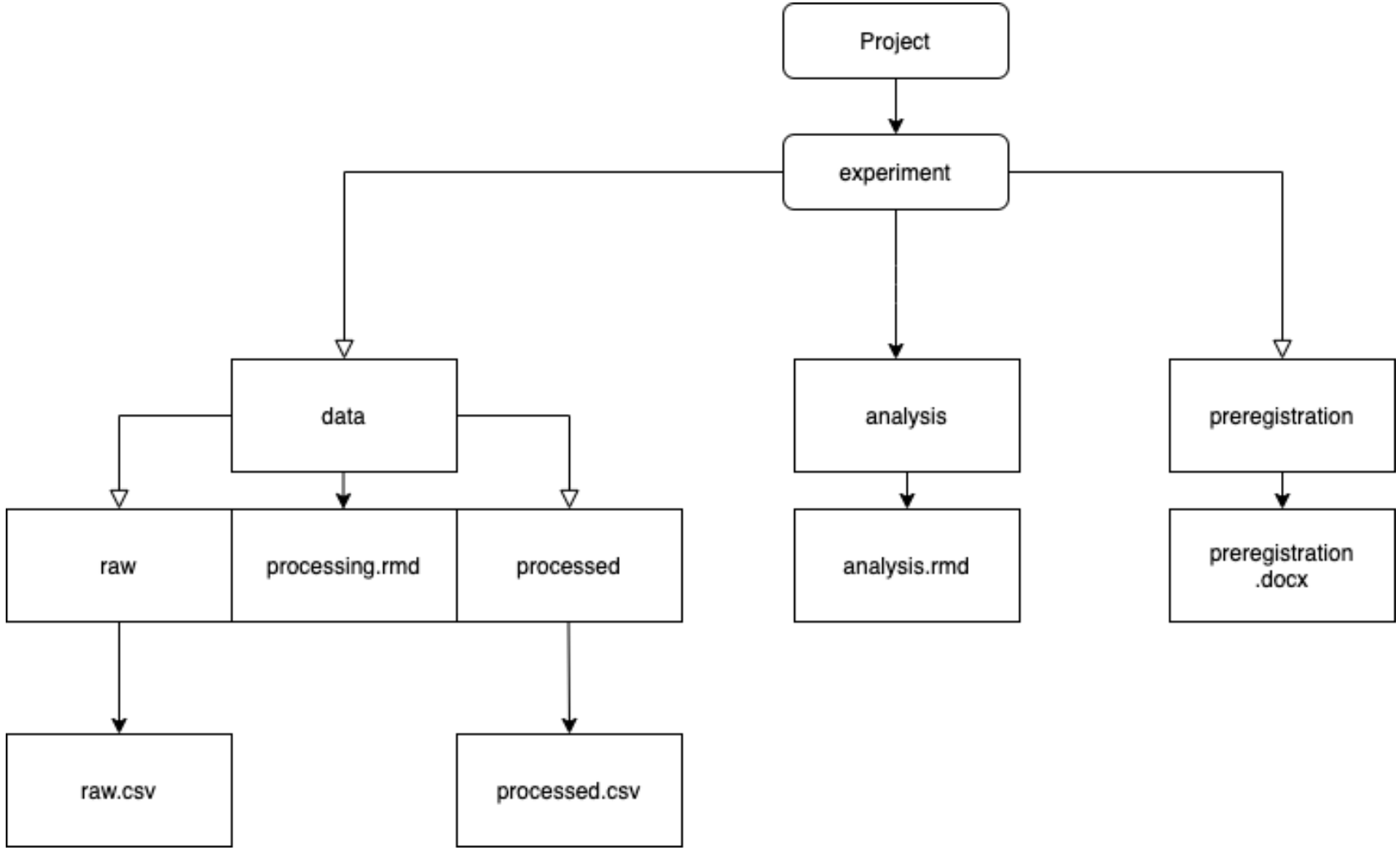
- No agreed-upon standard, but some efforts (e.g., psych-ds from SIPS; https://psych-ds.github.io/)
- General rule: data and processing in one folder, analysis in another



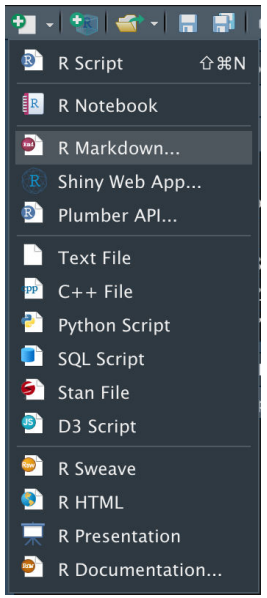


Folder structuring

- Consistent structuring with clear, human-readable folder names == more comprehensible and therefore reproducible
- Additionally: use a relative working directory
  - .r script vs. .rmd script
- Have you ever opened scripts from someone else, and the data are loaded in by specifying a really long wd name like:
  - /Users/jamiecummins/git/amp-awareness/experiment 1/data/processed
- Barrier to reproducibility
  - Also just a bit annoying



- Load raw data without relative wd in processing.rmd:
  - read_csv(“/Users/jcummins/git/Project/experiment/data/raw/raw.csv”)
- With relative wd:
  - read_csv(“raw/raw.csv”)
  - Read_csv(“../ data/ processed/ processed.csv”)



Untitled1* | Untitled2* | Untitled3* | Untitled4

Knit | Insert | Run

```
---
title: "Untitled"
author: "Jamie Cummins"
date: "9/22/2021"
output: html_document
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)
```

R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document. You can embed an R code chunk like this:

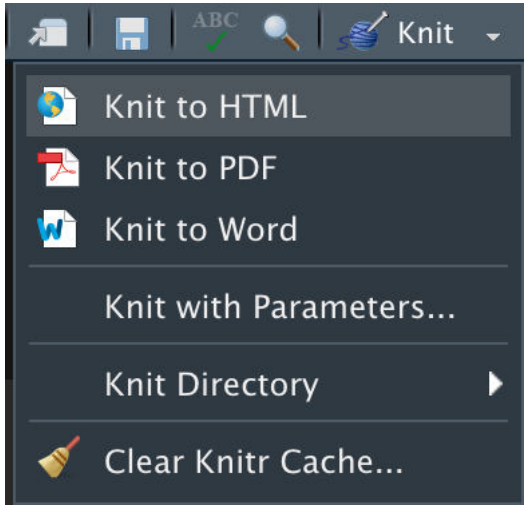
```{r cars}
summary(cars)
```

Including Plots

You can also embed plots, for example:

```{r pressure, echo=FALSE}
plot(pressure)
```

Note that the `echo = FALSE` parameter was added to the code chunk to prevent printing of the R code that generated the plot.
```


Knit
Knit to HTML
Knit to PDF
Knit to Word
Knit with Parameters...
Knit Directory
Clear Knitr Cache...

analysis

Jamie Cummins

21/04/2021

load in data

```
## # A tibble: 2 x 2
##   language_criterion     n
##                <dbl> <int>
## 1                  0     7
## 2                  1   257
```

Demographics:

In total, 271 individuals participated in this study. After exclusions, 136 participants were included in our analyses. This sample had a mean age of 21.6985294 with a SD of 2.2221773, with 74 male and 59 female participants and 3 identifying outside of the gender binary. 71 participants completed the English version, and 65 completed the Dutch version. 63 completed the evaluative IAT at follow-up, and 73 completed the identity IAT.

Measurement properties of the implicit measures

split-half reliability for the IATs

The evaluative IAT has a reliability of 0.7458576 with 95% CIs of 0.6939399 and 0.7866435. The identity IAT has a reliability of 0.6451784 with 95% CIs of 0.5728928 and 0.709405.

Internal consistency for the spontaneous measure

```
##      2.5%       50%     97.5%
## 0.5624052 0.6672350 0.7548883
```

The spontaneous attitude measure has an internal consistency of 0.667235, with 95% CIs of 0.5624052 and 0.7548883.

Journal of Statistical Software

August 2014, Volume 59, Issue 10. http://www.jstatsoft.org/

Tidy Data

Hadley Wickham
RStudio

Abstract

A huge amount of effort is spent cleaning data to get it ready for analysis, but there has been little research on how to make data cleaning as easy and effective as possible. This paper tackles a small, but important, component of data cleaning: data tidying. Tidy datasets are easy to manipulate, model and visualize, and have a specific structure: each variable is a column, each observation is a row, and each type of observational unit is a table. This framework makes it easy to tidy messy datasets because only a small set of tools are needed to deal with a wide range of un-tidy datasets. This structure also makes it easier to develop tidy tools for data analysis, tools that both input and output tidy datasets. The advantages of a consistent data structure and matching tools are demonstrated with a case study free from mundane data manipulation chores.

3 simple rules

- Every column is a variable
- Every row is an observation
- Every table is an observational unit

Every column is a variable

Messy:

religion	<$10k	$10–20k	$20–30k	$30–40k	$40–50k	$50–75k
Agnostic	27	34	60	81	76	137
Atheist	12	27	37	52	35	70
Buddhist	27	21	30	34	33	58
Catholic	418	617	732	670	638	1116
Don't know/refused	15	14	15	11	10	35
Evangelical Prot	575	869	1064	982	881	1486
Hindu	1	9	7	9	11	34
Historically Black Prot	228	244	236	238	197	223
Jehovah's Witness	20	27	24	24	21	30
Jewish	19	19	25	25	30	95

Tidy:

religion	income	freq
Agnostic	<$10k	27
Agnostic	$10–20k	34
Agnostic	$20–30k	60
Agnostic	$30–40k	81
Agnostic	$40–50k	76
Agnostic	$50–75k	137
Agnostic	$75–100k	122
Agnostic	$100–150k	109
Agnostic	>150k	84
Agnostic	Don't know/refused	96

Every table is an observational unit

Messy:

year	artist	time	track	date	week	rank
2000	2 Pac	4:22	Baby Don't Cry	2000-02-26	1	87
2000	2 Pac	4:22	Baby Don't Cry	2000-03-04	2	82
2000	2 Pac	4:22	Baby Don't Cry	2000-03-11	3	72
2000	2 Pac	4:22	Baby Don't Cry	2000-03-18	4	77
2000	2 Pac	4:22	Baby Don't Cry	2000-03-25	5	87
2000	2 Pac	4:22	Baby Don't Cry	2000-04-01	6	94
2000	2 Pac	4:22	Baby Don't Cry	2000-04-08	7	99
2000	2Ge+her	3:15	The Hardest Part Of ...	2000-09-02	1	91
2000	2Ge+her	3:15	The Hardest Part Of ...	2000-09-09	2	87
2000	2Ge+her	3:15	The Hardest Part Of ...	2000-09-16	3	92
2000	3 Doors Down	3:53	Kryptonite	2000-04-08	1	81
2000	3 Doors Down	3:53	Kryptonite	2000-04-15	2	70
2000	3 Doors Down	3:53	Kryptonite	2000-04-22	3	68
2000	3 Doors Down	3:53	Kryptonite	2000-04-29	4	67
2000	3 Doors Down	3:53	Kryptonite	2000-05-06	5	66

Tidy:

id	artist	track	time
1	2 Pac	Baby Don't Cry	4:22
2	2Ge+her	The Hardest Part Of ...	3:15
3	3 Doors Down	Kryptonite	3:53
4	3 Doors Down	Loser	4:24
5	504 Boyz	Wobble Wobble	3:35
6	98^0	Give Me Just One Nig...	3:24
7	A*Teens	Dancing Queen	3:44
8	Aaliyah	I Don't Wanna	4:15
9	Aaliyah	Try Again	4:03
10	Adams, Yolanda	Open My Heart	5:30
11	Adkins, Trace	More	3:05
12	Aguilera, Christina	Come On Over Baby	3:38
13	Aguilera, Christina	I Turn To You	4:00
14	Aguilera, Christina	What A Girl Wants	3:18
15	Alice Deejay	Better Off Alone	6:50

id	date	rank
1	2000-02-26	87
1	2000-03-04	82
1	2000-03-11	72
1	2000-03-18	77
1	2000-03-25	87
1	2000-04-01	94
1	2000-04-08	99
2	2000-09-02	91
2	2000-09-09	87
2	2000-09-16	92
3	2000-04-08	81
3	2000-04-15	70
3	2000-04-22	68
3	2000-04-29	67
3	2000-05-06	66

So, how do we do it in R?

- Using the *tidyverse* collection of *R* packages
- Main processing package: *dplyr*
- Tidyverse has its own syntax and grammar separate to base R

Base R:

```
crime.ny.2005 <- crime.ny.2005[, c("Type.of.Crime", "Count")]
```

dplyr:

```
crime.ny.2005 %>%
   select(Type.of.Crime, Count)
```

Tidyverse syntax

- %>% (pipe) operator is crucial; acts like “and then”
- E.g.:

```
df %>%
  select(participant_id, gender, iat_score) %>%
  group_by(gender) %>%
  summarise(mean = mean(iat_score))
```

- Reads like “take the data frame called df, and then select the three columns called participant_id, gender, and iat_score, and then group them by gender, and then get the summarised mean (which will be divided along the groups)”

Example:

```r
df <- data.frame(participant_id = c(1:20),
                 gender = sample(c("male", "female"), 20, replace = TRUE),
                 iat_score = rnorm(20, mean = 0.2, sd = .05))
```

```
> head(df)
  participant_id gender iat_score
1              1 female 0.2068610
2              2 female 0.2185307
3              3   male 0.2396678
4              4   male 0.2341369
5              5   male 0.1850788
6              6   male 0.2505791
```

```r
df %>%
  select(participant_id, gender, iat_score) %>%
  group_by(gender) %>%
  summarise(mean = mean(iat_score))
```

	gender	mean
1	female	0.2066172
2	male	0.2076018

Main tidyverse functions for data processing

- select()
- filter()
- mutate()
- group_by()
- summarise()
- rename()
- distinct()
- count()
- pivot_wider()/pivot_longer()

select()

- Takes column names as arguments; selects those columns named and drops all other columns
- Can use negations; df %>% select(-extra_column1, -extra_column2) would *drop* extra_column1 and extra_column2 from df and keep the rest
- Can also use indexing; df %>% select(1:3) would keep only the first 3 columns

filter()

- Used to keep only rows where a column’s value conform to specific values
- E.g., df %>% filter(gender == “male”) would keep only those rows where gender is the string “male”
- Can use negations; e.g., df %>% filter(gender != “male”)

mutate()

- Creates a new column (or alters an existing column) based on provided arguments
- E.g., df %>% mutate(gender = ifelse(gender == “male”, 1, 0))
- Can mutate multiple columns at once:

```
df %>%
        mutate(gender = ifelse(gender == “male”, 1, 0),
                        age_over_35 = ifelse(age > 35, 1, 0))
```

mutate() with ifelse() vs. case_when()

```
df %>%
        mutate(age_five_levels = ifelse(age < 20, 1,
                                ifelse(age < 30 & age >= 20, 2,
                                ifelse(… etc.
                                ))))))
df %>%
        mutate(age_five_levels = case_when(age < 20 ~ 1,
                                                        age < 30 & age >= 20 ~ 2,
                                                        age < 40 & age >= 30 ~ 3,
                                                        …
                                                        TRUE ~ 6))
```

summarise()/summarize()

- Similar to mutate(), but drops unmentioned columns and returns only unique values

```
> df %>%
+   mutate(mean_age = mean(age))
   participant_id gender iat_score age mean_age
1               1   male 0.2046557  40    40.75
2               2   male 0.1370771  60    40.75
3               3   male 0.2226200  49    40.75
4               4 female 0.1528026  29    40.75
5               5 female 0.2771575  44    40.75
6               6 female 0.2372081  28    40.75
7               7   male 0.2278126  22    40.75
8               8 female 0.1451481  65    40.75
9               9   male 0.2104898  26    40.75
10             10 female 0.2052989  59    40.75
11             11 female 0.2137251  39    40.75
12             12 female 0.2483404  20    40.75
13             13 female 0.2143609  41    40.75
14             14   male 0.1866856  23    40.75
15             15   male 0.1983894  62    40.75
16             16 female 0.2461770  47    40.75
17             17   male 0.3087670  46    40.75
18             18 female 0.2531526  43    40.75
19             19   male 0.2300986  27    40.75
20             20 female 0.1523299  45    40.75
```

```
> df %>%
+   summarise(mean_age = mean(age))
  mean_age
1    40.75
```

group_by()

- Doesn’t do anything major by itself, but can interface with subsequent functions

```
> df %>%
+   group_by(gender) %>%
+   mutate(mean_age = mean(age))
# A tibble: 20 x 5
# Groups:   gender [2]
   participant_id gender iat_score   age mean_age
            <int> <fct>      <dbl> <int>    <dbl>
 1              1 male       0.205    40     39.4
 2              2 male       0.137    60     39.4
 3              3 male       0.223    49     39.4
 4              4 female     0.153    29     41.8
 5              5 female     0.277    44     41.8
 6              6 female     0.237    28     41.8
 7              7 male       0.228    22     39.4
 8              8 female     0.145    65     41.8
 9              9 male       0.210    26     39.4
10             10 female     0.205    59     41.8
11             11 female     0.214    39     41.8
12             12 female     0.248    20     41.8
13             13 female     0.214    41     41.8
14             14 male       0.187    23     39.4
15             15 male       0.198    62     39.4
16             16 female     0.246    47     41.8
17             17 male       0.309    46     39.4
18             18 female     0.253    43     41.8
19             19 male       0.230    27     39.4
20             20 female     0.152    45     41.8
```

- Always remember to ungroup() after!

rename()

- Renames columns!
- e.g., df %>% rename(new_column_name = old_column_name)
- Note directionality; same as variable assignment direction
- Syntax can integrate into other functions, like:

```
df %>%
        select(subject = participant_id, gender, age)
```

distinct()

- Takes columns as argument, returns data frame with only unique combinations of those columns

subject	build	blocknum	trialnum	blockcode	trialcode	pretrialpause
217658675	4.0.10.0	1	2	demographics	age	0
217658675	4.0.10.0	1	3	demographics	gender	0
387783740	4.0.10.0	1	2	demographics	age	0
387783740	4.0.10.0	1	3	demographics	gender	0
187488134	4.0.10.0	1	2	demographics	age	0
187488134	4.0.10.0	1	3	demographics	gender	0
387783740	4.0.10.0	1	2	demographics	age	0
387783740	4.0.10.0	1	3	demographics	gender	0
608526962	4.0.10.0	1	2	demographics	age	0
608526962	4.0.10.0	1	3	demographics	gender	0
3545965	4.0.10.0	1	2	demographics	age	0
3545965	4.0.10.0	1	3	demographics	gender	0
689281609	4.0.10.0	1	2	demographics	age	0
689281609	4.0.10.0	1	3	demographics	gender	0

distinct()

```
df %>%
        distinct(subject, trialcode)
```

subject <int>	trialcode <fctr>
217658675	age
217658675	gender
387783740	age
387783740	gender
187488134	age
187488134	gender
608526962	age
608526962	gender
3545965	age
3545965	gender

```
df %>%
        distinct(subject, trialcode,
                 .keep_all = TRUE)
```

subject <int>	build <fctr>	blocknum <int>	trialnum <int>	blockcode <fctr>	trialcode <fctr>	pretrialpause <int>
217658675	4.0.10.0	1	2	demographics	age	0
217658675	4.0.10.0	1	3	demographics	gender	0
387783740	4.0.10.0	1	2	demographics	age	0
387783740	4.0.10.0	1	3	demographics	gender	0
187488134	4.0.10.0	1	2	demographics	age	0
187488134	4.0.10.0	1	3	demographics	gender	0
608526962	4.0.10.0	1	2	demographics	age	0

count()

```
# returns total n rows
df %>%
        count()

# returns n by gender
df %>%
        count(gender)

# equivalent to above
df %>%
        group_by(gender) %>%
        count()
```

pivot_wider()/pivot_longer()

- pivot_wider() updated version of spread()
- pivot_longer() updated version of gather()
- Changes data to different formats

pivot_wider()/pivot_longer()

Recall:

subject <int>	build <fctr>	blocknum <int>	trialnum <int>	blockcode <fctr>	trialcode <fctr>	pretrialpause <int>
217658675	4.0.10.0	1	2	demographics	age	0
217658675	4.0.10.0	1	3	demographics	gender	0
387783740	4.0.10.0	1	2	demographics	age	0
387783740	4.0.10.0	1	3	demographics	gender	0
187488134	4.0.10.0	1	2	demographics	age	0
187488134	4.0.10.0	1	3	demographics	gender	0
608526962	4.0.10.0	1	2	demographics	age	0

```
df %>%
    select(subject, trialcode, response) %>%
    pivot_wider(names_from = trialcode,
                values_from = response)
```

subject <chr>	age <chr>	gender <chr>
217658675	21	Male
387783740	41	male
187488134	29	male
608526962	29	male
3545965	63	female
689281609	27	Male
651077529	27	male
861999982	18	Male
142189929	27	female
252686048	25	female

```
wide_df <- df %>%
        select(subject, trialcode, response) %>%
        pivot_wider(names_from = trialcode,
                            values_from = response)
```

subject <chr>	age <chr>	gender <chr>
217658675	21	Male
387783740	41	male
187488134	29	male
608526962	29	male
3545965	63	female
689281609	27	Male
651077529	27	male
861999982	18	Male
142189929	27	female
252686048	25	female

