



Beyond significance tests

Professor Andy Field

WHOA
 @profandyfield

 www.youtube.com/user/ProfAndyField/

 www.discoveringstatistics.com

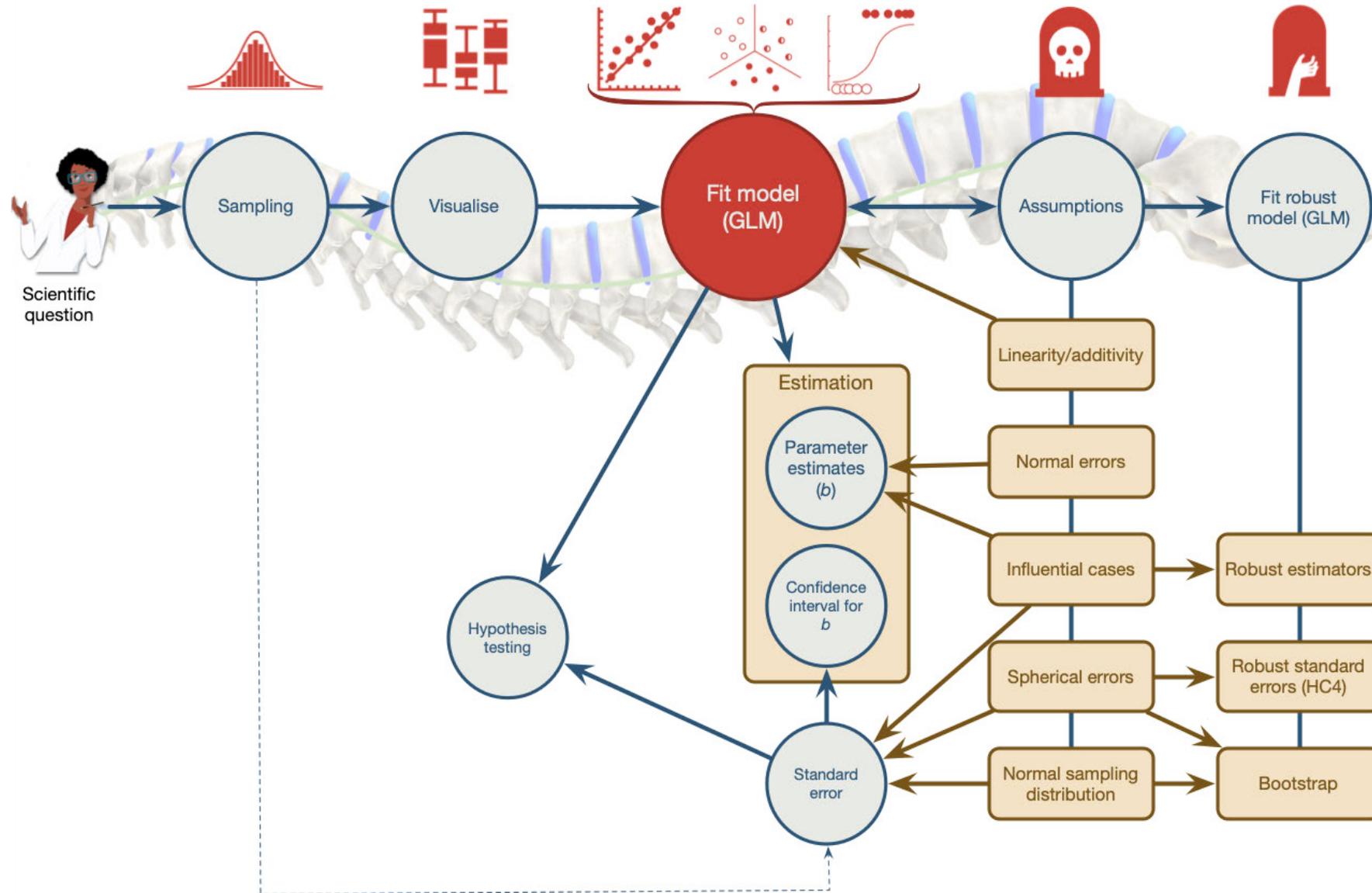
 www.milton-the-cat.rocks

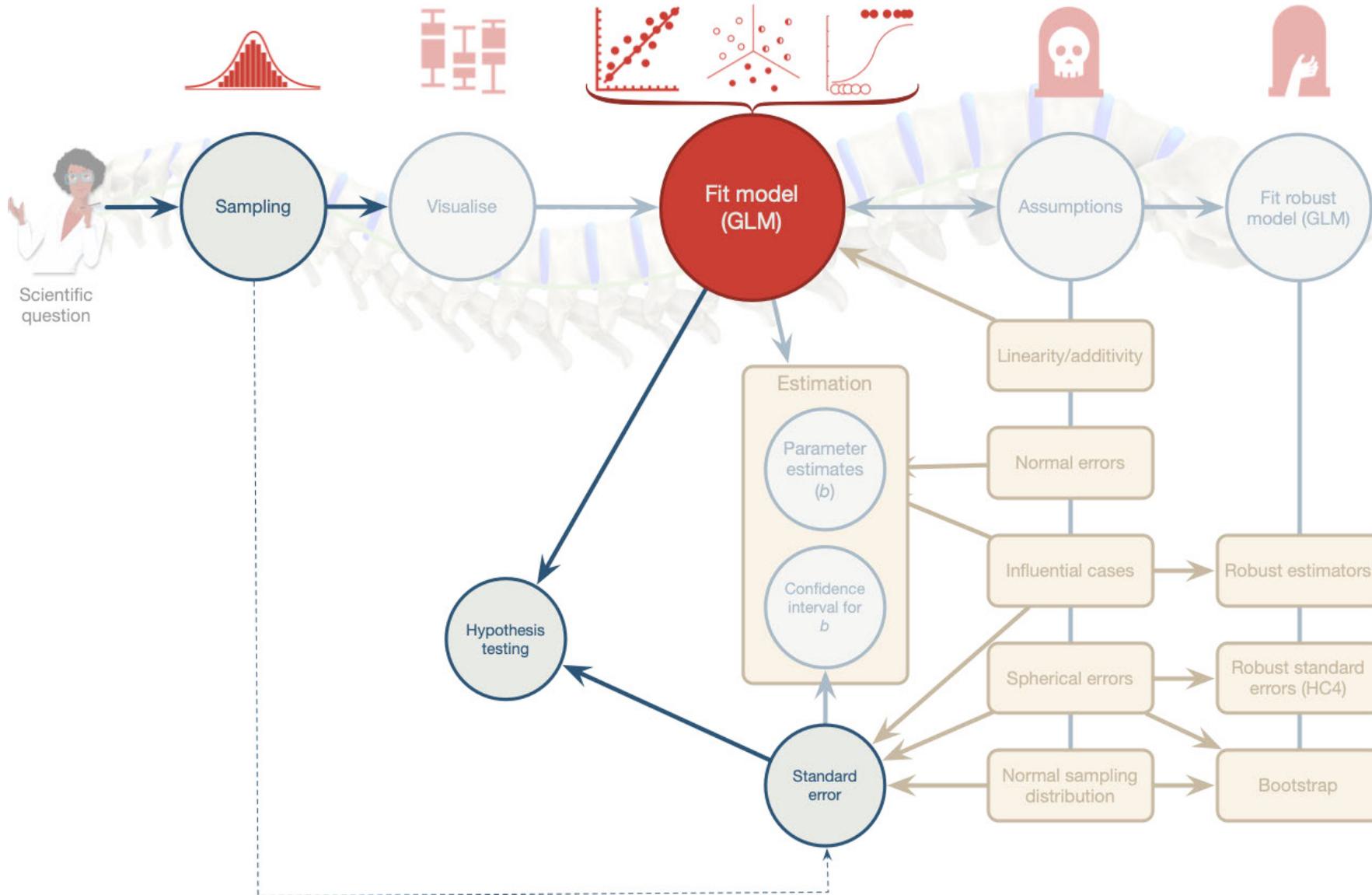
 www.discovr.rocks



ANDY FIELD







Learning outcomes

Effect sizes

- Understand what an effect size is and the relative merits of effect sizes

Bayes factors

- Articulate the principles of Bayesian approaches
- Define what a Bayes factor represents



ANDY FIELD



Problems with p - A recap

Tells us nothing about importance because p depends upon sample size.

Provides little evidence about the null (or alternative) hypothesis

Encourages all-or-nothing thinking

Based on long-run probabilities

- p is the frequency of the observed test statistic relative to all test statistics from an infinite number of identical experiments with the exact same a priori sample size.
- The type I error rate in a given study is either 0 or 1, but we don't know which.



