NB: If you haven't filled out the questionnaire yet, please do so!
(for link: see tutorial announcement email)

Bayesiansk statistik – ett alternativ till t-test och ANOVA?

Uppsala
24 Oct 2019

Ronald van den Berg
Department of Psychology
Uppsala University / Stockholm University

Bayesian statistics #1: Hypothesis testing

Somewhere in a digital cloud
17 June 2020

Ronald van den Berg
Department of Psychology
Stockholm University

Tutorial #1: hypothesis testing

Examples of hypothesis testing:

- Is drug *D* more effective than a placebo?
- Is there a correlation between age and mortality rate in disease *Y*?
- Does model *A* fit the data better than model *B*?
- Do my subjects have a non-zero guessing rate?

Tutorial #2 (next week): hypothesis testing

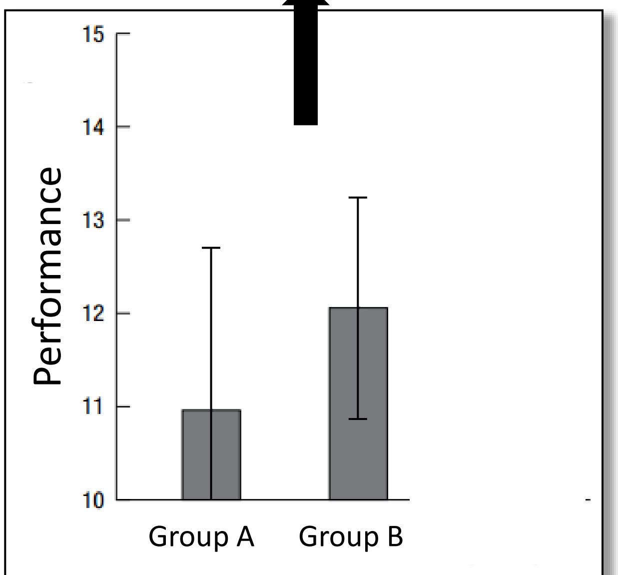
Examples of estimation:

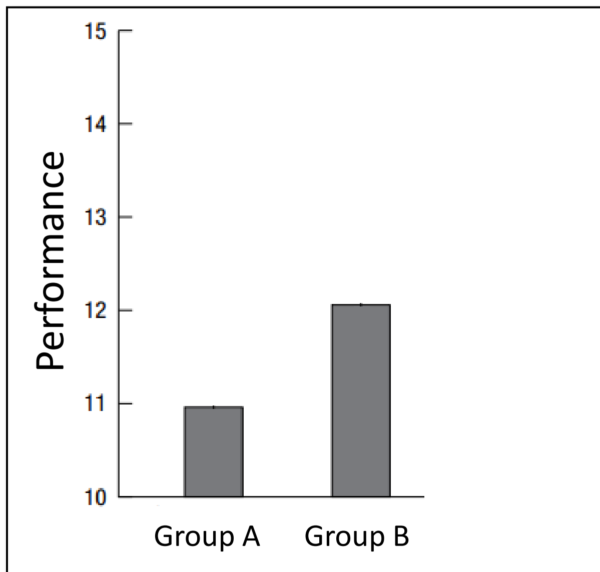
- On <u>what percentage</u> of people is this drug effective?
- <u>How strong</u> is the correlation between age and mortality rate in disease *Y*?
- <u>How much</u> better does model A fit the data than model B?
- <u>How frequently</u> did subjects guess in my experiment?

Why use statistics?

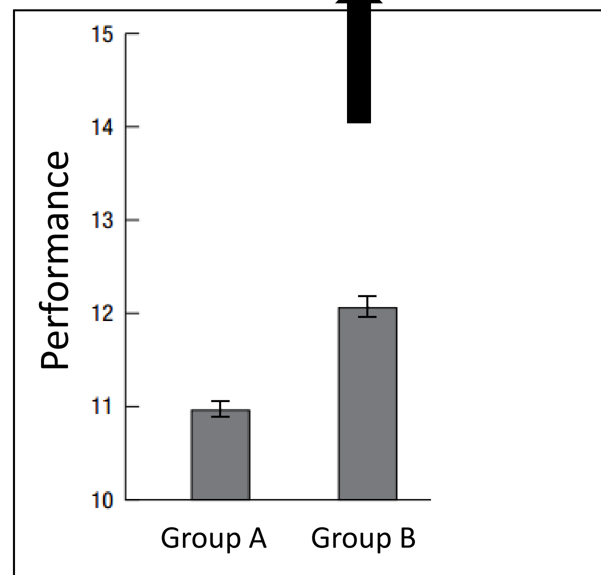
Why do we need statistical tests?

Differences are probably due to **random variation**





Differences are probably due to an **effect of group**

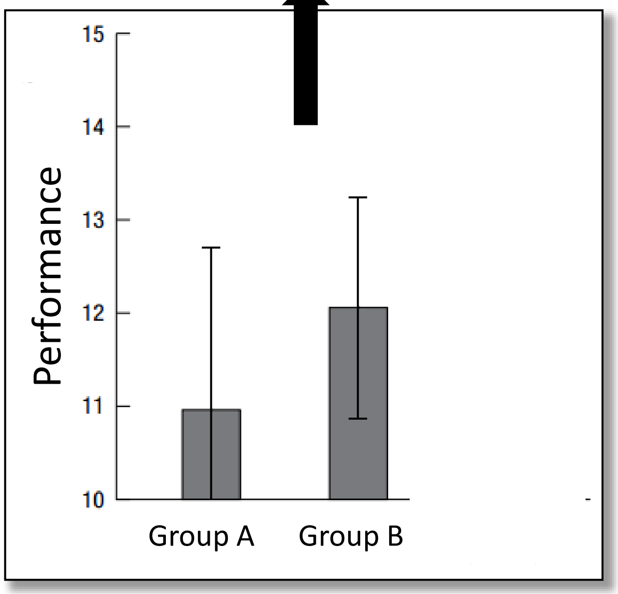


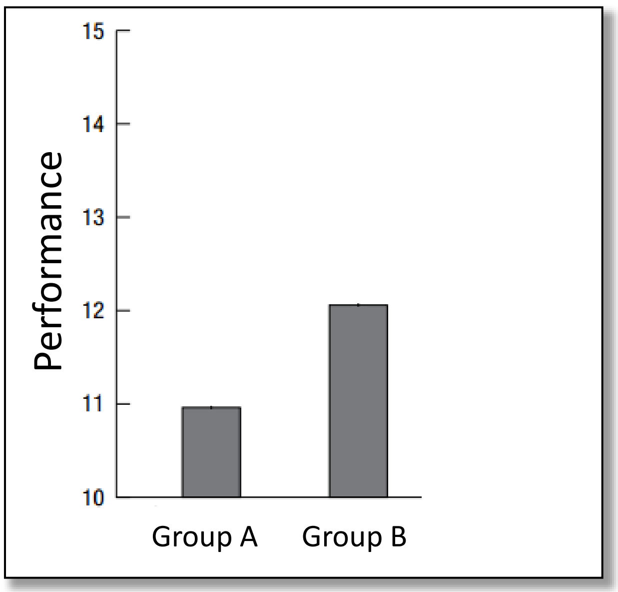
Why do we need statistical tests?

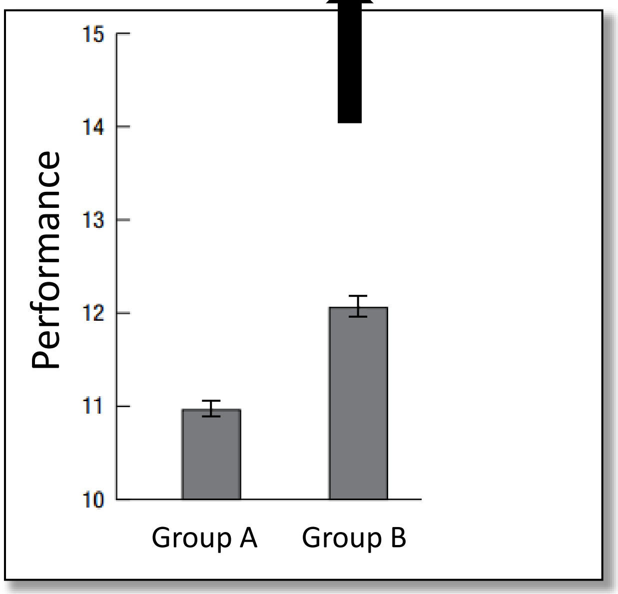
Task of statistics is to quantify this "probably"

Differences are probably due to **random variation**

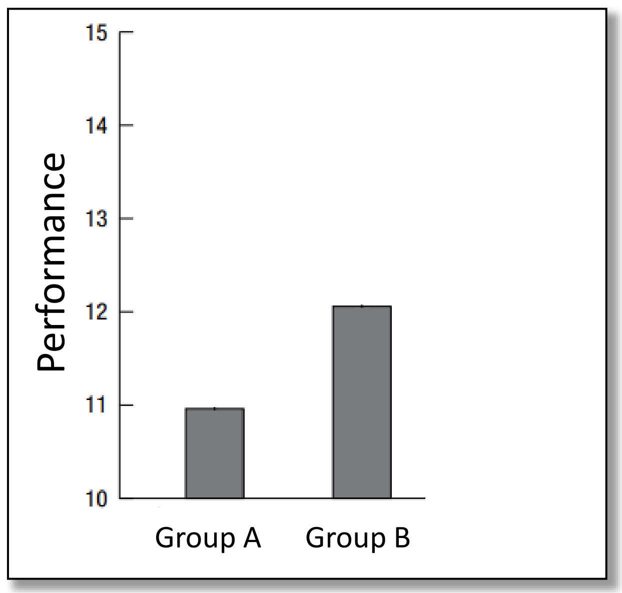
Differences are probably due to an **effect of group**







Is there an effect of group on performance?



H0: There is no effect of group on performance

H1: There is an effect of group on performance

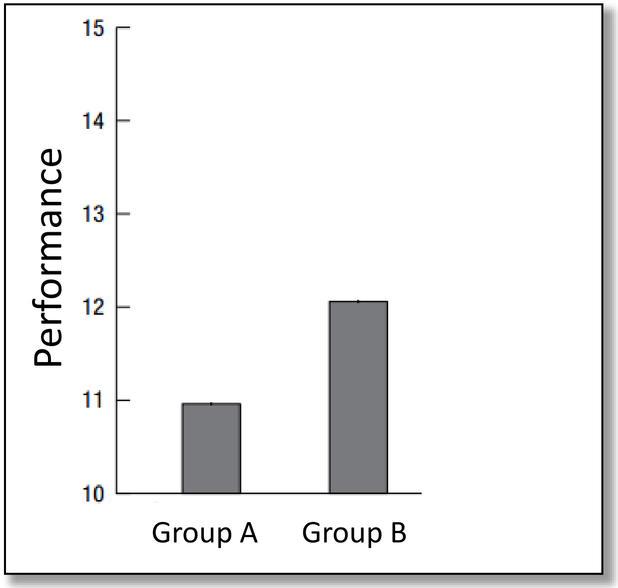
Is there an effect of group on performance?

Frequentist approach

Compute *p*(extremeness of the data | H0 is true)

Bayesian approach

Compute *p*(data | H0 is true) / p(data | H1 is true)



H0: There is no effect of group on performance

H1: There is an effect of group on performance

Frequentist approach

Note

Ther

Ron
(18

on
0)

ELSEVIER

The Journal of Socio-Economics 33 (2004) 587–606

The Journal of Socio-Economics

www.elsevier.com/locate/econbase

Mindless statistics

Gerd Gigerenzer*

Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

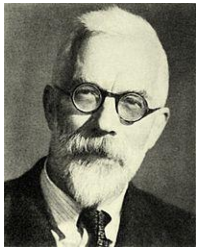
Abstract

Statistical rituals largely eliminate statistical thinking in the social sciences. Rituals are indispensable for identification with social groups, but they should be the subject rather than the procedure of science. What I call the "null ritual" consists of three steps: (1) set up a statistical null hypothesis, but do not specify your own hypothesis nor any alternative hypothesis, (2) use the 5% significance level for rejecting the null and accepting your hypothesis, and (3) always perform this procedure. I report evidence of the resulting collective confusion and fears about sanctions on the part of students and teachers, researchers and editors, as well as textbook writers.

Keywords: Rituals; Collective illusions; Statistical significance; Editors; Textbooks

The
incon

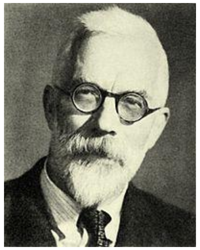
an
ect



Hypothesis testing: Fisher's approach

1. Formulate a null hypothesis, H_0
   E.g.: “the drug has no effect on recovery speed”

2. Compute *p*



Hypothesis testing: Fisher's approach

1. Formulate a null hypothesis, H_0
   E.g.: "the drug has no effect on recovery speed"

2. Compute *p*, i.e., the probability of observing your data or more extreme data *if* H_0 were true

3. A low *p* value implies that either something rare has occurred or H_0 is not true

- **Power analysis** has no place in this framework
- **High *p*** does not mean to accept H0

Reasoning:

the lower *p*, the more certain we can be that H0 is false

-> sounds reasonable, but ultimately a flawed way to test hypotheses

A *p*-roblem

Applying Fisher's approach to the case of Sally Clark

- 1996: Clark’s **1st son died** a few weeks after birth (SIDS?)
- 1998: Clark’s **2nd son died** a few weeks after birth (SIDS again????)
- 1999: Clark was **found guilty of murder** and given two life sentences

The conviction was partly based on the following statistical argument:

- H0: babies died from "Sudden Infant Death Syndrome" (SIDS) aka "crib death"
- SIDS occurence rate is 1 in 8,500
- The chance of this happening twice is 1 in 73 million, i.e., $p = 0.0000000137$
- Therefore, H0 is rejected
- Therefore, she must be guilty (double murder)

What is wrong with this line of reasoning?

Applying Fisher's approach to the case of Sally Clark

Even though H0 is unlikely, other hypotheses may be even more unlikely!!

The conviction was partly based on the following statistical argument:

- H0: babies died from "Sudden Infant Death Syndrome" (SIDS) aka "crib death"
- SIDS occurence rate is 1 in 8,500
- The chance of this happening twice is 1 in 73 million, i.e., $p = 0.0000000137$
- Therefore, H0 is rejected
- Therefore, she must be guilty (double murder)

What is wrong with this line of reasoning?

