Testing Bayesian Informative Hypotheses in 5 steps with JASP and R

Sara Garofalo, Ph.D.
University of Bologna - Department of Psychology

Psychometrics and Neuropsychology Lab
Center for studies and research in Cognitive Neuroscience

sara.garofalo@unibo.it

SIPS SOCIETY FOR THE IMPROVEMENT OF PSYCHOLOGICAL SCIENCE

OUTLINE

1. Bayes’ theorem & Bayes Factor
2. Informative hypotheses
3. Bayesian informative hypotheses testing
4. Example 1: 2x2 ANOVA
5. Example 2: Multiple linear regression

Initial remarks: two relevant papers

TUTORIAL PAPER AVAILABLE SOON!!

Garofalo S., Finotti G., Orsoni M., Giovagnoli S., Benassi M. (2024). Testing Bayesian Informative Hypotheses in 5 steps with R and JASP. ***Advances in Methods and Practices in Psychological Science***.
doi: 10.1177/25152459241260259

+

PUBLISH ABOUT

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Interaction effect: Are you doing the right thing?

Sara Garofalo, Sara Giovagnoli, Matteo Orsoni, Francesca Starita, Mariagrazia Benassi

Published: July 20, 2022 • https://doi.org/10.1371/journal.pone.0271668

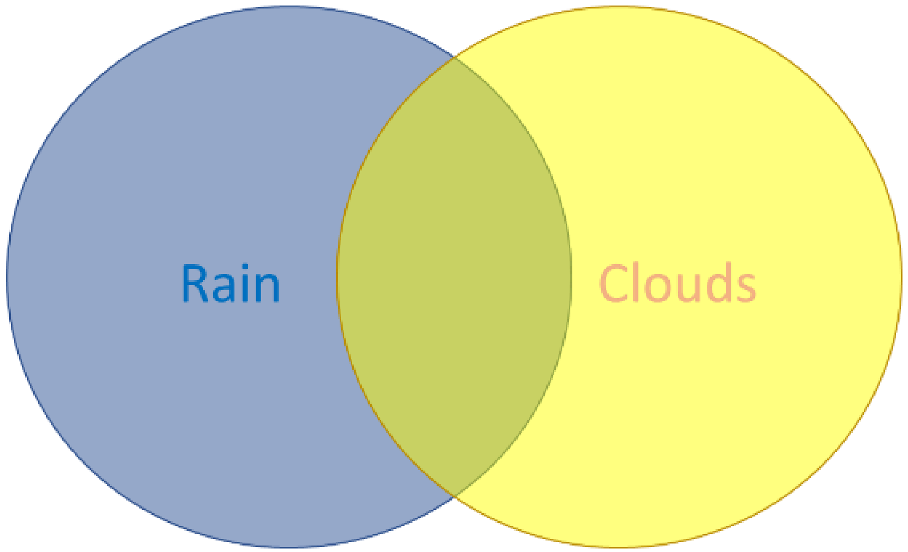
OSF

https://osf.io/ya9mp/

(Garofalo et al., 2022, PlosOne)