ANDY FIELD



Effect sizes

Raw effect size, the parameter (b)

Standardized version of the parameter

- β
- r
 - 0.1 = small, 0.5 = large
 - Try to avoid these thresholds
- Cohen's d
 - Difference between means divided by standard deviation (pooled or control group)
 - $d = 1$, the means are 1 standard deviation apart, $d = 0$, the means are the same
 - 0.2 = small, 0.8 = large
 - Try to avoid these thresholds



Cohen's d

Control group sd

$$\hat{d} = \frac{\bar{X}_{\text{Exp}} - \bar{X}_{\text{Control}}}{s_{\text{Control}}}$$

Pooled sd

$$\hat{d} = \frac{\bar{X}_1 - \bar{X}_2}{s_p}$$

$$s_p = \sqrt{\frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}}$$



ANDY FIELD



Cohen's d

Summary statistics for Study 2

group	n	Mean	Standard deviation
Teddy	100	12.89	5.63
Book	100	7.89	4.83

Control group sd

$$\begin{aligned}\hat{d} &= \frac{\bar{X}_{\text{Exp}} - \bar{X}_{\text{Control}}}{s_{\text{Control}}} \\ &= \frac{12.89 - 7.89}{4.83} \\ &= 1.04\end{aligned}$$

Cohen's d

Summary statistics for Study 2

group	n	Mean	Standard deviation
Teddy	100	12.89	5.63
Book	100	7.89	4.83

Pooled sd

$$s_p = \sqrt{\frac{(100 - 1)5.63^2 + (100 - 1)4.83^2}{100 + 100 - 2}} = 5.25$$

$$\hat{d} = \frac{12.89 - 7.89}{5.25} = 0.95$$

Same effects, different p s

Study 1: $n = 20$, $d = 0.77$

term	estimate	std.error	statistic	p.value
(Intercept)	12.89	0.525	24.565	0
groupBook	-5.00	0.742	-6.738	0

Study 2: $n = 200$, $d = 0.95$

term	estimate	std.error	statistic	p.value
(Intercept)	12.8	2.054	6.233	0.000
groupBook	-5.0	2.904	-1.722	0.102

