Making the most of your ANOVAs: From NHST to Bayesian analyses

Presentation at
Bayes @ Lund
22 April 2017

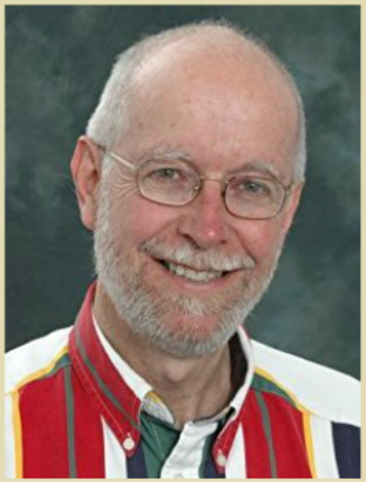
Stefan Wiens, PhD
Professor
Psychology
Stockholm University

Making the most of your ANOVAs: From NHST to Bayesian analyses

Presentation at
Bayes @ Lund
20 April 2017

Stefan Wiens, PhD
Professor
Psychology
Stockholm University

The "New Statistics"



The New Statistics: Why and How

Psychological Science
2014, Vol. 25(1) 7–29

DOI: 10.1177/0956797613504966
pss.sagepub.com
SAGE

Geoff Cumming
La Trobe University

Abstract
We need to make substantial changes to how we conduct research. First, in response to heightened concern that our published research literature is incomplete and untrustworthy, we need new requirements to ensure research integrity. These include prespecification of studies whenever possible, avoidance of selection and other inappropriate data-analytic practices, complete reporting, and encouragement of replication. Second, in response to renewed recognition of the severe flaws of null-hypothesis significance testing (NHST), we need to shift from reliance on NHST to estimation and other preferred techniques. *The new statistics* refers to recommended practices, including estimation based on effect sizes, confidence intervals, and meta-analysis. The techniques are not new, but adopting them widely would be new for many researchers, as well as highly beneficial. This article explains why the new statistics are important and offers guidance for their use. It describes an eight-step new-statistics strategy for research with integrity, which starts with formulation of research questions in estimation terms, has no place for NHST, and is aimed at building a cumulative quantitative discipline.

Focus on effect size

Focus on 95% confidence intervals

What about factorial designs?

Factorial designs

	2 wks before	1 wk before	1 wk after	2 wks after
Waiting Group				
CBT Group				
Meditation Group				

3 x 4
Therapy x Time

Factorial designs

	2 wks before	1 wk before	1 wk after	2 wks after
Waiting Group				
CBT Group				
Meditation Group				

Effects in ANOVA

Main effect of Therapy 2 dfs

Main effect of Time 3 dfs

Interaction (Therapy x Time) 2 x 3 = 6 dfs

non-specific

Effect size

(partial) eta-squared

informative?

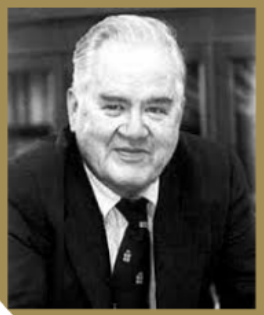
Effect size

Size of anything

384,400 km

23 g chocolate

Practical or clinical importance



Standarization? (Cohen's d, r)

"Being so disinterested in our variables that we do not care about their units can hardly be desirable" (Tukey, 1969, p. 89)