Bayes' theorem

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



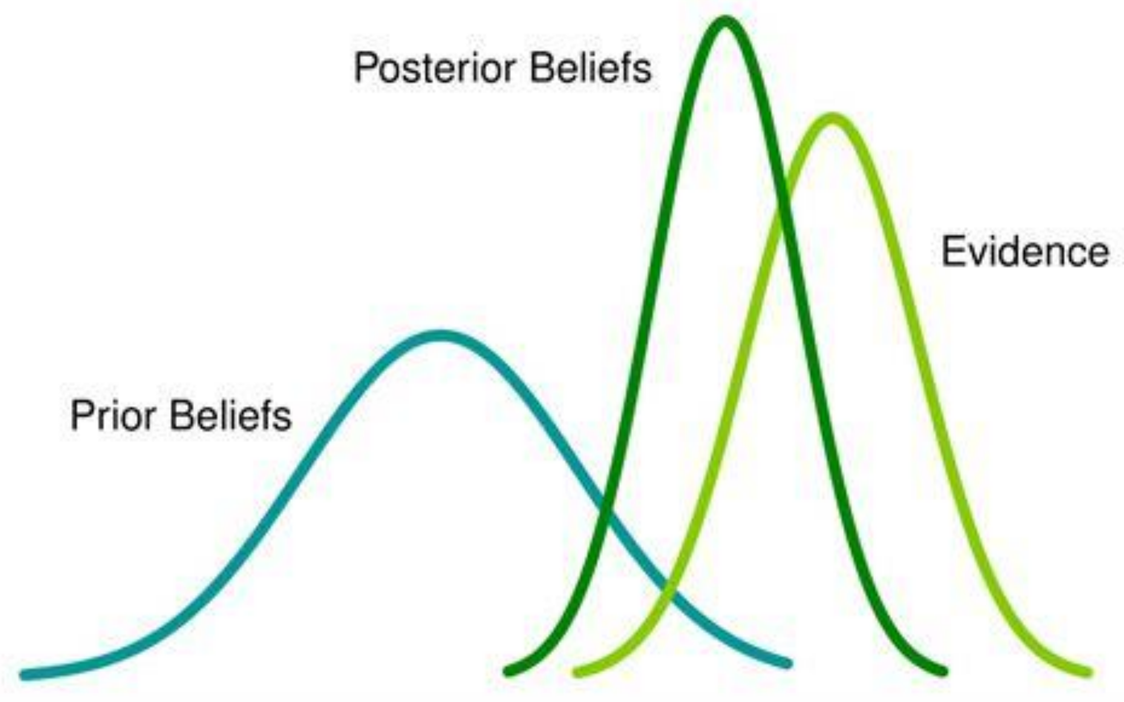
- A and B are two events
- B has a probability different from 0
- P(A|B) is the conditional probability of A occurring given that B is true (posterior probability)
- P(B|A) is the conditional probability of B occurring given that A is true (likelihood)
- P(A) is the unconditional (i.e., not secondary to other events) probability of observing A (prior probability)
- P(B) is the unconditional (i.e., not secondary to other events) probability of B (marginal probability)

Bayes' theorem for inference

Bayesian inference is a method of **statistical inference** in which **Bayes' theorem** is used to update the **probability of an hypothesis**, given a set of **evidence**

$$P(H|D) = \frac{P(\mathrm{D}|H)\,P(H)}{P(D)}$$

posterior ~ prior X likelihood



Bayes Factor

$$BF_{10} = \frac{P(data|\boldsymbol{H_1})}{P(data|\boldsymbol{H_0})}$$

More precisely

$$BF_{10} = \frac{P(D|H_1)}{P(D|H_0)} = \frac{P(H_1|D)/P(H_0|D)}{P(H_1)/P(H_0)}$$

Interpreting Bayes factors

strength of evidence (or predictive performance) of one hypothesis relative to another

< 0	Evidence for H0
= 1	No evidence
>1–3	Anecdotal evidence for H1
>3–10	Moderate evidence for H1
>10–30	Strong evidence for H1
>30–100	Very strong evidence for H1
>100	Extreme evidence for H1

"The data are N times more likely under one hypothesis than the other"

Theoretical background

H_0: $\mu_a = \mu_b \rightarrow \mu_a - \mu_b = 0$

H_1: $\mu_a \neq \mu_b \rightarrow \mu_a - \mu_b = \text{not } 0$

- **PROBLEMS WITH H_0**

It should only be used if it is a **plausible description** of the population

- **PROBLEMS WITH H_1**

Is this really your research hypothesis?

- **DICHOTOMOUS REASONING**

Experimental hypotheses in psychology

"I expect mean A to be bigger than means B and C"

"I expect that the relation between Y and both X1 and X2 is positive"

"I expect the relation between Y and X1 to be stronger than the relation between Y and X2"

Researchers formulate their expectations in terms of
equality or inequality constraints among the parameters
(e.g., means, effect sizes, etc)

↓

INFORMATIVE HYPOTHESES

(Hoijtink H., Chapman & Hall/CRC, 2012)

Experimental hypotheses in psychology

"I expect mean A to be bigger than means B and C"

$\mathbf{H_1}$: $\mu_a > (\mu_b , \mu_c)$ — $P(H_1|R)$

$\mathbf{H_2}$: $(\mu_a = \mu_b) > \mu_c$ — $P(H_2|R)$

$\mathbf{H_3}$: $\mu_a - \mu_b > \mu_a - \mu_c$ — $P(H_3|R)$

Informative hypotheses

Hypotheses that represent **research expectations** in terms of **inequality constraints** amongst parameters

(Hoijtink H., Chapman & Hall/CRC, 2012)

More complex informative hypotheses

$H_1 : \mu1 > \mu2 > \mu3$

Complete ordering of means between parameters

$H_2 : \mu1 > \mu2 \ \& \ \mu1 > \mu3$

Incomplete ordering of means between parameters

$H_3 : \mu11 - \mu12 > \mu21 - \mu22 \ \ \& \ \ \mu11 > \mu12 \ \& \ \mu11 > \mu21$