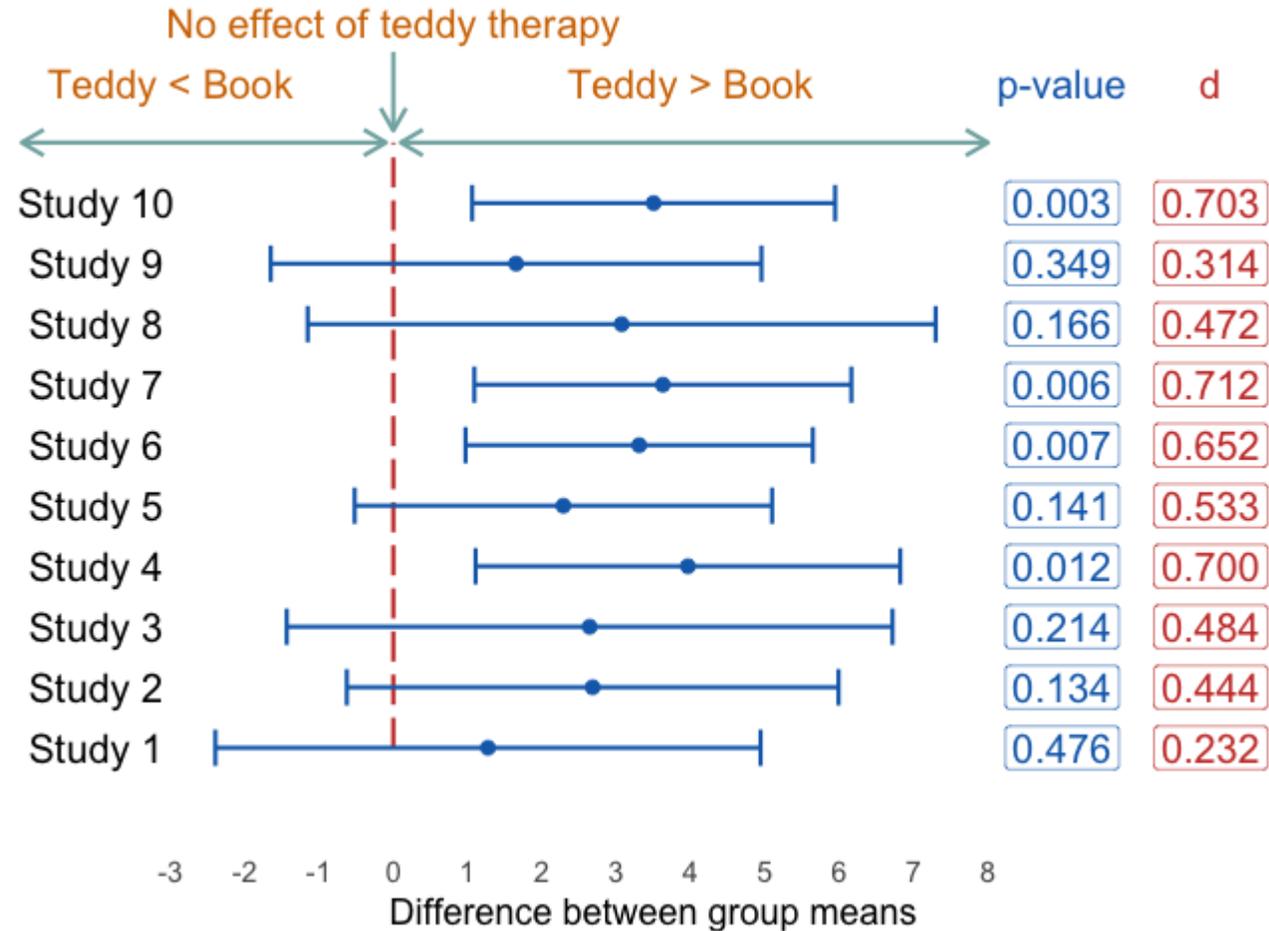
Zero effect (approx), significant p

Study 3: $n = 200,000$, $d = -0.01$

term	estimate	std.error	statistic	p.value
(Intercept)	12.113	0.018	660.082	0.000
groupBook	0.052	0.026	1.997	0.046