Applying Fisher's approach to the case of Sally Clark

Evidence is best treated as a <u>relative</u> concept

~~“How improbable is H0?”~~

“How (im)probable is H0, <u>relative to</u> H1?” ✓

- H1: double murder
- Infant murder rate in UK: approximately 1 in 33,000(*)
- The chance of this happening twice is 1 in 1.1 billion, i.e., p = 0.000000000918
- SIDS is 15 times more likely than murder!

(*) Marks, M. N., & Kumar, R. (1993). Infanticide in England and Wales. *Medicine, Science and the Law*, *33*(4), 329-339.

Applying Fisher's approach to the case of Sally Clark

How did it end for Clark?

- 1996: Clark's **first son died** suddenly within a few weeks of his birth
- 1998: Clark's **second son died** suddenly within a few weeks of his birth
- 1999: Clark was **found guilty of murder** and given two life sentences
- 2003: Clark is set free, yet highly traumatized
- 2007: Clark dies from alcohol poisoning

Applying Fisher's approach to the case of Sally Clark

The same kind of flawed reasoning was part of Lucia de Berk’s conviction in the Netherlands

The deeper problem here:

- Some events are unlikely under *any* hypothesis

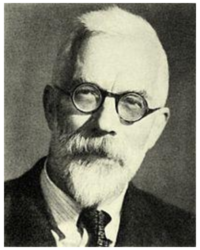
The deeper problem here:

- Some events are unlikely under *any* hypothesis
- Should we then reject them all and consider the event unexplainable?

Solution: lower the α value for rare events?

. . . no scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas.

Sir Ronald A. Fisher (1956)



The deeper problem here:

- Some events are unlikely under *any* hypothesis
- Should we then reject them all and consider the event unexplainable?

Solution: lower the α value for rare events?

However: how to do this without knowing the cause of the event??

The Bayes factor

Introduction to the Bayes Factor

$$\frac{p\left(H_0 \mid D\right)}{p\left(H_1 \mid D\right)}$$

← Probability of Hypothesis 0, given the data

← Probability of Hypothesis 1, given the data

Introduction to the Bayes Factor

$$\underbrace{\frac{p\left(H_0 \mid D\right)}{p\left(H_1 \mid D\right)}}_{\text{Posterior ratio}} = \underbrace{\frac{p\left(D \mid H_0\right)}{p\left(D \mid H_1\right)}}_{\boxed{\text{Bayes factor}}} \times \underbrace{\frac{p\left(H_0\right)}{p\left(H_1\right)}}_{\text{Prior ratio}}$$

Indicates how many times more likely the data are under H0 compared to H1

Introduction to the Bayes Factor

$$\underbrace{\frac{p(H_0 \mid D)}{p(H_1 \mid D)}}_{\text{Posterior ratio}} = \underbrace{\frac{p(D \mid H_0)}{p(D \mid H_1)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_0)}{p(H_1)}}_{\text{Prior ratio}}$$

Alternative interpretation:

BF indicates the change from prior odds to posterior odds brought about by the data

- By definition a *relative* measure
- Easy, pleasant interpretation(s)
- Allows to quantify evidence in favor of the null!
- Generalizes more easily than frequentist approach?

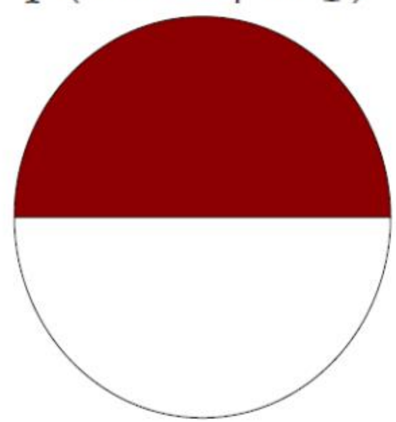
Introduction to the Bayes Factor

$$\underbrace{\frac{p(H_0 \mid D)}{p(H_1 \mid D)}}_{\text{Posterior ratio}} = \underbrace{\frac{p(D \mid H_0)}{p(D \mid H_1)}}_{\boxed{\text{Bayes factor}}} \times \underbrace{\frac{p(H_0)}{p(H_1)}}_{\text{Prior ratio}}$$

Visual interpretation of the Bayes factor

$\mathrm{BF}_{10} = 1$

$p(\mathrm{data} \mid \mathcal{H}_1)$



$p(\mathrm{data} \mid \mathcal{H}_0)$

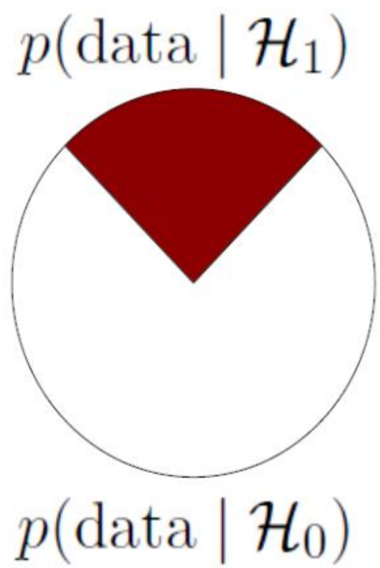
Introduction to the Bayes Factor

$$\underbrace{\frac{p(H_0 \mid D)}{p(H_1 \mid D)}}_{\text{Posterior ratio}} = \underbrace{\frac{p(D \mid H_0)}{p(D \mid H_1)}}_{\boxed{\text{Bayes factor}}} \times \underbrace{\frac{p(H_0)}{p(H_1)}}_{\text{Prior ratio}}$$

Visual interpretation of the Bayes factor

$$\mathrm{BF}_{10} = \frac{1}{3}$$

$$\mathrm{BF}_{01} = 3$$



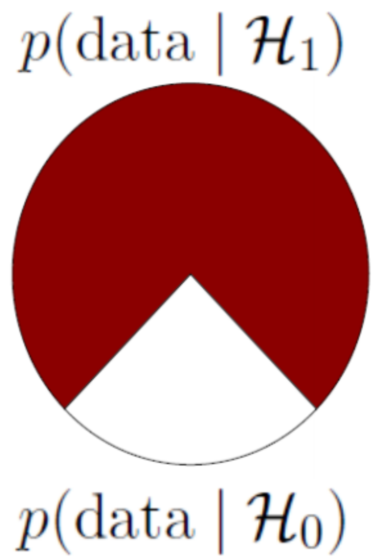
Introduction to the Bayes Factor

$$\underbrace{\frac{p(H_0 \mid D)}{p(H_1 \mid D)}}_{\text{Posterior ratio}} = \underbrace{\frac{p(D \mid H_0)}{p(D \mid H_1)}}_{\boxed{\text{Bayes factor}}} \times \underbrace{\frac{p(H_0)}{p(H_1)}}_{\text{Prior ratio}}$$

Visual interpretation of the Bayes factor

$$\mathrm{BF}_{10} = 3$$

$$\mathrm{BF}_{01} = \frac{1}{3}$$



Guideline for interpreting BF evidence strength

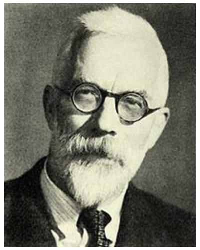
(source: Wagenmakers et al. 2016)

Bayes factor, BF_{10}	Evidence category
> 100	Extreme evidence for $\mathcal{H}_1$
30 - 100	Very strong evidence for $\mathcal{H}_1$
10 - 30	Strong evidence for $\mathcal{H}_1$
3 - 10	Moderate evidence for $\mathcal{H}_1$
1 - 3	Anecdotal evidence for $\mathcal{H}_1$
1	No evidence
1/3 - 1	Anecdotal evidence for $\mathcal{H}_0$
1/10 - 1/3	Moderate evidence for $\mathcal{H}_0$
1/30 - 1/10	Strong evidence for $\mathcal{H}_0$
1/100 - 1/30	Very strong evidence for $\mathcal{H}_0$
$< 1/100$	Extreme evidence for $\mathcal{H}_0$

The two approaches in 5 steps

	Frequentist approach (Fisher)	Bayesian approach
Step 1	Formulate a **single** hypothesis H0	Formulate **two or more** hypotheses (may or may not include "H0")
Step 2	Decide on **all** study factors **before** measuring a single data point (sample size, what to do with outliers, etc) – **revising these decisions later would invalidate the test**	Make some initial decisions, e.g. "collect data from 20 subjects" or "collect data until BF>10 or BF<1/10 – **may be revised later**
Step 3	Gather data	Gather data
Step 4	Compute *p*	Compute Bayes Factors
Step 5	If $p < 0.05$: reject H0 If $p > 0.05$: conclude nothing	Interpret the Bayes Factors as a **continuous** measure **in favor or against** the hypothesis

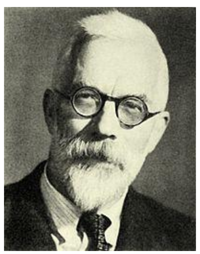
<u>Fisherian</u> vs <u>Bayesian</u> statistics:



p value

Bayes factor

Fisherian vs **Bayesian** statistics:



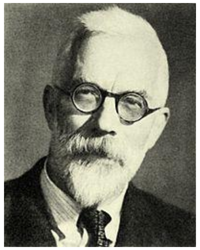
***p* value**

- Evidence is **absolute** (about single hypothesis)
- Can only **reject** hypotheses
- Tests are problem-specific?
- Confusing for non-statisticians

Bayes factor

- Evidence is always **relative** (w.r.t. alternative hypotheses)
- Can **reject** *and* **support** hypotheses
- Tests are general?
- Much less confusing

<u>Fisherian</u> vs <u>Bayesian</u> statistics:



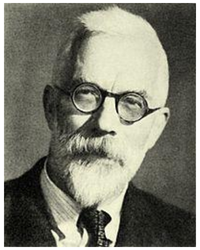
***p* value**

- Evidence is **absolute** (about single hypothesis)
- Can only **reject** hypotheses
- Tests are problem-specific?
- Confusing for non-statisticians

Bayes factor

- Evidence is always **relative** (w.r.t. alternative hypotheses)
- Can **reject** *and* **support** hypotheses
- Tests are general?
- Much less confusing

Fisherian vs **Bayesian** statistics:



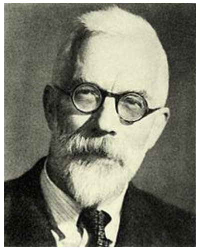
***p* value**

- Evidence is **absolute** (about single hypothesis)
- Can only **reject** hypotheses
- Tests are problem-specific
- Confusing for non-statisticians

Bayes factor

- Evidence is always **relative** (w.r.t. alternative hypotheses)
- Can **reject** *and* **support** hypotheses
- Tests are general
- Less confusing?

<u>Fisherian</u> vs <u>Bayesian</u> statistics:



***p* value**

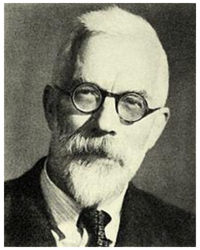
- Evidence is **absolute** (about single hypothesis)
- Can only **reject** hypotheses
- Tests are problem-specific?
- Confusing for non-statisticians

Bayes factor

- Evidence is always **relative** (w.r.t. alternative hypotheses)
- Can **reject** *and* **support** hypotheses
- Tests are general?
- Less confusing?

Why isn't everyone a Bayesian???

<u>Fisherian</u> vs <u>Bayesian</u> statistics:



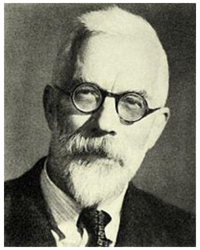
***p* value**

- Evidence is **absolute**
  (about single hypothesis)
- Can only **reject** hypotheses
- Tests are problem-specific?
- Confusing for non-statisticians

Bayes factor

- Evidence is always **relative**
  (w.r.t. alternative hypotheses)
- Can **reject** *and* **support** hypotheses
- Tests are general?
- Less confusing?
- Computationally expensive

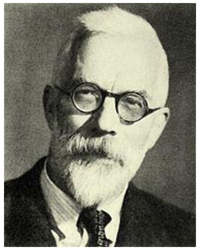
Fisherian vs Bayesian statistics:



***p* value**
- Evidence is **absolute** (about single hypothesis)
- Can only **reject** hypotheses
- Tests are problem-specific?
- Confusing for non-statisticians

Bayes factor
- Evidence is always **relative** (w.r.t. alternative hypotheses)
- Can **reject** *and* **support** hypotheses
- Tests are general?
- Less confusing?
- Computationally expensive
- Requires specification of **priors**

<u>Fisherian</u> vs <u>Bayesian</u> statistics:



***p* value**

- Evidence is **absolute** (about single hypothesis)
- Can only **reject** hypotheses
- Tests are problem-specific?
- Confusing for non-statisticians

"Objective"

Bayes factor

- Evidence is always **relative** (w.r.t. alternative hypotheses)
- Can **reject** *and* **support** hypotheses
- Tests are general?
- Less confusing?
- Computationally expensive
- Requires specification of **priors**

"Subjective"

Different philosophies

Bayesians quantify **degrees of belief**

-> highly subjective

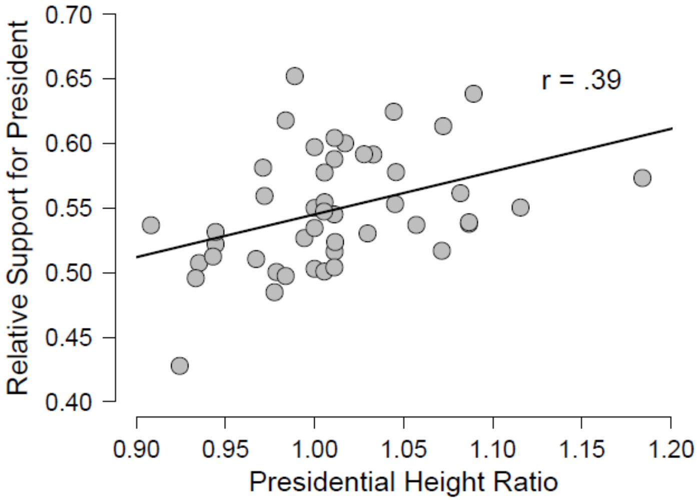
Frequentists quantify **long-term frequencies**

-> claimed to be fully objective

Example #1:

Correlation analysis

Correlation - example



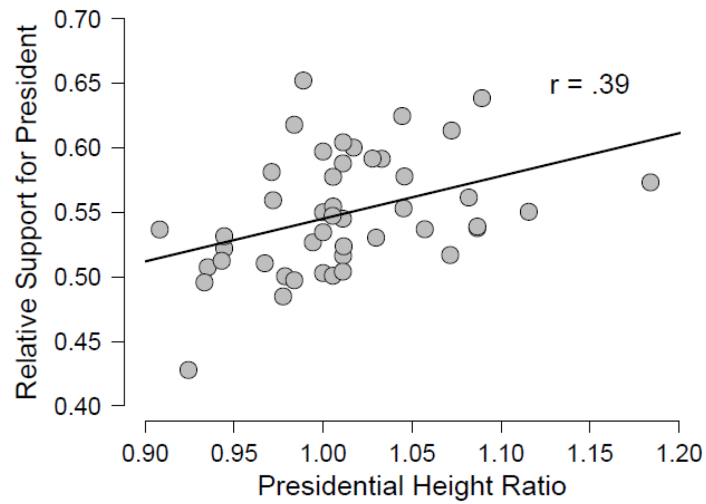
Two common questions:

1. Is the correlation "real"?
2. What is a plausible estimate of the strength of the “true” correlation?

Frequentist approach:

- Assume that data comes from a bivariate normal distribution
- Compute p value to answer first question
- Compute confidence interval to answer second question

Correlation - example



Intuitive way to think about the p-value:

$p \approx$ probability of finding $r_{\text{sample}} > 0.39$ if $r_{\text{population}} = 0$

Formally, however

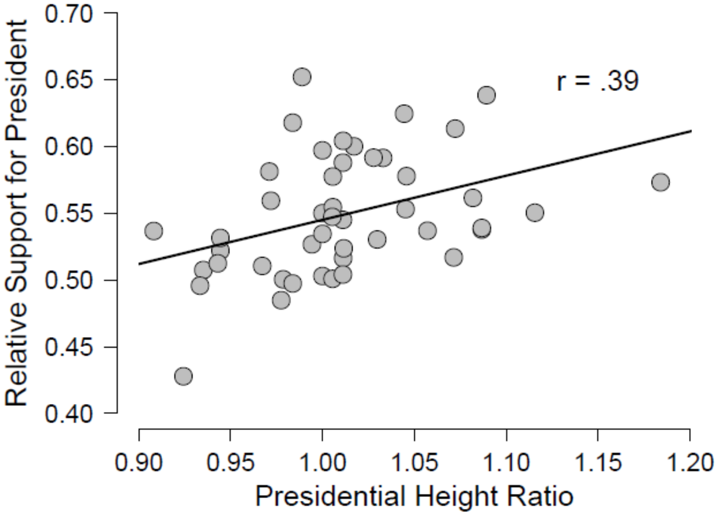
1. Compute t-statistic $t^* = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}}$

2. Compute p = $p(t^* > 0.39 \mid r_{\text{population}} = 0)$

Underlying logic:

If $r_{\text{population}}$=0, then t^* follows a t distribution with n-2 degrees of freedom

Correlation – frequentist results



H0: No correlation between height ratio and relative support

Frequentist results:

- *p* = 0.007
- CI = [.12; .62]

What have we learned from this analysis?

1. If the "true" (population-level) correlation were 0, we would have only 0.7% chance of finding data as extreme as our sample

2. We can be 95% confident that the "true" correlation is between .12 and .62

   Wrong! This is a Bayesian interpretation of a frequentist concept!

Correlation analysis:
a Bayesian approach

Bayesian correlation test

Same assumption

The data come from a bivariate normal distribution

Same question

Is there any evidence for a correlation at population level?

Different way to quantify this evidence

- Bayes factor instead of *p* value
- Credible interval instead of confidence interval

Bayesian correlation test

$$\underbrace{\frac{p(H_0 \mid D)}{p(H_1 \mid D)}}_{\text{Posterior ratio}} = \underbrace{\frac{p(D \mid H_0)}{p(D \mid H_1)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_0)}{p(H_1)}}_{\text{Prior ratio}}$$

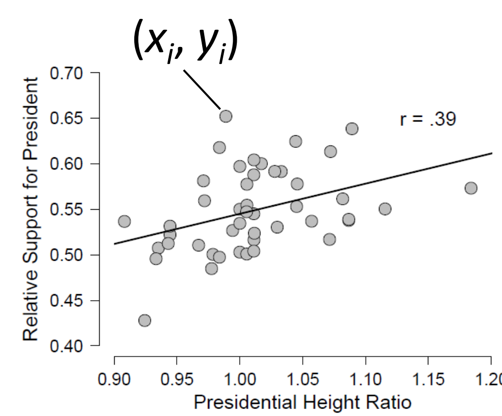
In the context of correlation analysis, we define:

H_0: $r = 0$

H_1: $r \neq 0$

Hence, we want to compute

$$\mathrm{BF}_{01} = \frac{p(D \mid r = 0)}{p(D \mid r \neq 0)}$$

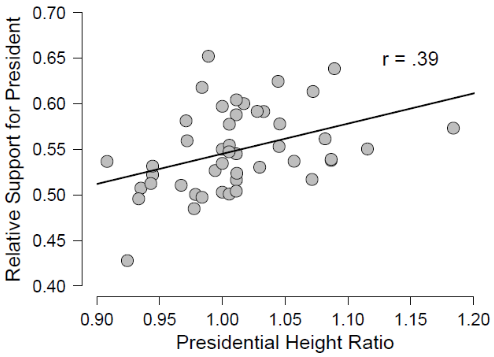


Bayesian correlation test

$$\mathrm{BF}_{01} = \frac{p(\mathbf{x}, \mathbf{y} \mid r = 0)}{p(\mathbf{x}, \mathbf{y} \mid r \neq 0)}$$

Hence, we want to compute

$$\mathrm{BF}_{01} = \frac{p(D \mid H_0)}{p(D \mid H_1)} = \frac{\boxed{p(\mathbf{x}, \mathbf{y} \mid r = 0)}}{\boxed{p(\mathbf{x}, \mathbf{y} \mid r \neq 0)}}$$

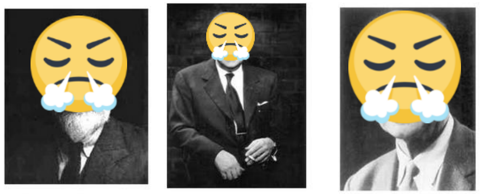


Bayesian correlation test

$$\mathrm{BF}_{01} = \frac{p(\mathbf{x},\mathbf{y} \mid r = 0)}{p(\mathbf{x},\mathbf{y} \mid r \neq 0)} = \frac{\int p(\mathbf{x},\mathbf{y} \mid r = 0, \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\int p(\mathbf{x},\mathbf{y} \mid r \neq 0, \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

Parameters of the assumed model

Prior over parameter values



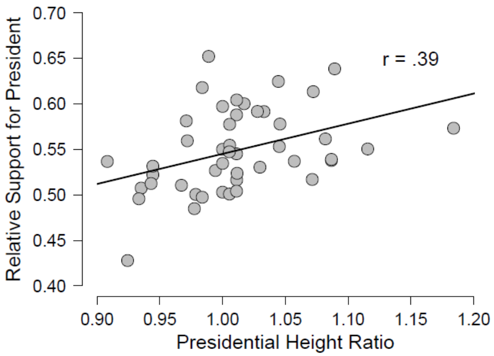
Ronald Fisher (1890 – 1962)

Jersey Neyman (1894 – 1981)

Egon Pearson (1895 – 1980)

Hence, we want to compute

$$\mathrm{BF}_{01} = \frac{p(D \mid H_0)}{p(D \mid H_1)} = \frac{\boxed{p(\mathbf{x},\mathbf{y} \mid r = 0)}}{\boxed{p(\mathbf{x},\mathbf{y} \mid r \neq 0)}}$$



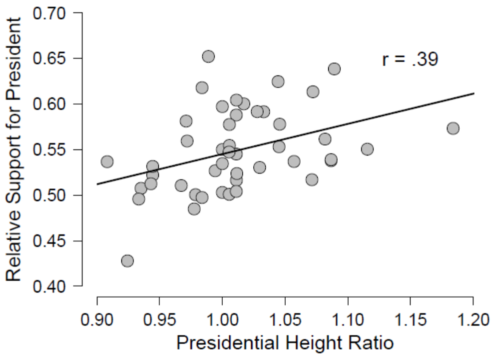
Bayesian correlation test

$$\mathrm{BF}_{01} = \frac{p(\mathbf{x},\mathbf{y} \mid r = 0)}{p(\mathbf{x},\mathbf{y} \mid r \neq 0)} = \frac{\int p(\mathbf{x},\mathbf{y} \mid r = 0, \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\int p(\mathbf{x},\mathbf{y} \mid r \neq 0, \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

Need to specify what we mean here

Hence, we want to compute

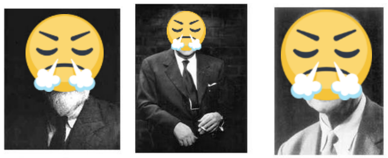
$$\mathrm{BF}_{01} = \frac{p(D \mid H_0)}{p(D \mid H_1)} = \frac{\boxed{p(\mathbf{x},\mathbf{y} \mid r = 0)}}{\boxed{p(\mathbf{x},\mathbf{y} \mid r \neq 0)}}$$



Bayesian correlation test

$$\mathrm{BF}_{01} = \frac{p(\mathbf{x},\mathbf{y} \mid r = 0)}{p(\mathbf{x},\mathbf{y} \mid r \neq 0)} = \frac{\int p(\mathbf{x},\mathbf{y} \mid r = 0, \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\int p(\mathbf{x},\mathbf{y} \mid r \neq 0, \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}$$

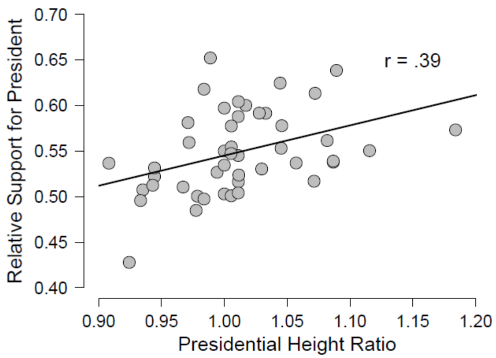
$$\mathrm{BF}_{01} = \frac{p(\mathbf{x},\mathbf{y} \mid r = 0)}{p(\mathbf{x},\mathbf{y} \mid r \neq 0)} = \frac{\int p(\mathbf{x},\mathbf{y} \mid r = 0, \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\int p(\mathbf{x},\mathbf{y} \mid r, \boldsymbol{\theta})\, p(r)\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}\, dr}$$



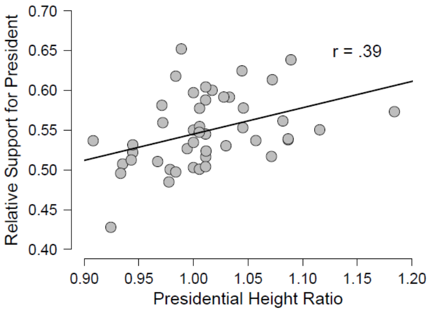
Ronald Fisher (1890 – 1962) | Jersey Neyman (1894 – 1981) | Egon Pearson (1895 – 1980)

Hence, we want to compute

$$\mathrm{BF}_{01} = \frac{p(D \mid H_0)}{p(D \mid H_1)} = \frac{\boxed{p(\mathbf{x},\mathbf{y} \mid r = 0)}}{\boxed{p(\mathbf{x},\mathbf{y} \mid r \neq 0)}}$$



Bayesian correlation test



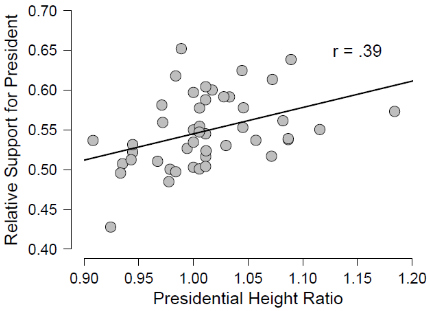
$$\mathrm{BF}_{01} = \frac{p(\mathbf{x},\mathbf{y} \mid r = 0)}{p(\mathbf{x},\mathbf{y} \mid r \neq 0)} = \frac{\int p(\mathbf{x},\mathbf{y} \mid r = 0, \boldsymbol{\theta})\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int p(\mathbf{x},\mathbf{y} \mid r, \boldsymbol{\theta})\, p(r)\, p(\boldsymbol{\theta}) d\boldsymbol{\theta} dr}$$

How to proceed from here?

Naive approach

1. Plug in bivariate normal distribution
2. Specify prior over *r*
3. Specify prior over **θ** = $\{\mu_1, \mu_2, \sigma_1, \sigma_2\}$

Bayesian correlation test



$$\mathrm{BF}_{01} = \frac{p(\mathbf{x},\mathbf{y} \mid r = 0)}{p(\mathbf{x},\mathbf{y} \mid r \neq 0)} = \frac{\int p(\mathbf{x},\mathbf{y} \mid r = 0, \boldsymbol{\theta})\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}}{\int p(\mathbf{x},\mathbf{y} \mid r, \boldsymbol{\theta})\, p(r)\, p(\boldsymbol{\theta})\, d\boldsymbol{\theta}\, dr}$$

How to proceed from here?

Smarter approach: ask the internet

[HTML] A default **Bayesian** hypothesis **test** for correlations and partial correlations [HTML] springer.com

R Wetzels, EJ Wagenmakers - Psychonomic bulletin & review, 2012 - Springer

... We illustrate the use of the **Bayesian correlation test** with three examples from the psychological literature ... It should be noted that Jeffreys (1961) also proposed a **Bayesian correlation test**, one that differs slightly from the one outlined here ...

Cited by 334 Related articles All 20 versions

Psychon Bull Rev (2012) 19:1057–1064
DOI 10.3758/s13423-012-0295-x

BRIEF REPORT

A default Bayesian hypothesis test for correlations and partial correlations

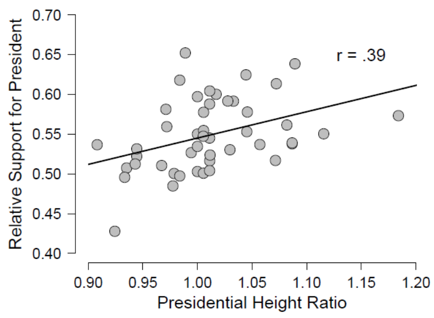
Ruud Wetzels · Eric-Jan Wagenmakers

In order to calculate the Bayes factor for the JZS (partial) correlation test, we conceptualize these Bayesian tests as a comparison between two regression models, such that the test becomes equivalent to a variable selection test for linear regression (i.e., a test of whether or not the regression coefficient β should be included in the model). This conceptualization allows us to exploit the JZS prior distribution. Computer code for calculating the JZS Bayes factors is presented in the Appendix.