Directional description of an interaction effect (combination of parameters)

$H_4 : \mu1 > \mu2 + .2\hat{}\sigma$

Effect size: the first mean is at least .2 standard deviations larger than the second mean

Bayesian Informative Hypotheses testing

1) Specify the expected relations between parameters

$H_1 : \mu 1 > \mu 2 > \mu 3$
$H_2 : \mu 1 > \mu 2 \ \& \ \mu 1 > \mu$
$H_3 : \mu 1 - \mu 2 > \mu 1 - \mu 3$

2) Direct comparison between hypotheses

Model comparison
(relative index of evidence)

≠ from information criteria (e.g. AIC BIC)

2a) Posterior model probabilities

$$PMP_{H1} = \frac{BF_{1\mathrm{u}}}{(BF_{1\mathrm{u}} + BF_{2\mathrm{u}} + BF_{3\mathrm{u}})}$$

2b) Bayes Factors

$$BF_{12} = \frac{P(D|H_1)}{P(D|H_1)} \qquad BF_{1u} = \frac{P(D|H_1)}{P(D|H_u)}$$

(van de Schoot et al., 2011)

One step back: having a fail-safe tests

UNCONSTRAINED HYPOTHESIS

$$H_u : \mu_1, \mu_2, \mu_3$$

A model that contains **all possible sets of relationships** between the parameters.

As such, it contains **also the informative hypotheses defined** by the researcher.

COMPLEMENT HYPOTHESIS

$$H_c : \text{ not } H_i$$

A model that contains **all possible sets of relationships** between the parameters **except the one represented by the hypothesis** being tested.

For example, the complement hypothesis of H_1 is not-H_1.

Comparing Posterior Model Probabilities

The ratio of two Posterior Model Probabilities (PMP, the posterior odds) can be computed using the BF and the prior odds via:

$$\frac{PMP(H_i|\text{data})}{PMP(H_c|\text{data})} = \text{BF}_{ic} \times \frac{PRI(H_i)}{PRI(H_c)}$$

POSTERIOR MODEL PROBABILITY

Interpreted as a **conditional error probability**

What is the probability of making an error after observing the data?

Degree of support by weighting *fit* and *complexity*

PRIOR

Support for the hypotheses before observing the data

Informative

- Info from the means implied by each hypothesis
- Distribution of the data (bounds)

NB: sensitivity analysis

Uninformative

If PRIs are equal for all hypotheses, PMP's convey the same information as the BF

Comparing Bayes Factors

The Bayes Factor quantifies the relative support in the data for two hypotheses

OUR RESEARCH HYPOTHESIS

$$H_i : \mu_1 > \mu_2 > \mu_3$$

VS

- **OTHER HYPOTHESES** $H_{i'} : \mu_1 = \mu_2 = \mu_3$

$$BF_{ii'} = \frac{f_i}{c_i} / \frac{f_{i'}}{c_{i'}} = BF_{iu} / BF_{i'u}$$

- **UNCONSTRAINED HYPOTHESIS** $H_u : \mu_1, \mu_2, \mu_3$

$$BF_{iu} = \frac{f_i}{c_i}$$

- **COMPLEMENT HYPOTHESIS** $H_c : \text{ not } H_i$

$$BF_{ic} = \frac{f_i}{c_i} / \frac{1 - f_i}{1 - c_i} = BF_{iu} / BF_{cu}$$

Comparing what?

RESEARCH HYPOTHESIS

$$H_i : \mu_1 > \mu_2 > \mu_3$$

UNCONSTRAINED HYPOTHESIS

$$H_u : \mu_1, \mu_2, \mu_3$$

$$BF_{iu} = \frac{f_i}{c_i} = \frac{\text{fit } H_i}{\text{complexity } H_i}$$

→ proportion of the **posterior** supported by the model

→ proportion of the **prior** supported by the model

Balancing Fit and Complexity

Fit and complexity are determined relative to Hu

(Savage-Dickey method for BF - see, Wagenmakers et al 2010; Wetzels et al 2010)

Complexity (simplified!)

The complexity of a hypothesis is the **proportion of the prior distribution** supported by the hypothesis at hand.

It has a value **between 0 and 1** where smaller values denote a less complex (i.e., more parsimonious) hypothesis.