Contrast analysis

Specific

Informative

Unstandardized

effect size

Perform *t* test

8 kg

-1

13 kg

+1

-8 + 13 = 5 kg

Contrast analysis

Mean difference between sets of means

Before	After 6 wks	After 12 wks
8 kg	13 kg	15 kg
-1	+0.5	+0.5

-8 + 6.5 + 7.5 = 6 kg

Contrast score is effect size

Contrast analysis

A Observed raw means

Alcohol \ Antabuse	yes	no
yes	16.38 a	4.20 b
no	4.22 c	4.10 d

E Weights for Synergistic

Alcohol \ Antabuse	yes	no
yes	+1	-1/3
no	-1/3	-1/3

B Weights for Column effect

Alcohol \ Antabuse	yes	no
yes	+0.5	-0.5
no	+0.5	-0.5

C Weights for Row effect

Alcohol \ Antabuse	yes	no
yes	+0.5	+0.5
no	-0.5	-0.5

D Weights for Interaction

Alcohol \ Antabuse	yes	no
yes	+0.5 a	-0.5 b
no	-0.5 c	+0.5 d

4.22 c	4.10 d

A

no	-1/3

B **Weights for Column effect**

Alcohol \ Antabuse	yes	no
yes	+0.5	-0.5
no	+0.5	-0.5

C **Weights for Row effect**

Alcohol \ Antabuse	yes	no
yes	+0.5	+0.5
no	-0.5	-0.5

D **Weights for Interaction**

Alcohol \ Antabuse	yes	no
yes	+0.5 a	-0.5 b
no	-0.5 c	+0.5 d

Contrast analysis

A Observed raw means

Alcohol \ Antabuse	yes	no
yes	16.38 a	4.20 b
no	4.22 c	4.10 d

E Weights for Synergistic

Alcohol \ Antabuse	yes	no
yes	+1	-1/3
no	-1/3	-1/3

B Weights for Column effect

Alcohol \ Antabuse	yes	no
yes	+0.5	-0.5
no	+0.5	-0.5

C Weights for Row effect

Alcohol \ Antabuse	yes	no
yes	+0.5	+0.5
no	-0.5	-0.5

D Weights for Interaction

Alcohol \ Antabuse	yes	no
yes	+0.5 a	-0.5 b
no	-0.5 c	+0.5 d

Contrast analysis

Compare between treatments after vs. before

	2 wks before	1 wk before	1 wk after	2 wks after
Waiting Group	0	0	0	0
CBT Group	-0.5	-0.5	+0.5	+0.5
Meditation Group	+0.5	+0.5	-0.5	-0.5

Compare treatments with waiting, after vs. before

	2 wks before	1 wk before	1 wk after	2 wks after
Waiting Group	+0.50	+0.50	-0.50	-0.50
CBT Group	-0.25	-0.25	+0.25	+0.25
Meditation Group	-0.25	-0.25	+0.25	+0.25

What does NHST compare?

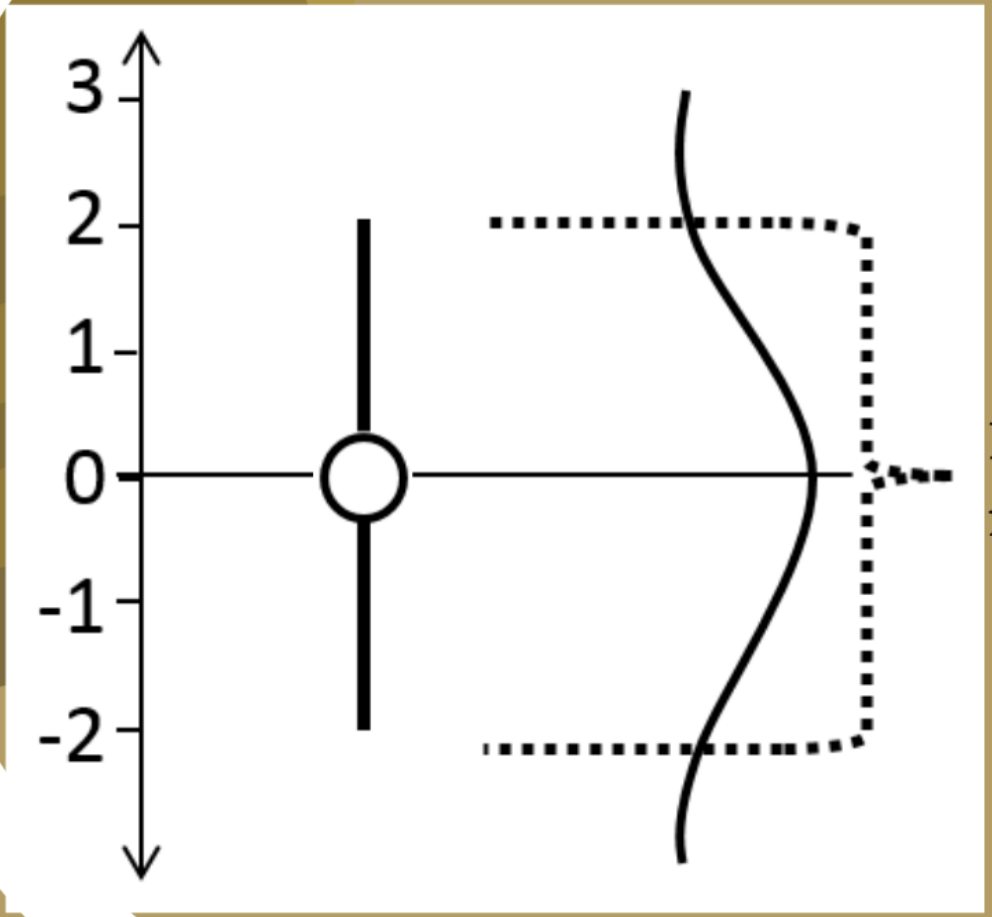
https://www.flickr.com/photos/23868780@N00/12559689854/