Keywords Bayesian inference · Correlation · Statistical evidence

result is compelling, nor may they continue data collection when the fixed sample size result is ambiguous (Edwards et al., 1963). These drawbacks are not merely theoretical but have

Bayesian correlation test



$$\mathrm{BF}_{01} = \frac{p(\mathbf{x},\mathbf{y} \mid r = 0)}{p(\mathbf{x},\mathbf{y} \mid r \neq 0)} = \frac{\int p(\mathbf{x},\mathbf{y} \mid r = 0, \boldsymbol{\theta})\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}}{\int p(\mathbf{x},\mathbf{y} \mid r, \boldsymbol{\theta})\, p(r)\, p(\boldsymbol{\theta}) d\boldsymbol{\theta} dr}$$

How to proceed from here?

Wetzels & Wagenmaker’s approach:

1. Assume a JZS prior on *r* [an “uninformative” prior]
2. Now the BF can be computed analytically and depends only on r_{sample} and *n*.

Bayesian stats in action

JASP:

- Free
- Similar interface as SPSS
- Bayesian and frequentist tests
- Powered by BayesFactor for R

BayesFactor for R

- Free
- Gives much more control over what you're doing than JASP

Bayesian correlation test results

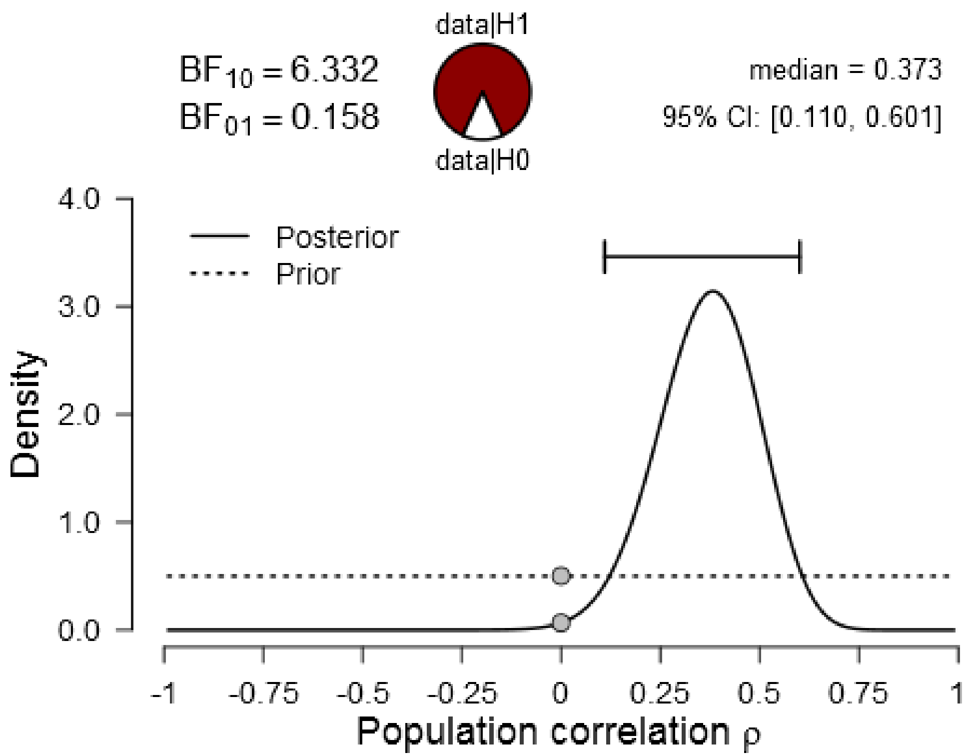
Frequentist approach:
- p = 0.007
- CI = [.12; .62]

(CONFIDENCE interval)

Bayesian approach:
- BF_{10} = 6.33
- CI = [.11; .60]

(CREDIBLE interval)

JASP result:



Bayesian correlation test results

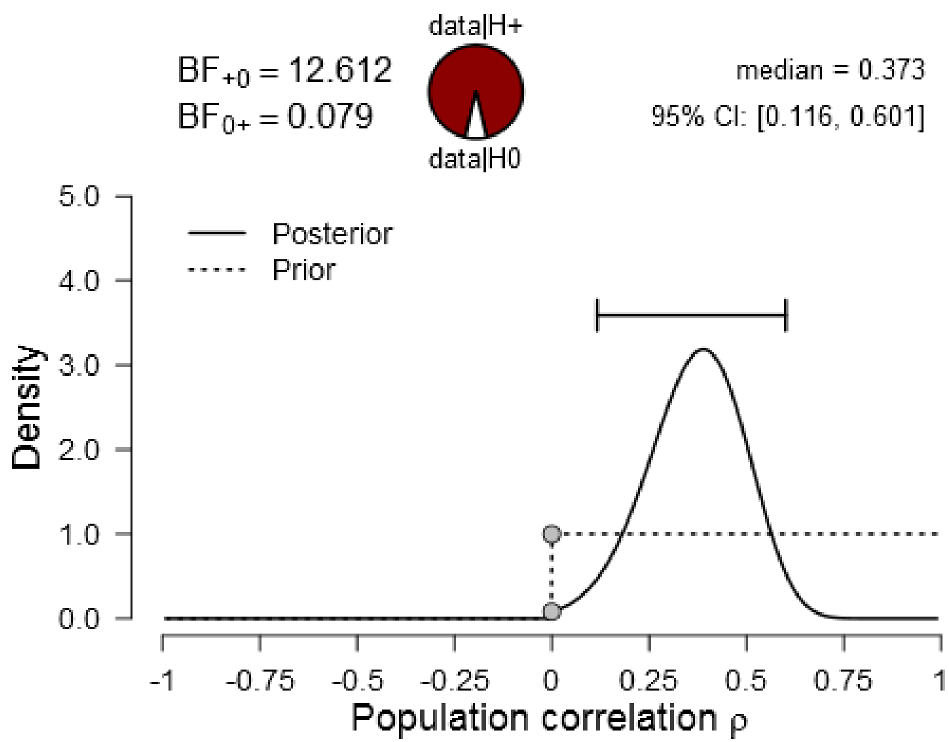
Test #2: prior belief is that *r* is **positive**

Frequentist approach:
- p = 0.003
- CI = [.16; 1.0]

(CONFIDENCE interval)

Bayesian approach:
- BF_{+0} = 12.61
- CI = [.11; .60]

(CREDIBLE interval)



Bayesian correlation test results

Test #3: prior belief is that *r* is **negative**

Frequentist approach:

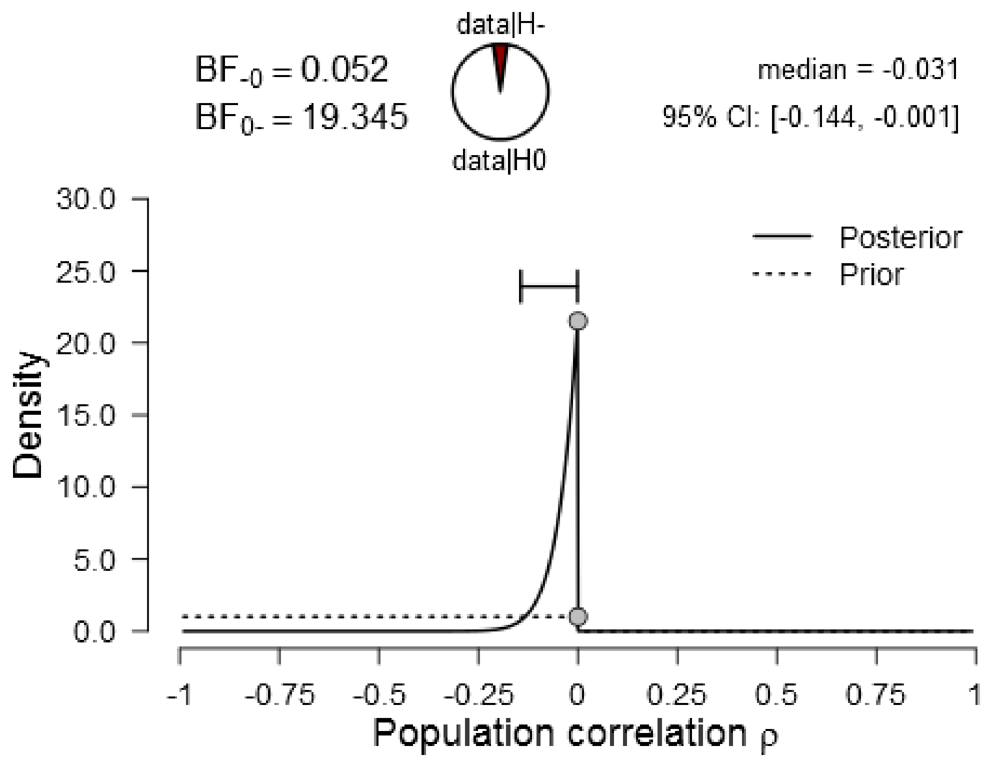
- p = 0.997
- CI = [-1, .58]

(CONFIDENCE interval)

Bayesian approach:

- BF_{-0} = 0.052
- CI = [-.14; -.001]

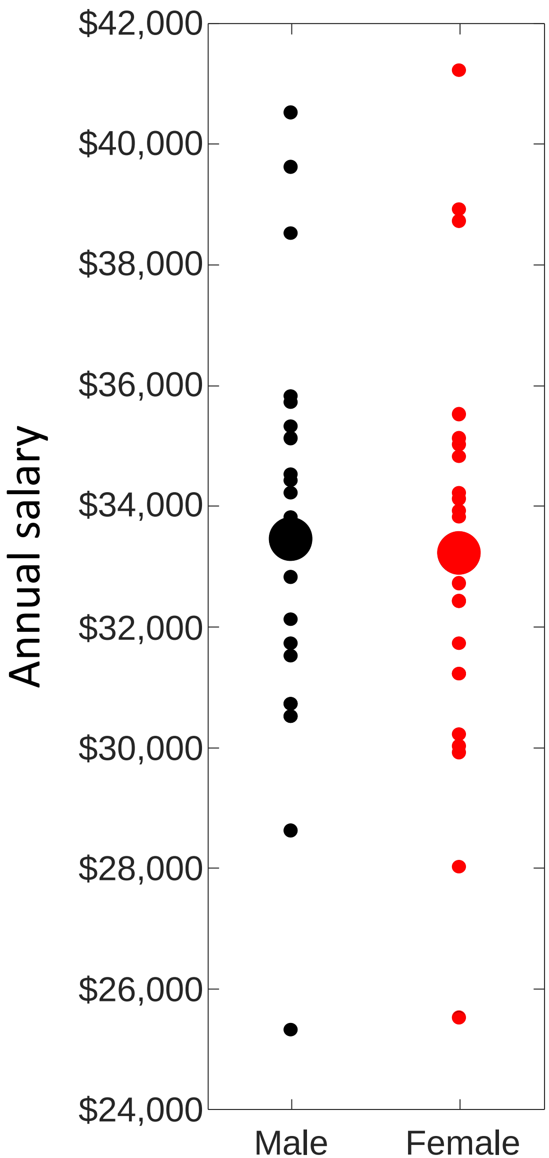
(CREDIBLE interval)



Example #2:

t-test

T-test: frequentist approach



H0: δ = 0

No difference in salary between men and women

Frequentist approach:

1. Compute t-statistic
2. Compute p value (based on *t* and *n*)

Result: p = 0.21

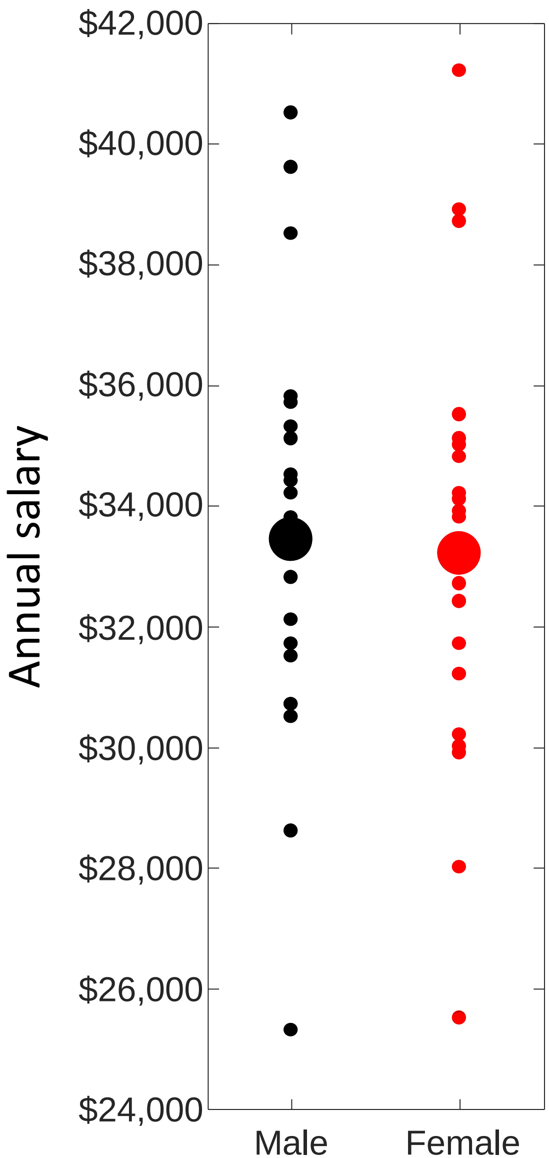
Interpretation:

"Assuming H0 is true, we would find a test statistics as extreme (or more extreme) as in our sample in 21% of samples drawn from this population"