H1 : 1 > 2 > 3
contains 1 ordering of three means, 1-2-3, thus is
parsimonious

H2 : 1 > 2; 3
contains 2 orderings of three means, 1-2-3 and 1-3-2, thus is
less parsimonious

Hu : 1; 2; 3
contains all six possible orderings of three means, thus is
not parsimonious

Evaluating a lot of hypotheses (is not good!)

How many hypotheses?

The Bayesian **error** probability associated with a preference for a given H is given by the **sum of the posterior probabilities of the other hypotheses**

Thus, **the larger the number of hypotheses** under consideration,
the larger the error, that is the probability of preferring the wrong hypothesis

TIP: <u>only include hypotheses that are plausible</u> and represent the main (competing) expectations of a research question. For example, $\mathbf{H_0}$ is not necessarily meaningful.

Only **include hypotheses** that:

- Are **plausible**
- Represent the actual **expectations**
- Are **parsimonious**

OUTLINE

1. Bayes' theorem & Bayes Factor
2. Informative hypotheses
3. Bayesian informative hypotheses testing
4. **Example 1: 2x2 ANOVA**
5. **Example 2: Multiple linear regression**

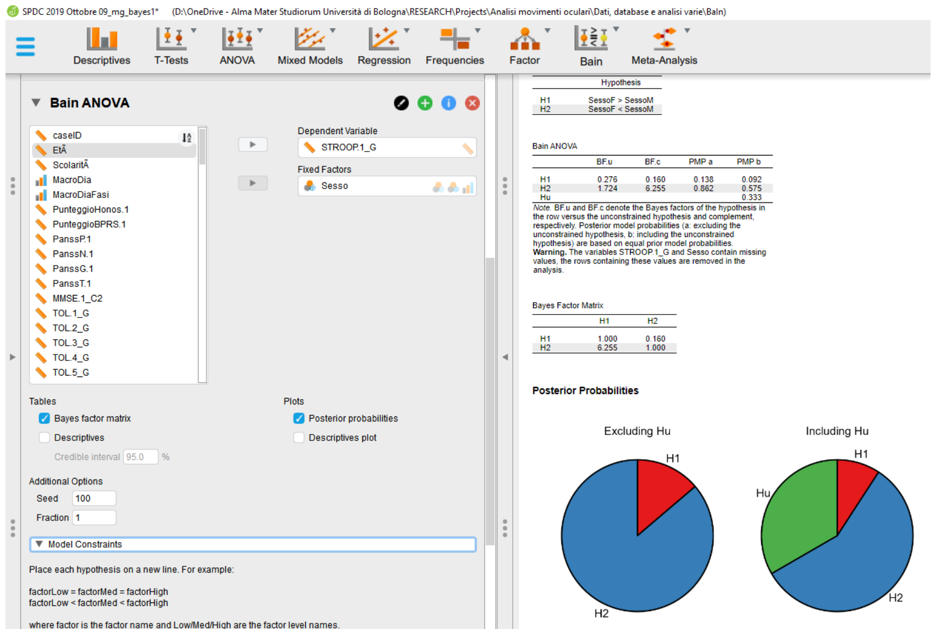
How to

The **“bain” package** (short for Bayesian informative hypothesis)

`bain: Bayes Factors for Informative Hypotheses`

Computes approximated adjusted fractional Bayes factors for equality, inequality, and about equality constrained hypotheses. For a tutorial on this method, see Hoijtink, Mulder, van Lissa, & Gu, (2019) <doi:10.31234/osf.io/v3shc>. For applications in structural equation modeling, see: Van Lissa, Gu, Mulder, Rosseel, Van Zundert, & Hoijtink, (2021) <doi:10.1080/10705511.2020.1745644>. For the statistical underpinnings, see Gu, Mulder, and Hoijtink (2018) <doi:10.1111/bmsp.12110>; Hoijtink, Gu, & Mulder, J. (2019) <doi:10.1111/bmsp.12145>; Hoijtink, Gu, Mulder, & Rosseel, (2019) <doi:10.31234/osf.io/q6h5w>.