$p < .05$: Hypothesis myopia

What does NHST compare?

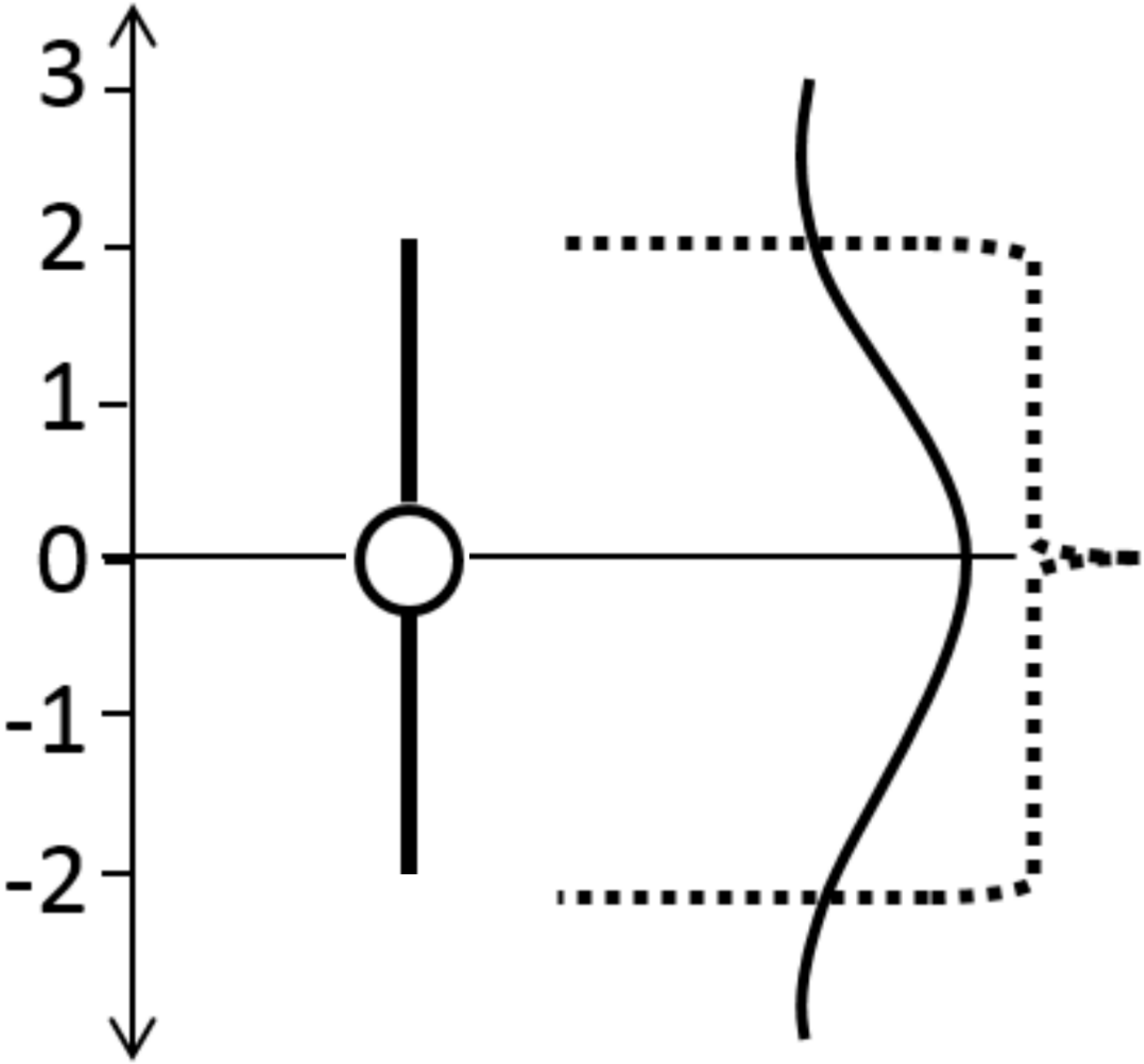
Beyond NHST

Confidence interval: 95% CI [LL, UL]