Conclusion

None – high *p* value does not imply H0 to be true

T-test: Bayesian approach



H0: δ = 0
H1: δ ≠ 0

Psychonomic Bulletin & Review
2009, 16 (2), 225-237
doi:10.3758/PBR.16.2.225

Bayesian *t* tests for accepting and rejecting the null hypothesis

Jeffrey N. Rouder, Paul L. Speckman, Dongchu Sun, and Richard D. Morey
University of Missouri, Columbia, Missouri

and

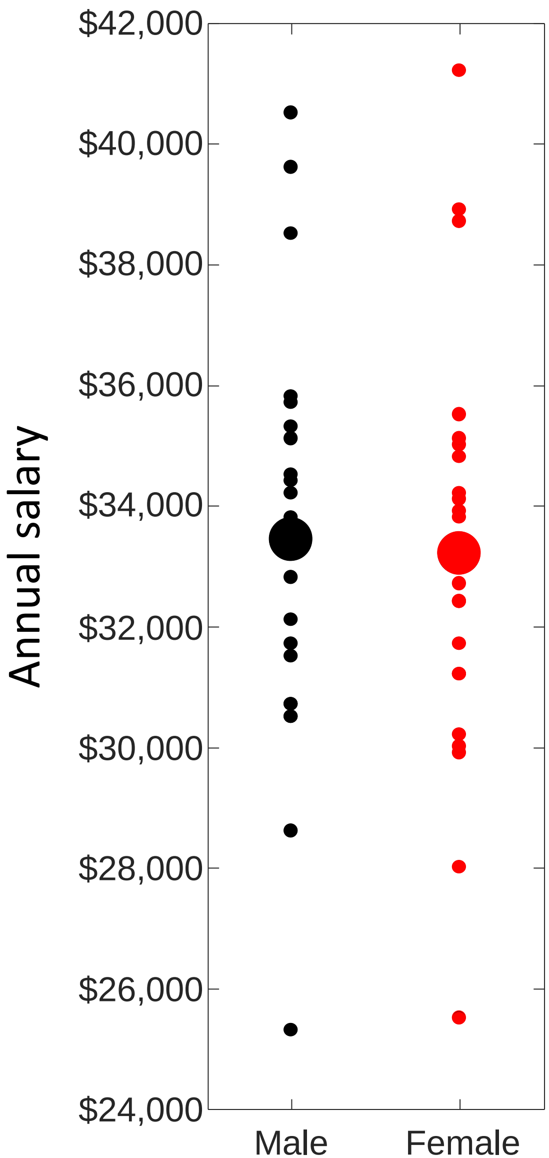
Geoffrey Iverson
University of California, Irvine, California

Progress in science often comes from discovering invariances in relationships among variables; these invariances often correspond to null hypotheses. As is commonly known, it is not possible to state evidence for the null hypothesis in conventional significance testing. Here we highlight a Bayes factor alternative to the conventional *t* test that will allow researchers to express preference for either the null hypothesis or the alternative. The Bayes factor has a natural and straightforward interpretation, is based on reasonable assumptions, and has better properties than other methods of inference that have been advocated in the psychological literature. To facilitate use of the Bayes factor, we provide an easy-to-use, Web-based program that performs the necessary calculations.

Advances in science often come from identifying *invariances*—those elements that stay constant when others change. Kepler, for example, described the motion of planets. From an Earth-bound vantage point, planets seem to have strange and variable orbits. Not only do they differ in their speeds and locations, they even appear to back-

Shibley Hyde, 2005, 2007). To believe that only effects of genders, rather than invariances across genders, will appear in performance strikes us as an extreme position. A second example comes from the domain of subliminal priming (see, e.g., Dehaene et al., 1998): To prove that subliminal priming occurs, it must be shown that detection

T-test: Bayesian approach



H0: $\delta = 0$
H1: $\delta \neq 0$

$$\mathrm{BF}_{01} = \frac{p\left(D \mid H_0\right)}{p\left(D \mid H_1\right)} = \frac{p\left(D \mid \delta = 0\right)}{p\left(D \mid \delta \neq 0\right)}$$

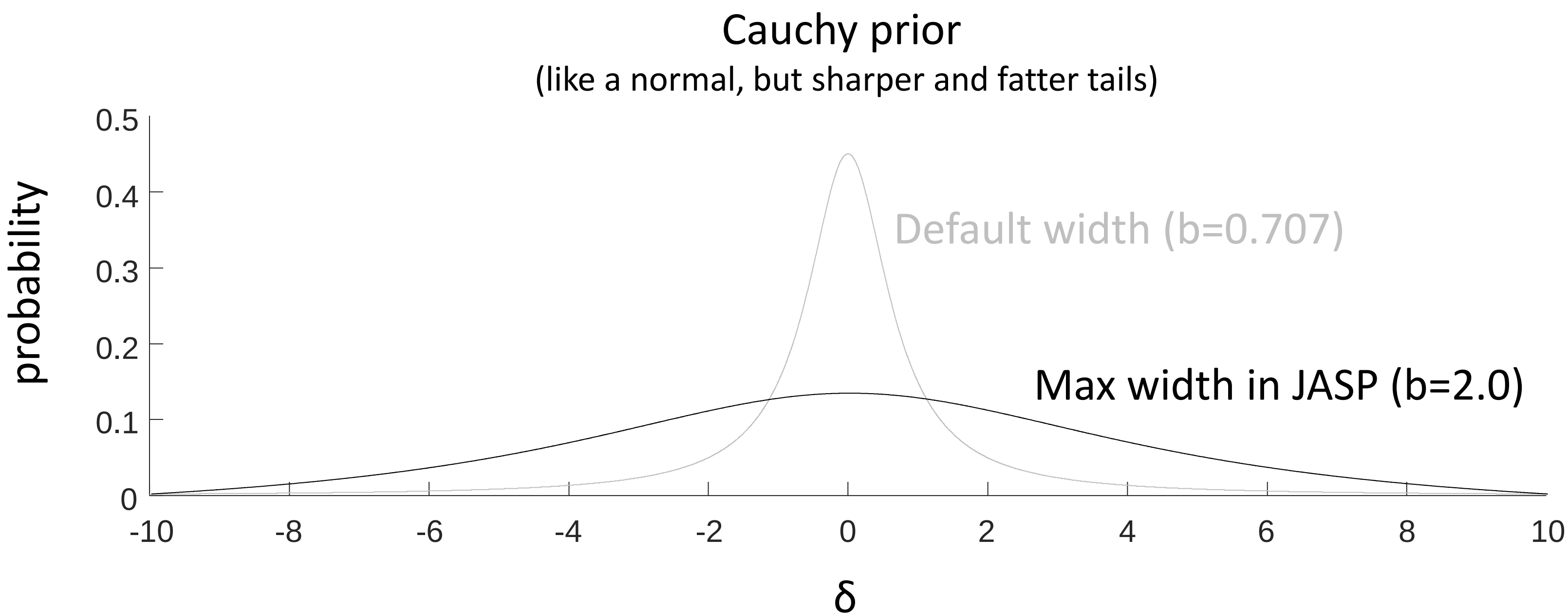
Approach

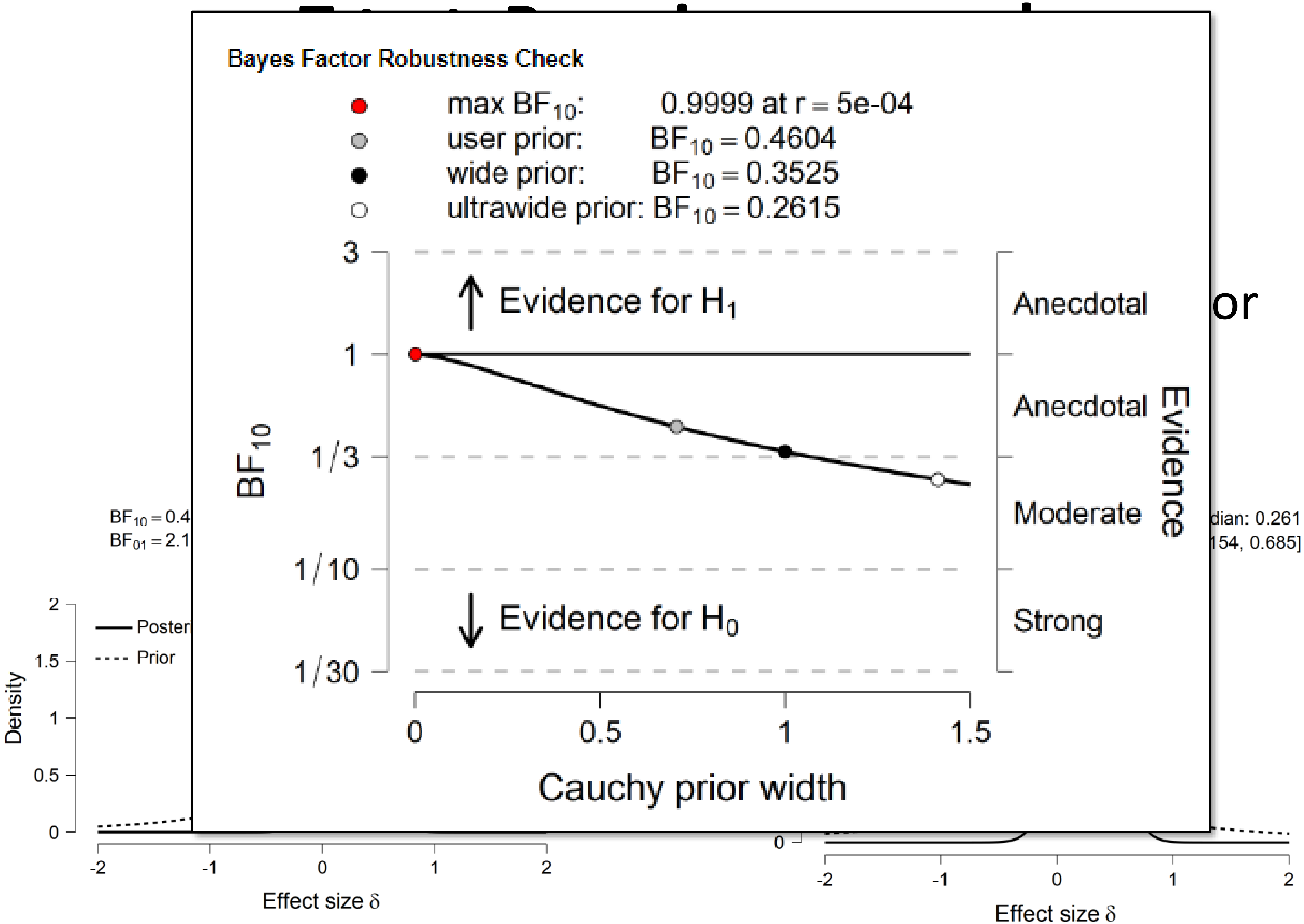
- Assume Cauchy prior on effect size
- Assume Jeffreys prior on variance, $p(\sigma^2) \propto 1/\sigma^2$
- Compute BF as follows:

$$B_{01} = \frac{\left(1 + \frac{t^2}{\nu}\right)^{-(\nu+1)/2}}{\int_0^\infty (1+Ng)^{-1/2} \left(1 + \frac{t^2}{(1+Ng)\nu}\right)^{-(\nu+1)/2} (2\pi)^{-1/2} g^{-3/2} e^{-1/(2g)} dg}$$

t = t statistic, N = #measurements, ν = #DoF = N-1

T-test: Bayesian approach




Bayes Factor Robustness Check
max BF10: 0.9999 at r = 5e-04
user prior: BF10 = 0.4604
wide prior: BF10 = 0.3525
ultrawide prior: BF10 = 0.2615
Evidence for H1
Evidence for H0
BF10
3
1
1/3
1/10
1/30
Anecdotal
Anecdotal
Moderate
Strong
Evidence
0
0.5
1
1.5
Cauchy prior width
BF10 = 0.4
BF01 = 2.1
Poster
Prior
Density
Effect size δ
dian: 0.261
154, 0.685]
Effect size δ
or

Example #3:

ANOVA & Regression

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*

ARTICLE INFO

Article history:
Received 14 December 2011
Received in revised form
3 July 2012
Available online 31 August 2012

Keywords:
Bayes factor
Model selection
Bayesian statistics
Linear models

ABSTRACT

Bayes factors have been advocated as superior to *p*-values for assessing statistical evidence in data. Despite the advantages of Bayes factors and the drawbacks of *p*-values, inference by *p*-values is still nearly ubiquitous. One impediment to the adoption of Bayes factors is a lack of practical development, particularly a lack of ready-to-use formulas and algorithms. In this paper, we discuss and expand a set of default Bayes factor tests for ANOVA designs. These tests are based on multivariate generalizations of Cauchy priors on standardized effects, and have the desirable properties of being invariant with respect to linear transformations of measurement units. Moreover, these Bayes factors are computationally convenient, and straightforward sampling algorithms are provided. We cover models with fixed, random, and mixed effects, including random interactions, and do so for within-subject, between-subject, and mixed designs. We extend the discussion to regression models with continuous covariates. We also discuss how these Bayes factors may be applied in nonlinear settings, and show how they are useful in differentiating between the power law and the exponential law of skill acquisition. In sum, the current development makes the computation of Bayes factors straightforward for the vast majority of designs in experimental psychology.