Version:	0.2.8
Depends:	R (≥ 3.0.0), stats
Imports:	lavaan
Suggests:	MASS, testthat, knitr, rmarkdown
Published:	2021-12-06
Author:	Xin Gu [aut], Herbert Hoijtink [aut], Joris Mulder [aut], Caspar J van Lissa [aut, cre], Van Zundert Camiel [ctb], Jeff Jones [ctb], Niels Waller [ctb]
Maintainer:	Caspar J van Lissa <c.j.vanlissa at uu.nl>
BugReports:	https://github.com/cjvanlissa/bain/
License:	GPL (≥ 3)
URL:	https://informative-hypotheses.sites.uu.nl/software/bain/



Supported statistical models

The **“bain” package** (short for Bayesian informative hypothesis)

R / RStudio

- Welch’s t-test (paired and one sample)
- Analysis of variance (ANOVA, within-subjects and between-subjects designs)
- Analysis of covariance (ANCOVA)
- Logistic regression
- Linear regression
- Structural equation modelling

JASP

- Welch’s t-test (paired and one sample)
- Analysis of variance (ANOVA, between-subjects design only)
- Analysis of covariance (ANCOVA)
- Linear regression
- Structural equation modelling

(Gu et al., 2014, 2019; Hoijtink, Gu, et al., 2019; Hoijtink, Mulder, et al., 2019; Van Lissa et al., 2021)

2x2 Anova Example

AIM: test the efficacy of a new anxiety treatment

SAMPLE: 200 volunteers suffering from anxiety

EXP. DESIGN: 2x2 design

- INDEPENDEND VARIABLES:
  - **treatment** (Drug/Placebo)
  - **symptoms** (High/Low)
- DEPENDENT VARIABLE: **treatment efficacy index**

Materials

OSF: https://osf.io/sz39j/

or **Landing page**

OR

JASP ?

Define and test informative hypotheses with BaIn

Let's contrast the following hypotheses:

H_1: $(\mu_{Drug.High}, \mu_{Drug.Low}) > 0$ & $(\mu_{Drug.High} - \mu_{Placebo.High}) = (\mu_{Drug.Low} - \mu_{Placebo.Low})$

the two groups receiving a real treatment (Drug) show a positive increase (> 0) in the treatment efficacy index and (&) that the strength of such treatment, intended as the difference between the two groups (Drug/Placebo), is similar regardless of symptomatology level (High/Low)

H_2: $(\mu_{Drug.High}, \mu_{Drug.Low}) > 0$ & $(\mu_{Drug.High} - \mu_{Placebo.High}) > (\mu_{Drug.Low} - \mu_{Placebo.Low})$

the two groups receiving a real treatment (Drug) show a positive increase (> 0) in the treatment efficacy index and (&) that the strength of such treatment, intended as the difference between the two (Drug/Placebo), is higher for the group with High symptoms, as compared to the group with Low symptoms

H_3: $\mu_{Drug.High} = \mu_{Placebo.High} = \mu_{Drug.Low} = \mu_{Placebo.Low}$

there are comparable scores regardless of treatment and symptomatology level

Step-by-step tutorial in JASP

1. Adding the bain module to JASP
2. Load the dataset “dataset_anova.txt”
3. Fit the model
4. Define and test informative hypotheses with bain

(groupDrug.High, groupDrug.Low) > 0 & groupDrug.High - groupPlacebo.High = groupDrug.Low - groupPlacebo.Low

(groupDrug.High, groupDrug.Low) > 0 & groupDrug.High - groupPlacebo.High > groupDrug.Low - groupPlacebo.Low

groupDrug.High = groupPlacebo.High = groupDrug.Low = groupPlacebo.Low

Step-by-step tutorial in

- **Install and load bain**

```
install.packages("bain")
library(bain)
```

- **Load the dataset "dataset_anova.txt"**

- **Fit the model**

```
fit = lm(score ~  group - 1, data = dataset_anova)
```

Step-by-step tutorial in

- **Define and test informative hypotheses**

```
set.seed(123)

results <- bain(x = fit,
    hypothesis =
   "(groupDrug.High, groupDrug.Low) > 0 & groupDrug.High - groupPlacebo.High = groupDrug.Low - groupPlacebo.Low;
      groupDrug.High, groupDrug.Low) > 0 & groupDrug.High - groupPlacebo.High > groupDrug.Low - groupPlacebo.Low;
      groupPlacebo.High = groupDrug.High = groupPlacebo.Low = groupDrug.Low")

print(results)

results$BFmatrix
```

Main results

	BF_U	BF_C	PMP_A	PMP_B	PMP_C
H_1	2.21	2.21	0.221	0.201	0.221
H_2	7.778	438.875	0.779	0.708	0.777
H_3	3.20×10^{-238}	3.20×10^{-238}	3.20×10^{-239}	2.91×10^{-239}	3.20×10^{-239}
H_U				0.091	
H_C	0.018				0.002

Posterior model probabilities (PMP)