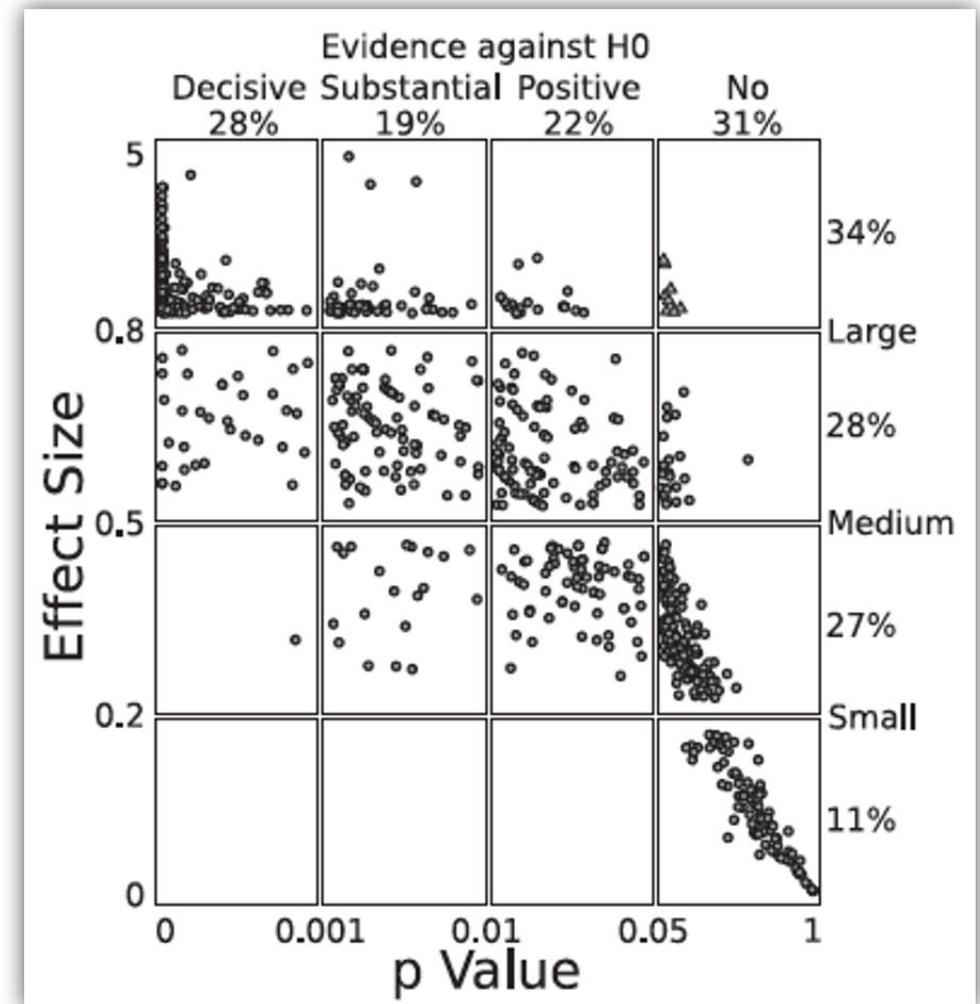


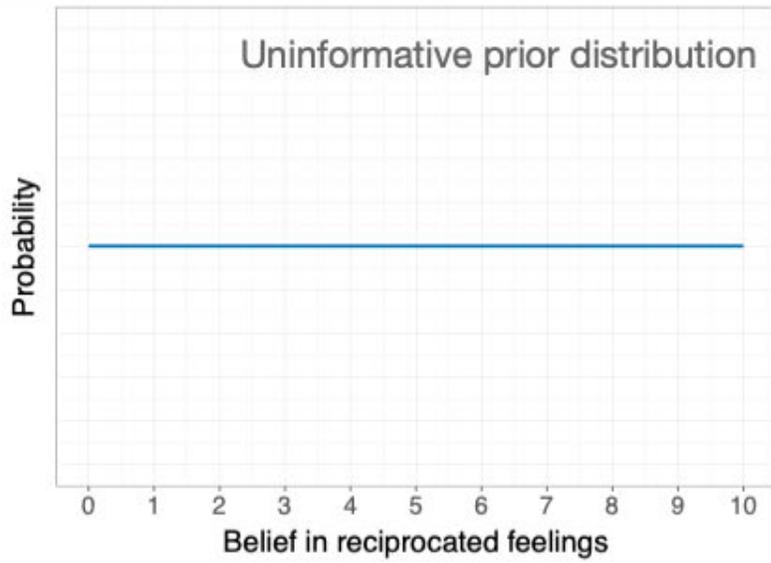
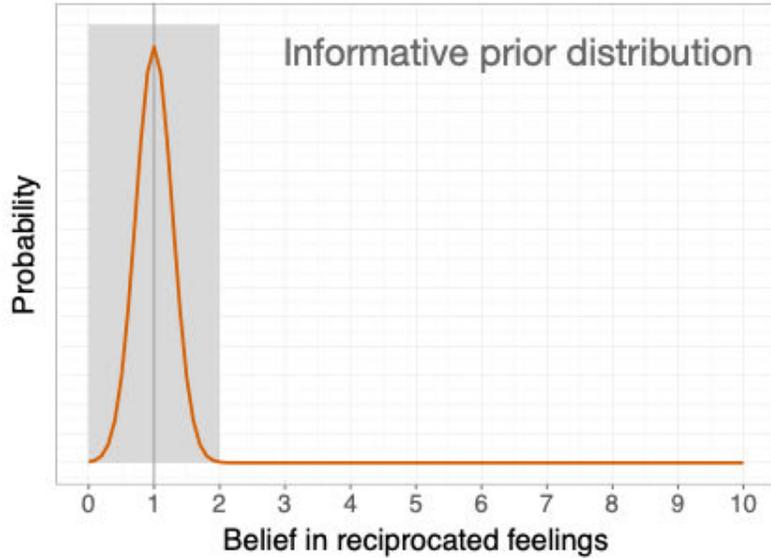
All or nothing thinking