Multivariate Behavioral Research, 47:877–903, 2012

ISSN: 0027-3171 print/1532-7906 online
DOI: 10.1080/00273171.2012.734737

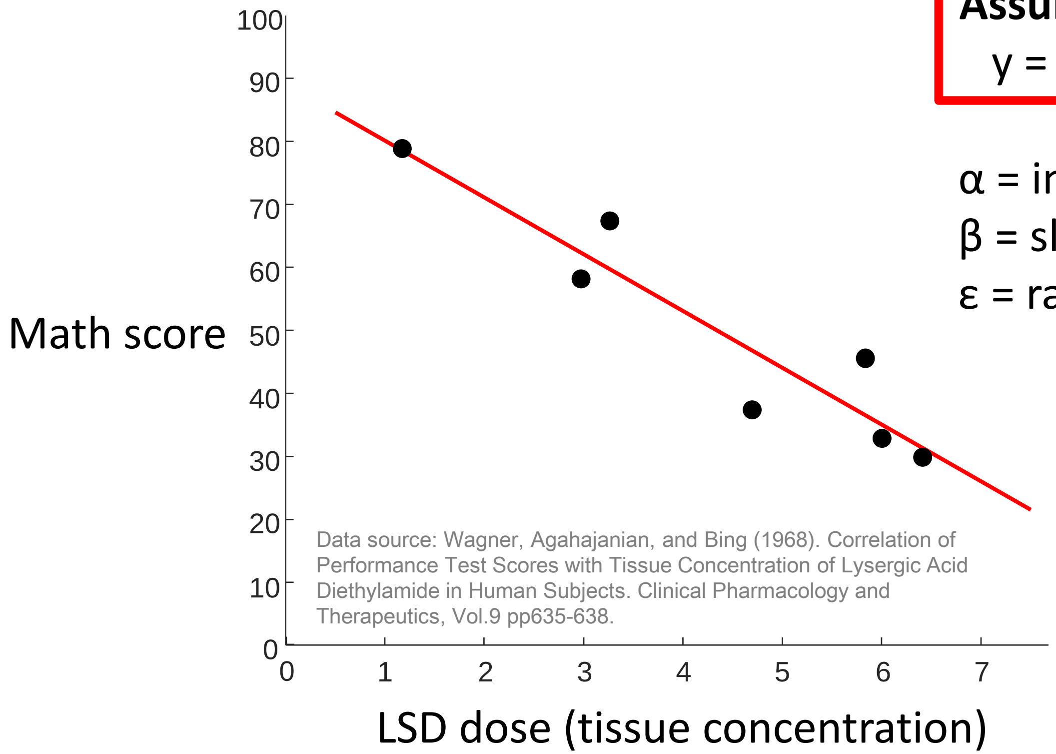
Psychology Press
Taylor & Francis Group

Default Bayes Factors for Model Selection in Regression

Jeffrey N. Rouder
University of Missouri

Richard D. Morey
University of Groningen

Bayesian approach to simple linear regression



Assumed model

$$y = \alpha + \beta x + \varepsilon$$

α = intercept
β = slope
ε = random error (Gaussian)

Frequentist vs Bayesian approach

- Same assumed underlying model
- Same questions/hypotheses
- Different way of quantifying evidence

Bayesian approach to simple linear regression

$$\underbrace{\frac{p(H_0 \mid D)}{p(H_1 \mid D)}}_{\text{Posterior ratio}} = \underbrace{\frac{p(D \mid H_0)}{p(D \mid H_1)}}_{\text{Bayes factor}} \times \underbrace{\frac{p(H_0)}{p(H_1)}}_{\text{Prior ratio}}$$

Assumed model

y = α + βx + ε

The hypotheses are:

H_0: β = 0

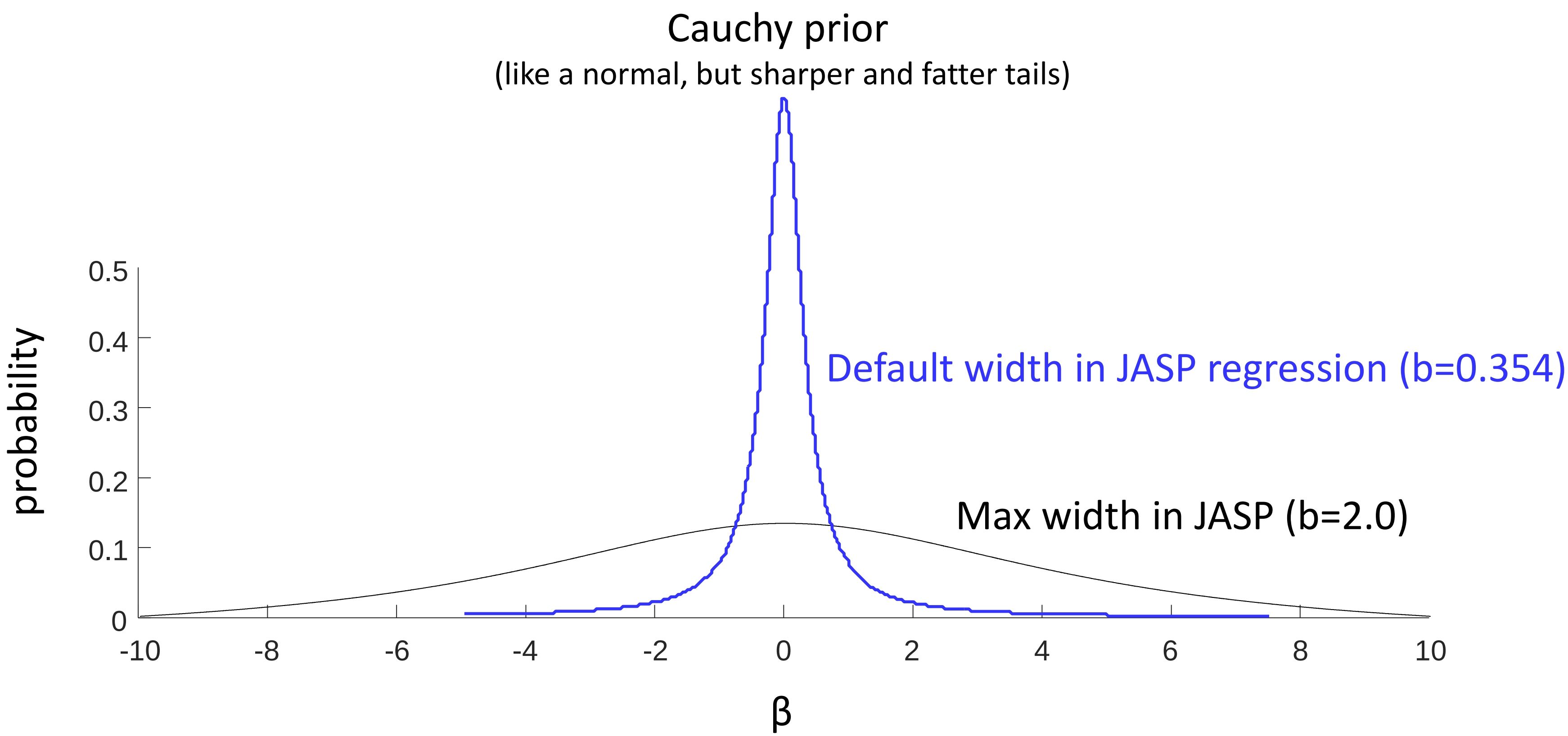
H_1: β ≠ 0

$$\mathrm{BF}_{01} = \frac{p(D \mid H_0)}{p(D \mid H_1)} = \frac{\boxed{p(D \mid \beta = 0)}}{\boxed{p(D \mid \beta \neq 0)}}$$

Computable

Uncomputable unless we specify what we mean with "β≠0"
-> Cauchy prior

Bayesian approach to simple linear regression

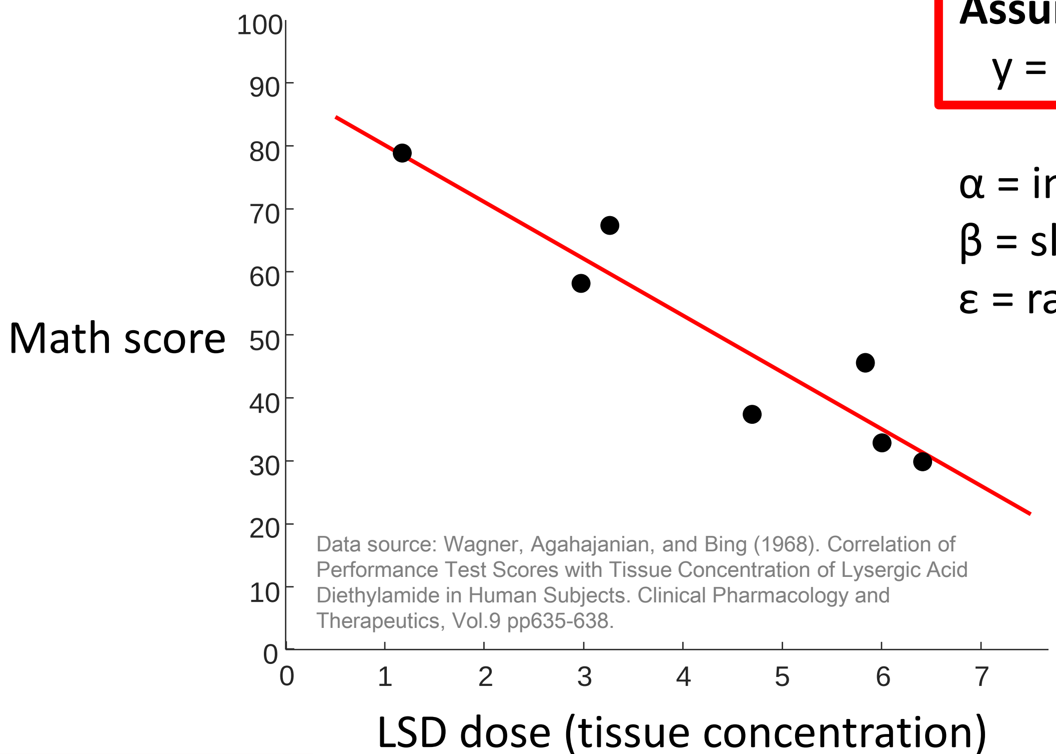


$$\mathrm{BF}_{01} = \frac{p(D \mid H_0)}{p(D \mid H_1)} = \frac{\boxed{p(D \mid \beta = 0)}}{\boxed{p(D \mid \beta \neq 0)}}$$

Computable

Uncomputable unless we specify what we mean with “β≠0”
-> Cauchy prior

Bayesian approach to simple linear regression



Assumed model

$y = \alpha + \beta x + \varepsilon$

α = intercept

β = slope

ε = random error (Gaussian)

Model Comparison - Math score

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.500	0.046	0.048	1.000	
LSD dose	0.500	0.954	20.852	20.852	0.003

Bayesian approach to simple linear regression

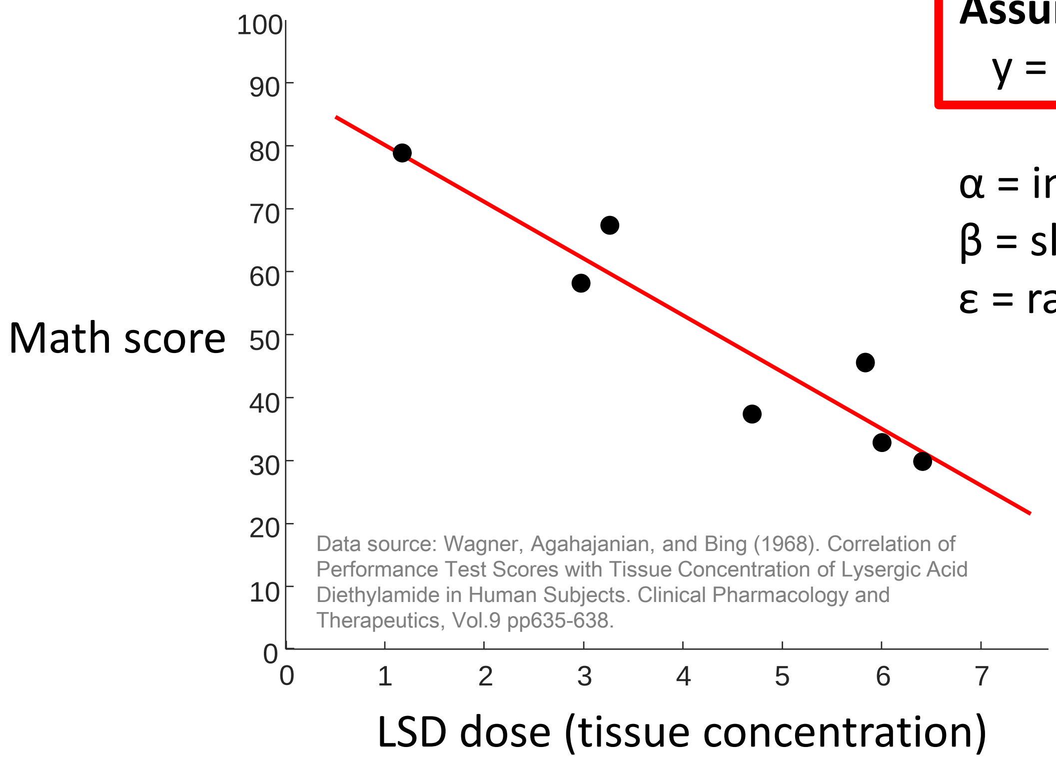
Assumed model

$y = \alpha + \beta x + \varepsilon$

α = intercept

β = slope

ε = random error (Gaussian)

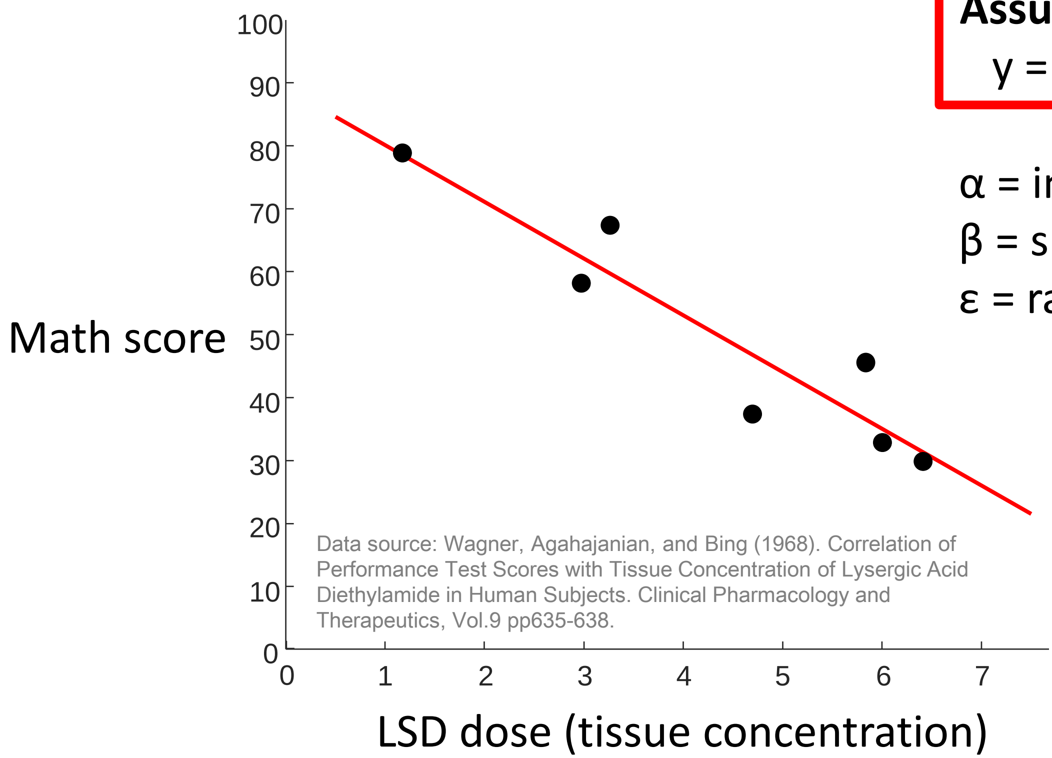


Model Comparison - Math score

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.500	0.046	0.048	1.000	
LSD dose	0.500	0.954	20.852	20.852	0.003