```
wide_df %>%
        pivot_longer(age:gender,
                            names_to = “trialcode”,
                            values_to = “response”)
```

subject <chr>	trialcode <chr>	respons <chr>
217658675	age	21
217658675	gender	Male
387783740	age	41
387783740	gender	male
187488134	age	29
187488134	gender	male

Practical example: Inquisit data, EC study

- Basic study:
  - EC procedure -> IAT -> self-report
  - Between-subjects design (varied identity of CS+ and CS-)
  - Counterbalanced order of IAT blocks
- We will have 6 data files: demographics, EC procedure (2 separate files), self-report, and IAT (2 separate files for block ordering)
- We have two observational units: group-level data (where each row == one participant) and trial-level data (where each row == 1 trial from the IAT)
- So we want 2 corresponding processed data files

Best research practices in the Open Science and (post)reproducibility crisis era

Marco Perugini,

Università degli Studi di Milano-Bicocca

marco.perugini@unimib.it

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

Outline

- **Replicability in Psychology**
- **(Some) Tips for getting it right**

UNIVERSITÀ DEGLI STUDI DI MILANO BICOCCA

Replicability in Psychology

2011: A year to remember

- The year 2011 has been an *annus horribilis* for Psychology
- Three main events:

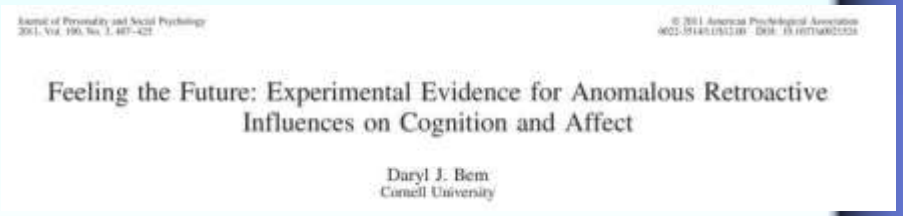
Bem

Stapel

Simmons (et al)

Bem (JPSP, 2011)

- Nine experiments showing ESP
- Strong reactions
- Initial article not replicating results was refused by JPSP
- Hard questions on the modal way of analyzing data and on "cherry-picking" results



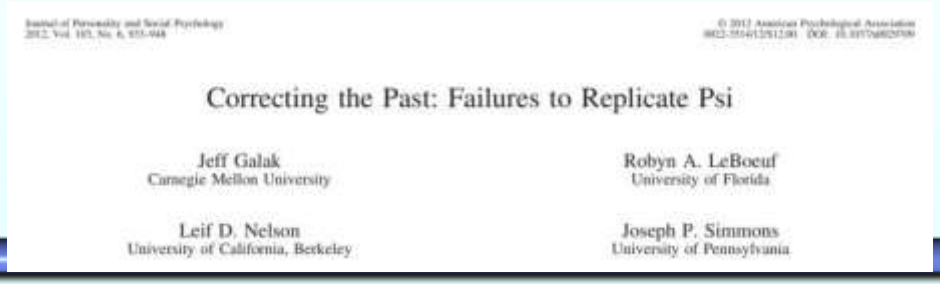
Feeling the Future: Experimental Evidence for Anomalous Retroactive Influences on Cognition and Affect

Daryl J. Bem
Cornell University

Why Psychologists Must Change the Way They Analyze Their Data: The Case of Psi: Comment on Bem (2011)

Eric–Jan Wagenmakers, Ruud Wetzels, Denny Borsboom, and Han L. J. van der Maas
University of Amsterdam

- Galak et al (2012): 7 failed replication attempts (n=3289)



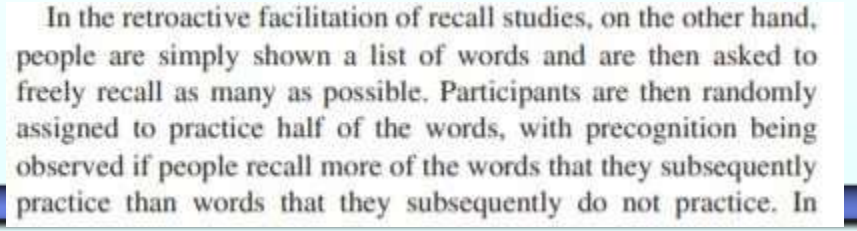
Correcting the Past: Failures to Replicate Psi

Jeff Galak
Carnegie Mellon University

Robyn A. LeBoeuf
University of Florida

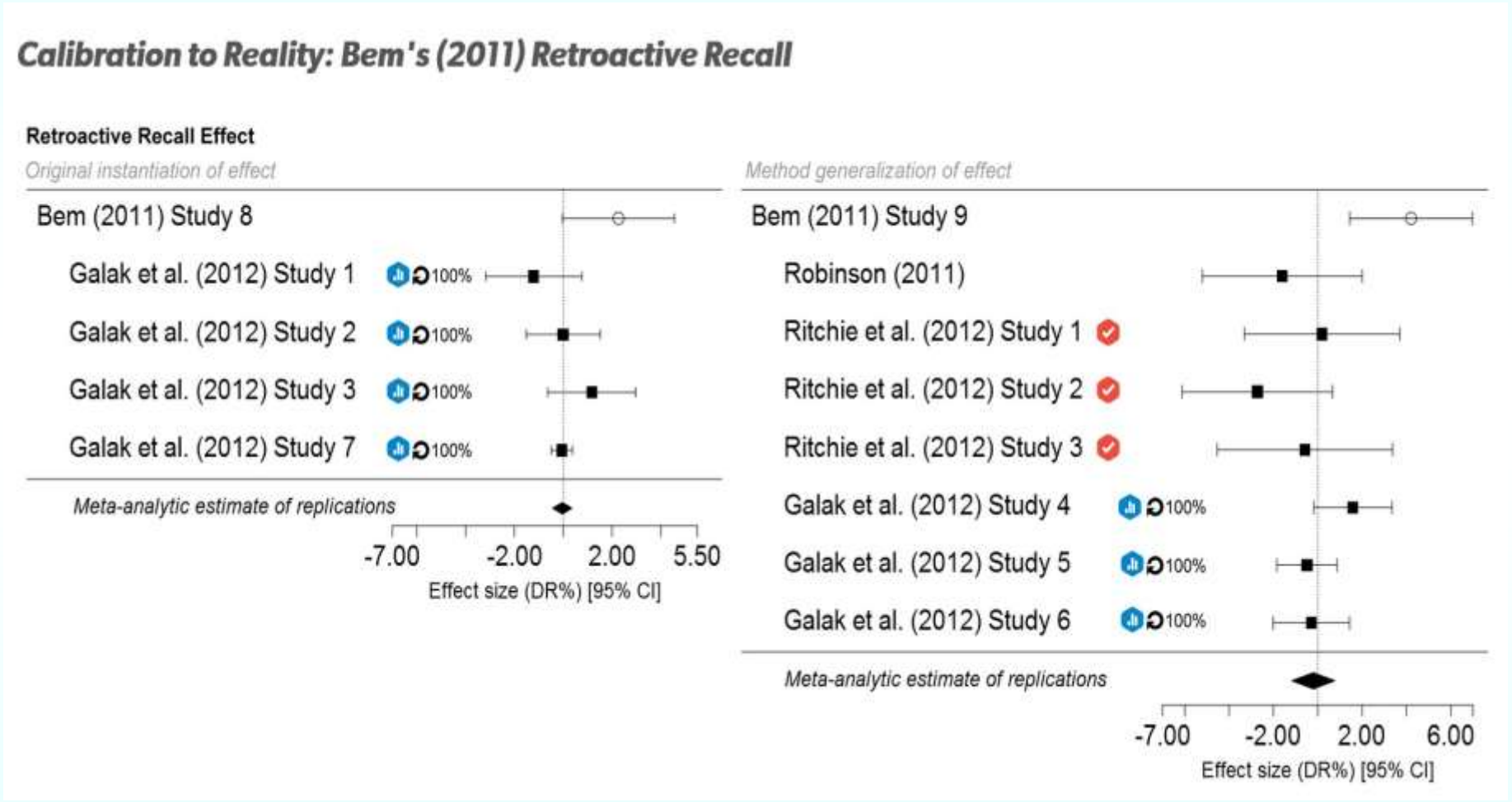
Leif D. Nelson
University of California, Berkeley

Joseph P. Simmons
University of Pennsylvania



In the retroactive facilitation of recall studies, on the other hand, people are simply shown a list of words and are then asked to freely recall as many as possible. Participants are then randomly assigned to practice half of the words, with precognition being observed if people recall more of the words that they subsequently practice than words that they subsequently do not practice. In

Bem (2011)


Calibration to Reality: Bem's (2011) Retroactive Recall
Retroactive Recall Effect
Original instantiation of effect
Bem (2011) Study 8
Galak et al. (2012) Study 1
Galak et al. (2012) Study 2
Galak et al. (2012) Study 3
Galak et al. (2012) Study 7
100%
Meta-analytic estimate of replications
-7.00
-2.00
2.00
5.50
Effect size (DR%) [95% CI]
Method generalization of effect
Bem (2011) Study 9
Robinson (2011)
Ritchie et al. (2012) Study 1
Ritchie et al. (2012) Study 2
Ritchie et al. (2012) Study 3
Galak et al. (2012) Study 4
Galak et al. (2012) Study 5
Galak et al. (2012) Study 6
Meta-analytic estimate of replications
-7.00
-2.00
2.00
6.00
Effect size (DR%) [95% CI]

Stapel (September 2011)

- Resigned from Dean at Tilburg University (NL)
- Faked data: 53 retracted papers and 10 PhD thesis with invented or dubious data
- Levelt report (2012): proofs beyond doubts of faked data and strong criticisms to the scientific community
- Huge media impact

Simmons et al (PS, 2011)

False-Positive Psychology: Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant

Psychological Science
22(11) 1359–1366

Joseph P. Simmons[1], Leif D. Nelson[2], and Uri Simonsohn[1]
[1]The Wharton School, University of Pennsylvania, and [2]Haas School of Business, University of California, Berkeley

- They show that commonly used questionable research practices can allow to provide empirical evidence even for null effects (***false positive***)

  Huge scientific impact (2813 citations, one of the most cited papers from 2011 in all Psychology). Not all solutions are convincing, but they make many good points and suggestions

The replicability crisis

- From 2011 increasing appreciation of problems in published research in top journals in Psychology
- Cases of not replicated results and outright frauds
- Hard questions on the modal way of analyzing data and on “cherry-picking” results
- Fraud is a problem, but it is not only about that
- Important advances in research methodology
- Rapid changes in standards for research and for publishing

Replicability

- If a result is not replicated, it is not valid
- To be replicated, it needs to be replicable

Replicability

- A key concept in Science
- Almost forgotten in Psychology
- Now at the forefront
- What is replicability?

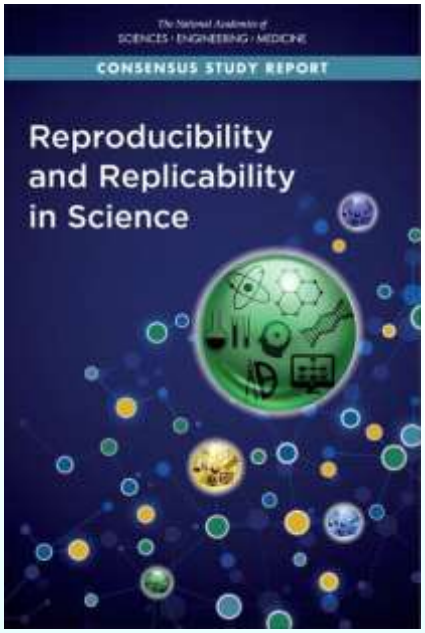
Replicability

European Journal of Personality, *Eur. J. Pers.* **27**: 108–119 (2013)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/per.1919

Recommendations for Increasing Replicability in Psychology†

JENS B. ASENDORPF[1]*, MARK CONNER[2], FILIP DE FRUYT[3], JAN DE HOUWER[4], JAAP J. A. DENISSEN[5], KLAUS FIEDLER[6], SUSANN FIEDLER[7], DAVID C. FUNDER[8], REINHOLD KLIEGL[9], BRIAN A. NOSEK[10], MARCO PERUGINI[11], BRENT W. ROBERTS[12], MANFRED SCHMITT[13], MARCEL A. G. VANAKEN[14], HANNELORE WEBER[15] and JELTE M. WICHERTS[5]

[1]*Department of Psychology, Humboldt University Berlin, Berlin, Germany*
[2]*Institute of Psychological Sciences, University of Leeds, Leeds, UK*
[3]*Department of Developmental, Personality and Social Psychology, Ghent University, Ghent, Belgium*
[4]*Department of Experimental Clinical and Health Psychology, Ghent University, Ghent, Belgium*
[5]*School of Social and Behavioral Sciences, Tilburg University, Tilburg, The Netherlands*
[6]*Department of Psychology, University of Heidelberg, Heidelberg, Germany*
[7]*Max Planck Institute for Research on Collective Goods, Bonn, Germany*
[8]*Department of Psychology, University of California at Riverside, Riverside, CA USA*
[9]*Department of Psychology, University of Potsdam, Potsdam, Germany*
[10]*Department of Psychology, University of Virginia, Charlottesville, VA USA*
[11]*Department of Psychology, University of Milano-Bicocca, Milan, Italy*
[12]*Department of Psychology, University of Illinois, Chicago, IL USA*
[13]*Department of Psychology, University of Koblenz–Landau, Landau, Germany*
[14]*Department of Psychology, Utrecht University, Utrecht, The Netherlands*
[15]*Department of Psychology, University of Greifswald, Greifswald, Germany*



The National Academies of SCIENCES ENGINEERING MEDICINE

Reproducibility and Replicability in Science (2019)

Consensus Study Report

Definitions

- **Reproducibility**: Obtaining consistent computational results using the same input data, computational steps, methods, and code, and conditions of analysis

- **Replicability**: obtaining consistent results across studies aimed at answering the same scientific question, each of which has obtained its own data

Conditions for Replicability

- The study should be described in a way such that everyone qualified can replicate it
- This implies a very detailed method section, with information that often is not disclosed
- Also, the data should be publicly available (at a minimum upon request) to replicate the results using appropriate analyses (*reproducibility*)


Behav Res (2016) 48:1205–1226
DOI 10.3758/s13428-015-0664-2

The prevalence of statistical reporting errors in psychology (1985–2013)

Michèle B. Nuijten[1] · Chris H. J. Hartgerink[1] · Marcel A. L. M. van Assen[1] · Sacha Epskamp[2] · Jelte M. Wicherts[1]

- Transparency in research

Replicability as a pre-condition for validity

- If an effect is not replicable, it cannot be valid (according to scientific standards)
- If an effect is replicable, it may or may not be valid
- Validity assumes but goes beyond replicability
- Analogy with psychometric measures: Reliability is a necessary but insufficient condition for validity of a measure

Replicability: a continuum

Generalizibility

Validity

Highly similar ← *Replication continuum* → *Highly dissimilar*

Direct replication | Conceptual replication

Design facet	**Exact replication** (Everything controllable the same)	**Very close replication** (Procedure or context is different)	**Close replication** (IV or DV stimuli are different)	**Far replication** (IV or DV operationalization is different)	**Very Far replication** (Everything can be different)
IV operationalization	same	same	same	different	
DV operationalization	same	same	same	different	
IV stimuli	same	same	different		
DV stimuli	same	same	different		
Procedural details	same	different			
Physical setting	same	different			
Contextual variables	different				
⋮	⋮				

Figure 1. A simplified replication taxonomy to guide the classification of relative methodological similarity of a replication study to an original study. "Same" ("different") indicates the design facet in question is the same (different) compared to an original study. IV = independent variable. DV = dependent variable. "Everything controllable" indicates design facets over which a researcher has control. Procedural details involve minor experimental particulars (e.g., task instruction wording, font, font size, etc.). See the online article for the color version of this figure.

The Reproducibility Project

RESEARCH ARTICLE

PSYCHOLOGY

Estimating the reproducibility of psychological science

Open Science Collaboration*†

Reproducibility is a defining feature of science, but the extent to which it characterizes current research is unknown. We conducted replications of 100 experimental and correlational studies published in three psychology journals using high-powered designs and original materials when available. Replication effects were half the magnitude of original effects, representing a substantial decline. Ninety-seven percent of original studies had statistically significant results. Thirty-six percent of replications had statistically significant results; 47% of original effect sizes were in the 95% confidence interval of the replication effect size; 39% of effects were subjectively rated to have replicated the original result; and if no bias in original results is assumed, combining original and replication results left 68% with statistically significant effects. Correlational tests suggest that replication success was better predicted by the strength of original evidence than by characteristics of the original and replication teams.

- Open Science Center
- 270 volunteers, 64 universities, 11 countries, 100 replicated studies from 3 main journals (JPSP, PS, JEP:LMC) in 2008.
- Started in 2011, published in 2015 in *Science*
- **What results one can expect?**

Main results

Table 1. Summary of reproducibility rates and effect sizes for original and replication studies overall and by journal/discipline. *df/N* refers to the information on which the test of the effect was based (for example, *df* of *t* test, denominator *df* of *F* test, sample size −3 of correlation, and sample size for *z* and χ^2). Four original results had *P* values slightly higher than 0.05 but were considered positive results in the original article and are treated that way here. Exclusions (explanation provided in supplementary materials, A3) are "replications $P < 0.05$" (3 original nulls excluded; $n = 97$ studies); "mean original and replication effect sizes" (3 excluded; $n = 97$ studies); "meta-analytic mean estimates" (27 excluded; $n = 73$ studies); "percent meta-analytic ($P < 0.05$)" (25 excluded; $n = 75$ studies); and, "percent original effect size within replication 95% CI" (5 excluded, $n = 95$ studies).