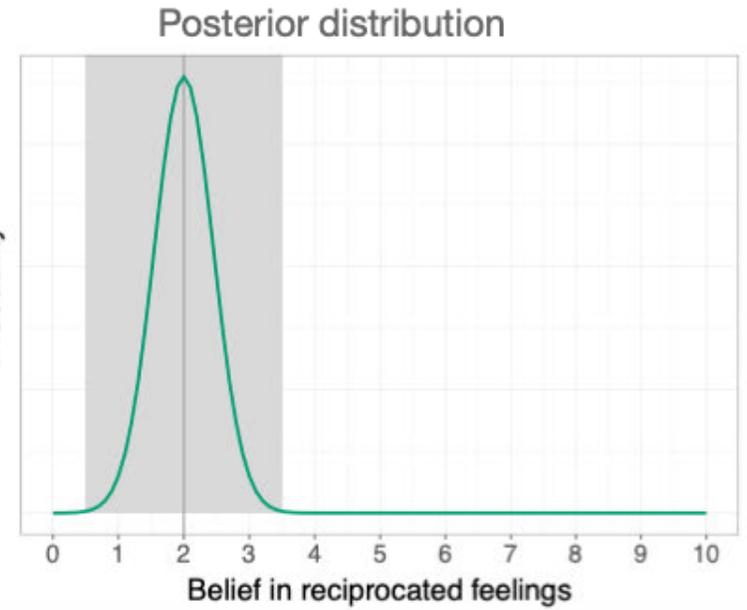
Effect sizes and p

- Wetzels et al. (2011). Statistical Evidence in Experimental Psychology: An Empirical Comparison Using 855 t -tests. *Perspectives on Psychological Science*, 6, 291–298. <https://doi.org/10.1177/1745691611406923>





Data



Bayes theorem

$$p(A|B) = \frac{p(B|A) \times p(A)}{p(B)}$$

$$p(\text{model}|\text{data}) = \frac{p(\text{data}|\text{model}) \times p(\text{model})}{p(\text{data})}$$

$$\text{posterior probability} = \frac{\text{liklihood} \times \text{prior probability}}{\text{marginal liklihood}}$$



Null hypothesis: You're human



ANDY FIELD



		DNA sample matches alien DNA		
		Match	No match	Total
Accused	Green lizard alien	1	0	1
	Human	99	1900	1999
Total		100	1900	2000

$$p(\text{hypothesis}|\text{match}) = \frac{p(\text{match}|\text{hypothesis}) \times p(\text{hypothesis})}{p(\text{match})}$$

$$\text{posterior probability} = \frac{\text{liklihood} \times \text{prior probability}}{\text{marginal liklihood}}$$



		DNA sample matches alien DNA		
		Match	No match	Total
Accused	Green lizard alien	1	0	1
	Human	99	1900	1999
Total		100	1900	2000

$$\text{prior probability} = p(\text{alien}) = \frac{1}{2000} = 0.0005$$

$$\text{marginal likelihood} = p(\text{match}) = \frac{100}{2000} = 0.05$$

$$\text{likelihood} = p(\text{match}|\text{alien}) = \frac{p(\text{alien} \cap \text{match})}{p(\text{alien})} = \frac{\frac{1}{2000}}{\frac{1}{2000}} = 1$$

Alternative hypothesis (you're alien)

$$\begin{aligned} p(\text{alien}|\text{match}) &= \frac{p(\text{match}|\text{alien}) \times p(\text{alien})}{p(\text{match})} \\ &= \frac{1 \times 0.0005}{0.05} \\ &= 0.01 \end{aligned}$$



		DNA sample matches alien DNA		
		Match	No match	Total
Accused	Green lizard alien	1	0	1
	Human	99	1900	1999
	Total	100	1900	2000

$$\text{prior probability} = p(\text{human}) = \frac{1999}{2000} = 0.9995$$

$$\text{marginal likelihood} = p(\text{match}) = \frac{100}{2000} = 0.05$$

$$\text{likelihood} = p(\text{match}|\text{human}) = \frac{p(\text{human} \cap \text{match})}{p(\text{human})} = \frac{\frac{99}{2000}}{\frac{1999}{2000}} = 0.0495$$