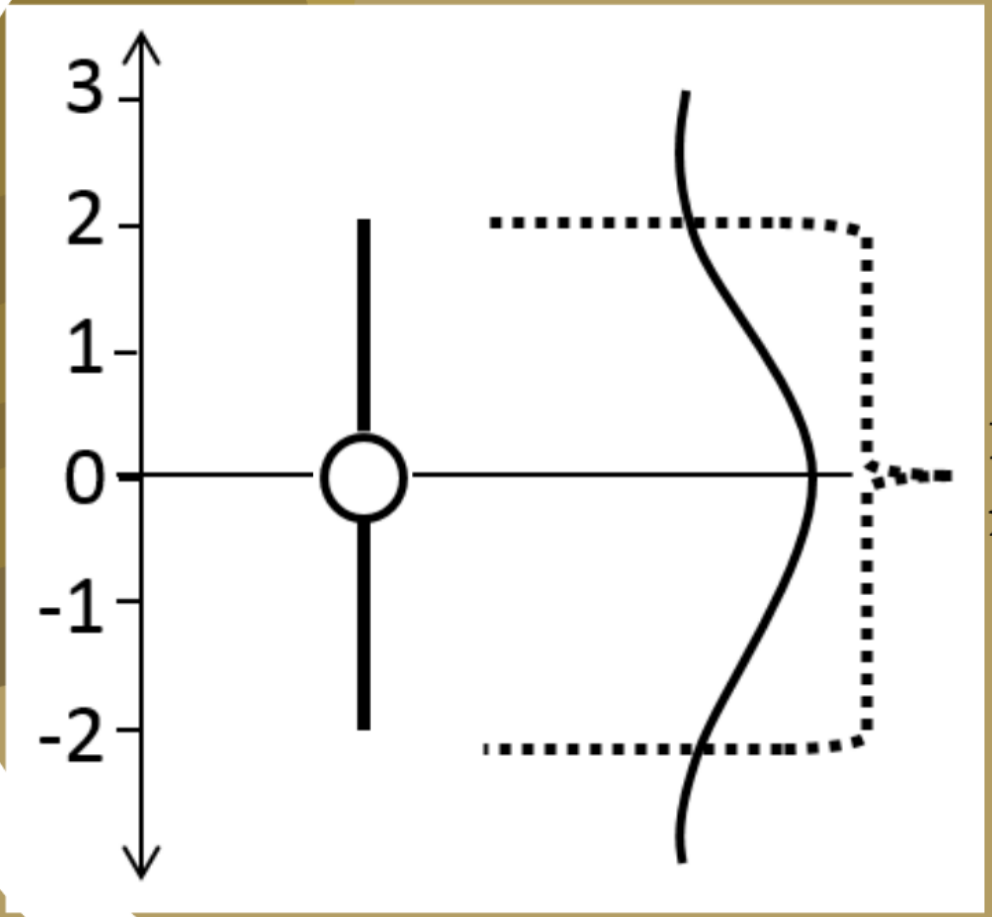
Beyond NHST

Confidence interval: 95% CI [LL, UI

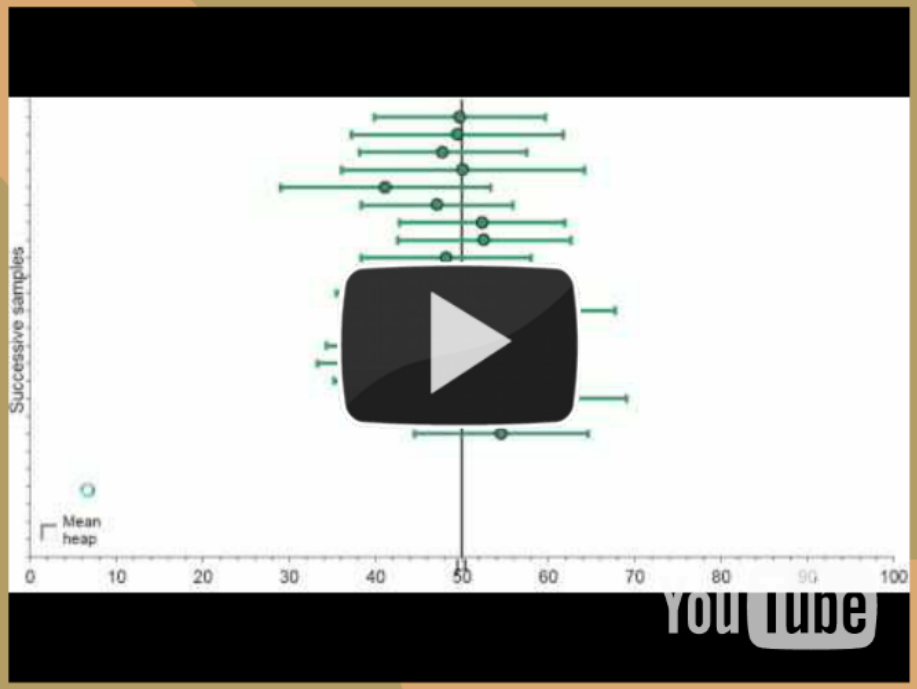


Beyond NHST

Confidence interval: 95% CI [LL, UL]



"Confidence" in Confidence Intervals



Cumming:

ESCI chapters 1-4 Jul 4 2011

http://rpsychologist.com/d3/CI/

Misinterpretation of Confidence Intervals

Psychon Bull Rev
DOI 10.3758/s13423-015-0947-8

THEORETICAL REVIEW

The fallacy of placing confidence in confidence intervals

Richard D. Morey[1] · Rink Hoekstra[2] · Jeffrey N. Rouder[3] · Michael D. Lee[4] · Eric-Jan Wagenmakers[5]