			Effect size comparison					Original and replication combined			
	Replications $P < 0.05$ in original direction	Percent	Mean (SD) original effect size	Median original *df/N*	Mean (SD) replication effect size	Median replication *df/N*	Average replication power	Meta-analytic mean (SD) estimate	Percent meta-analytic ($P < 0.05$)	Percent original effect size within replication 95% CI	Percent subjective "yes" to "Did it replicate?"
Overall	35/97	36	0.403 (0.188)	54	0.197 (0.257)	68	0.92	0.309 (0.223)	68	47	39
JPSP, social	7/31	23	0.29 (0.10)	73	0.07 (0.11)	120	0.91	0.138 (0.087)	43	34	25
JEP:LMC, cognitive	13/27	48	0.47 (0.18)	36.5	0.27 (0.24)	43	0.93	0.393 (0.209)	86	62	54
PSCI, social	7/24	29	0.39 (0.20)	76	0.21 (0.30)	122	0.92	0.286 (0.228)	58	40	32
PSCI, cognitive	8/15	53	0.53 (0.2)	23	0.29 (0.35)	21	0.94	0.464 (0.221)	92	60	53

Effect size comparison

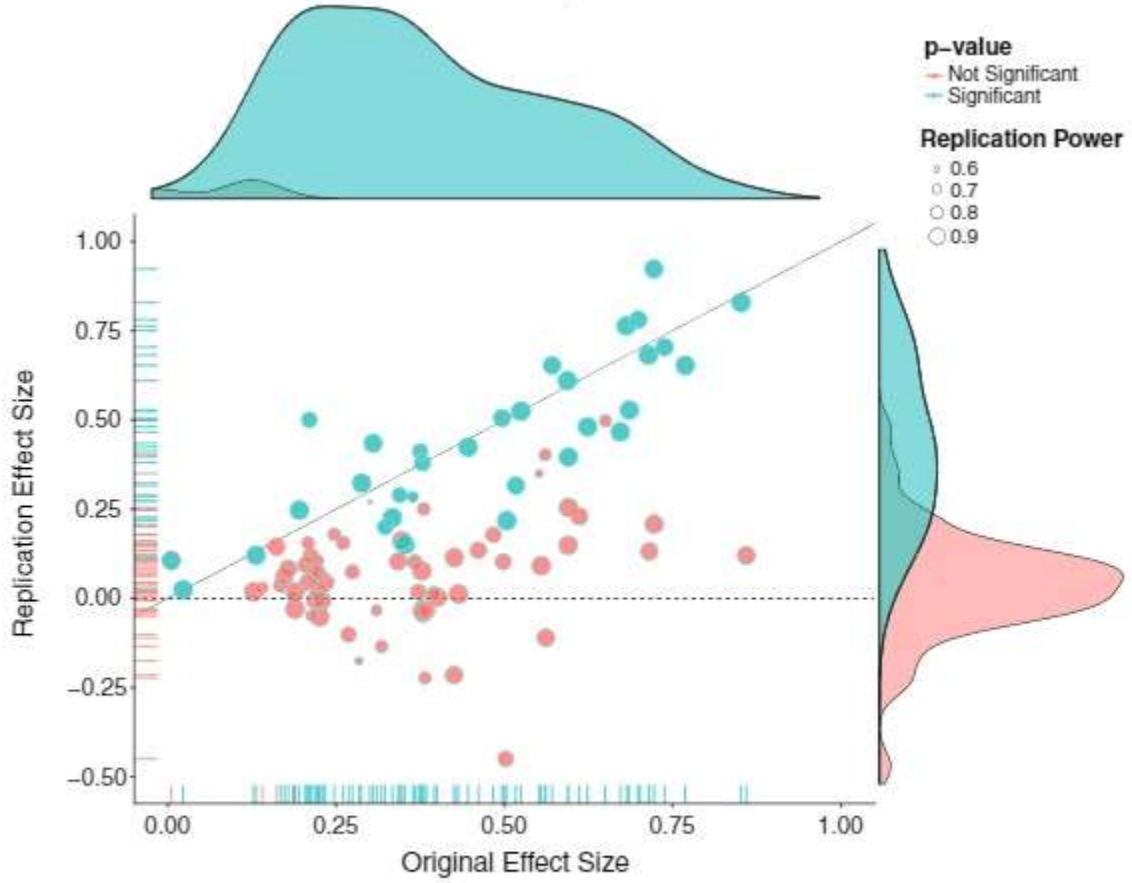


Fig. 3. Original study effect size versus replication effect size (correlation coefficients). Diagonal line represents replication effect size equal to original effect size. Dotted line represents replication effect size of 0. Points below the dotted line were effects in the opposite direction of the original. Density plots are separated by significant (blue) and nonsignificant (red) effects.

r(spearman) = 0.51

Most replications smaller effect size than original

(d=0.20 vs. 0.40)

Summing up RP main results

- 36% replicate at p<.05 (simple answer)
- Effect size are half (publication bias, file drawer effect)
- Less likely to replicate if weaker evidence in original study (p<.05 worse than p<.001)
- Milestone achievement of Psychology
- Followed by RPs in other scientific domains and many other RPs in Psychology
  (over 1000 replication studies from 2011)

Is the problem unique to Psychology?

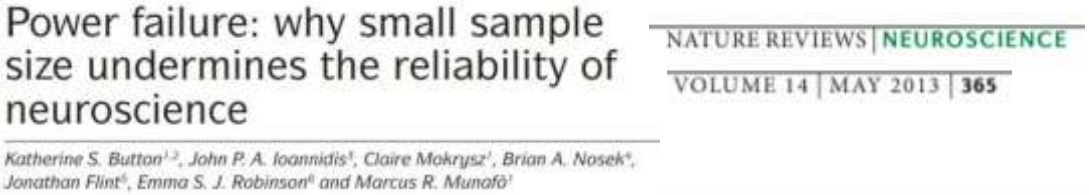
- **NO !!** (Ioannidis, 2005)


Why Most Published Research Findings Are False

John P. A. Ioannidis

PLoS Medicine | www.plosmedicine August 2005 | Volume 2 | Issue 8 | e124

- Average power in **Neuroscience**: .21 (Button et al., 2013)
  This means around 1/5 chance of positive findings (which means there must be many published false-positive findings…)


Power failure: why small sample size undermines the reliability of neuroscience

Katherine S. Button[1,2], John P. A. Ioannidis[3], Claire Mokrysz[1], Brian A. Nosek[4], Jonathan Flint[5], Emma S. J. Robinson[6] and Marcus R. Munafò[1]

NATURE REVIEWS | NEUROSCIENCE

VOLUME 14 | MAY 2013 | 365

- **Cancer Biology**: Replication rate of main results from pre-clinical trails (Begley & Ellis, 2012): from 11% to 25%. Recent reproducibility study ongoing (?)


IN FOCUS NEWS

CANCER BIOLOGY

Reproducibility project yields muddy results

An ambitious effort to replicate cancer studies is provoking controversy.

BY MONYA BAKER AND ELIE DOLGIN


CRITIQUE & DEBATE

National Science Review
5: 619–624, 2018
doi: 10.1093/nsr/nwy021
Advance access publication 2 February 2018

MOLECULAR BIOLOGY & GENETICS

On the low reproducibility of cancer studies

Haijun Wen[1], Hurng-Yi Wang[2], Xionglei He[1] and Chung-I Wu[1,3,*]

Replicability Projects

RESEARCH ARTICLE

SCIENCE

PSYCHOLOGY

Estimating the reproducibi psychological science

Open Science Collaboration*†

Reproducibility is a defining feature of science, but the extent to wh current research is unknown. We conducted replications of 100 exp

Social Psychology 2014; Vol. 45(3):142–152
DOI: 10.1027/1864-9335/a000178

Replication

ECONOMICS

Evaluating replicability of laboratory experiments in economics

Colin F. Camerer,[1*†] Anna Dreber,[2†] Eskil Forsell,[2†] Teck-Hua Ho,[3,4†] Jürgen Huber,[5†] Magnus Johannesson,[2†] Michael Kirchler,[5,6†] Johan Almenberg,[7] Adam Altmejd,[2] Taizan Chan,[8] Emma Heikensten,[2] Felix Holzmeister,[5] Taisuke Imai,[1] Siri Isaksson,[2] Gideon Nave,[1] Thomas Pfeiffer,[9,10] Michael Razen,[5] Hang Wu[4]

Review of Philosophy and Psychology
pp 1–36 | Cite as

Estimating the Reproducibility of Experimental Philosophy

Authors | Authors and affiliations

Florian Cova, Brent Strickland, Angela Abatista, Aurélien Allard, James Andow, Mario Attie, James Be

Renatas Bernūnas, Jordane Boudesseul, Matteo Colombo, Fiery Cushman, Rodrigo Diaz, Noah N'Djaye

nature human behaviour

LETTERS

https://doi.org/10.1038/s41562-018-0399-z

Evaluating the replicability of social science experiments in *Nature* and *Science* between 5

reber[2,16], Felix Holzmeister[3,16], Teck-Hua Ho[4,16], Jürgen Huber[3,16],

Psychological Science
2019, Vol. 30(5) 711–727

How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project

TC

Christopher J. Soto
Department of Psychology, Colby College

Registered Replication Report

ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Many Labs 2: Investigating Variation in Replicability Across Samples and Settings

Richard A. Klein[1], Michelangelo Vianello[2], Fred Hasselman[3,4],

Advances in Methods and Practices in Psychological Science
2018, Vol. 1(4) 443–490

DOI: 10.1177/2515245918810225
www.psychologicalscience.org/AMPPS
SAGE

RP Experimental Economics

ECONOMICS

Evaluating replicability of laboratory experiments in economics

Colin F. Camerer,[1*†] Anna Dreber,[2†] Eskil Forsell,[2†] Teck-Hua Ho,[3,4†] Jürgen Huber,[5†] Magnus Johannesson,[2†] Michael Kirchler,[5,6†] Johan Almenberg,[7] Adam Altmejd,[2] Taizan Chan,[8] Emma Heikensten,[2] Felix Holzmeister,[5] Taisuke Imai,[1] Siri Isaksson,[2] Gideon Nave,[1] Thomas Pfeiffer,[9,10] Michael Razen,[5] Hang Wu[4]

The replicability of some scientific findings has recently been called into question. To contribute data about replicability in economics, we replicated 18 studies published in the *American Economic Review* and the *Quarterly Journal of Economics* between 2011 and 2014. All of these replications followed predefined analysis plans that were made publicly available beforehand, and they all have a statistical power of at least 90% to detect the original effect size at the 5% significance level. We found a significant effect in the same direction as in the original study for 11 replications (61%); on average, the replicated effect size is 66% of the original. The replicability rate varies between 67% and 78% for four additional replicability indicators, including a prediction market measure of peer beliefs.

SCIENCE sciencemag.org

25 MARCH 2016 • VOL 351 ISSUE 6280 1433

61% studies replicated (significant p value)

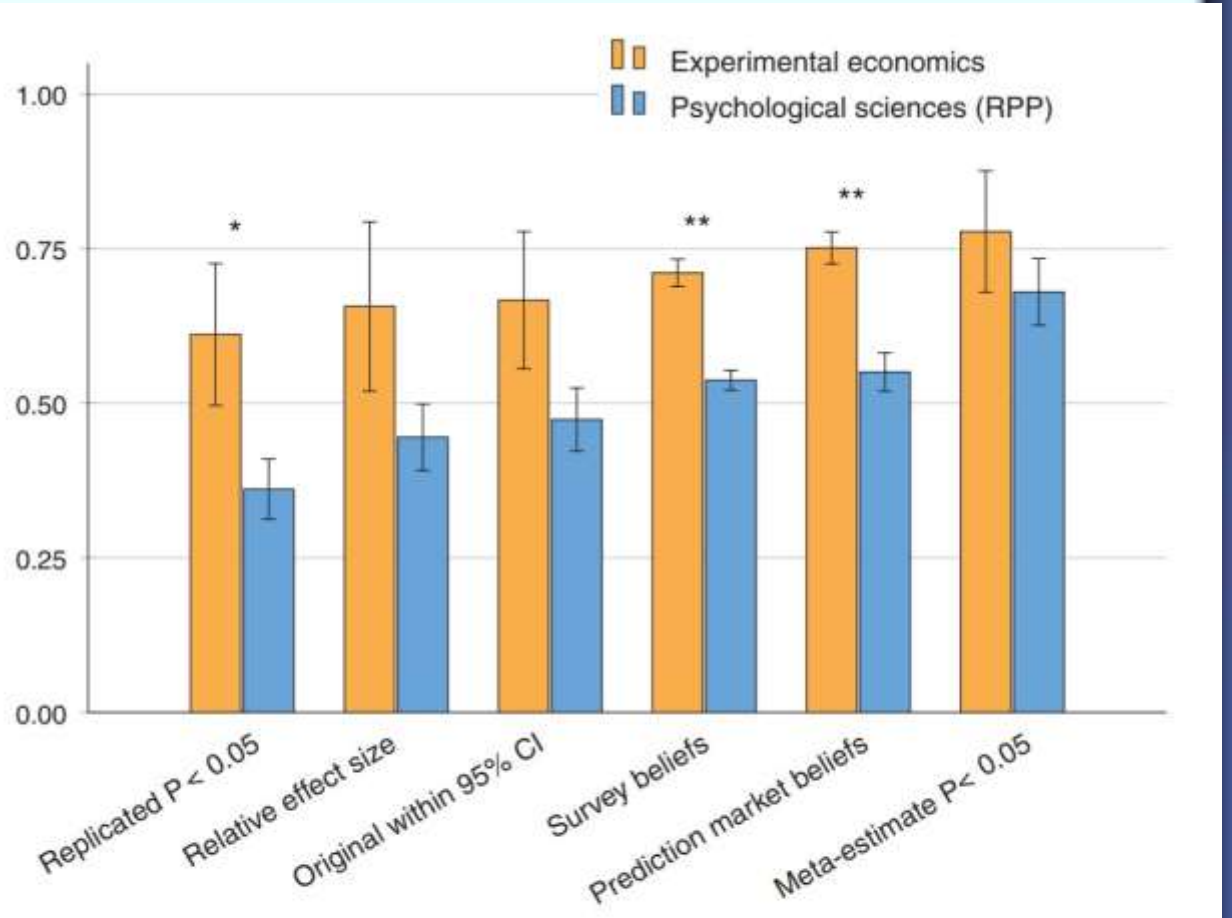


Fig. 4. A comparison of replicability indicators in experimental economics (this study) and psychological sciences (RPP). The graph shows means ± SE for replicability indicators. All six replicability indicators are higher for experimental economics; this difference is significant for three of the replicability indicators. The average difference in replicability across the six indicators is 19 percentage points. Details about the statistical tests are included in the supplementary materials. *$P < 0.05$; **$P < 0.01$.

RP Experimental Philosophy

Review of Philosophy and Psychology
pp 1–36 | Cite as

Estimating the Reproducibility of Experimental Philosophy

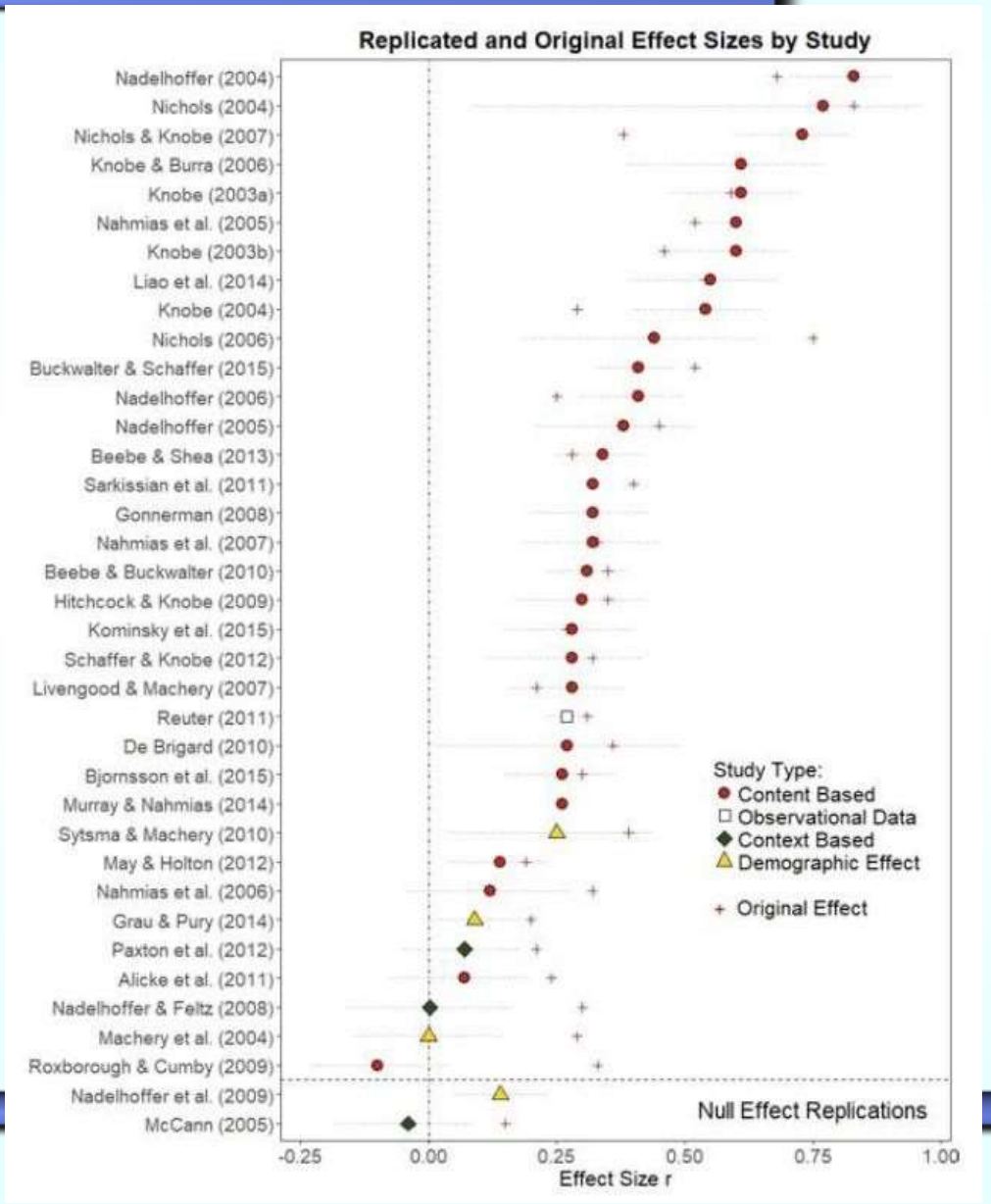
Authors | Authors and affiliations

Florian Cova, Brent Strickland, Angela Abatista, Aurélien Allard, James Andow, Mario Attie, James Be
Renatas Berniūnas, Jordane Boudesseul, Matteo Colombo, Fiery Cushman, Rodrigo Diaz, Noah N'Djaye
Vilius Dranseika, Brian D. Earp, show 27 more

Article
First Online: 14 June 2018

16 Shares | 450 Downloads | 1 Citations

75% studies replicated (significant p value)



RP Social Science in N & S

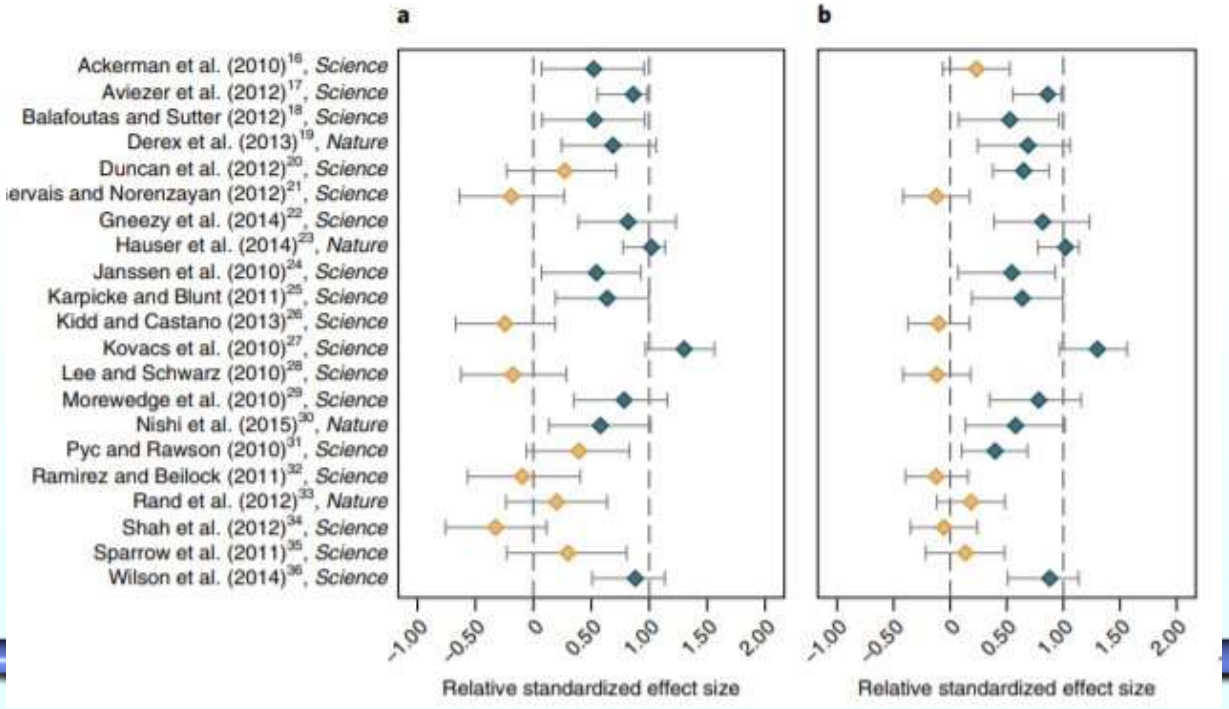
nature human behaviour

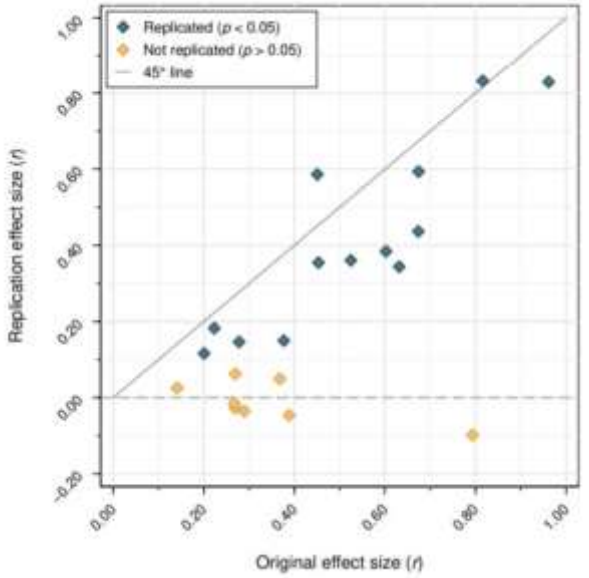
LETTERS

https://doi.org/10.1038/s41562-018-0399-z

Evaluating the replicability of social science experiments in *Nature* and *Science* between 2010 and 2015

Colin F. Camerer[1,16], Anna Dreber[2,16], Felix Holzmeister[3,16], Teck-Hua Ho[4,16], Jürgen Huber[3,16], Magnus Johannesson[2,16], Michael Kirchler[3,5,16], Gideon Nave[6,16], Brian A. Nosek[7,8,16*], Thomas Pfeiffer[9,16], Adam Altmejd[2], Nick Buttrick[7,8], Taizan Chan[10], Yiling Chen[11], Eskil Forsell[12], Anup Gampa[7,8], Emma Heikensten[2], Lily Hummer[8], Taisuke Imai[13], Siri Isaksson[2], Dylan Manfredi[6], Julia Rose[3], Eric-Jan Wagenmakers[14] and Hang Wu[15]





62% studies replicated
(significant p value)

Effect Size is around half
(d=0.51 vs. 1.04)

Many Labs 2 (2018)

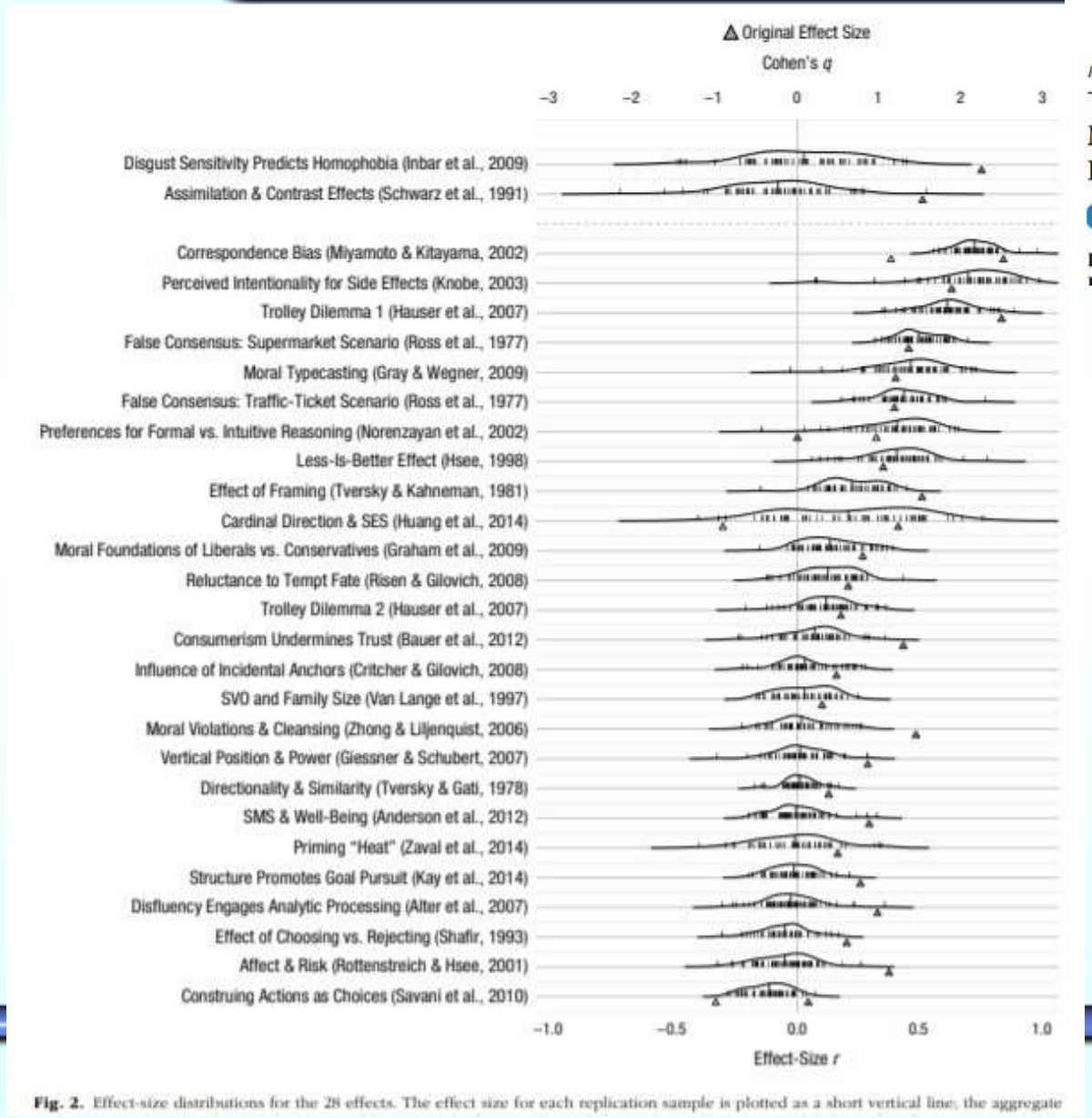


Fig. 2. Effect-size distributions for the 28 effects. The effect size for each replication sample is plotted as a short vertical line; the aggregate

Registered Replication Report

Many Labs 2: Investigating Variation in Replicability Across Samples and Settings

Richard A. Klein[1], Michelangelo Vianello[2], Fred Hasselman[3,4], Byron G. Adams[5,6], Reginald B. Adams, Jr.[7], Sinan Alper[8]

Advances in Methods and Practices in Psychological Science
2018, Vol. 1(4) 443–490
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2515245918810225
www.psychologicalscience.org/AMPPS

14/28 (50%) sig. effect

1/28 (4%) weak effect

13/28 (46%) no or opposite effect

Meta-analysis vs. Replication

nature human behaviour

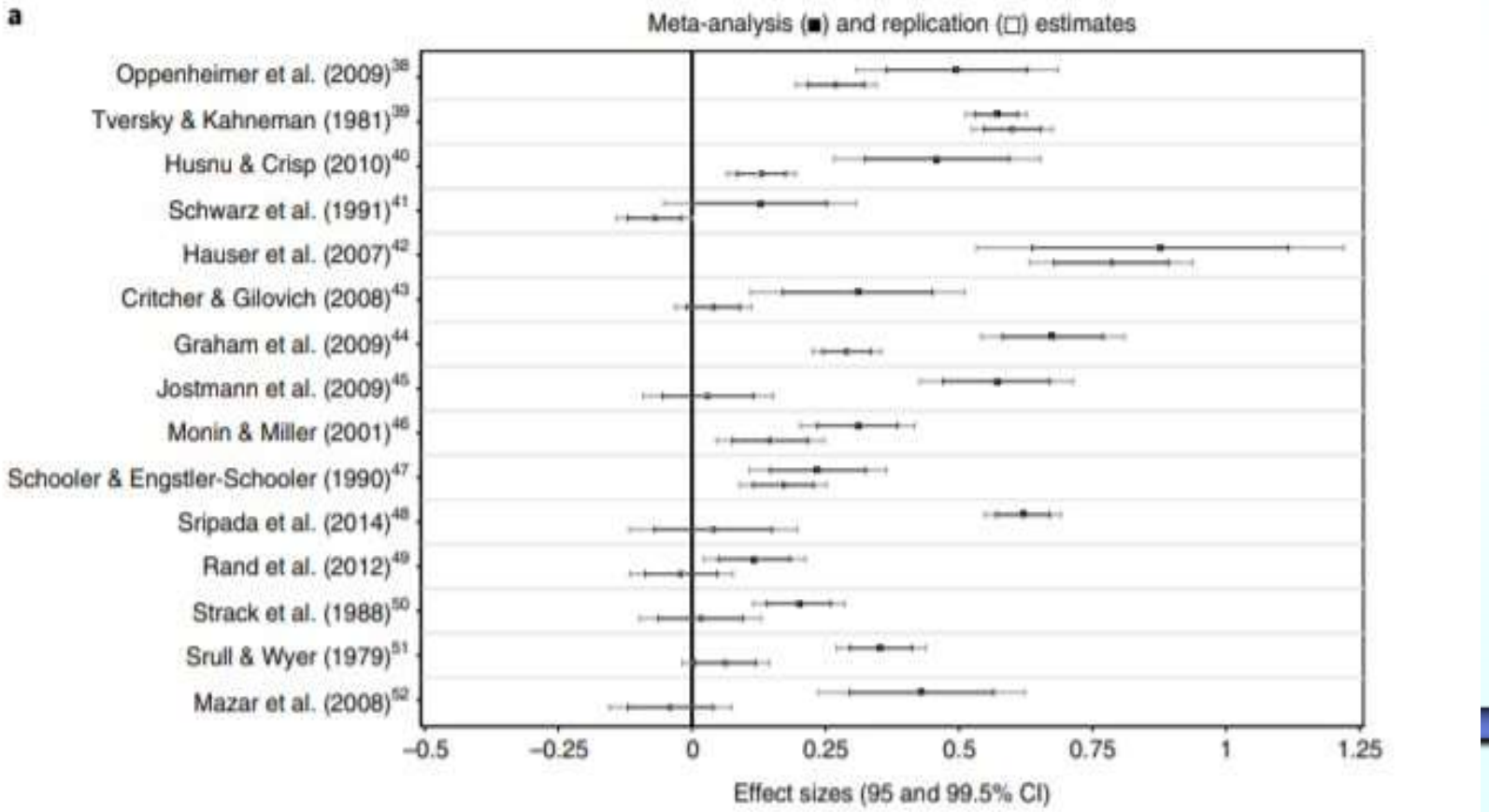
ARTICLES

https://doi.org/10.1038/s41562-019-0787-z

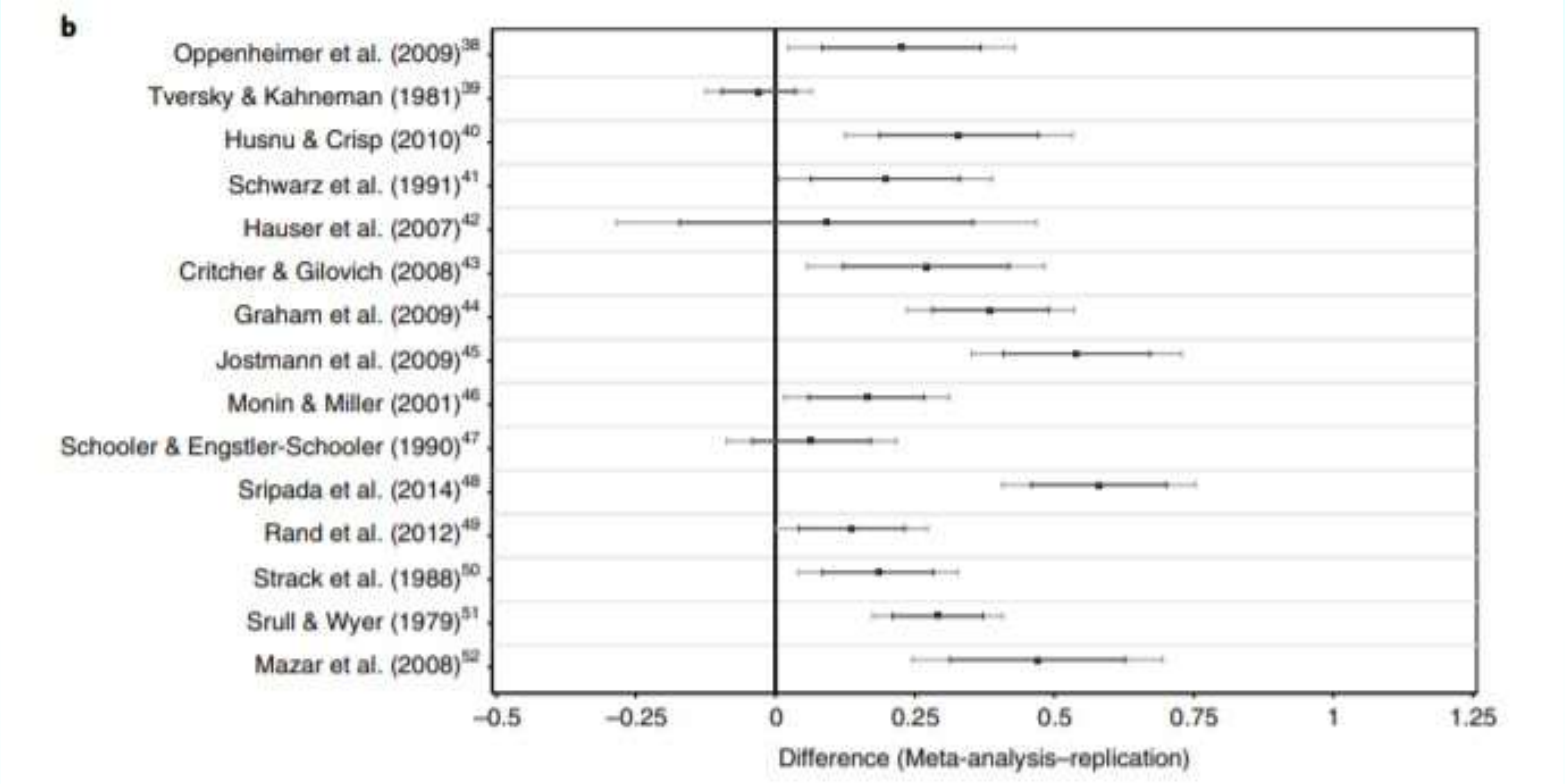
Corrected: Author correction

Comparing meta-analyses and preregistered multiple-laboratory replication projects

Amanda Kvarven[1,3], Eirik Strømland[1,3] and Magnus Johannesson[2*]



Meta-analysis vs. Replication



7/15 (47%) replicated significant effect

12/15 (80%) smaller ES

Average ES (Cohen's d): 0.16 vs. 0.42

Money Priming ! (Vohs et al)

Journal of Experimental Psychology: General
2013, Vol. 142, No. 2, 301–306

0096-3445/13/$12.00 DOI: 10.1037/a0029288

BRIEF REPORT

Mere Exposure to Money Increases Endorsement of Free-Market Systems and Social Inequality

Eugene M. Caruso
University of Chicago

Kathleen D. Vohs
University of Minnesota

Brittani Baxter
University of Chicago

Adam Waytz
Northwestern University

CURRENT DIRECTIONS IN PSYCHOLOGICAL SCIENCE

Merely Activating the Concept of Money Changes Personal and Interpersonal Behavior

Kathleen D. Vohs,[1] Nicole L. Mead,[2] and Miranda R. Goode[3]

[1]Department of Marketing, Carlson School of Management, University of Minnesota, [2]Department of Psychology, Florida State University, and [3]Sauder School of Business, University of British Columbia

REPORTS 17 NOVEMBER 2006 VOL 314 SCIENCE www.sciencemag.org

The Psychological Consequences of Money

Kathleen D. Vohs,[1]* Nicole L. Mead,[2] Miranda R. Goode[3]

Money has been said to change people's motivation (mainly for the better) and their behavior toward others (mainly for the worse). The results of nine experiments suggest that money brings about a self-sufficient orientation in which people prefer to be free of dependency and dependents. Reminders of money, relative to nonmoney reminders, led to reduced requests for help and reduced helpfulness toward others. Relative to participants primed with neutral concepts, participants primed with money preferred to play alone, work alone, and put more physical distance between themselves and a new acquaintance

Money Priming example

Journal of Experimental Psychology: General
2013, Vol. 142, No. 2, 301–306

0096-3445/13/$12.00 DOI: 10.1037/a0029288

BRIEF REPORT

Mere Exposure to Money Increases Endorsement of Free-Market Systems and Social Inequality

Eugene M. Caruso
University of Chicago

Kathleen D. Vohs
University of Minnesota

Brittani Baxter
University of Chicago

Adam Waytz
Northwestern University

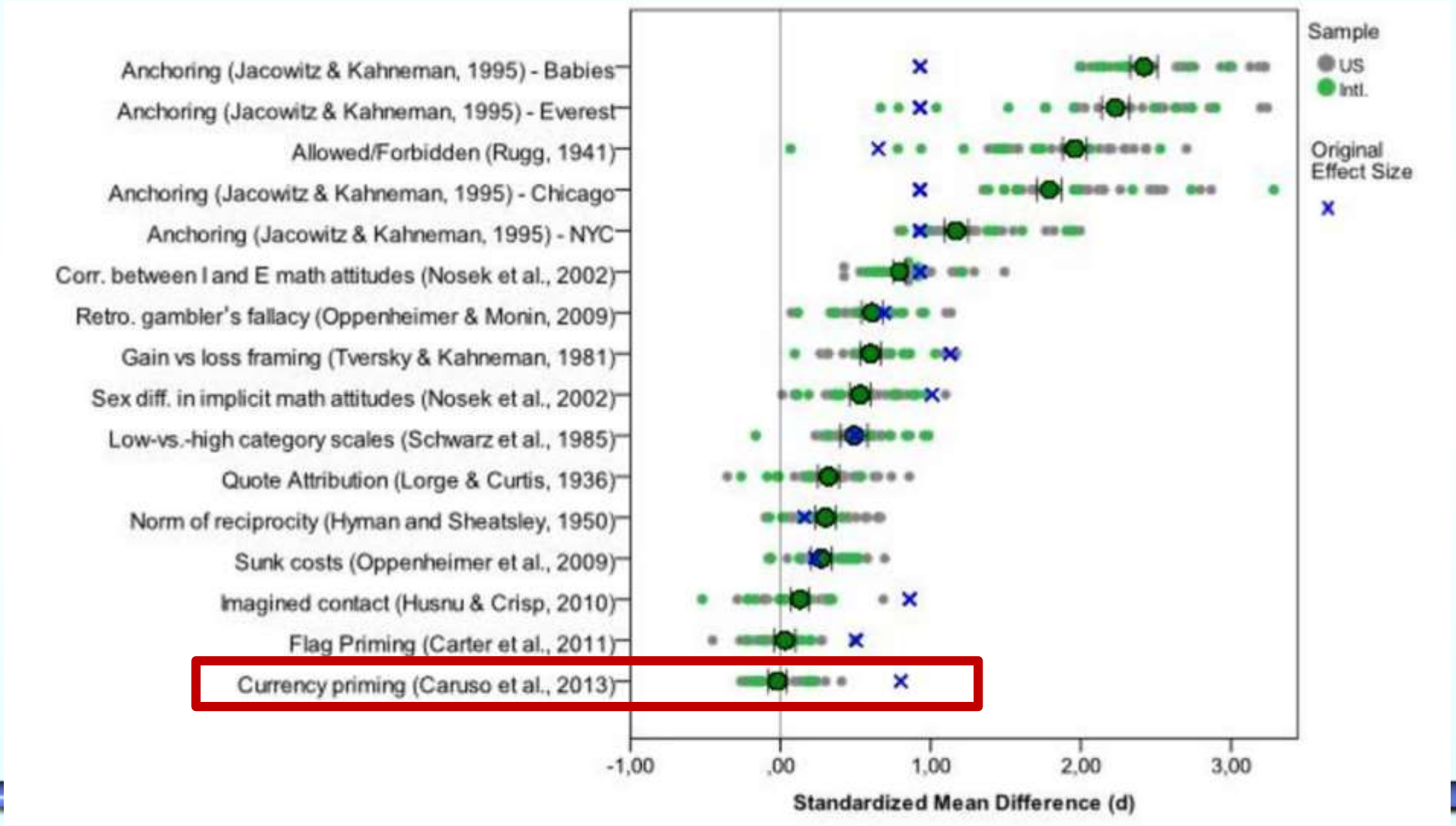