Null hypothesis (you're human)

$$\begin{aligned} p(\text{human}|\text{match}) &= \frac{p(\text{match}|\text{human}) \times p(\text{human})}{p(\text{match})} \\ &= \frac{0.0495 \times 0.9995}{0.05} \\ &= 0.99 \end{aligned}$$



$$\text{posterior odds} = \frac{p(\text{hypothesis 1}|\text{data})}{p(\text{hypothesis 2}|\text{data})} = \frac{p(\text{alien}|\text{match})}{p(\text{human}|\text{match})} = \frac{0.01}{0.99} = 0.01$$

$$\frac{p(\text{alternative}|\text{data})}{p(\text{null}|\text{data})} = \frac{\frac{p(\text{data}|\text{alternative}) \times p(\text{alternative})}{p(\text{data})}}{\frac{p(\text{data}|\text{null}) \times p(\text{null})}{p(\text{data})}} = \frac{p(\text{alien}|\text{match})}{p(\text{human}|\text{match})} = \frac{0.01}{0.99} = 0.01$$

$$\frac{p(\text{alternative}|\text{data})}{p(\text{null}|\text{data})} = \frac{p(\text{data}|\text{alternative})}{p(\text{data}|\text{null})} \times \frac{p(\text{alternative})}{p(\text{null})}$$

posterior odds

Bayes factor

prior odds



$$\underbrace{\frac{p(\text{alternative}|\text{data})}{p(\text{null}|\text{data})}}_{\text{posterior odds}} = \underbrace{\frac{p(\text{data}|\text{alternative})}{p(\text{data}|\text{null})}}_{\text{Bayes factor}} \times \underbrace{\frac{p(\text{alternative})}{p(\text{null})}}_{\text{prior odds}}$$

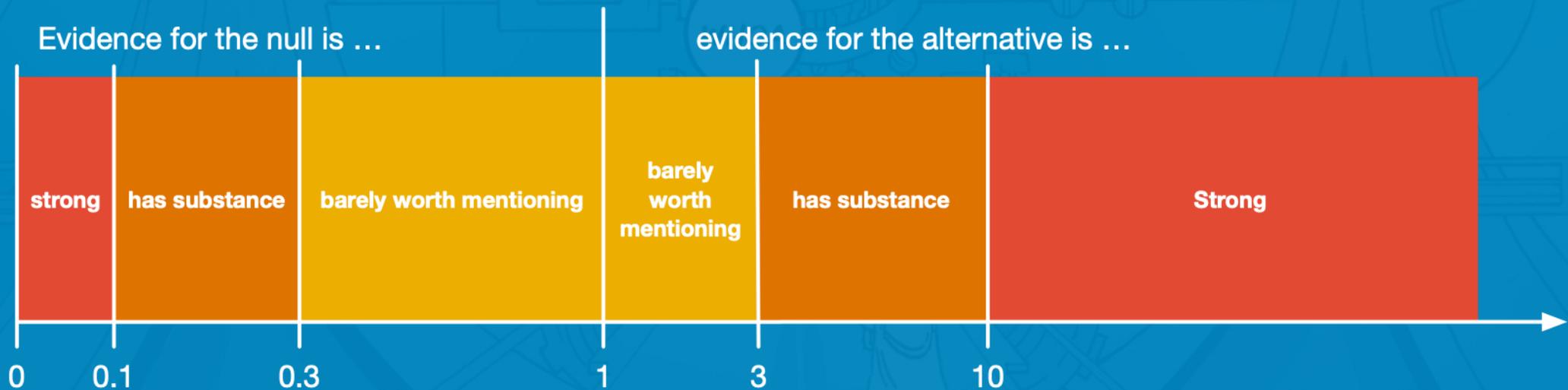
$$\frac{0.01}{0.99} = \frac{1}{0.0495} \times \frac{0.0005}{0.9995}$$

$$0.01 = 20.20 \times 0.0005$$

Given a DNA match, we should shift our belief towards that person being an alien by a factor of about 20

Bayes factor (BF_{10})