Do not provide likelihood and precision

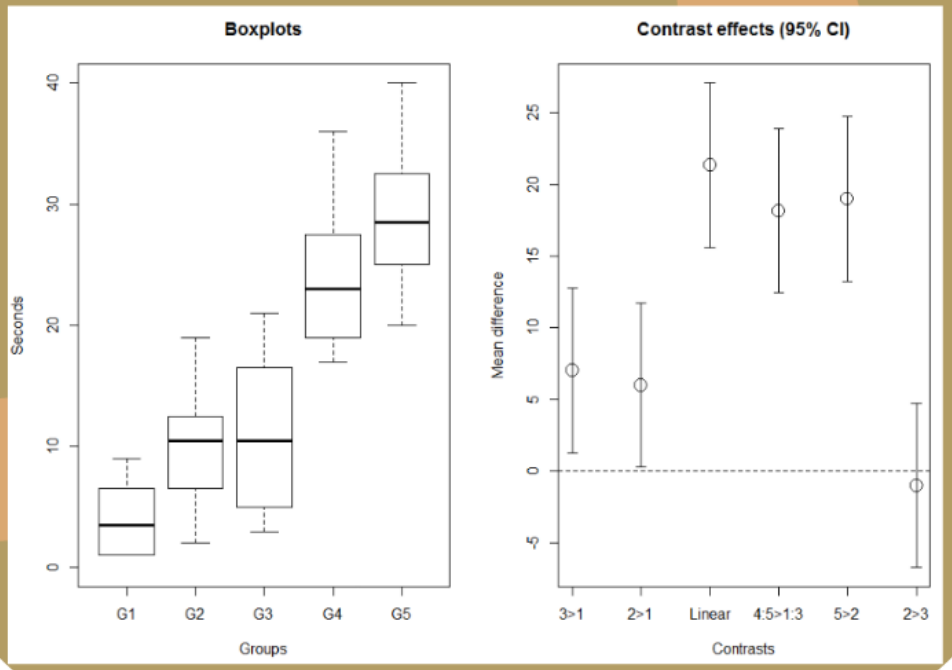
Use Credible intervals

R scripts

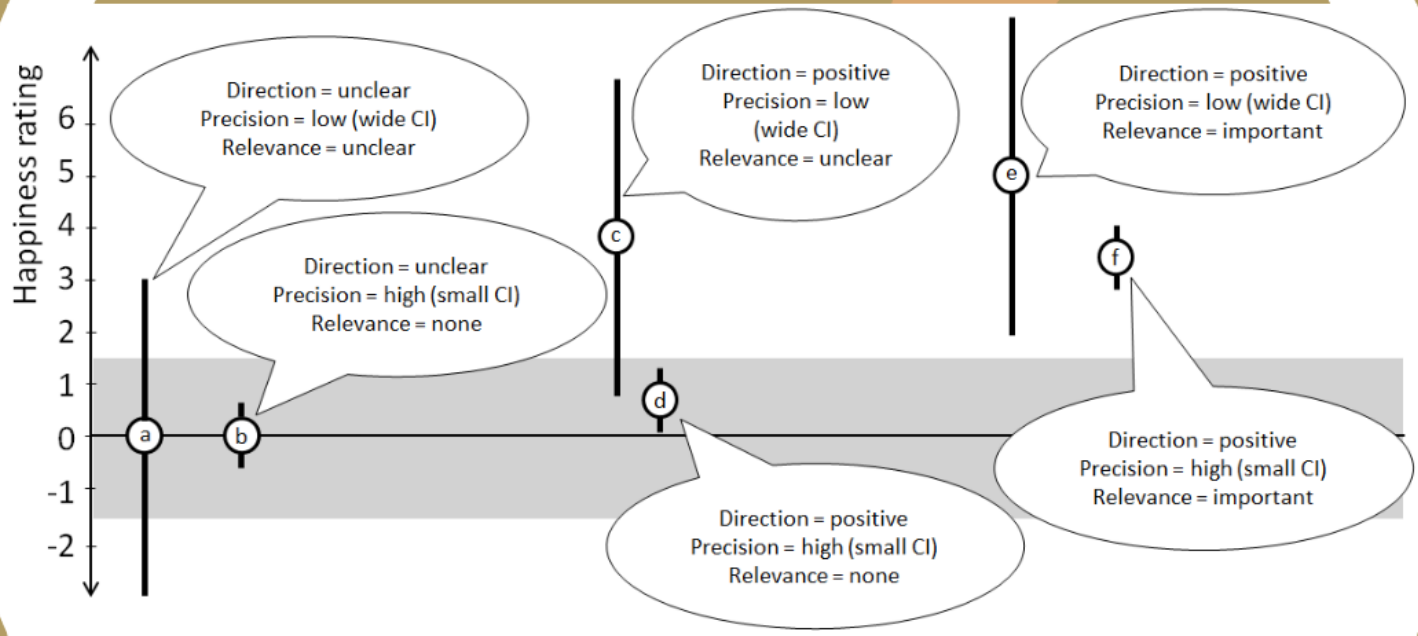
Wiens, S., & Nilsson, M. E. (2016). Performing contrast analysis in factorial designs: From NHST to confidence intervals and beyond. Educational and Psychological Measurement, in press. doi: 10.1177/0013164416668950

https://osf.io/p5b2r/

Contrast_name	w1	w2	w3	w4	w5	CI_LL	Contrast_	CI_UL	MOE	MS_error	df_error	MS_contrast	F_contrast	t_contrast	p
3>1	-1,000	0,000	1,000	0,000	0,000	1,258	7,000	12,742	5,742	32,000	35,000	196,000	6,125	2,475	0,018
2>1	-1,000	1,000	0,000	0,000	0,000	0,258	6,000	11,742	5,742	32,000	35,000	144,000	4,500	2,121	0,041
Linear	-0,667	-0,333	0,000	0,333	0,667	15,591	21,333	27,075	5,742	32,000	35,000	3276,800	102,400	10,119	0,000
4:5>1:3	-0,333	-0,333	-0,333	0,500	0,500	12,425	18,167	23,909	5,742	32,000	35,000	3168,267	99,008	9,950	0,000
5>2	0,000	-1,000	0,000	0,000	1,000	13,258	19,000	24,742	5,742	32,000	35,000	1444,000	45,125	6,718	0,000
2>3	0,000	1,000	-1,000	0,000	0,000	-6,742	-1,000	4,742	5,742	32,000	35,000	4,000	0,125	0,354	0,726