Background-image conditions. Participants assigned to the *background-image/money condition* saw a faint image of $100 bills in the background of the initial instruction screen, whereas participants assigned to the *background-image/control condition* saw a blurred version of this image, such that the bills were unrecognizable (Caruso et al., 2013; for a similar manipulation, see Kushlev, Dunn, & Ashton-James, 2012).

Money Prime

Control Prime

Figure 1. Images used for the money prime condition and control condition (Experiments 1, 4, and 5).

Many Labs (SP, 2014)



Money Priming ???

Show Me the Money: A Systematic Exploration of Manipulations, Moderators, and Mechanisms of Priming Effects

Psychological Science
2017, Vol. 28(8) 1148–1159
© The Author(s) 2017
Reprints and permissions:
sagepub.com/journalsPermissions.nav
DOI: 10.1177/0956797617706161
www.psychologicalscience.org/PS
SAGE

Eugene M. Caruso[1], Oren Shapira[2], and Justin F. Landy[1]
[1]Booth School of Business, University of Chicago, and [2]Department of Medicine, Stony Brook University

Abstract

A major challenge for accumulating knowledge in psychology is the variation in methods and participant populations across studies in a single domain. We offer a systematic approach to addressing this challenge and implement it in the domain of money priming. In three preregistered experiments (N = 4,649), participants were exposed to one of a number of money manipulations before completing self-report measures of money activation (Study 1); engaging in a behavioral-persistence task (Study 3); completing self-report measures of subjective wealth, self-sufficiency, and communion-agency (Studies 1–3); and completing demographic questions (Studies 1–3). Four of the five manipulations we tested activated the concept of money, but, contrary to what we expected based on the preponderance of the published literature, no manipulation consistently affected any dependent measure. Moderation by sociodemographic characteristics was sparse and inconsistent across studies. We discuss implications for theories of money priming and explain how our approach can complement recent efforts to build a reproducible, cumulative psychological science.

But sometimes most effects get replicated…

Personality (2019)

Psychological Science
2019, Vol. 30(5) 711–727
© The Author(s) 2019
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/0956797619831612
www.psychologicalscience.org/PS
SAGE

How Replicable Are Links Between Personality Traits and Consequential Life Outcomes? The Life Outcomes of Personality Replication Project

Christopher J. Soto
Department of Psychology, Colby College

66/76 (82%) sig. effects
ES: d=0.47 vs 0.61
<u>Most original outcomes based on large samples rather than on searching for p<.05</u>

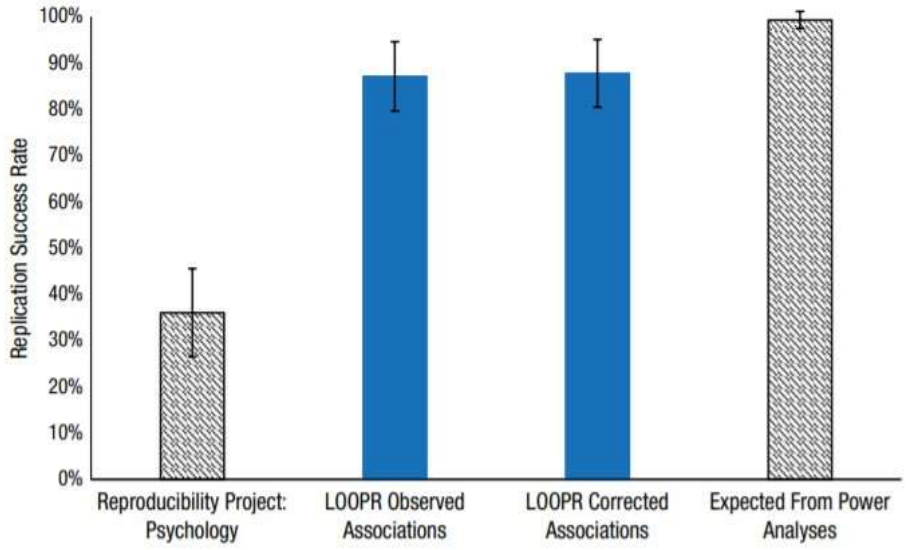


Fig. 1. Replication success rates obtained in the Life Outcomes of Personality Replication (LOOPR) Project, compared with the rate expected from power analyses of the original effect size and replication sample size and with the rate obtained in the Reproducibility Project: Psychology. A successful replication was defined as a statistically significant effect (i.e., two-tailed $p < .05$) in the hypothesized direction. Corrected associations were partially disattenuated to correct for the unreliability of abbreviated outcome measures. Error bars represent 95% confidence intervals.

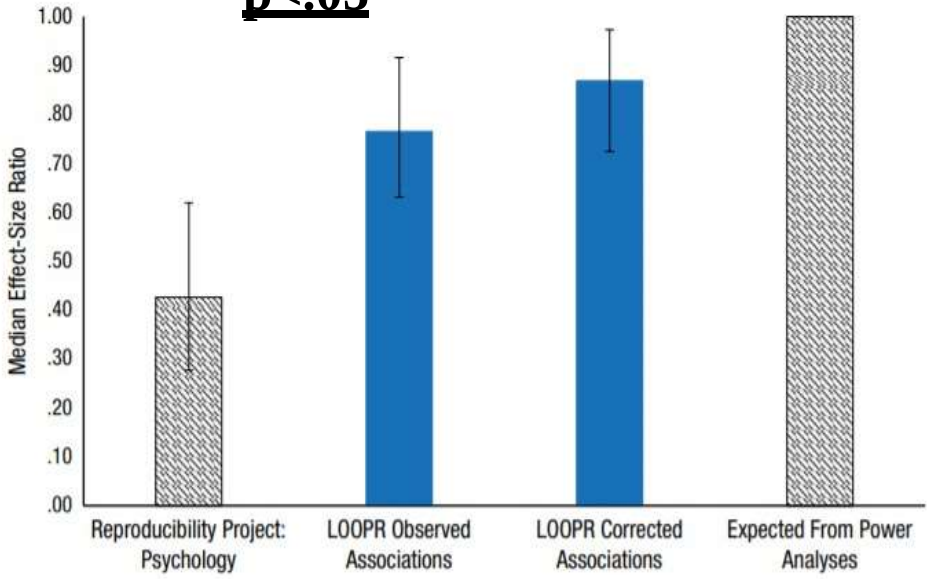


Fig. 2. Median effect-size ratios obtained in the Life Outcomes of Personality Replication (LOOPR) Project, compared with the ratio expected if all original effect sizes represented true effects and with the median ratio obtained in the Reproducibility Project: Psychology. Effect-size ratios were computed as the ratio of the z-transformed replication effect size to the transformed original effect size. Corrected associations were partially disattenuated to correct for the unreliability of abbreviated outcome measures. Error bars represent 95% confidence intervals.

Classic findings in Cognitive Psychology

Psychon Bull Rev (2018) 25:1968–1972
DOI 10.3758/s13423-017-1348-y

BRIEF REPORT

Participant Nonnaiveté and the reproducibility of cognitive psychology

Rolf A. Zwaan[1] · Diane Pecher[1] · Gabriele Paolacci[2] · Samantha Bouwmeester[1] · Peter Verkoeijen[1,3] · Katinka Dijkstra[1] · René Zeelenberg[1]

All nine effects replicated (highly significant effects)

<u>All within Ss studies</u>:

Table 1 Brief descriptions of and references to all replicated experiments

Number	Task	Description	Reference
1	*Simon task*	*Choice-reaction time task that measures spatial compatibility. Responses are faster when a visual target (a red square is presented on the left of the screen) is spatially compatible with the response (pressing the left button) than when the target is spatially incompatible with the response (presented on the right of the screen).*	Craft and Simon (1970)
2	*Flanker task*	*Response inhibition task in which relevant information is selected and inappropriate responses in a certain context are suppressed. Responses are faster for congruent trials in which compatible distractors flank a central target (AAAAA) than for incongruent trials in which incompatible distractors flank a central target (AAEAA).*	Eriksen and Eriksen (1974)
3	*Motor priming (a = masked, b = unmasked)*	*A task with a priming procedure in which responses to stimuli (arrow probes <<) are required that are primed by presented compatible (<<) or incompatible (>>) items. Responses are slower for compatible items when primes are masked but faster when primes are visible.*	Forster and Davis (1984)
4	*Spacing effect*	*Learning task in which learning (of words) is spaced over time. Recall of words is higher for spaced item repetitions with intervening items than for massed items immediately repeated after their first presentation.*	Greene (1989)
5	*False memories*	*Memory task that assesses false memory of recognition performance of items that have not been presented before in a word list but tend to be recognized as presented before because they are semantically related to the words in the list.*	Roediger and McDermott (1995)
6	*Serial position (a = primacy, b = recency)*	*Memory task that examines recall probability based on a word's position in a list. Recall is higher for the first and last words in the list and lowest for items in the middle of the list.*	Murdock (1962)
7	*Associative priming*	*Implicit memory task which requires a response to a target word that is preceded by prime word. Responses are faster when the prime is related than when the prime is unrelated.*	Meyer and Schvaneveldt (1971)
8	*Repetition priming (a = low frequency, b = high frequency)*	*Implicit memory task in which speed of response depends on previous exposure to an item and the word frequency of that item. Responses are faster for repeated than for new items. This repetition effect is larger for low frequency words than high frequency words.*	Forster and Davis (1984)
9	*Shape simulation*	*Sentence-verification task that requires a response on whether the object in a picture was present in the previous sentence. Yes responses are faster when the picture matches the implied shape mentioned in sentence than when it mismatches.*	Zwaan, Yaxley, and Stanfield (2002)

Some initial take-home messages

- Just because something has been shown once, it does not mean it is a consolidated fact
- There are robust and replicable effects and elusive effects
- Contextual variations (e.g., country, online vs. lab) seem to matter less than the effect as such
- Conceptual replication is not the same as direct replication, meta-analysis is not the same as multi-lab direct replication

Consequences

- Replicability is starting to have a stable place in Psychology (and in other sciences too)
- Beliefs in some effects is currently low (e.g., ego-depletion) but in others is high (e.g., anchoring)
- Research and publication standards are changing
- More awareness of false-positives and QRPs
- More attention to methodological issues
- More sophisticated statistical approaches

(Some) Tips for getting it right

The signal and the noise

- As scientists, we all want to get something right
- If we get it right, it is replicable and will be replicated. But what does it mean “to get it right”?
- Positive consequences of the replicability crisis
- Increase the **signal**, decrease the **noise**

Some pointers on how to get it right

1st pointer: Power

- Design your study with adequate power (***probability of finding an effect if it does exist***)

- Underpowered studies produce conflicting evidence and false negatives **but also false positives** (Maxwell, 2004; Ioannidis, 2005).
  More power means less overall inference errors

- Direct effect on **False Negatives** but also indirect effect on **False Positives**
  (**False Discovery Rate /True False Positives**)

Why many effects are not replicated?

- A mix of different factors and possible explanations
- Two main factors

  a) Low power and **b) Publication bias**

- Under these conditions, it is predictable that the literature will contain many false positives (results that seems significant but are not) and artificially boosted effect sizes
- Hence effects will be difficult to replicate

1.a Low power

Is a real problem for ψ?

PLOS | BIOLOGY

META-RESEARCH ARTICLE

Empirical assessment of published effect sizes and power in the recent cognitive neuroscience and psychology literature

Denes Szucs[1]*, John P. A. Ioannidis[2]

1 Department of Psychology, University of Cambridge, Cambridge, United Kingdom, 2 Meta-Research Innovation Center at Stanford (METRICS) and Department of Medicine, Department of Health Research and Policy, and Department of Statistics, Stanford University, Stanford, California, United States of America

We have empirically assessed the distribution of published effect sizes and estimated power by analyzing 26,841 statistical records from 3,801 cognitive neuroscience and psychology papers published recently. The reported median effect size was D = 0.93 (interquartile range: 0.64–1.46) for nominally statistically significant results and D = 0.24 (0.11–0.42) for nonsignificant results. Median power to detect small, medium, and large effects was 0.12, 0.44, and 0.73, reflecting no improvement through the past half-century. This is so because sample sizes have remained small. Assuming similar true effect sizes in both disci-

Psychological Science
2016, Vol. 27(8) 1069–1077

DOI: 10.1177/0956797616647519
pss.sagepub.com
SAGE

Researchers' Intuitions About Power in Psychological Research

Marjan Bakker[1], Chris H. J. Hartgerink[1], Jelte M. Wicherts[1], and Han L. J. van der Maas[2]

[1]Department of Methodology and Statistics, Tilburg School of Social and Behavioral Sciences, Tilburg University, and [2]Department of Psychology, Psychological Methods, University of Amsterdam

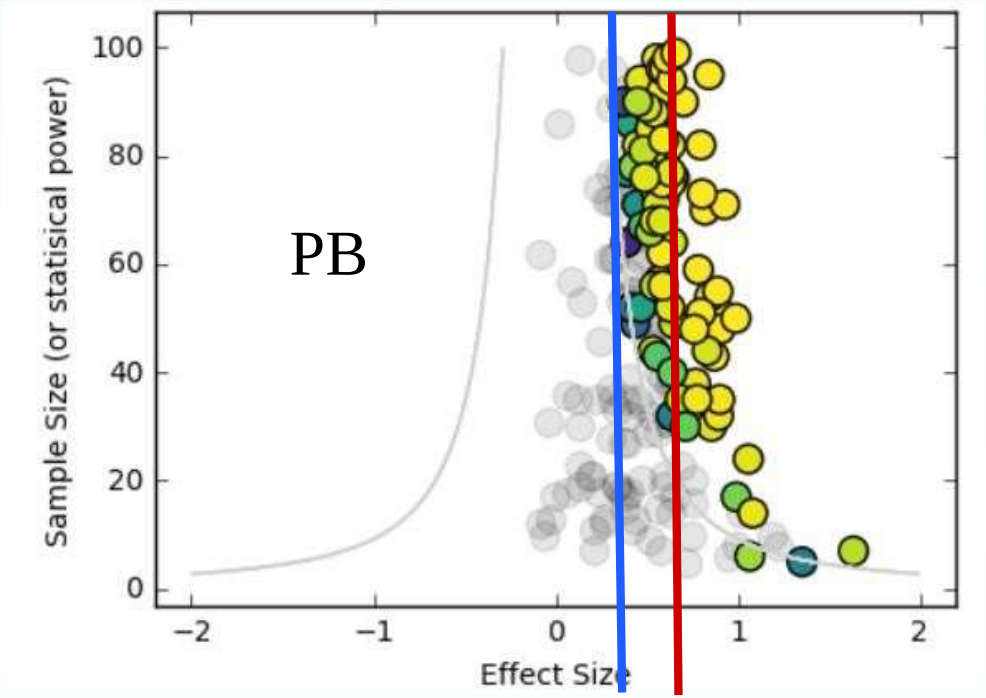
1990; Maxwell, 2004). Specifically, given the typical effect sizes (ESs) and sample sizes reported in the psychological literature, the statistical power of a typical two-group between-subjects design has been estimated to be less than .50 (Cohen, 1990) or even .35 (Bakker et al., 2012). These low power estimates appear to con-

Yes!

1.b Publication bias

- Tendency to publish mainly significant results (and to submit for publication mainly studies with significant results)
- There are sometimes understandable reasons (unclear evidence, contradictory support, pilot studies, tentative paradigms, etc.)
- But often is a by-product of confirmation/positivity biases and insufficient culture of cumulative knowledge in a scientific field

Publication bias



The ES will be overestimated. How much depends on the extent of PB and on the prevalence of small samples.

A reader will think that Cohen's d=0.60 but in fact is d=0.30

Publication bias, Effect Sizes, underpowered studies

ES: Cohen's d=0.60 (vs. d= 0.30)

N for power:

80%	**90%**
72 Ss (vs. 278)	98 Ss (vs. 382)

Suppose we run a study with 98 Ss.