Bayes Factor of each hypothesis versus the **unconstrained hypothesis**

Bayes Factor of each hypothesis versus its **complement hypothesis**

Contains the PMP of **each hypothesis**

Contains the PMP of each **hypothesis including hu**

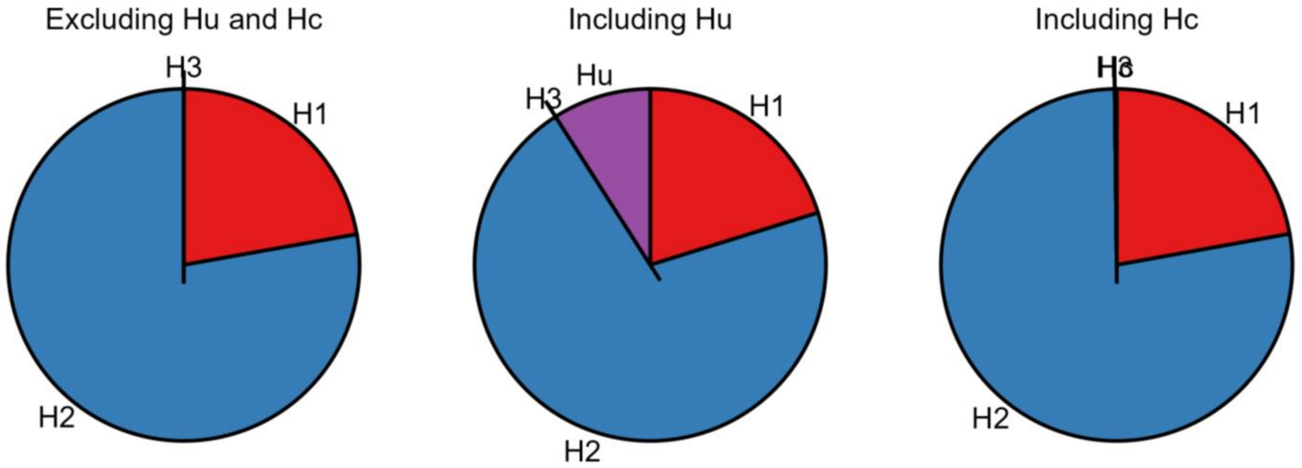
Contains the PMP of each hypothesis **including the complement hypothesis (hc)**

Results

	BF_U	BF_C	PMP_A	PMP_B	PMP_C
H_1	2.21	2.21	0.221	0.201	0.221
H_2	7.778	438.875	0.779	0.708	0.777
H_3	3.20×10^{-238}	3.20×10^{-238}	3.20×10^{-239}	2.91×10^{-239}	3.20×10^{-239}
H_U				0.091	
H_C	0.018				0.002

Bayes Factor of each hypothesis versus the **unconstrained hypothesis**

Bayes Factor of each hypothesis versus its **complement hypothesis**



Which questions can you answer?

	BF_U	BF_C	PMP_A	PMP_B	PMP_C
H_1	2.21	2.21	0.221	0.201	0.221
H_2	7.778	438.875	0.779	0.708	0.777
H_3	3.20×10^{-238}	3.20×10^{-238}	3.20×10^{-239}	2.91×10^{-239}	3.20×10^{-239}
H_U				0.091	
H_C	0.018				0.002

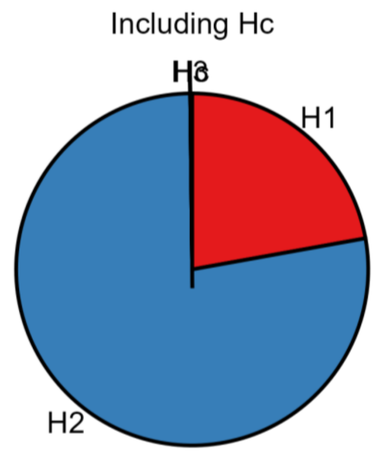
2) How much more likely an hypothesis is relative to other explanations?

quantify the strength of support relative to the complement hypotheses

Bayes Factor of each hypothesis versus its **complement hypothesis**

1) Which of a set of hypotheses is the best?

greatest proportion of PMP



Main results

Bayes Factor Matrix

	H_1	H_2	H_3
H_1	1	0.284	6.91×10^{-237}
H_2	3.519	1	2.43×10^{-238}
H_3	1.45×10^{-238}	4.11×10^{-239}	1

3) How much more likely a given hypothesis is relative to another specific hypothesis?