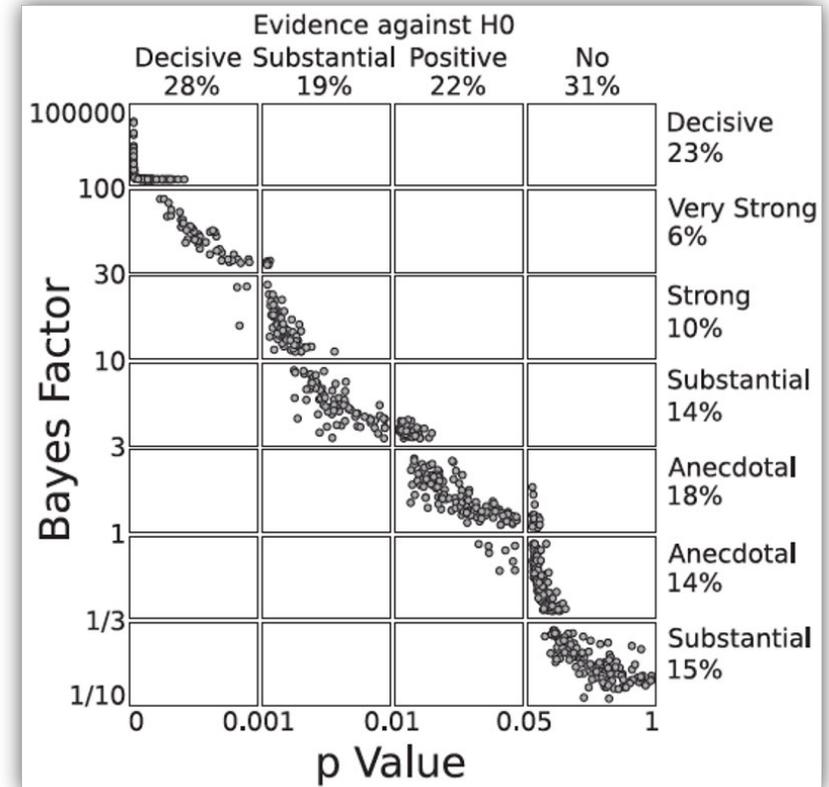
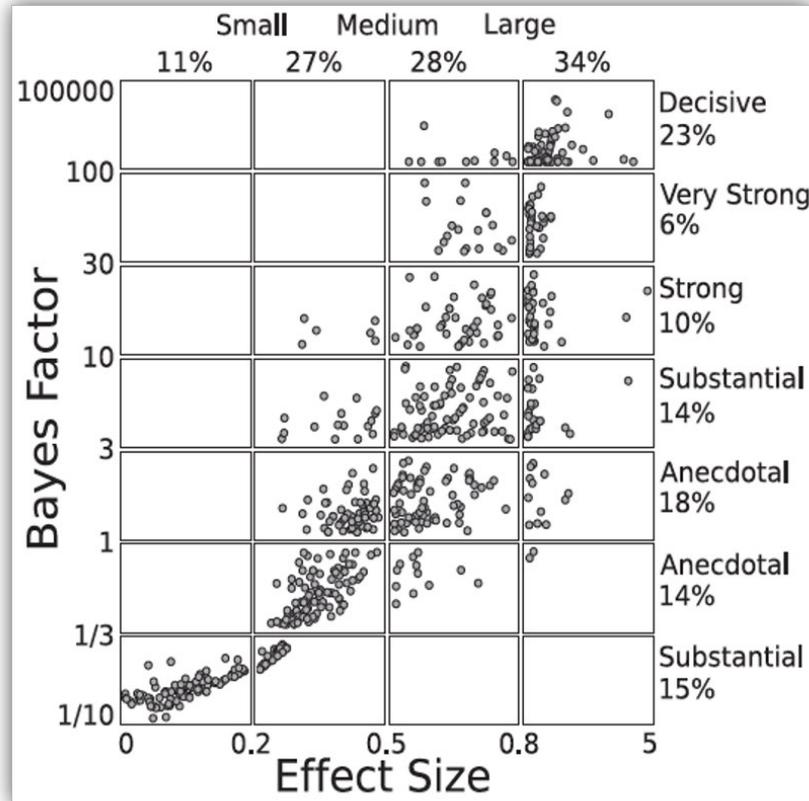
- The probability of the data given the alternative hypothesis relative to the probability of the data given the null.
- The extent to which you should change your beliefs about the alternative hypothesis relative to the null
- You sometimes see Bayes factors expressed the opposite way around (BF_{01})



ANDY FIELD



Bayes Factors, p and effect sizes¹



Wetzels et al. (2011). *Perspectives on Psychological Science*, 6, 291–298. <https://doi.org/10.1177/1745691611406923>



ANDY FIELD



Summary

- We can go beyond p to evaluate the plausibility of a hypothesis
- Other methods address more useful questions, are less dependent on sample sizes, and avoid all-or-nothing thinking
- Effect sizes quantify the size of the effect
 - Encourage thinking about effects on a continuum
- Bayes factors quantify the relative probability of the data given the null and alternative hypothesis



ANDY FIELD