Expected power is 0.90 but **real power will be 0.43**

Vicious cycle: PB leads to overestimated ES leading to underpowered studies leading to non replicated effects, **even assuming that the effects are true and the researchers do not "cheat"**

Publication bias, Effect Sizes, sample sizes

Without publication bias, there should be **no** relation (r=0)

September 2014 | Volume 9 | Issue 9 | e105825

OPEN ACCESS Freely available online

PLOS | ONE

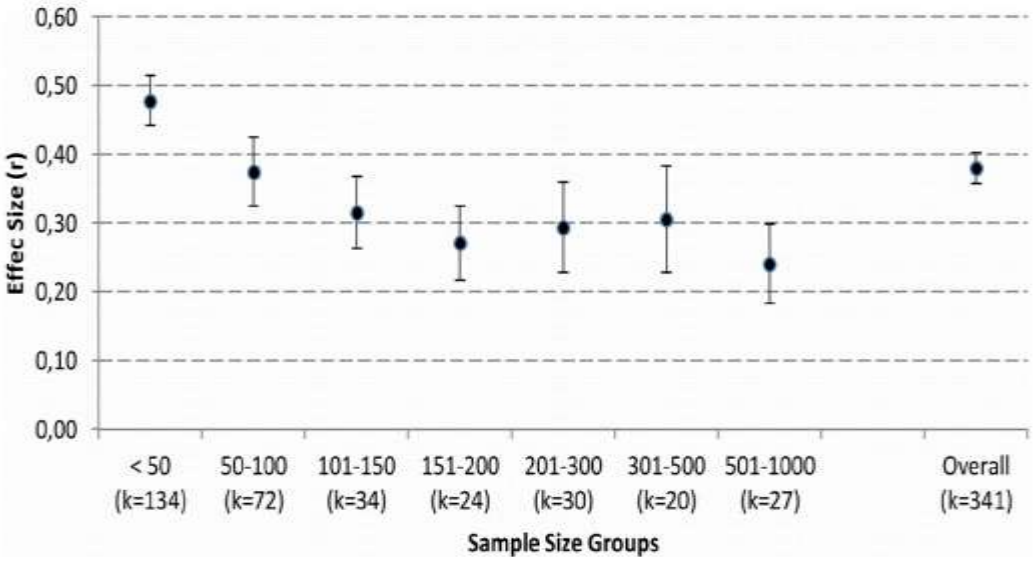
Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size

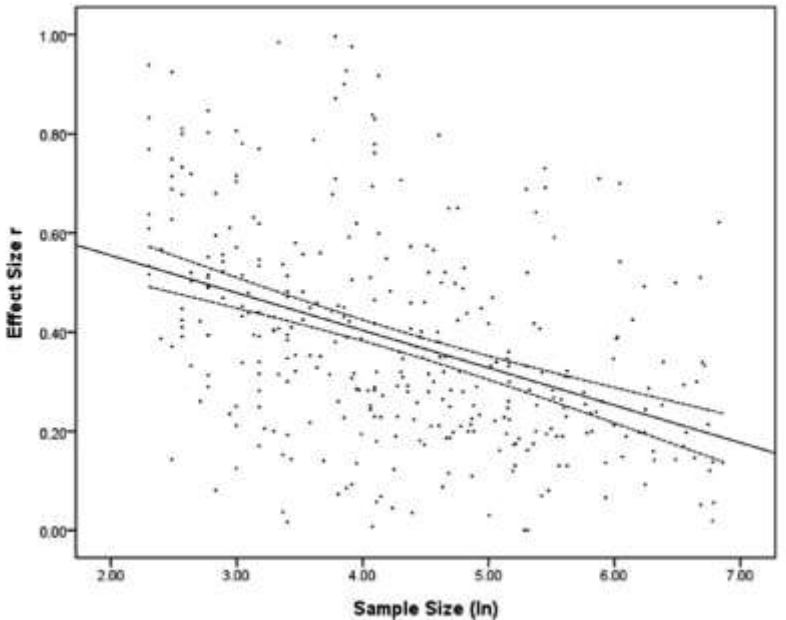
Anton Kühberger[1,2]*, Astrid Fritz[3], Thomas Scherndl[1]

1 Department of Psychology, University of Salzburg, Salzburg, Austria, 2 Centre for Cognitive Neuroscience, University of Salzburg, Salzburg, Austria, 3 Österreichisches

Methods: We investigate whether effect size is independent from sample size in psychological research. We randomly sampled 1,000 psychological articles from all areas of psychological research. We extracted *p* values, effect sizes, and sample sizes of all empirical papers, and calculated the correlation between effect size and sample size, and investigated the

r = .54 !





1st pointer: Power

2nd pointer: Confirmatory studies

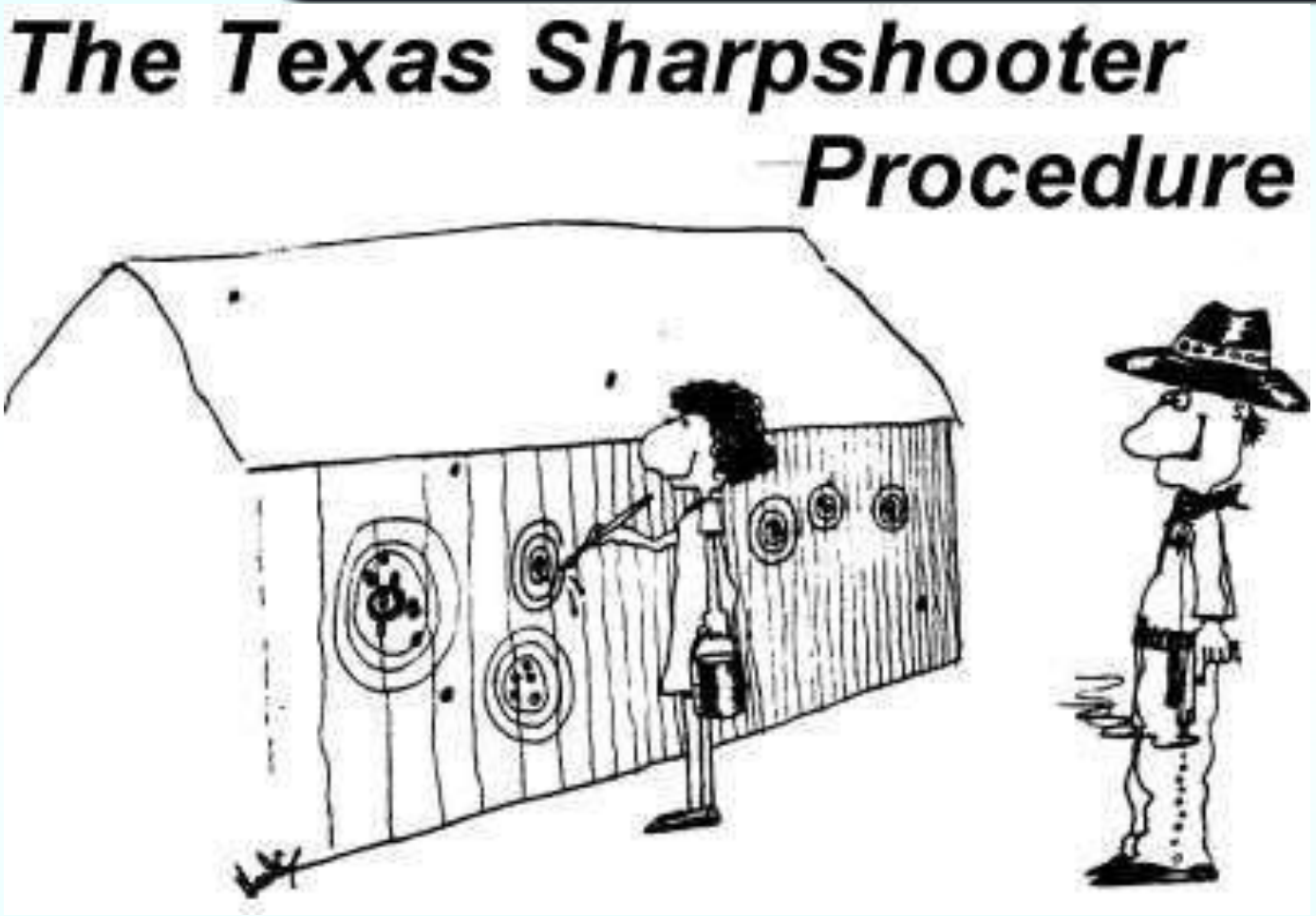
- Distinguish between exploratory and confirmatory studies
- If you find a “surprising effect”, confirm it with another well powered study before building on it
- Results can be significant simply out of random sampling
- Post-hoc ≠ pre-hoc
- Consider pre-registration (and other “badges” too)

Badges

- Encouraged practice for some journals (e.g., Psychological Science)

- More info in a later talk by Juliette and Cristina
- See also https://osf.io/tvyxz/wiki/home/ and https://aspredicted.org/

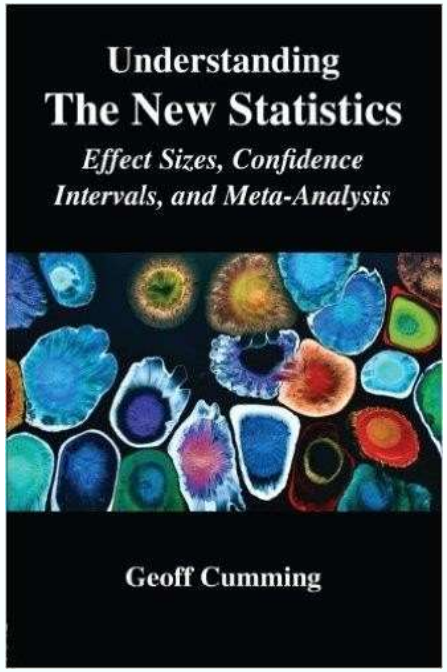
Why pre-registration?



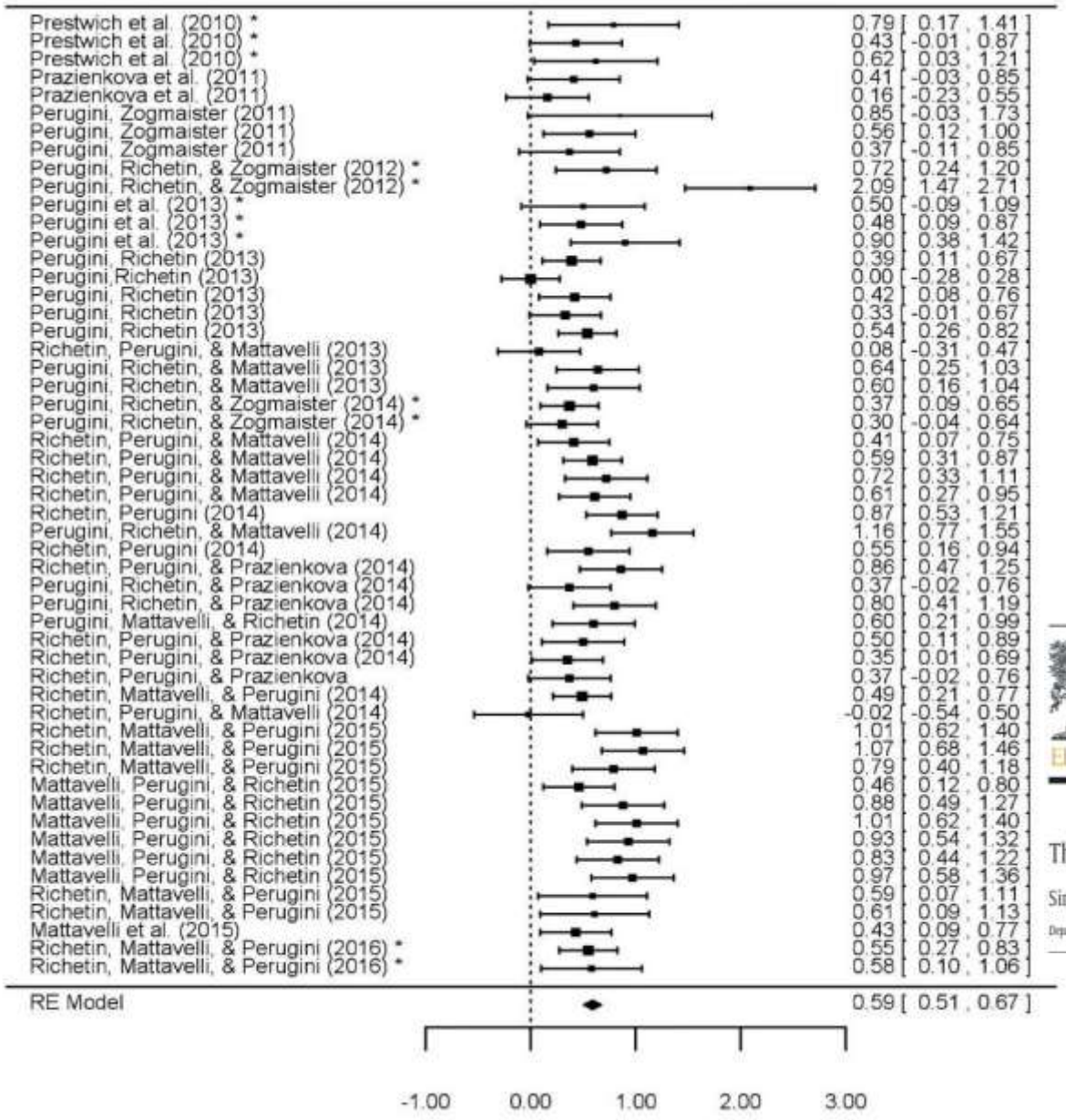
I always avoid prophesying beforehand, because it is a much better policy to prophesy after the event has already taken place (Winston Churchill, 1943)

3rd pointer: Meta-analytic

- Use meta-analytic **mind-set**
- Do not over-interpret significant or non-significant results in single studies
- The dance of p values
- There must be some studies that fail to replicate a real effect!
- Example with my own research



Real data are a bit messy…
Self-Referencing (Implicit; k=53)



d=0.59
[0.51, 0.67]


Journal of Experimental Social Psychology 71 (2017) 68–82

Contents lists available at ScienceDirect

Journal of Experimental Social Psychology

journal homepage: www.elsevier.com/locate/jesp

The Self-Referencing task: Theoretical overview and empirical evidence

Simone Mattavelli *, Juliette Richetin, Marcello Gallucci, Marco Perugini

Department of Psychology, University of Milan-Bicocca, Piazza dell'Ateneo Nuovo, 1, 20126 Milan, Italy

4th pointer: Meta-conditional

- Avoid dichotomous thinking (and, if you can, also dichotomous theorizing…)
- Everything happens under some circumstances
- Try to identify these circumstances and understand whether they are robust
- Meta-conditional approach: what, how, when, for whom, how much something happens

THE AMERICAN STATISTICIAN
2019, VOL. 73, NO. S1, 271–280: Statistical Inference in the 21st Century
https://doi.org/10.1080/00031305.2018.1518266

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

The New Statistics for Better Science: Ask How Much, How Uncertain, and What Else Is Known

Robert J. Calin-Jageman[a] and Geoff Cumming[b]

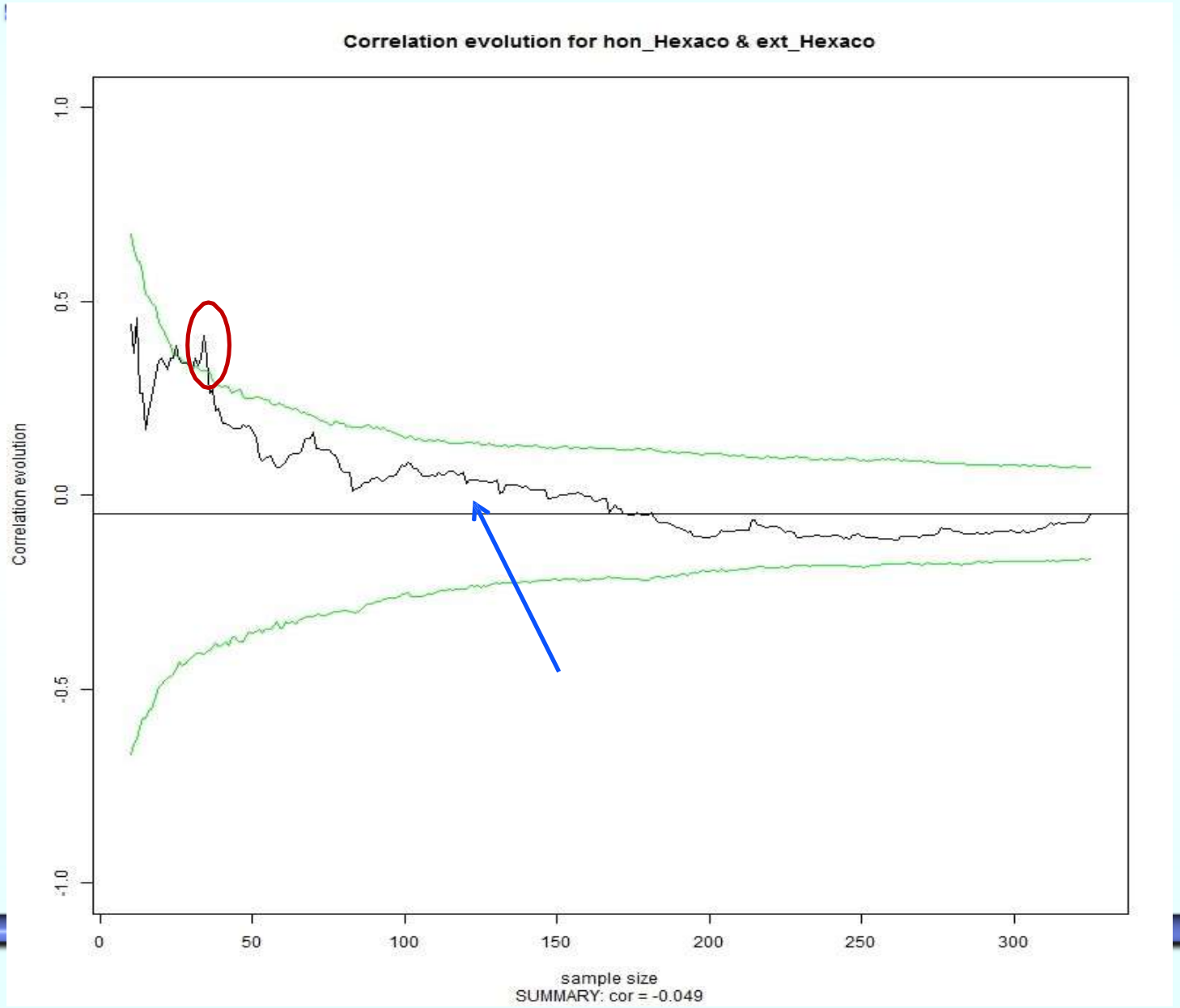
5th pointer: Stability

- Results stabilize with bigger sample sizes
- Try to have a decent sample size
- Sometimes results can be significant in opposite directions with small sample sizes
- For example, stability of correlation coefficients (cf. Schonbrodt & Perugini, 2013)

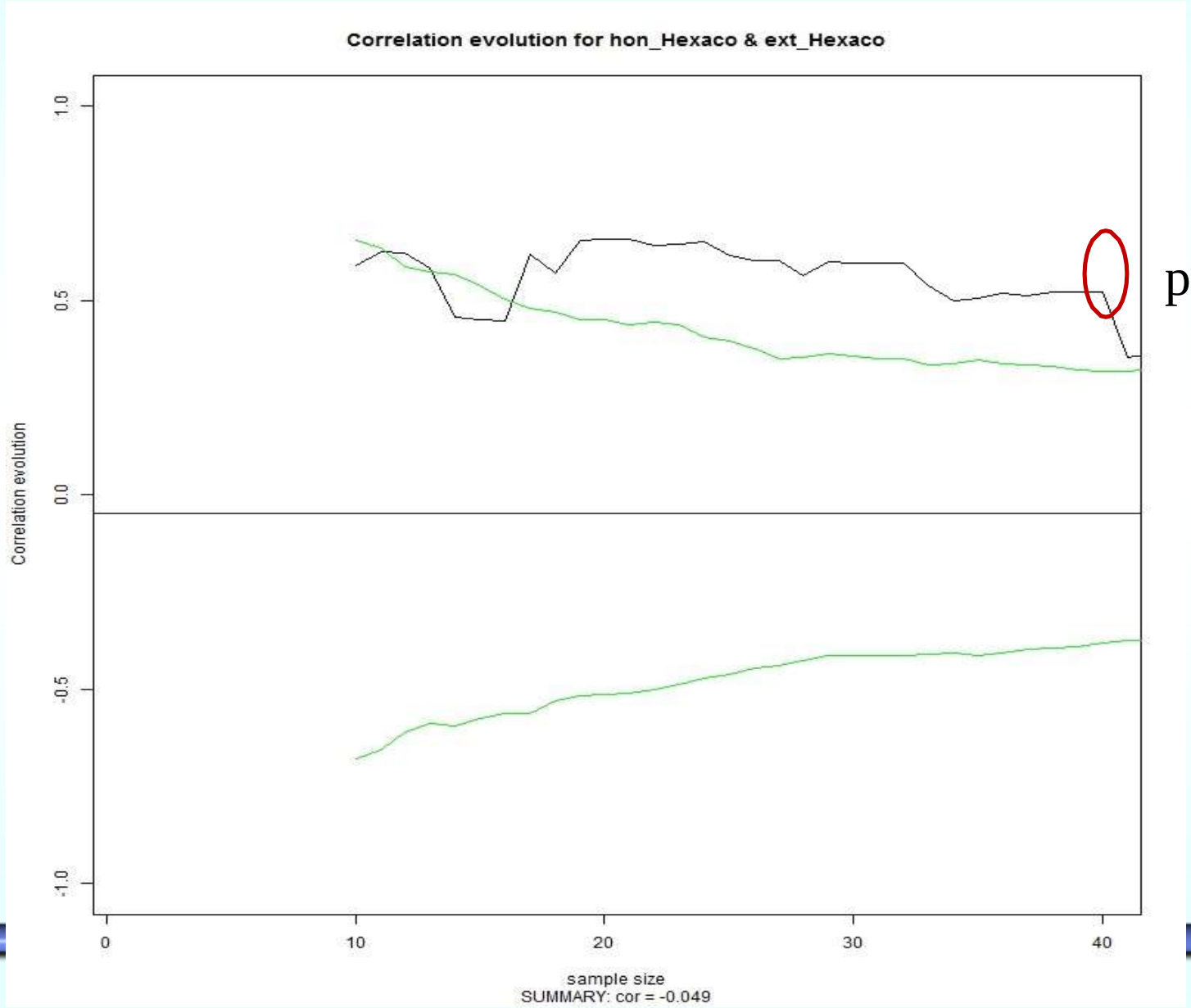
Stability of correlations

- Example from my own data
- One effect that is trivial in the full sample (r between H/H_{quest} and Ext_{quest}=-.05)
- Correlations calculated adding Ss at each step starting from N=10 to full sample (evolution of r)
- Real Ss order
- Boostrapped (s=1000) CI 95%

H vs. E


Correlation evolution for hon_Hexaco & ext_Hexaco
Correlation evolution
1.0
0.5
0.0
-0.5
-1.0
0
50
100
150
200
250
300
sample size
SUMMARY: cor = -0.049

H vs. E


Correlation evolution for hon_Hexaco & ext_Hexaco
Correlation evolution
1.0
0.5
0.0
-0.5
-1.0
0
10
20
30
40
sample size
SUMMARY: cor = -0.049

p<.001!

Contents lists available at SciVerse ScienceDirect

Journal of Research in Personality

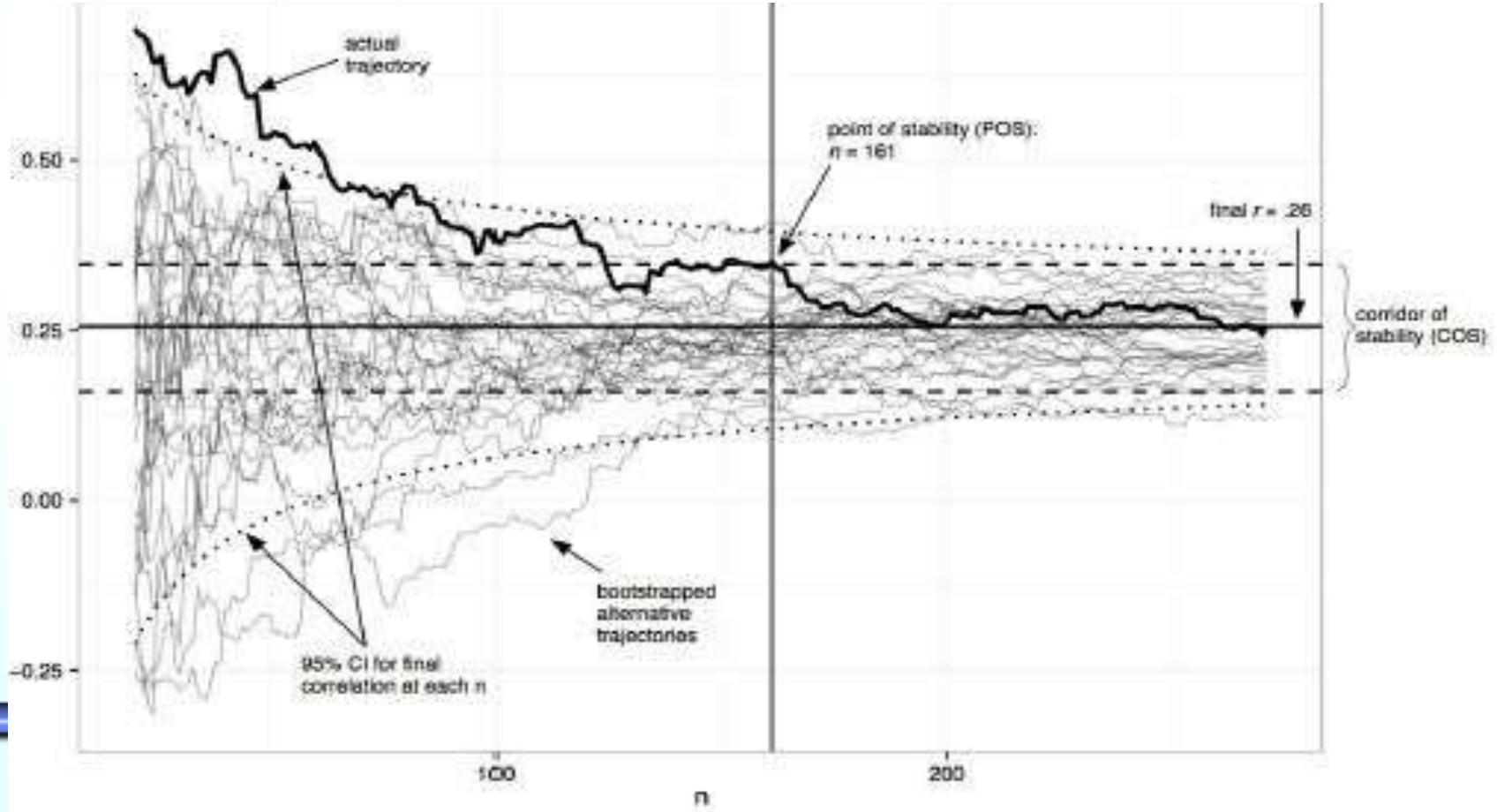
journal homepage: www.elsevier.com/locate/jrp

Brief Report

At what sample size do correlations stabilize?

Felix D. Schönbrodt [a,*], Marco Perugini [b]

[a] Department of Psychology, Ludwig-Maximilians-Universität, Leopoldstr. 13, 80802 München, Germany
[b] Department of Psychology, University of Milan, Bicocca, Piazza dell'Ateneo Nuovo 1 (U6), 20126 Milan, Italy



Implications

- Sequential effects can be devastating for small samples (e.g., N≤60)
- Estimates start to stabilize for N≥150 (but it depends on the expected correlation and desired width; e.g., with w=.1, N≈180 for r = .4 & N ≈65 for r=.7)
- Small samples (N≤60) can give many false positives/negatives, especially for small effects. But there are appropriate sequential approaches

European Journal of Social Psychology, Eur. J. Soc. Psychol. **44**, 701–710 (2014)
Published online in Wiley Online Library (wileyonlinelibrary.com) **DOI**: 10.1002/ejsp.2023

Special issue article: Methods and statistics in social psychology: Refinements and new developments

Performing high-powered studies efficiently with sequential analyses

DANIËL LAKENS*
Human Technology Interaction Group, Eindhoven University of Technology, Eindhoven, The Netherlands

Psychological Methods
2015, Vol. 22, No. 2, 322–339

Sequential Hypothesis Testing With Bayes Factors: Efficiently Testing Mean Differences

Felix D. Schönbrodt
Ludwig-Maximilians-Universität München

Eric-Jan Wagenmakers
University of Amsterdam

Michael Zehetleitner
Ludwig-Maximilians-Universität München

Marco Perugini
University of Milan–Bicocca

6th pointer: Reduce variability

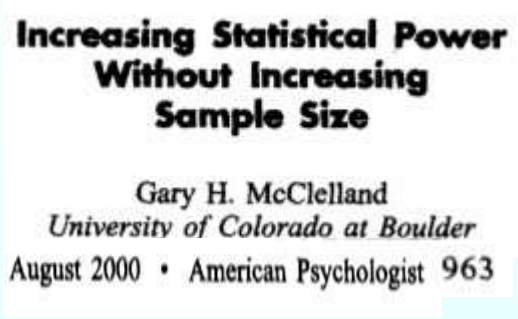
- Results stabilize with smaller standard errors
- Standard errors depend on N and SD
- Smaller SD means smaller SE
- SD can be reduced (*ceteris paribus*) with more reliable measures, more precise experimental designs, less Ss variability
- Plan your design as simple and as clean as possible

$$SE = \sqrt{\frac{S^2}{n}}$$

Distinguish conceptually between unnecessary ("added noise") and necessary ("natural") variance

Improve your design. Optimize it.

Reduce the **noise**! Increase the **signal**!



Increasing Statistical Power Without Increasing Sample Size

Gary H. McClelland
University of Colorado at Boulder

August 2000 • American Psychologist 963

Increasing the Power of Your Study by Increasing the Effect Size

TOM MEYVIS
STIJN M. J. VAN OSSELAER

Journal of Consumer Research. Feb2018, Vol. 44 Issue 5, p1157-1173.

7th pointer:
Parameter estimate vs. Statistical inference

- Parameter estimate ≠ Statistical inference
- More data points are **always** better than less data points
- Parameters are estimated more accurately/precisely, error is reduced
- It is a basic statistical principle
- NHST or Bayes (NHBT) are formal tools for statistical inference
- Do not mix them up with parameter estimate!!

Precision vs. Power

- They have different aims

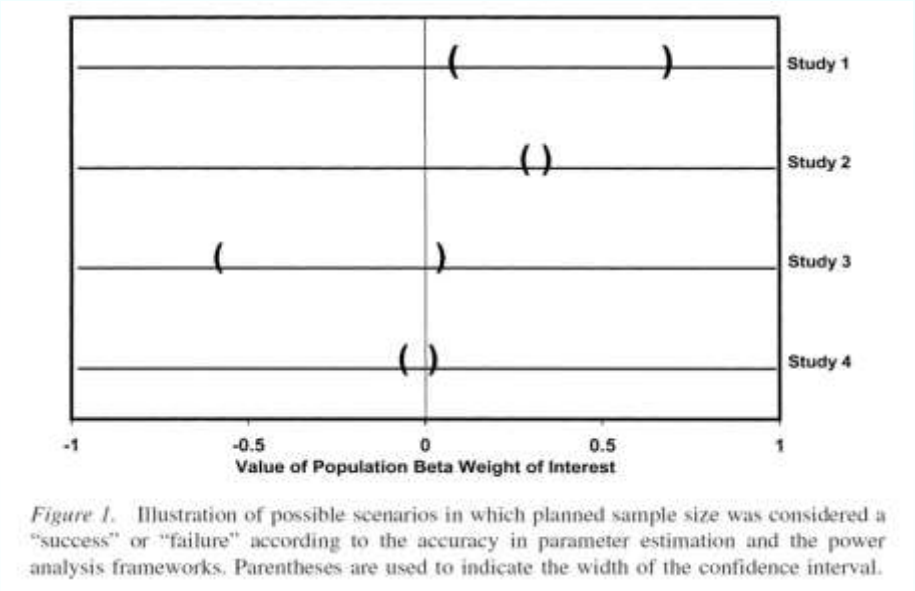
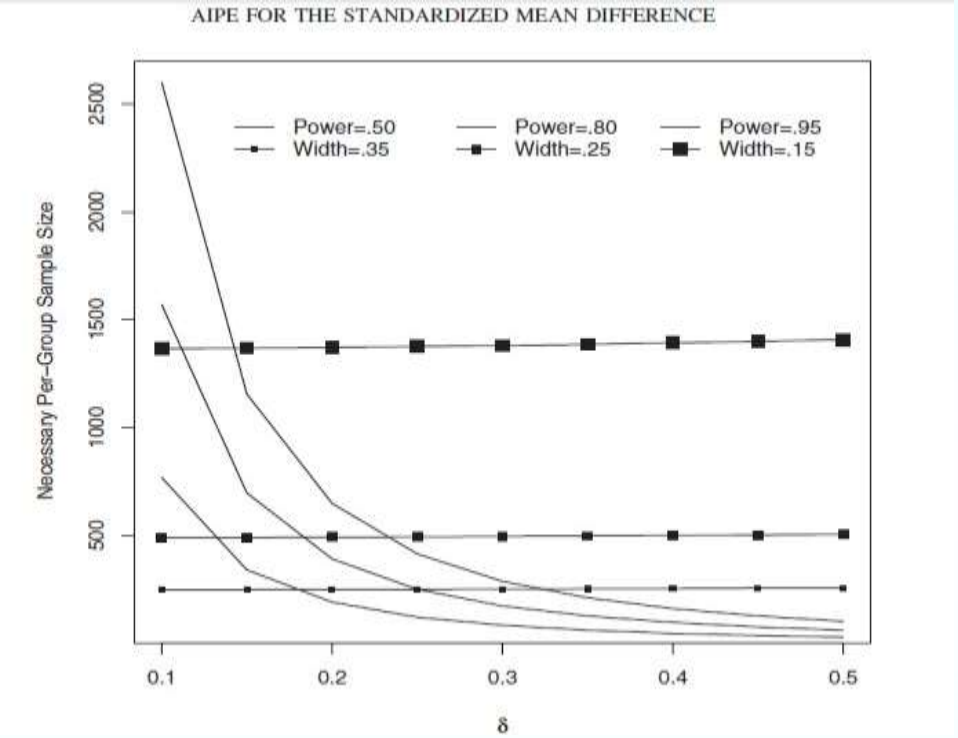


Figure 1. Illustration of possible scenarios in which planned sample size was considered a "success" or "failure" according to the accuracy in parameter estimation and the power analysis frameworks. Parentheses are used to indicate the width of the confidence interval.



- Precision is valuable no matter everything else, but…

  BIG sample sizes are needed for precise estimates no matter the effect size

8th pointer: Don't get it personal

- Get it right ≠ I am right
- Try to plan studies pitting against different hypotheses and predictions (e.g., also use of Bayes Factors)
- Try to depersonalize your preferred theoretical explanation (strong inference, Platt 1964)
- Consider adversarial collaborations
- Try to start from "truth-seeking" and add good aesthetic standards (Giner Sorolla, 2012)

9th pointer: Don't QRPs

- Avoid Questionable Research Practices (QRP)
- Do not cherry-pick DVs among many that you have, do not exclude cases as is, do not make multiple interim analyses to decide whether to collect additional Ss, correct for multiple testing (FDR) when study is exploratory
- Read Simmons et al. (2011): not every recommendation is perfect, but they do give many good ones

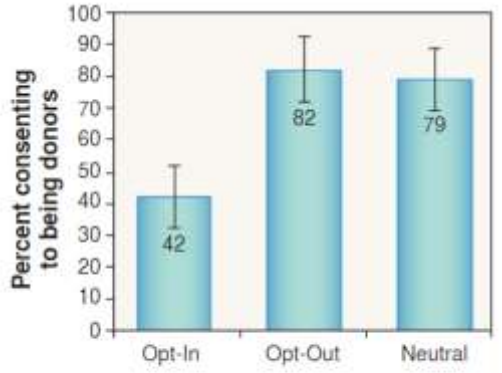
10th pointer: Be wary of "easy fixes"

- Be wary of easy quick fixes with subtle "psychological" manipulations to big problems
- Some seems to **work** (e.g., Default organ donor)

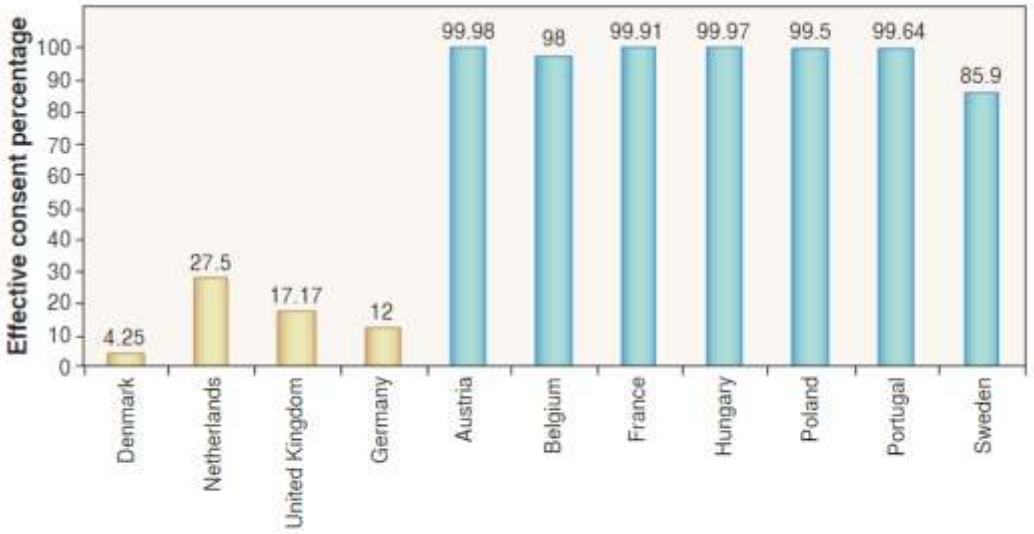
Do Defaults Save Lives?

Eric J. Johnson* and Daniel Goldstein

21 NOVEMBER 2003 VOL 302 SCIENCE



Effective consent rates, online experiment, as a function of default.



Effective consent rates, by country. Explicit consent (opt-in, gold) and presumed consent (opt-out, blue).

…but others don't

Signing at the beginning makes ethics salient and decreases dishonest self-reports in comparison to signing at the end

PNAS | September 18, 2012 | vol. 109 | no. 38 | 15197–15200

Lisa L. Shu[a], Nina Mazar[b,1], Francesca Gino[c], Dan Ariely[d], and Max H. Bazerman[c]

Experiment 3: Participants and Procedure. We conducted a field experiment with an insurance company in the southeastern United States asking some of their existing customers to report their odometer reading.

We compared the reported current odometer mileage on 13,488 completed policy forms for 20,741 cars to the latest records of each car's odometer mileage to calculate its use (number of miles driven). Customers who signed at the beginning on average revealed higher use (M = 26,098.4, SD = 12,253.4) than those who signed at the end [M = 23,670.6, SD = 12,621.4; $F(1, 13{,}485)$ = 128.63, $P < 0.001$]. The difference was 2,427.8 miles per car. That is, asking customers to sign at the beginning of the form led to a 10.25% increase in implied miles driven (based on reported odometer readings) over the current practice of asking for a signature at the end. Follow-up analyses suggested that the higher

Signing at the beginning versus at the end does not decrease dishonesty

PNAS | March 31, 2020 | vol. 117 | no. 13 | 7103–7107

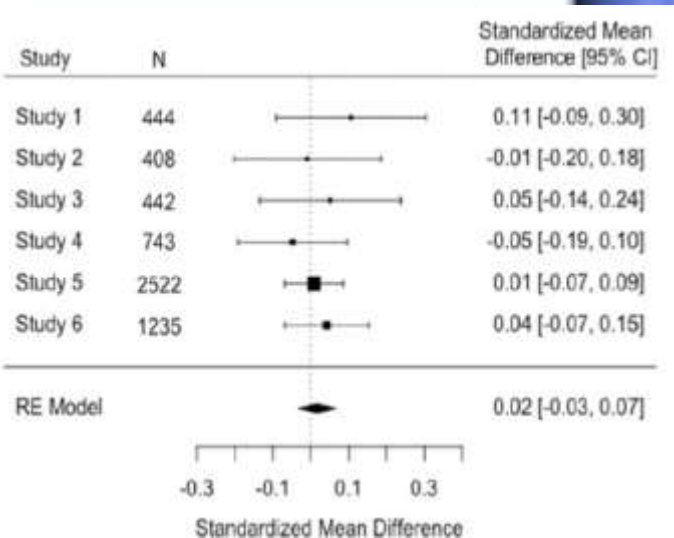
Ariella S. Kristal[a], Ashley V. Whillans[a], Max H. Bazerman[a], Francesca Gino[a,1], Lisa L. Shu[b], Nina Mazar[c], and Dan Ariely[d]

To test this hypothesis in the original PNAS paper (1), there were two laboratory experiments (n = 101 and n = 60, respectively) and one field experiment (n = 13,488). Across the two laboratory ex-

Table 2. Effect sizes of the experiments in the current and original investigation demonstrating the effect of having people sign a veracity statement attesting to their honest reporting placed before versus after reporting

Study	Sample size	Number of conditions	Cheating task	Population	Average performance reported effect size (d) [95% CI]*
This study					
Study 1	444	6	Die rolling	Community laboratory	0.11 [−0.09, 0.30]
Study 2	408	4	Anagrams	Community laboratory	−0.01 [−0.20, 0.18]
Study 3	442	2	Anagrams	MTurk	0.05 [−0.14, 0.24]
Study 4	743	3	Anagrams	MTurk	−0.05 [−0.19, 0.10]
Study 5	2,522	2	Anagrams	Naive MTurk	0.01 [−0.07, 0.09]
Study 6 (direct replication of PNAS study 1)	1,235	2	Paper matrix; self-reported travel expenses	Community laboratory	−0.04 [−0.07, 0.15]†
Shu et al. (1) study					
Study 1	101	3	Paper matrix; self-reported travel expenses	Students	−1.05 [−1.55, −0.53]†