Direct comparison between each pair of informative hypotheses tested by the researcher

(thus, not considering Hu and Hc)

Now it’s your turn!

Multiple linear regression

AIM: investigate if factors other than symptom severity can impact its effectiveness

SAMPLE: 100 volunteers suffering from anxiety

EXP. DESIGN: 2x2 design

- INDEPENDEND VARIABLES:
  - **Age**
  - **Dosage**
  - **Sympthoms**
- DEPENDENT VARIABLE: **treatment efficacy index**

Materials

OSF: https://osf.io/sz39j/

or **Landing page**

OR

JASP ?

Define and test informative hypotheses with BaIn

Let's contrast the following hypotheses:

The first hypothesis poses that dosage has a higher impact than symptoms, which has a higher impact than age on treatment effect:

$$\mathbf{H_1: dosage > symptoms > age}$$

The second hypothesis poses that symptoms have a higher impact than dosage, which has a higher impact than age on treatment effect:

$$\mathbf{H_2: symptoms > dosage > age}$$

The third hypothesis poses that dosage and symptoms have a comparably higher impact than age on treatment effect:

$$\mathbf{H_3: (dosage, symptoms) > age}$$

Now you answer these questions!

1) **Which of a set of hypotheses is the best?**

1) **How much more likely a given hypothesis is relative to other possible explanations?**

1) **How much more likely a given hypothesis is relative to another specific hypothesis?**

Well done!!

Final remarks: two relevant papers

Testing Bayesian Informative Hypotheses in 5 steps with JASP and R

These materials are based on the following tutorial paper, currently in press:

Garofalo S., Finotti G., Orsoni M., Giovagnoli S., Benassi M. (2024) Testing Bayesian Informative Hypotheses in 5 steps with R and JASP. Advances in Methods and Practices in Psychological Science. doi: 10.1177/25152459241260259

Please refer to the full paper for more information.

+

Interaction effect: are you doing the right thing?

PLOS ONE

PUBLISH ABOUT

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

Interaction effect: Are you doing the right thing?

Sara Garofalo, Sara Giovagnoli, Matteo Orsoni, Francesca Starita, Mariagrazia Benassi

Published: July 20, 2022 • https://doi.org/10.1371/journal.pone.0271668

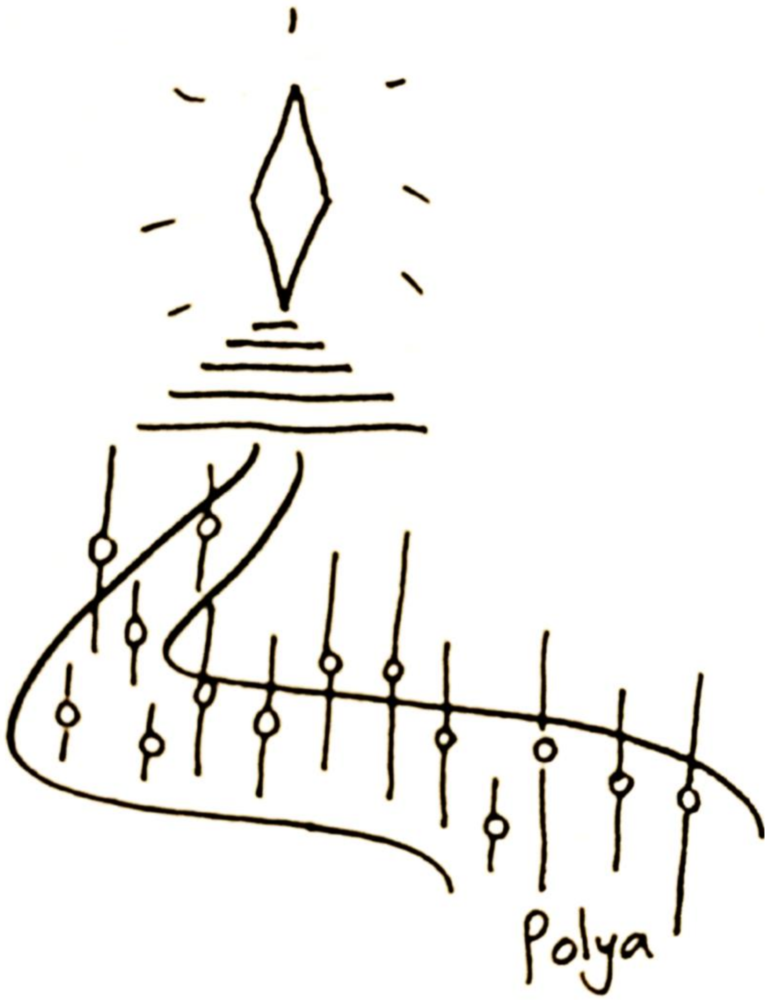
https://osf.io/ya9mp/

(Garofalo et al., 2022, PlosOne)

Take home message

1. Choose the **appropriate logic** to test hypothesis:
    - Is your hypothesis **dichotomous**?
    - Is it **informative**?