Informativeness of CI



Direction, Precision, Relevance

Hypothesis testing

$$\frac{P(H_1|D) = P(D|H_1)}{P(H_o|D) = P(D|H_o)} * \frac{P(H_1)}{P(H_o)}$$

Posterior confidence

Bayes Factor

Prior confidence in H_1 rather than H_o

A **Observed raw means**

Alcohol \ Antabuse	yes	no
yes	16.38 a	4.20 b
no	4.22 c	4.10 d

E **Weights for Synergistic**

Alcohol \ Antabuse	yes	no
yes	+1	-1/3
no	-1/3	-1/3

Observed raw means

Alcohol \ Antabuse	yes	no
yes		4.20
no	4.22	4.10

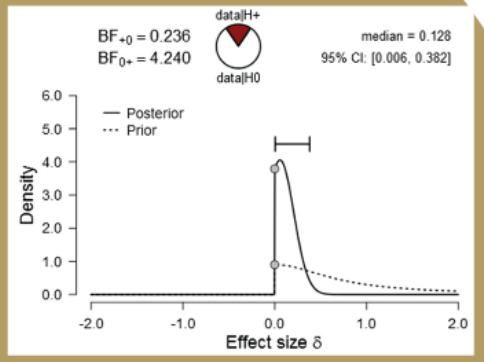
Bayes Factor

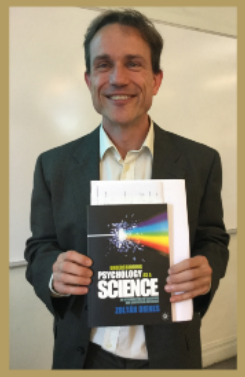
Eric-Jan Wagenmakers

Default prior

JASP

https://jasp-stats.org/





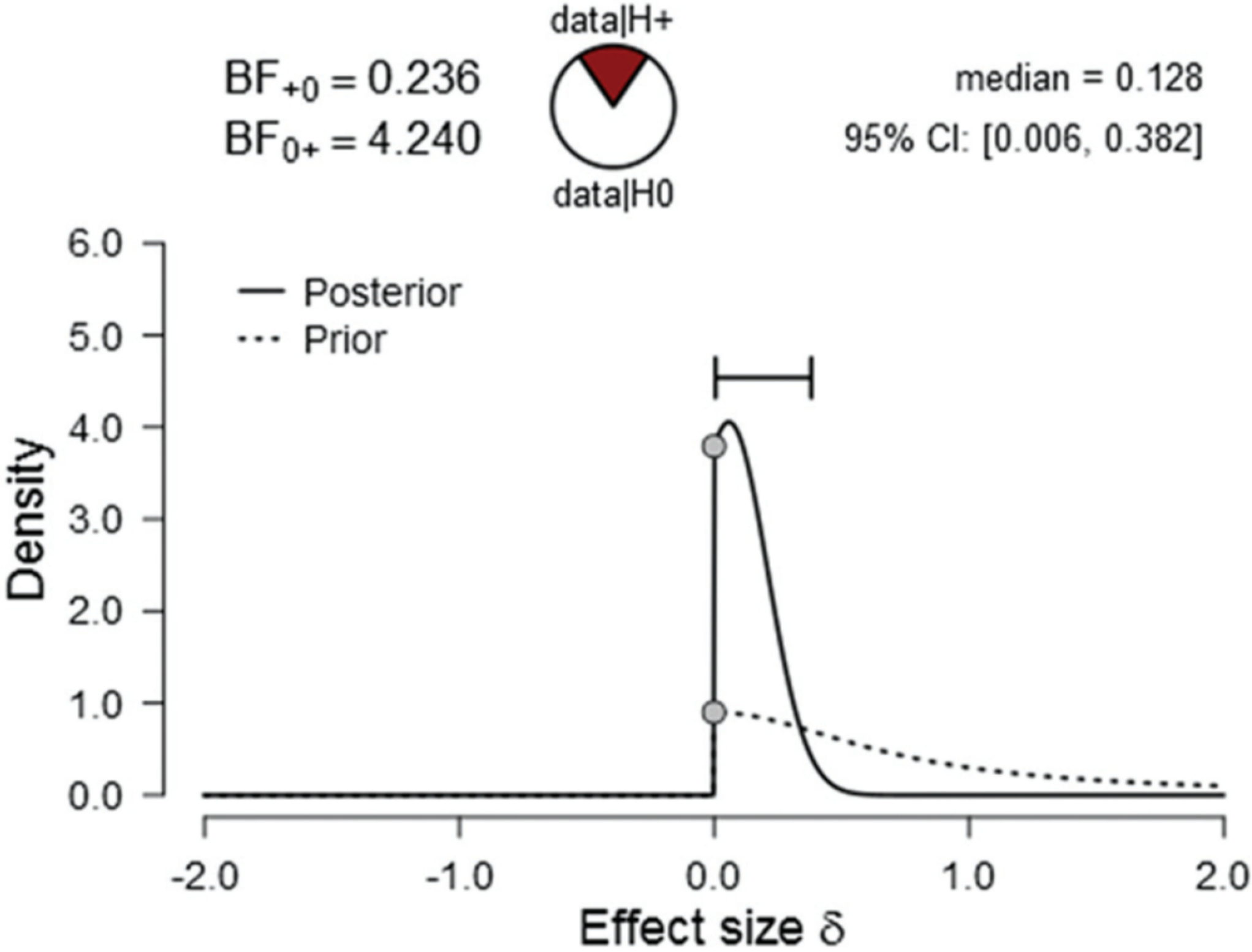
Zoltan Dienes

Choice of prior

Uniform
Half-normal
Normal

Online calculator

http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm


data|H+
BF+0 = 0.236
BF0+ = 4.240
median = 0.128
95% CI: [0.006, 0.382]
data|H0
Posterior
Prior
Density
6.0
5.0
4.0