Study 2	60	2	Paper matrix; self-reported travel expenses	Students	−0.53 [−1.04, −0.01]†
Study 3	13,488	2	Odometer reading reported on audit form	Automobile insurance clients	−0.20 [−0.16, −0.23]

*For all tasks, effect sizes are reported for the differences in total amounts reported between conditions. Negative effect size indicates reduction in cheating
†Effect sizes reported in the last column are based on the paper matrix performance only, not the claimed travel expenses.



…with a twist

SCIENTIFIC INTEGRITY

Honesty study was based on fabricated data

Made-up data set raises questions about behavioral scientist Dan Ariely

By **Cathleen O'Grady**

27 AUGUST 2021 • VOL 373 ISSUE 6558 SCIENCE

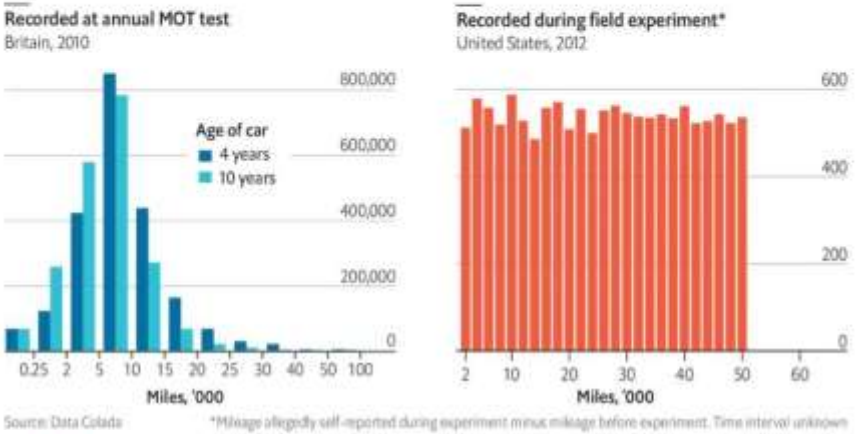
Daily chart

A study on dishonesty was based on fraudulent data

The numbers were clearly faked. No one will admit to faking them

Lies, damned lies and faked statistics

Distribution of miles driven, number of cars



Source: Data Colada

*Mileage allegedly self-reported during experiment minus mileage before experiment. Time interval unknown

[98] Evidence of Fraud in an Influential Field Experiment About Dishonesty

Posted on August 17, 2021 by Uri, Joe, & Leif

This post is co-authored with a team of researchers who have chosen to remain anonymous. They uncovered most of the evidence reported in this post. These researchers are not connected in any way to the papers described herein.

NEWS

DukeWeek: Star Duke professor Dan Ariely faces allegations of research fraud

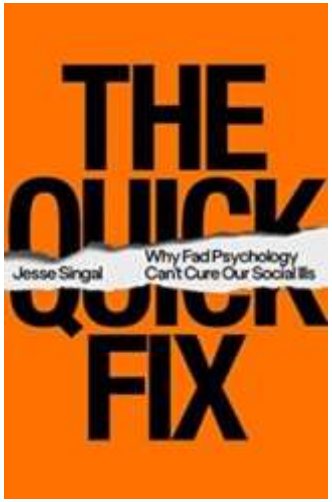
10th pointer: Be wary of "easy fixes"

NATURE HUMAN BEHAVIOUR | VOL 4 | NOVEMBER 2020 | 1092–1094 |

Use caution when applying behavioural science to policy

Social and behavioural scientists have attempted to speak to the COVID-19 crisis. But is behavioural research on COVID-19 suitable for making policy decisions? We offer a taxonomy that lets our science advance in 'evidence readiness levels' to be suitable for policy. We caution practitioners to take extreme care translating our findings to applications.

Hans IJzerman, Neil A. Lewis Jr., Andrew K. Przybylski, Netta Weinstein, Lisa DeBruine, Stuart J. Ritchie, Simine Vazire, Patrick S. Forscher, Richard D. Morey, James D. Ivory and Farid Anvari



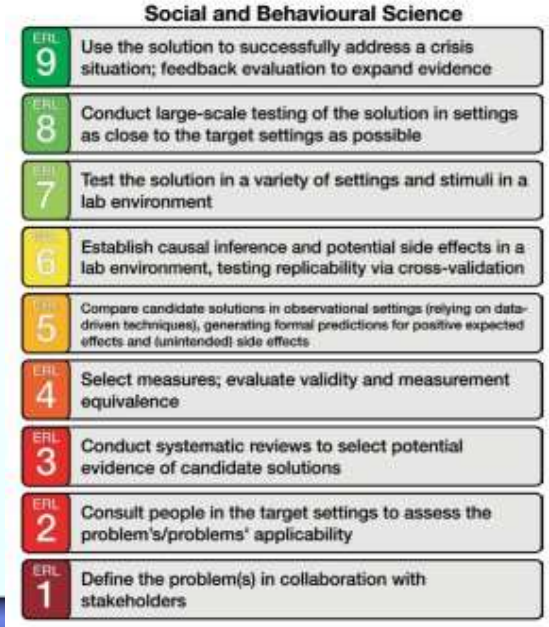


Fig. 2 | Proposed social and behavioural sciences evidence readiness levels.

11th pointer: Think about α

- Consider using p<.005 as significant (instead of p<.05, suggestive) for **NOVEL** findings

NATURE HUMAN BEHAVIOUR | www.nature.com/nathumbehav

comment

Redefine statistical significance

We propose to change the default *P*-value threshold for statistical significance from 0.05 to 0.005 for claims of new discoveries.

Daniel J. Benjamin, James O. Berger, Magnus Johannesson, Brian A. Nosek, E.-J. Wagenmakers, Richard Berk, Kenneth A. Bollen, Björn Brembs, Lawrence Brown, Colin Camerer, David Cesarini, Christopher D. Chambers, Merlise Clyde, Thomas D. Cook, Paul De Boeck, Zoltan Dienes, Anna Dreber, Kenny Easwaran, Charles Efferson, Ernst Fehr, Fiona Fidler, Andy P. Field, Malcolm Forster, Edward I. George, Richard Gonzalez, Steven Goodman, Edwin Green, Donald P. Green, Anthony Greenwald, Jarrod D. Hadfield, Larry V. Hedges, Leonhard Held, Teck Hua Ho, Herbert Hoijtink, Daniel J. Hruschka, Kosuke Imai, Guido Imbens, John P. A. Ioannidis, Minjeong Jeon, James Holland Jones, Michael Kirchler, David Laibson, John List, Roderick Little, Arthur Lupia, Edouard Machery, Scott E. Maxwell, Michael McCarthy, Don Moore, Stephen L. Morgan, Marcus Munafó, Shinichi Nakagawa, Brendan Nyhan, Timothy H. Parker, Luis Pericchi, Marco Perugini, Jeff Rouder, Judith Rousseau, Victoria Savalei, Felix D. Schönbrodt, Thomas Sellke, Betsy Sinclair, Dustin Tingley, Trisha Van Zandt, Simine Vazire, Duncan J. Watts, Christopher Winship, Robert L. Wolpert, Yu Xie, Cristobal Young, Jonathan Zinman and Valen E. Johnson

nature human behaviour

Comment | Published: 26 February 2018

Justify your alpha

Daniel Lakens, Federico G. Adolfi, Casper J. Albers, Farid Anvari, Matthew A. J. Apps, Shlomo E. Argamon, Thom Baguley, Raymond B. Becker, Stephen D. Benning, Daniel E. Bradford, Erin M. Buchanan, Aaron R. Caldwell, Ben Van Calster, Rickard Carlsson, Sau-Chin Chen, Bryan Chung, Lincoln J. Colling, Gary S. Collins, Zander Crook, Emily S. Cross, Sameera Daniels, Henrik Danielsson, Lisa DeBruine, Daniel J. Dunleavy, Brian D. Earp, Michele I. Feist, Jason D. Ferrell, James G. Field, Nicholas W. Fox, Amanda Friesen, Caio Gomes, Monica Gonzalez-Marquez, James A. Grange, Andrew P. Grieve, Robert Guggenberger, James Grist, Anne-Laura van Harmelen, Fred Hasselman, Kevin D. Hochard, Mark R. Hoffarth, Nicholas P. Holmes, Michael Ingre, Peder M. Isager, Hanna K. Isotalus, Christer Johansson, Konrad Juszczyk, David A. Kenny, Ahmed A. Khalil, Barbara Konat, Junpeng Lao, Erik Gahner Larsen, Gerine M. A. Lodder, Jiří Lukavský, Christopher R. Madan, David Manheim, Stephen R. Martin, Andrea E. Martin, Deborah G. Mayo, Randy J. McCarthy, Kevin McConway, Colin McFarland, Amanda Q. X. Nio, Gustav Nilsonne, Cilene Lino de Oliveira, Jean-Jacques Orban de Xivry, Sam Parsons, Gerit Pfuhl, Kimberly A. Quinn, John J. Sakon, S. Adil Saribay, Iris K. Schneider, Manojkumar Selvaraju, Zsuzsika Sjoerds, Samuel G. Smith, Tim Smits, Jeffrey R. Spies, Vishnu Sreekumar, Crystal N. Steltenpohl, Neil Stenhouse, Wojciech Świątkowski, Miguel A. Vadillo, Marcel A. L. M. Van Assen, Matt N. Williams, Samantha E. Williams, Donald R. Williams, Tal Yarkoni, Ignazio Ziano & Rolf A. Zwaan - Show fewer authors

In response to recommendations to redefine statistical significance to P ≤ 0.005, we propose that researchers should transparently report and justify all choices they make when designing a study, including the alpha level.

arXiv.org > stat > arXiv:1709.07588

Statistics > Methodology

Abandon Statistical Significance

Blakeley B. McShane, David Gal, Andrew Gelman, Christian Robert, Jennifer L. Tackett

(Submitted on 22 Sep 2017 (v1), last revised 10 Apr 2018 (this version, v2))

Why?

- Four main possible reasons:

1) Findings with p<.005 were twice more likely to be replicated than with p<.05 (in both Psychology and Experimental Economics)
2) Your intuition tells you something about .05 that is not true (BF vs. 95%)
3) The base rate of correct hypotheses in Psychology is low
4) It is less draconian than it might seem at first (about 70% extra sample size)

12th pointer: Look around you

- Some of this stuff is already implemented in top level journals

DISCLOSURE QUESTIONS:

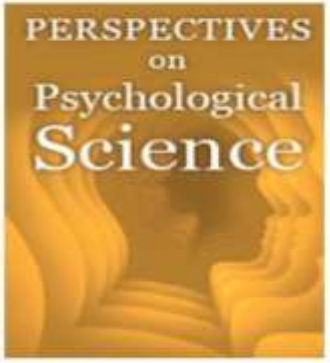
For all studies in your recently published article titled [publication title], please endorse the following statements: (please type an X to indicate your answer)

We reported the total number of observations which were excluded (if any) and the criterion for doing so. (If no observations excluded, please indicate Yes)
Yes: ___ No: ___
If no, please report this information here (e.g., data from 3 participants in Study 2 excluded due to computer malfunction; 4 participants in Study 1 excluded for not following instructions):

We reported all tested experimental conditions, including failed manipulations. Yes: ___ No: ___
If no, please provide brief explanation for not reporting this information (e.g., critical software implementation error; editorial request):

We reported all administered measures/items. Yes: ___ No: ___
If no, please provide brief explanation for not reporting this information (e.g., measures not related to research question; scores from unreported measure insufficiently reliable):

We reported (a) how we determined our sample size **and** (b) our data collection stopping rule. Yes: ___ No: ___
If no, please describe (a) the basis for the sample sizes used and (b) how you decided to stop collecting data (e.g., decided ahead of time to collect data until minimum sample size achieved and this was followed; sample size determined by power analysis but did not achieve it by the end of term):



*

*

- Some journals have a registered reports option
- Some journals have a replication study option also with registered reports
- Some journals provide badges
- Most journals accepts or ask for SM
- Increased collaborative efforts (team research)
- Before 2011 little if any of this was present

Current scenario

-Good practices and methodological advances

Journal of Economic Psychology 75 (2019) 102117

Contents lists available at ScienceDirect

Journal of Economic Psychology

journal homepage: www.elsevier.com/locate/joep

Predicting replication outcomes in the Many Labs 2 study

Eskil Forsell[a,1], Domenico Viganola[b,1], Thomas Pfeiffer[c], Johan Almenberg[d], Brad Wilson[e], Yiling Chen[f], Brian A. Nosek[g,h], Magnus Johannesson[b], Anna Dreber[b,i,*]

nature human behaviour

PERSPECTIVE

PUBLISHED: 10 JANUARY 2017 | VOLUME: 1 | ARTICLE NUMBER: 0021

OPEN

A manifesto for reproducible science

Marcus R. Munafò[1,2*], Brian A. Nosek[3,4], Dorothy V. M. Bishop[5], Katherine S. Button[6], Christopher D. Chambers[7], Nathalie Percie du Sert[8], Uri Simonsohn[9], Eric-Jan Wagenmakers[10], Jennifer J. Ware[11] and John P. A. Ioannidis[12,13,14]

The Psychological Science Accelerator: Advancing Psychology Through a Distributed Collaborative Network

Advances in Methods and Practices in Psychological Science
2018, Vol. 1(4) 501–515
© The Author(s) 2018
Article reuse guidelines:
sagepub.com/journals-permissions
DOI: 10.1177/2515245918797607
www.psychologicalscience.org/AMPPS

SAGE

Hannah Moshontz[1], Lorne Campbell[2], Charles R. Ebersole[3], Hans IJzerman[4], Heather L. Urry[5], Patrick S. Forscher[6],

Current scenario

2600–2606 | PNAS | March 13, 2018 | vol. 115 | no. 11

The preregistration revolution

Brian A. Nosek[a,b,1], Charles R. Ebersole[b], Alexander C. DeHaven[a], and David T. Mellor[a]

[a]Center for Open Science, Charlottesville, VA 22903; and [b]Department of Psychology, University of Virginia, Charlottesville, VA 22904

Putting the Self in Self-Correction: Findings from the Loss-of-Confidence Project

Julia Rohrer,[1,2] Warren Tierney,[3] Eric L. Uhlmann,[3] Lisa M. DeBruine,[4] Tom Heyman,[5,6] Benedict Jones,[7] Stefan C. Schmukle,[2] Raphael Silberzahn,[8] Rebecca M. Willén,[9] Rickard Carlsson,[10] Richard E. Lucas,[11] Julia Strand,[12] Simine Vazire,[13] Jessica K. Witt,[14] Thomas R. Zentall,[15] Christopher F. Chabris,[16] Tal Yarkoni[17]

Comment

Argue about what a replication means before you do it

ROYAL SOCIETY OPEN SCIENCE

royalsocietypublishing.org/journal/rsos

Research

Are replication rates the same across academic fields? Community forecasts from the DARPA SCORE programme

Cite this article: Gordon M *et al.* 2020 Are replication rates the same across academic fields? Community forecasts from the DARPA SCORE programme. *R. Soc. Open Sci.* **7**: 200566. http://dx.doi.org/10.1098/rsos.200566

The Defense Advanced Research Projects Agency (DARPA) programme 'Systematizing Confidence in Open Research and Evidence' (SCORE) aims to generate confidence scores for a large number of research claims from empirical studies in the social and behavioural sciences. The confidence scores will provide a quantitative assessment of how likely a claim will hold up in an independent replication. To create the scores, we follow earlier approaches and use prediction markets and surveys to forecast replication outcomes. Based on an initial set of forecasts for the overall replication rate in SCORE

Registered reports

cos.io/rr/

CENTER FOR OPEN SCIENCE

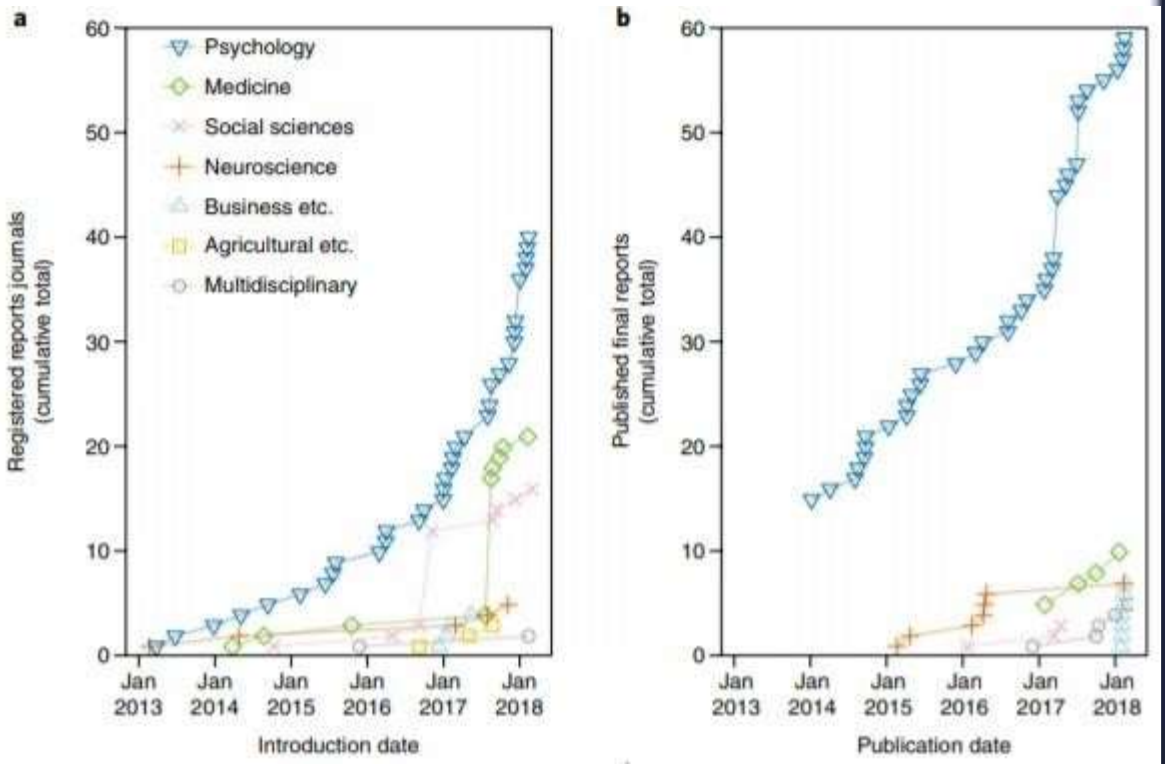
Currently, **209** journals use the Registered Reports

nature human behaviour

Comment | Published: 01 October 2018

Mapping the universe of registered reports

Tom E. Hardwicke & John P. A. Ioannidis



Registered reports and null findings

PLOS BIOLOGY

PERSPECTIVE

Open science challenges, benefits and tips in early career and beyond