2. Only **include hypotheses** that:
    - Are **plausible**
    - Represent the actual **expectations**
    - Are **parsimonious**



If there are specific expectations

- Define your actual **research expectations** in terms of specific ordering of means
- Use **Bayesian informative hypothesis** to contrast models

Psychometrics & neuropsychology lab

Department of Psychology, University of Bologna

Mariagrazia Benassi

Sara Giovagnoli

Matteo Orsoni

Gianluca Finotti

Sara Garofalo | **sara.garofalo@unibo.it**

References

- Béland, S., Klugkist, I., Raîche, G., Magis, D., 2012. A short introduction into Bayesian evaluation of informative hypotheses as an alternative to exploratory comparisons of multiple group means. Tutor. Quant. Methods Psychol. 8, 122–126. https://doi.org/10.20982/tqmp.08.2.p122
- Gu, X., Mulder, J., Hoijtink, H., 2018. Approximated adjusted fractional Bayes factors: A general method for testing informative hypotheses. Br. J. Math. Stat. Psychol. 71, 229–261. https://doi.org/10.1111/bmsp.12110
- Kruschke, J.K., Liddell, T.M., 2018. The Bayesian New Statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. Psychon. Bull. Rev. 25, 178–206. https://doi.org/10.3758/s13423-016-1221-4
- **Hoijtink, H., Mulder, J., van Lissa, C., Gu, X., 2019. A Tutorial on Testing Hypotheses Using the Bayes Factor. Psychol. Methods. https://doi.org/10.1037/met0000201**
- **Hoijtink, H., 2012. Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists.**
- **Hoijtink, H., van Kooten, P., Hulsker, K., 2016. Why Bayesian Psychologists Should Change the Way They Use the Bayes Factor. Multivariate Behav. Res. 51, 2–10. https://doi.org/10.1080/00273171.2014.969364**
- Okada, K., 2015. Bayesian meta-analysis of Cronbach's coefficient alpha to evaluate informative hypotheses. Res. Synth. Methods 6, 333–346. https://doi.org/10.1002/jrsm.1155
- "Piled Higher and Deeper" by Jorge Cham www.phdcomics.com

References

- Garofalo S, Giovagnoli S, Orsoni M, Starita F, Benassi M (2022) Interaction effect: Are you doing the right thing? PLoS ONE 17(7): e0271668. https://doi.org/10.1371/journal.pone.0271668
- Hoijtink, H., Mulder, J., van Lissa, C., & Gu, X. (2019). A Tutorial on Testing Hypotheses Using the Bayes Factor. *Psychological Methods*. https://doi.org/10.1037/met0000201
- Hoijtink, H. (2012). *Informative Hypotheses: Theory and Practice for Behavioral and Social Scientists*. Chapman & Hall/CRC.
- Rosnow, R., & Rosenthal, R. (1995). “Some things you learn aren’t so”: Cohen’s paradox, Asch’s paradigm, and the interpretation of interaction. *Psychological Science*.
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., van Langen, J., & Kievit, R. A. (2021). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 63. https://doi.org/10.12688/wellcomeopenres.15191.2
- Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., & Kievit, R. A. (2019). Raincloud plots: A multi-platform tool for robust data visualization. *Wellcome Open Research*, *4*, 1–41. https://doi.org/10.12688/wellcomeopenres.15191.1
- Cumming, G., & Finch, S. (2005). Inference by eye confidence intervals and how to read pictures of data. *American Psychologist*, *60*(2), 170–180. https://doi.org/10.1037/0003-066X.60.2.170
- Cumming, G. (2007). Inference by eye: Pictures of confidence intervals and thinking about levels of confidence. *Teaching Statistics*, *29*(3), 89–93. https://doi.org/10.1111/j.1467-9639.2007.00267.x
- Cumming, G. (2009). Inference by eye: Reading the overlap of independent confidence intervals. *Statistics in Medicine*, *28*(2), 205–220. https://doi.org/10.1002/sim.3471