Prior model evidence

Bayesian approach to simple linear regression



Assumed model
$y = \alpha + \beta x + \varepsilon$

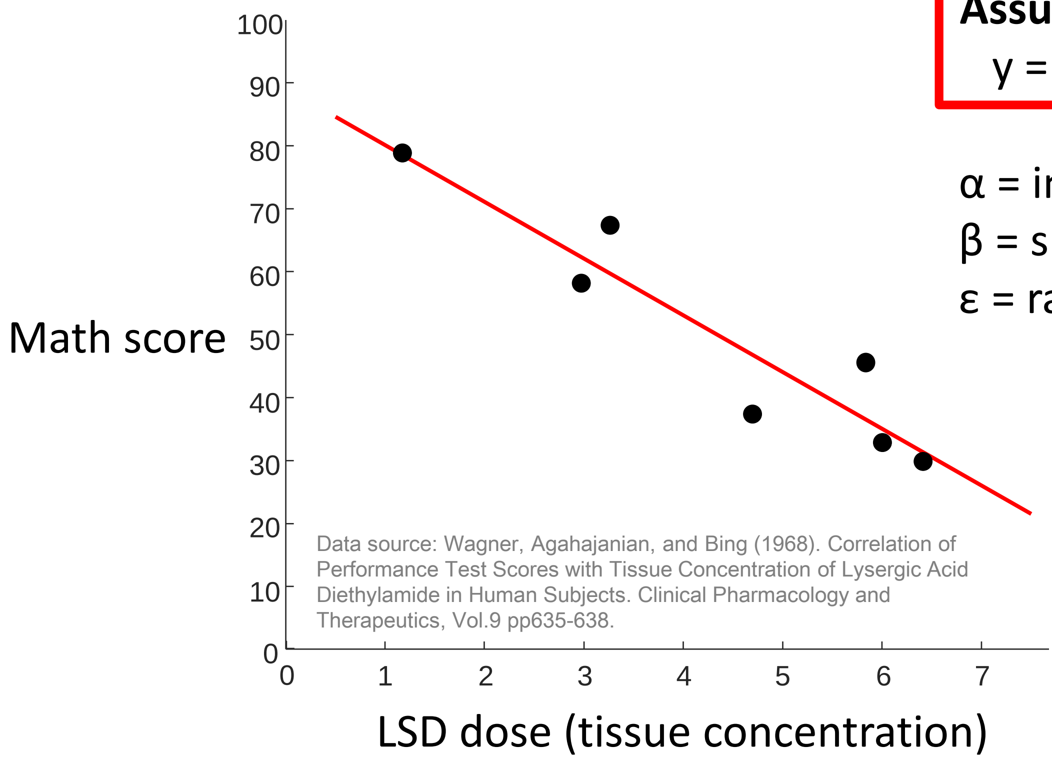
α = intercept
β = slope
ε = random error (Gaussian)

Model Comparison - Math score

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.500	0.046	0.048	1.000	
LSD dose	0.500	0.954	20.852	20.852	0.003

Posterior model evidence

Bayesian approach to simple linear regression



Assumed model

$y = \alpha + \beta x + \varepsilon$

α = intercept

β = slope

ε = random error (Gaussian)

Model Comparison - Math score

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.500	0.046	0.048	1.000	
LSD dose	0.500	0.954	20.852	20.852	0.003

Change from prior to posterior odds

(=Bayes factor of model Mx relative to **all** others)

Bayesian approach to simple linear regression

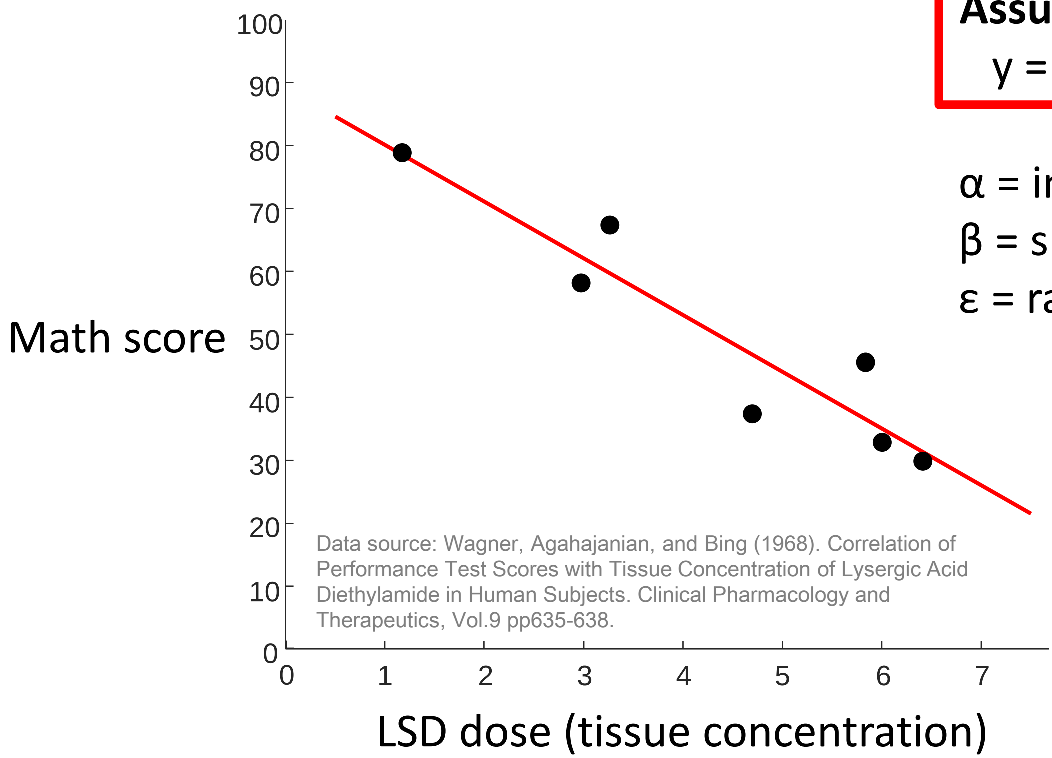
Assumed model

$y = \alpha + \beta x + \varepsilon$

α = intercept

β = slope

ε = random error (Gaussian)

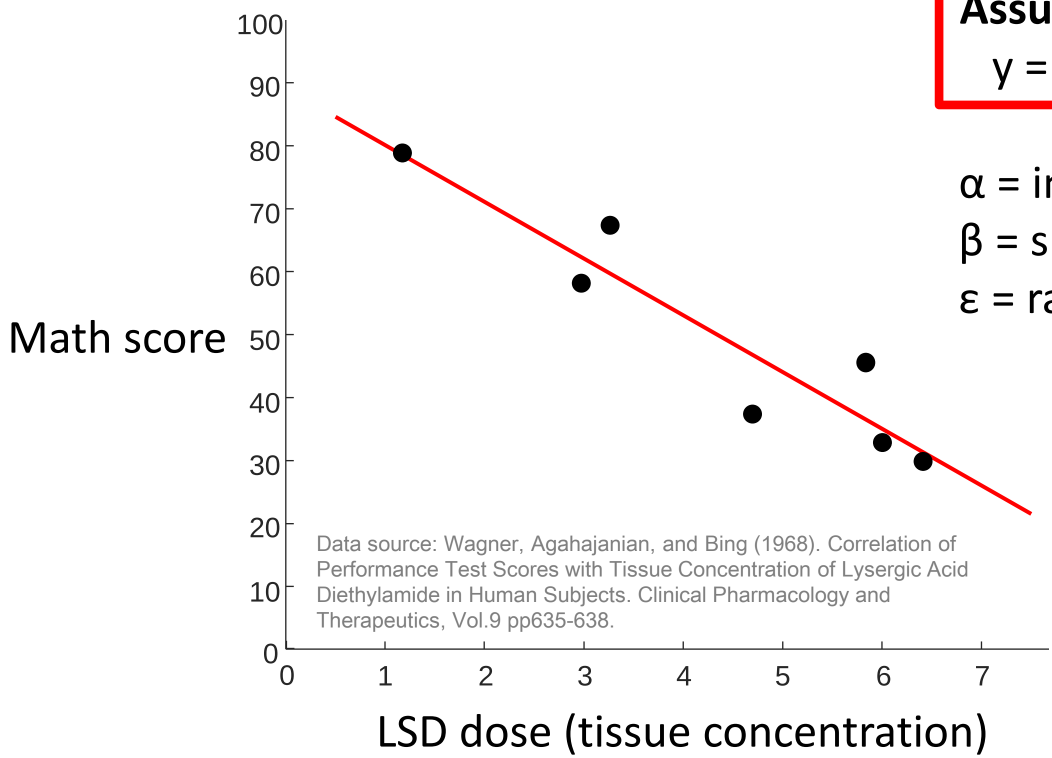


Model Comparison - Math score

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.500	0.046	0.048	1.000	
LSD dose	0.500	0.954	20.852	20.852	0.003

Bayes factor of Mx relative to M0

Bayesian approach to simple linear regression



Assumed model

$y = \alpha + \beta x + \varepsilon$

α = intercept
β = slope
ε = random error (Gaussian)

Model Comparison - Math score

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.500	0.046	0.048	1.000	
LSD dose	0.500	0.954	20.852	20.852	0.003

BF estimation error

Example with multiple regressors (aka covariates)

Example with multiple regressors

Data (Source: R. Higgs (1971). "Race, Skills, and Earnings: American Immigrants in 1909", The Journal of Economic History)

A	B	C	D	E
Origin	**Avg weekly wage ($)**	**English speaking (%)**	**Literate (%)**	**>5 years in US (%)**
Armenian	9.73	54.9	92.1	54.6
Bohemian/Moravian	13.07	66.0	96.8	71.2
Bulgarian	10.31	20.3	78.2	8.5
Canadian (French)	10.62	79.4	84.1	86.7
Canadian (Other)	14.15	100.0	99.0	90.8
Croation	11.37	50.9	70.7	38.9
Danish	14.32	96.5	99.2	85.4
Dutch	12.04	86.1	97.9	81.9
English	14.13	100.0	98.9	80.6
Finnish	13.27	50.3	99.1	53.6
Flemish	11.07	45.6	92.1	32.9
French	12.92	68.6	94.3	70.1
German	13.63	87.5	98.0	86.4
Greek	8.41	33.5	84.2	18.0
Hebrew (Russian)	12.71	74.7	93.3	57.1
Hebrew (Other)	14.37	79.5	92.8	73.8
Irish	13.01	100.0	96.0	90.6
Italian (Northern)	11.28	58.8	85.0	55.2
Italian (Southern)	9.61	48.7	69.3	47.8

	Dependent variable	Covariate #1	Covariate #2	Covariate #3

Assumed model: $y = \alpha + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \varepsilon$

Example with multiple regressors

Dependent variable: average weekly salary

Covariates: (1) english speaking (%), (2) literate (%), (3) >5 years in US (%)

FREQUENTIST RESULT

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
1	intercept	2.576	1.312		1.964	0.059
	English speaking (%)	0.041	0.024	0.484	1.733	0.093
	Literate (%)	0.079	0.020	0.497	3.930	< .001
	>5 years in US (%)	−0.003	0.021	−0.037	−0.149	0.882

Example with multiple regressors

Dependent variable: average weekly salary

Covariates: (1) english speaking (%), (2) literate (%), (3) >5 years in US (%)

FREQUENTIST RESULT

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
1	intercept	2.576	1.312		1.964	0.059
	English speaking (%)	0.041	0.024	0.484	1.733	0.093
	Literate (%)	0.079	0.020	0.497	3.930	< .001
	>5 years in US (%)	−0.003	0.021	−0.037	−0.149	0.882

BAYESIAN RESULT

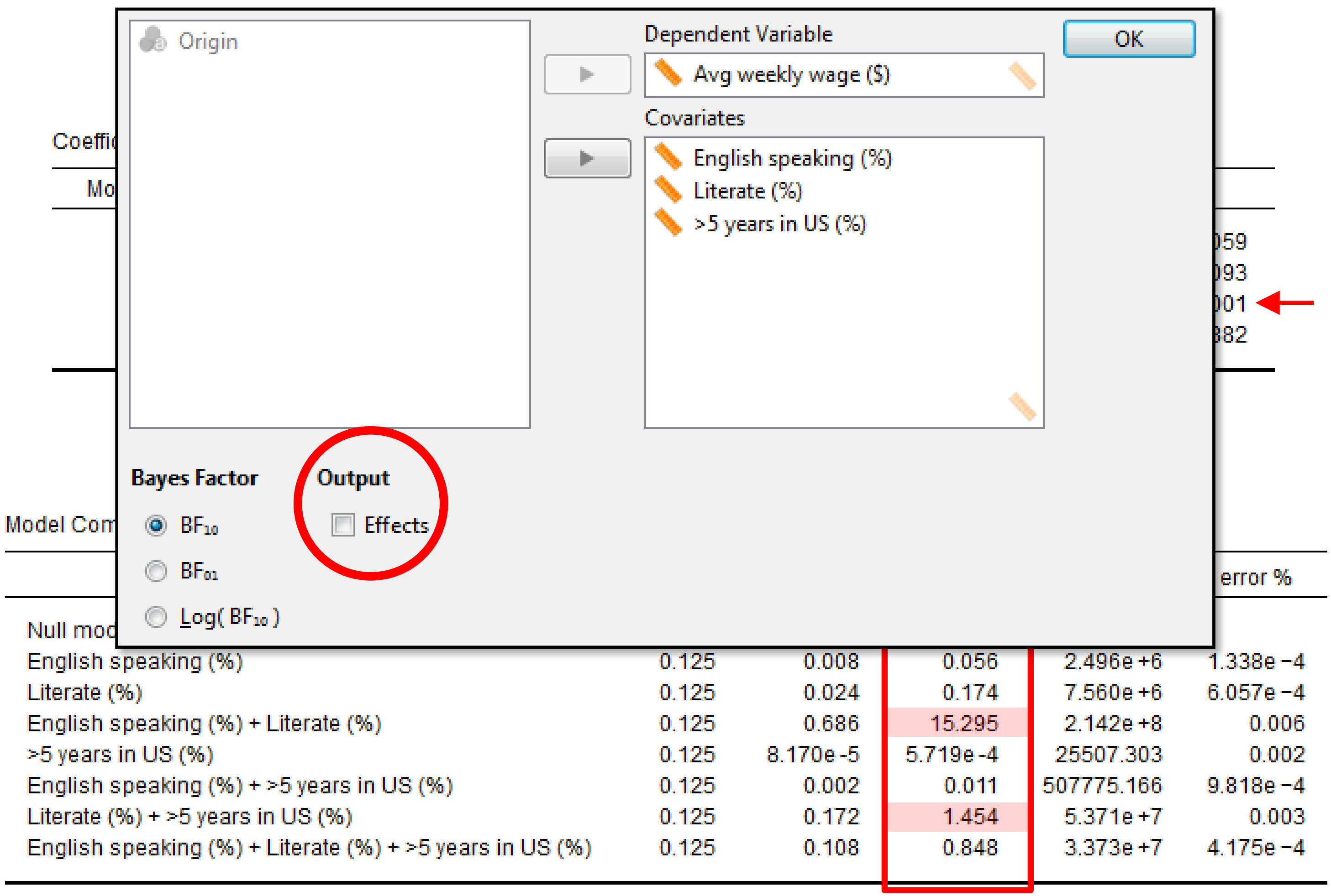
Model Comparison - Avg weekly wage ($)