Christopher Allen[1]*, **David M. A. Mehler**[1,2]*

1 Cardiff University Brain Research Imaging Centre (CUBRIC), Wales, United Kingdom, 2 Department of Psychiatry, University of Muenster, Germany

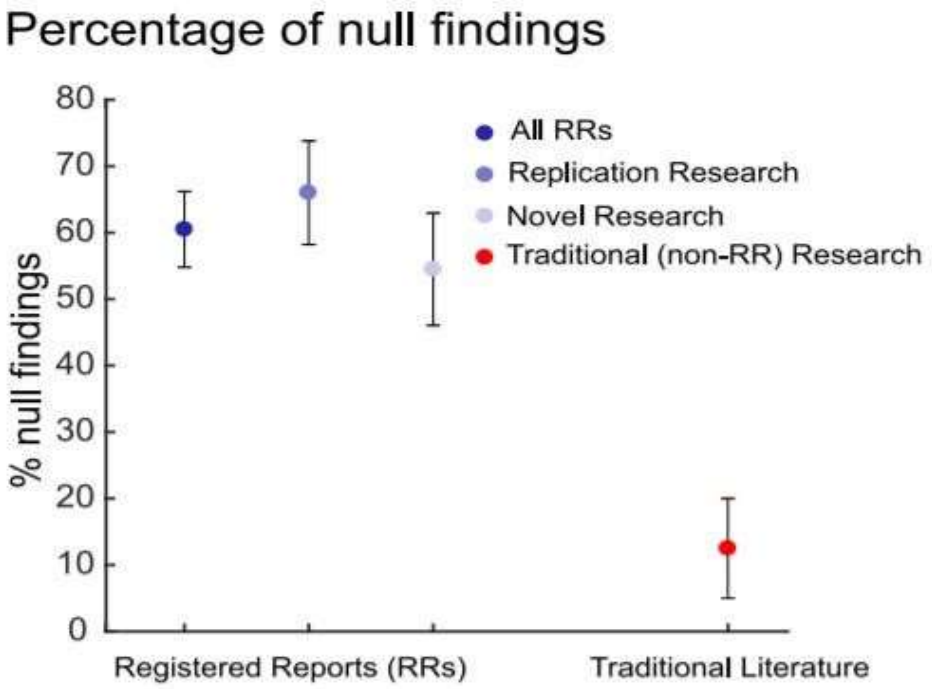


Fig 1. Percentages of null findings among RRs and traditional (non-RR) literature [46,47], with their respective 95% confidence intervals. In total, we extracted $n = 153$ hypotheses from RRs that were declared as replication attempts and $n = 143$ hypotheses that were declared as original research. The bounds of the confidence intervals shown for traditional literature were based on estimates (5% and 20%, respectively) of null findings that have been previously reported for traditional literature [46,47]. Data is available on the Open Science Framework (https://osf.io/wy2ek/) and in S1 Data. RR, registered report.

Registered reports, publication bias and meta-analyses

Journal of Experimental Psychology: General

A Comprehensive Meta-Analysis of Money Priming

Paul Lodder and How Hwee Ong
Tilburg University

Raoul P. P. P. Grasman
University of Amsterdam

Jelte M. Wicherts
Tilburg University

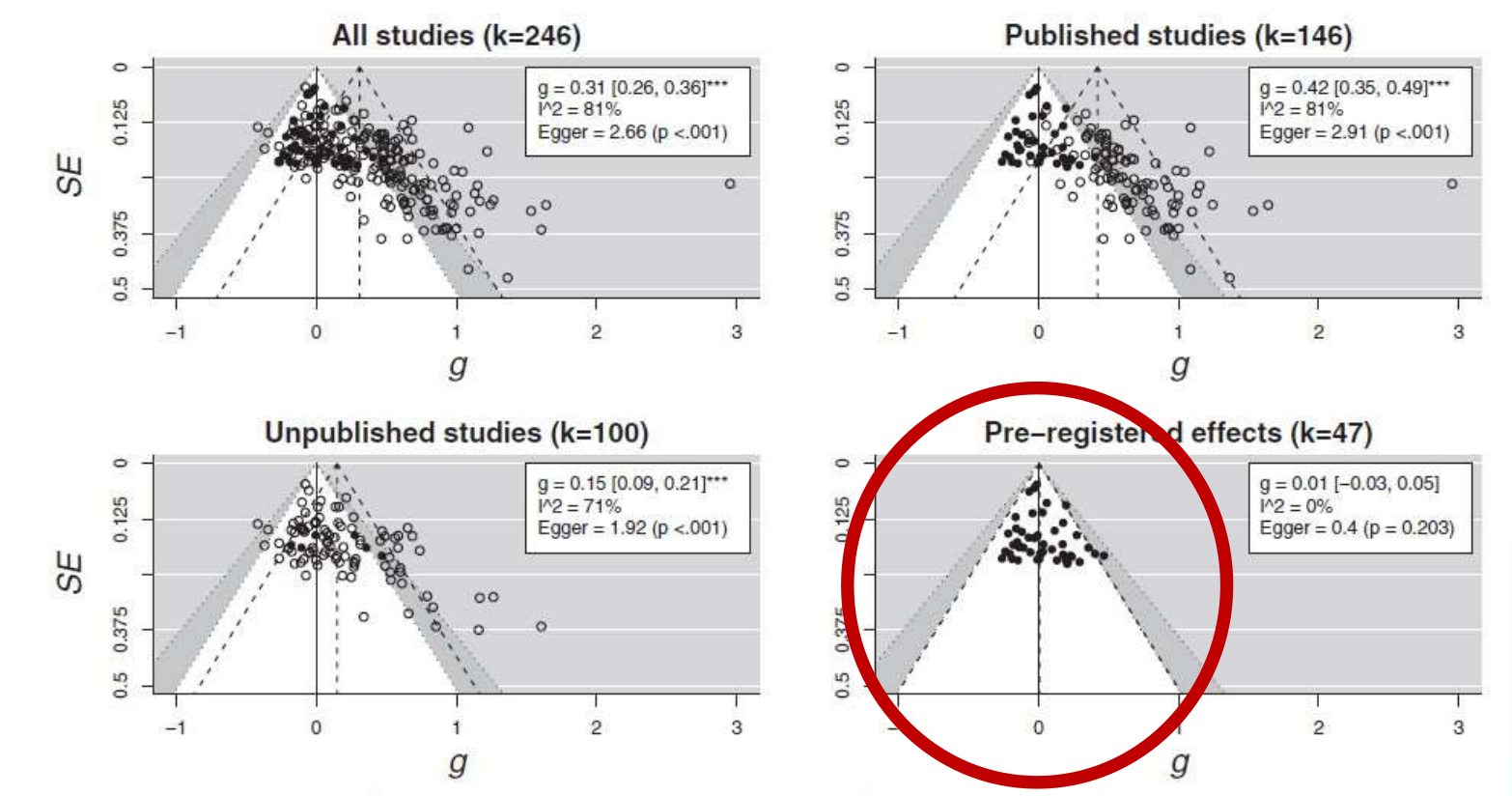


Figure 1. Funnel plots of all studies, published studies, unpublished studies, preregistered studies, main effects, and interaction effects. k = number of included effects. g = Hedges' g random effects model estimate (center of dotted funnel), including 95% confidence interval. I^2 = heterogeneity measure; Egger = Egger's test regression coefficient and p value; The white- and gray funnel represent a 95% and 99% confidence level, respectively. Black dots represent preregistered studies. *** $p < .001$.

Pointers Recap (1 for each month)

- **Power**
- **Confirmatory** vs. Exploratory studies
- **Meta-analytic** thinking
- **Meta-conditional** theorizing
- **Increase sample size** for stable estimates (signal ↑)
- Simple/clean design to **reduce error variability** (noise ↓)
- **Parameter estimate ≠** statistical inference
- **Get it right** ≠ I am right
- **Don't QRPs**
- **Be wary of easy fixes**
- Think about lowering α **to .005 (for novel findings)**
- Look around you (**things are happening now**)

Conclusions

- Increase sample size if you want to get it right
- Decrease **noise** and increase **signal** in the study
- It is not just a statistical issue (Bayes is no miracle cure)
- Clear methodological thinking matters a lot
- Dichotomous thinking does not help
- Ask what, when, how much, how something happens
- Get it right ≠ I am right
- To get it right means to reduce False positives (Type I error), False negatives (Type II error) and to have reasonably precise estimates

We don’t want to fall in the nine circles of scientific hell..

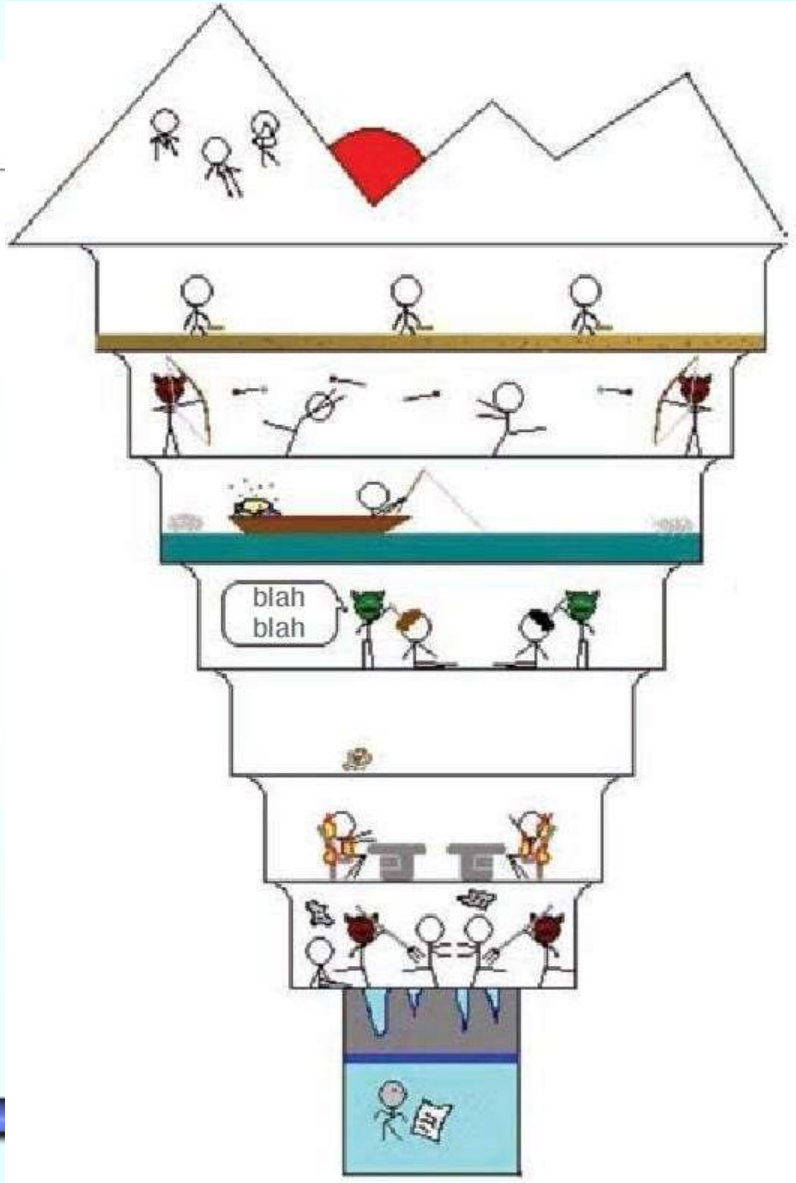
The Nine Circles of Scientific Hell

Neuroskeptic[1]

aps ASSOCIATION FOR PSYCHOLOGICAL SCIENCE

Perspectives on Psychological Science
7(6) 643–644

DOI: 10.1177/1745691612459519
http://pps.sagepub.com
SAGE



I	Limbo
II	Overselling
III	Post-Hoc Storytelling
IV	P-Value Fishing
V	Creative Outliers
VI	Plagiarism
VII	Non-Publication
VIII	Partial Publication
IX	Inventing Data

We want to move from the Middle Age…

…to the Renaissance

ANNUAL REVIEWS

Annual Review of Psychology

Psychology's Renaissance

Leif D. Nelson,[1] Joseph Simmons,[2] and Uri Simonsohn[2]

[1]Haas School of Business, University of California, Berkeley, California 94720; email: Leif_Nelson@haas.berkeley.edu

[2]The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104; email: jsimmo@upenn.edu, urisohn@gmail.com

Annu. Rev. Psychol. 2018. 69:511–34

INTRODUCTION

If a team of research psychologists were to emerge today from a 7-year hibernation, they would not recognize their field. Authors voluntarily posting their data. Top journals routinely publishing replication attempts, both failures and successes. Hundreds of researchers preregistering their studies. Crowded methods symposia at many conferences. Enormous increases in sample sizes. Some top journals requiring the full disclosure of measures, conditions, exclusions, and the rules for determining sample sizes. Several multilab replication efforts accepted for publication before any data were collected. Overall, an unprecedented focus on replicability. What on earth just happened?

The different steps before and after data collection

The importance of a checklist

Cristina Zogmaister,
University of Milano-Bicocca
cristina.zogmaister@unimib.it

Juliette Richetin,
University of Milano-Bicocca
juliette.richetin@unimib.it

With the precious help from Marine Rougier, Jan De Houwer, Jamie Cummins, Giulio Costantini, Daniele Romano, Erica Casini, Emanuele Preti, Rossella Di Pierro, & Marco Perugini

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.

IDEA

writing-down method

Preparation of materials

ethics committee approval request

Pre-registration*

data collection

(deviations from prereg?)

Data analysis

(deviations from prereg?)

Writing

Submission

discussion with coauthors & feedback

BEFORE...

Once you have some ideas about the study you want to run...

Write-up a method document including

- aim
- procedure
- hypotheses,
- analyses plan
- sample size determination/power analysis
- inclusion/exclusion criteria (sample size needed after exclusion - e.g., as a function of a manipulation check).
- timeline for all steps including data collection and analysis

It will help:

- to clarify the aim of the study
- to check for potential incongruencies between what you wanted (aim) and what you planned (way to achieve the aim)
- to confront your ideas and design with collaborators
- to realize that some statistical analyses might be more complex than expected -or even impossible with traditional ANOVAs, i.e., without knowledge on more specific techniques)
- for the Ethics approval phase
- for the preregistration phase
- to write-up papers about the study

Once you have defined the specific protocol...

Submit a request for Ethics approval from your institution

Requirement for many journals (together with informed consent from participants)

After elaborating the material for your study

- Recruit mock participants to check whether data are stored properly and in the format you need (You can even run mock analyses to elaborate a transformation and analyses script)
- Store your script

Once you have created your study protocol...

Pre-registration

For reproducibility and publication purposes

- Registered report

- Pre-registration badge

In case of registered report, perhaps it might make sense to request to Ethics committee after the final version of the RR is accepted? It depends on the degree of flexibility of the request to the Ethics committee, as the reviewers might ask for changes to the protocol.

Different ways of proceeding serving different purposes.

2 examples

- Fast way with As Predicted (study protocol)
- More elaborate way with OSF (from protocol to analyses and data storage)

Preregistration - Fast way (one example)

https://aspredicted.org/

Create a new pre-registration CREATE

See your pre-registrations SEE ALL
(e.g., to share with reviewers or make public)

Creating New AsPredicted

☐ I am just trying things out. (Check the box and the submission will self destruct within 24 hours) 

Participating Authors (Up to 5)

Order	First	Last	email	Affiliation
1	Juliette	Richetin	juliette.richetin@unimib.it	University of Milano-Bicocca
2				
3				
4				
5				

AsPredicted Questions

(version 2.00)

This **blog post** on how to answer pre-registration questions may be a useful resource.

1) Data collection. Have any data been collected for this study already?

- (●) Yes, we already collected the data.
- ( ) No, no data have been collected for this study yet.
- ( ) It's complicated. We have already collected some data but explain in Question 8 why readers may consider this a valid **pre**-registration nevertheless.

(Note: *"Yes"* is not an accepted answer.)

2) Hypothesis. What's the main question being asked or hypothesis being tested in this study?

Example: A month-long academic summer program for disadvantaged kids will reduce the drop in academic performance that occurs during the summer

3) Dependent variable. Describe the key dependent variable(s) specifying how they will be measured.

Example: Simple average GPA across all courses during the first semester after the intervention.

4) Conditions. How many and which conditions will participants be assigned to?

Example 1: Two conditions: Offering summer program: yes vs no.
Example 2: 12 conditions in a mixed design lab study. Participants will be assigned to one of four conditions: math training, verbal training, memory task, or control (4 between-subject conditions). Each participant will complete a math test, a verbal test, and a memory test (3 within-subject conditions).

5) Analyses. Specify exactly which analyses you will conduct to examine the main question/hypothesis.

Example. Linear regression predicting the simple average GPA in the semester after the intervention with a dummy variable indicating whether the participant was offered the summer program or not (intention-to-treat analysis). We will also conduct the same regression controlling for simple average GPA during the semester before the intervention, gender, & household income (an 8-point scale ranging from 1 = below $20,000 and 8 = above $150,000).

6) Outliers and Exclusions. Describe exactly how outliers will be defined and handled, and your precise rule(s) for excluding observations.

Example 1. We will compute the overall mean and standard deviation across all conditions, and winsorize at 2.5 SD above/below the mean.
Example 2: We will exclude participants who incorrectly answer at least 2 of our 3 attention check questions.
Example 3. We will exclude any participants who complete the survey in less than 30 seconds.

7) Sample Size. How many observations will be collected or what will determine sample size?
No need to justify decision, but be precise about exactly how the number will be determined.

Example: We will offer the program until 500 people have agreed to participate in it or until June 30, 2016 (whichever comes first).

8) Other. Anything else you would like to pre-register?
(e.g., secondary analyses, variables collected for exploratory purposes, unusual analyses planned?)

Example: We will include a battery of questions for exploratory purposes, including life satisfaction, amount of videogame playing, and family activity. We will also provide an additional survey with 24 questions assessing achievement orientation.We will not report the results of those analyses for the project being pre-registered.
NOTE: If you leave this blank it will read 'Nothing else to pre-register.'

9) Name. Give a title for this AsPredicted pre-registration
Suggestion: use the name of the project, followed by study description.

Example: SUMMER PROGRAMS - GPA performance, Chicago, July 2018

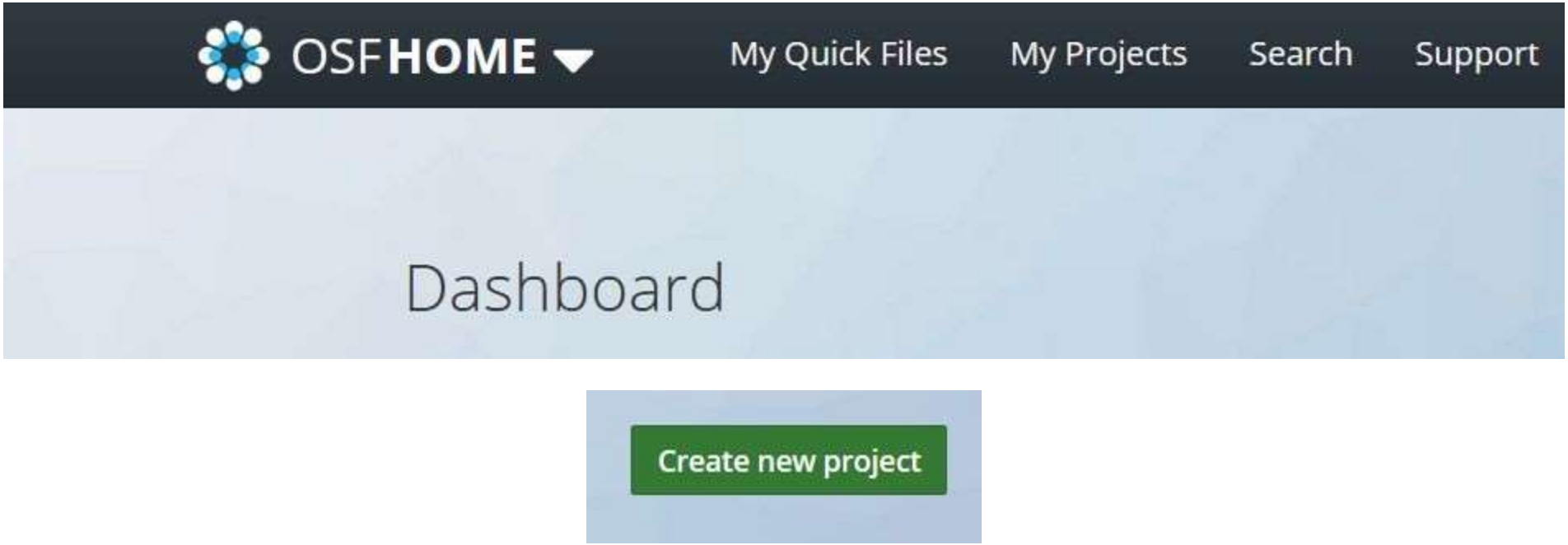
Finally. For record keeping purposes, please tell us the type of study you are pre-registering.

- ○ Class project or assignment
- ○ Experiment
- ○ Survey
- ○ Observational/archival study
- ○ Other:

PREVIEW

Preregistration - OSF

https://osf.io/



https://www.cos.io/initiatives/prereg

OSF preregistration in Social Psychology

Metadata

Title

Example of preregistration for the 2021 Timisoara Summer School LEARNVUL

Description

fictitious

Contributors

Cristina Zogmaister

Category

Uncategorized

Affiliated institutions

No affiliated institutions

License

CC-By Attribution 4.0 International

Subjects

Psychology | Social and Behavioral Sciences

Tags

fictitious

A. Hypotheses - Essential elements

Description of essential elements

Describe the (numbered) hypotheses in terms of directional relationships between your (manipulated or measured) variables.