3.0
2.0
1.0
0.0
-2.0
-1.0
0.0
1.0
2.0
Effect size δ

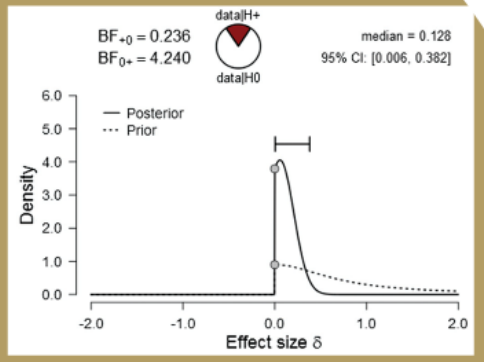
Bayes Factor

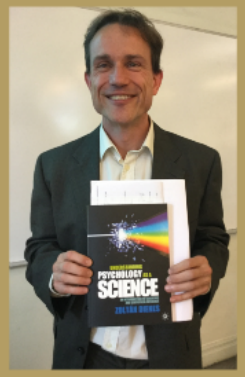
Eric-Jan Wagenmakers

Default prior

JASP

https://jasp-stats.org/





Zoltan Dienes

Choice of prior

Uniform
Half-normal
Normal

Online calculator

http://www.lifesci.sussex.ac.uk/home/Zoltan_Dienes/inference/Bayes.htm

Summary

Factorial designs

Contrast analysis (t test)
specific, informative
unstandardized effect size

Confidence intervals

Credible intervals

Hypothesis testing (p)

Bayes Factor

Making the most of your ANOVAs: From NHST to Bayesian analyses

Presentation at
Bayes @ Lund
20 April 2017

Stefan Wiens, PhD
Professor
Psychology
Stockholm University

Thank you!

Wiens, S., & Nilsson, M. E. (2016). Performing contrast analysis in factorial designs: From NHST to confidence intervals and beyond. Educational and Psychological Measurement.

https://osf.io/p5b2r/