Models	P(M)	P(M|data)	BF_M	BF_{10}	error %
Null model	0.125	3.203e-9	2.242e-8	1.000	
English speaking (%)	0.125	0.008	0.056	2.496e+6	1.338e−4
Literate (%)	0.125	0.024	0.174	7.560e+6	6.057e−4
English speaking (%) + Literate (%)	0.125	0.686	15.295	2.142e+8	0.006
>5 years in US (%)	0.125	8.170e-5	5.719e-4	25507.303	0.002
English speaking (%) + >5 years in US (%)	0.125	0.002	0.011	507775.166	9.818e−4
Literate (%) + >5 years in US (%)	0.125	0.172	1.454	5.371e+7	0.003
English speaking (%) + Literate (%) + >5 years in US (%)	0.125	0.108	0.848	3.373e+7	4.175e−4

Example with multiple regressors

Dependent variable: average weekly salary

Covariates: (1) english speaking (%), (2) literate (%), (3) >5 years in US (%)



Example with multiple regressors

Dependent variable: average weekly salary

Covariates: (1) english speaking (%), (2) literate (%), (3) >5 years in US (%)

FREQUENTIST RESULT

Coefficients

Model		Unstandardized	Standard Error	Standardized	t	p
1	intercept	2.576	1.312		1.964	0.059
	English speaking (%)	0.041	0.024	0.484	1.733	0.093
	Literate (%)	0.079	0.020	0.497	3.930	< .001
	>5 years in US (%)	−0.003	0.021	−0.037	−0.149	0.882

BAYESIAN RESULT

Analysis of Effects - Avg weekly wage ($)

Effects	P(incl)	P(incl|data)	$BF_{Inclusion}$
English speaking (%)	0.500	0.804	4.094
Literate (%)	0.500	0.990	102.066
>5 years in US (%)	0.500	0.282	0.392

Take-home points

#1

‘NHST’ is a widespread but flawed approach

(*) NHST=Null Hypothesis Significance Testing

Take-home points

#2

Evidence is best treated as a relative concept

- ❑ The Bayes Factor is by definition a relative measure
- ❑ The p-value is an absolute measure

Take-home points

#3

Ideally we want to be able to both reject and accept hypotheses

- The Bayes Factor can quantify evidence in both directions
- The p-value can only reject
- Disregard of “null results” is a main driver behind the replication crisis

Take-home points

#4

Ideally we want statistical evidence to be conditioned only on data

- The Bayes Factor has this property
- The p-value depends on data collection stopping rule!

Take-home points

#5

The Bayesian approach requires specifying priors

- Some see this as a curse
- Others see this as an opportunity to include prior knowledge

Take-home points

#6

Bayesians quantify belief, frequentists compute long-run frequencies

Take-home points

#7

Above all: make sure you know what you are doing!

Mindful Bayesian
>
Mindful frequentist
>>>>>>
Mindless Bayesian
>
Mindless Frequentist


RECOMMENDED
READING

ELSEVIER

The Journal of Socio-Economics 33 (2004) 587–606

The Journal of Socio-Economics

www.elsevier.com/locate/econbase

Mindless statistics

Gerd Gigerenzer*

Max Planck Institute for Human Development, Lentzeallee 94, 14195 Berlin, Germany

Abstract

Statistical rituals largely eliminate statistical thinking in the social sciences. Rituals are indispensable for identification with social groups, but they should be the subject rather than the procedure of science. What I call the "null ritual" consists of three steps: (1) set up a statistical null hypothesis, but do not specify your own hypothesis nor any alternative hypothesis, (2) use the 5% significance level for rejecting the null and accepting your hypothesis, and (3) always perform this procedure. I report evidence of the resulting collective confusion and fears about sanctions on the part of students and teachers, researchers and editors, as well as textbook writers.

Keywords: Rituals; Collective illusions; Statistical significance; Editors; Textbooks

> . . . no scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his evidence and his ideas.
>
> Sir Ronald A. Fisher (1956)

I once visited a distinguished statistical textbook author, whose book went through many editions, and whose name does not matter. His textbook represents the relative best in the social sciences. He was not a statistician; otherwise, his text would likely not have been used in a p ... Bayesian statistics, a ... ment in statistical t ... tion the existence o ... unheard

* Tel.: +49 ...

E-mail a...

1053-5357/$ ...

doi:10.1016/j...

Special Issue:
Bayesian Probability and Statistics
in Management Research

Journal of Management
Vol. 41 No. 2, February 2015 421–440
DOI: 10.1177/0149206314547522

Editorial Commentary

Surrogate Science: The Idol of a Universal Method for Scientific Inference

Gerd Gigerenzer
Max Planck Institute for Human Development
Julian N. Marewski
University of Lausanne

The application of statistics to science is not a neutral act. Statistical tools have shaped and were also shaped by its objects. In the social sciences, statistical methods fundamentally changed research practice, making statistical inference its centerpiece. At the same time, textbook writers in the social sciences have transformed rivaling statistical systems into an apparently monolithic method that could be used mechanically. The idol of a universal method for scientific inference has been worshipped since the "inference revolution" of the 1950s. Because no such method has ever been found, surrogates have been created, most notably the quest for significant p *values. This form of surrogate science fosters delusions and borderline cheating and has done much harm, creating, for one, a flood of irreproducible results. Proponents of the "Bayesian revolution" should be wary of chasing yet another chimera: an apparently universal inference procedure. A better path would be to promote both an understanding of the various devices in the "statistical toolbox" and informed judgment to select among these.*

Keywords: *research methods; regression analysis; psychometrics; Bayesian methods*

No scientific worker has a fixed level of significance at which from year to year, and in all circumstances, he rejects hypotheses; he rather gives his mind to each particular case in the light of his e

956: 42)

Acknowledg *do (2011)*
case of mind *values in*
the Academ *Dimov for*
helping with

Correspondi *Adaptive*
Behavior an

E-mail: gige

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]

[1] University of Amsterdam
[2] Masaryk University
[3] University of Missouri
[4] Cardiff University

Correspondence concerning this article should be addressed to:
Eric-Jan Wagenmakers
University of Amsterdam, Department of Psychological Methods
Weesperplein 4, 1018 XA Amsterdam, The Netherlands
E-Mail should be sent to EJ.Wagenmakers@gmail.com.

Abstract

Bayesian parameter estimation and Bayesian hypothesis testing present attractive alternatives to classical inference using confidence intervals and p values. In part I of this two-part series we outline ten prominent advantages of the Bayesian approach. Many of these advantages translate to concrete opportunities for pragmatic researchers. For instance, Bayesian hypothesis testing allows researchers to quantify evidence and monitor its progression as data come in, without needing to know the intention with which the data were collected. We end by countering several objections to Bayesian hypothesis testing. Part II of this series discusses JASP, a free and open source software program that makes it easy to conduct Bayesian estimation and testing for a range of popular statistical scenarios (Love et al., this issue).

Keywor ... or
distributi

Theoretica ... *f coherent statistics.* Dennis

Bayesian Inference for Psychology. Part II: Example Applications with JASP

Eric-Jan Wagenmakers[1], Jonathon Love[1], Maarten Marsman[1], Tahira Jamil[1], Alexander Ly[1], Josine Verhagen[1], Ravi Selker[1], Quentin F. Gronau[1], Damian Dropmann[1], Bruno Boutin[1], Frans Meerhoff[1], Patrick Knight[1], Akash Raj[2], Erik-Jan van Kesteren[1], Johnny van Doorn[1], Martin Šmíra[3], Sacha Epskamp[1], Alexander Etz[4], Dora Matzke[1], Jeffrey N. Rouder[5], Richard D. Morey[6]
[1] University of Amsterdam
[2] Birla Institute of Technology and Science
[3] Masaryk University
[4] University of California at Irvine
[5] University of Missouri
[6] Cardiff University

Correspondence concerning this article should be addressed to:
Eric-Jan Wagenmakers
University of Amsterdam, Department of Psychological Methods
Nieuwe Achtergracht 129-B, 1018 VZ Amsterdam, The Netherlands
E-Mail should be sent to EJ.Wagenmakers@gmail.com.

Abstract

Bayesian hypothesis testing presents an attractive alternative to p value hypothesis testing. Part I of this series outlined several advantages of Bayesian hypothesis testing, including the ability to quantify evidence and the ability to monitor and update this evidence as data come in, without the need to know the intention with which the data were collected. Despite these and other practical advantages, Bayesian hypothesis tests are still reported relatively rarely. An important impediment to the widespread adoption of Bayesian tests is arguably the lack of user-friendly software for the run-of-the-mill statistical problems that confront psychologists for the analysis of almost every experiment: the t-test, ANOVA, correlation, regression, and contingency tables. In Part II of this series we introduce JASP
(ja ... ical
sof ... ests
for ... sian
an ... R.
Ar ... ing
are

Ke ... rior
dis

Some extra slides

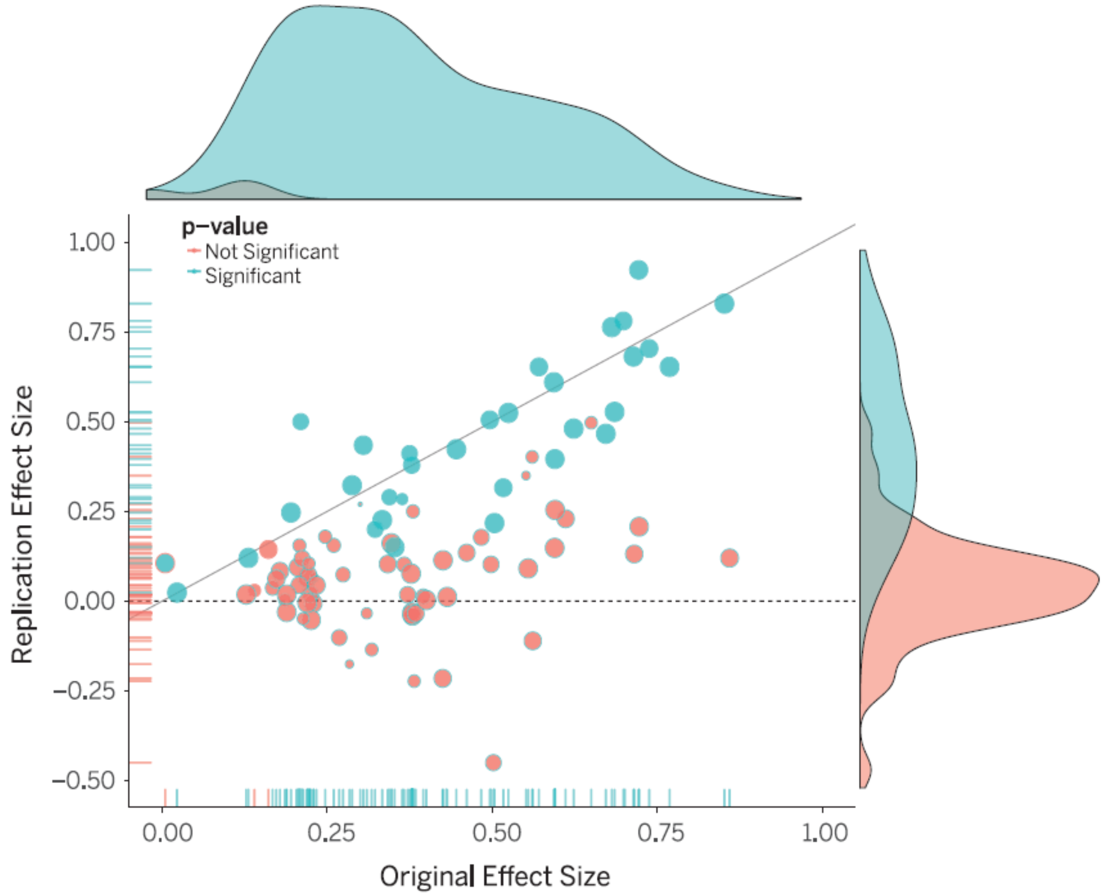
Fisher vs Neyman-Pearson

Fisher's approach	**Neyman-Pearson's approach**
Outcome: significant / non-significant	Outcome: accept / reject
p is a measure of evidence against H0	*p* is NOT a measure of evidence and should not be interpreted
An alternative hypothesis **cannot** be specified	An alternative hypothesis **must** be specified
Does not have a concept of "power"	Power has to be specified prior to the experiment
A single rejection of H0 is the start, not the end, of an investigation. Replication needed and meta-analyses are useful	A single rejection is meaningless – the framework only guarantees long-term type-1 and type-2 error rates but does not allow to make inference about a single case.

Presently, much statistical testing in psychology research is an "*inconsistent hybrid that every decent statistician would reject*" (Gigerenzer, 2004)

Why should we bother about statistical literacy?

Open Science Collaboration (2015), *Estimating the reproducibility of psychological science.* Science, *349*(6251)



Main findings

1) Only 36% of significant results replicated

2) Effect sizes shrunk by ~50% in the replications

What caused the crisis?

A **toxic mix** of the following:

- Publication pressure
- Disregard for “null findings”

… which incentivizes **poor methodological hygiene**:

- Hide null findings (file drawer problem)
- Test many variables, report few (fishing)
- Try many tests, report few (p-hacking)
- Post-hoc hypothesizing (HARK-ing)
- …

Bayesian stats is not a miracle cure, but understanding the Bayesian approach will make you a more insightful consumer of statistics – which will likely lead to better statistical practices even if you stick to the frequentist methods.