H1: Effect of TYPE OF COMMUNICATION - Participants who receive communication FEAR will display higher scores in the variable Interest than those receiving communication NO-FEAR.
H2: Effect of MODALITY: Participants will show a higher interest in the VISUAL than in the AUDITORY modality.
H3: An interaction between the TYPE OF COMMUNICATION and MODALITY will emerge.

For interaction effects, describe the expected shape of the interactions.

We expect that communication FEAR will be more effective than communication NO-FEAR both in the AUDITORY and in the VISUAL modality.
However, this difference between the impact of the two communications will be
VISUAL than in the AUDITORY modality.

**If you are manipulating a variable, make predictions for successful check
explain why no manipulation check is included.**

Type of communication: participants will be asked to judge the degree to which
arouses in them four different emotions: Joy, Fear, Anger, Excitement. These ju
on 7-point scales (from not-at-all to extremely).
We expect that participants receiving the FEAR communication will show higher
response to the Fear emotion.

Recommended elements

Recommended elements

A figure or table may be helpful to describe complex interactions; this facilitates correct specification of the ordering of all group means.

No files selected

For original research, add rationales or theoretical frameworks for why a certain hypothesis is tested.

(Here a few references would be added for why the three hypotheses were developed)

If multiple predictions can be made for the same IV-DV combination, describe what outcome would be predicted by which theory.

No response

B. Methods - Essential elements

Description of essential elements

Design

List, based on your hypotheses from section A:

Independent variables with all their levels
a. whether they are within- or between-participant
b. the relationship between them (e.g., orthogonal, nested).

Type of communication: Fear vs. No-fear (between participants)
Modality: visual vs. auditory
The IVs are orthogonal.

List dependent variables, or variables in a correlational design

Self-reported interest for the message, measured with the 'Interest for the message' questionnaire (Zogmaister, 2021)

Third variables acting as covariates or moderators.

None

Planned Sample

If applicable, describe pre-selection rules.

University students between the age of 18 and the age of 25.

Indicate where, from whom and how the data will be collected.

The study will be administered online, through the Inquisit software and the Millisecond website. We will initially recruit participants from our University subject pool and by inviting acquaintances and other volunteers through social networks. University students will receive course credit for their participation.
in case of difficulty in reaching the sample size set in this way, we will resort to a platform for crowdsourcing (e.g. MTurk or prolific)

Justify planned sample size

We plan to collect N = 212 participants, based on ... (here I would add the description of the power analysis I conducted)

Note that here you can upload a file related to your power analysis here (e.g., a protocol of power analyses from G*Power, a script, a screenshot, etc.).

If applicable, you can upload a file related to your power analysis here (e.g., a protocol of power analyses from G*Power, a script, a screenshot, etc.).

No files selected

Describe data collection termination rule.

The planned sample size is 212. We will initially collect 212 participants.
After achieving this sample, sampling will stop and the data will be checked for exclusions. Exact additional participants will be recruited. This process will continue until the target sample size is met. No hypothesis testing will be conducted until the target analytic sample is met.

Exclusion Criteria

Describe anticipated specific data exclusion criteria. For example:

a) missing, erroneous, or overly consistent responses;
b) failing check-tests or suspicion probes;
c) demographic exclusions;
d) data-based outlier criteria;
e) method-based outlier criteria (e.g. too short or long response times).

Data from participants:
- with missing responses
- who self-report a level of fear below 4 (the midpoint of the scale) in the manipulation check
- who fail outside the age limit 18-25
will be excluded from the analysis.

Procedure

Describe all manipulations, measures, materials and procedures including the order of presentation and the method of randomization and blinding (e.g., single or double blind), as in a published Methods section.

Manipulations: type of message (fear vs. no-fear) and administration modality (visual vs. auditory).
Each participant will receive only one message (i.e., this is a between participants design).
Measures: Self reported interest (Zogmaister 2021), Manipulation check on the emotions aroused by the message.
The messages can be found here: <>
Procedure: Participants will receive one message (either written or through headphones), then they will answer the manipulation check and the interest self-report.
The experiment administration software will randomly assign participants to the four conditions based on the two IVs.
Blinding: Participants will not be aware of the presence of the different experimental conditions.

Recommended elements

Recommended elements

Procedure

Set fail-safe levels of exclusion at which the whole study needs to be stopped, altered, and restarted. You may pre-determine what proportion of excluded participants will cause the study to be stopped and restarted.

- message administration
- manipulation check
- interest questionnaire.

All participants who do not answer all questions will be excluded from the analysis.

<>

If applicable, you can upload any files related to your methods and procedure here (e.g., a paper describing a scale you are using, experimenter instructions, etc.)

No files selected

C. Analysis plan - Essential elements

Confirmatory Analyses

Describe the analyses that will test the first main prediction from the hypotheses section. Include:

the relevant variables and how they are calculated;

Relevant variable: Interest.
Calculation: summed score of the 5 items of the questionnaire

the statistical technique;

Factorial ANOVA
<< here I would also add the software that I will use>>

each variable's role in the technique (e.g., IV, DV, moderator, mediator, covariate);

IVs:
- modality
- type
DV:
interest

rationale for each covariate used, if any;

Not Applicable

if using techniques other than null hypothesis testing (for example, Bayesian statistics), describe your criteria and inputs toward making an evidential conclusion, including prior values or distributions.

A significant main effect of the factor TYPE is expected.
critical p-value = 0.05
The average level of interest will be higher for the FEAR than the NO-FEAR level.

Second Prediction

Describe the analyses that will test the second main prediction from the hypotheses section. Include:

the relevant variables and how they are calculated;

same as for the first prediction

the statistical technique;

same as for the first prediction

each variable's role in the technique (e.g., IV, DV, moderator, mediator, covariate);

same as for the first prediction

rationale for each covariate used, if any;

same as for the fi

if using techniqu
describe your cr
values or distrib

A significant mair
The average level
critical p-value = (

Third Prediction

Describe the analyses that will test the third main prediction from the hypotheses section. Include:

the relevant variables and how they are calculated;

same as for the first prediction

the statistical technique;

same as for the first prediction

each variable's role in the technique (e.g., IV, DV, moderator, mediator, covariate);

same as for the first prediction

rationale for each covariate used, if any;

same as for the first prediction

if using techniques other than null hypothesis testing (for example, Bayesian statistics), describe your criteria and inputs toward making an evidential conclusion, including prior values or distributions.

A significant interaction effect is expected.
critical p-value = 0.05

Final questions

Has data collection begun for this project?

No, data collection has not begun

If data collection has begun, have you looked at the data?

No

The (estimated) start and end dates for this project are

Start date: September 13, 2021
End date: December 21, 2021

Any additional comments before I pre-register this project

No response

Other preregistrations for clinical studies

PROSPERO for clinical systematic reviews and meta-analyses

NIHR | National Institute for Health Research

PROSPERO
International prospective register of systematic reviews

In case of clinical trials, there are WHO approved registries (https://www.who.int/clinical-trials-registry-platform/network/primary-registries)

World Health Organization

Health Topics | Countries | Newsroom | Emergencies

Home / International Clinical Trials Registry Platform (ICTRP) / ICRTP Registry Network / Primary registries

the EU Clinical Trials Register (EU-CTR: https://www.clinicaltrialsregister.eu/ctr-search/search)

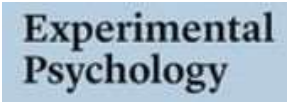
EU Clinical Trials Register

Registered Reports

Some journals propose the option of Registered Reports

Proposal of a study protocol into a standard review process

Once the study protocol accepted, paper accepted independently from the results



see: https://www.cos.io/initiatives/registered-reports

AFTER...

Once data collection is complete

Data Storage

Some journals require access to data at the moment of the submission or publication

Different possibilities:

- OSF
- University (Gent?)
- The importance to be FAIR - https://www.go-fair.org/fair-principles/ (Findable, Accessible, Interoperable, Reusable)

Once you start playing with data

Keeping track of

- **data transformation**
- **data cleaning,**
- **and analyses - syntax/script**

if this was not part of the preregistration

Writing-up

- Discuss authorship with co-authors. Credit taxonomy might be useful https://casrai.org/credit/

- Report all protocol deviations!

- All minor deviations from the pre-registrations (pending that major deviations are reported in the main text)

Submit & Preprint

Check on sherpa/romeo whether the journal you chose accepts preprints
https://v2.sherpa.ac.uk/romeo/

Use https://psyarxiv.com/ to share your preprint

Useful Checklist materials

TRANSPARENCY CHECKLIST
http://www.shinyapps.org/apps/TransparencyChecklist/
Divided into 4 sections (preregistration; methods; results and discussion; data code and materials availability)
Aczel *et al.* (2020) *Nat Hum Behav*

CHECKLIST FOR REPLICATION STUDIES
https://doi.org/10.1016/j.metip.2021.100045

REPORTING GUIDELINES FROM THE EQUATOR NETWORK (Enhancing the QUAlity and Transparency Of health Research)
https://www.equator-network.org

CHECKLIST FOR REVIEWERS
Empowering peer reviewers with a checklist to improve transparency
Parker et al. (2018) *Nature ecology & evolution*

YOUR TURN

Starting from an existing project or a future project

Elaborate your own checklist for one or more of the following steps

- Write-up a method document
- Elaborate study material,
- Recruit mock participants to check whether data are stored properly and in the format you need
- Run mock analyses to elaborate a transformation and analyses script
- Store your script
- Preregistration
- Data storage
- Data Transformation and Cleaning, Analyses script storage
- Writing-up paper

Ask around for advices and feedback

IDEA

writing-down method

Preparation of materials

ethics committee approval request

Pre-registration*

data collection

(deviations from prereg?)

Data analysis

(deviations from prereg?)

Writing

Submission

discussion with coauthors & feedback

Thank you for your attention!

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No. 952464.