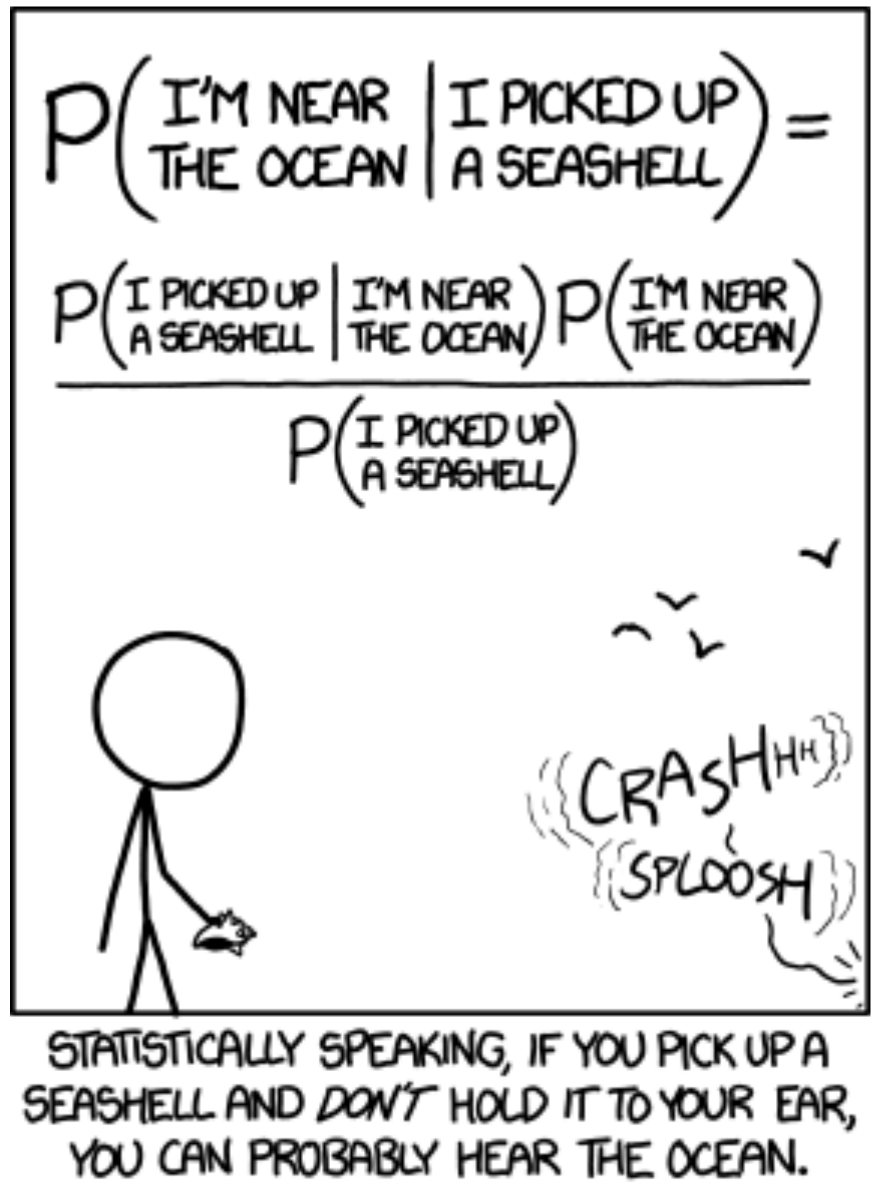
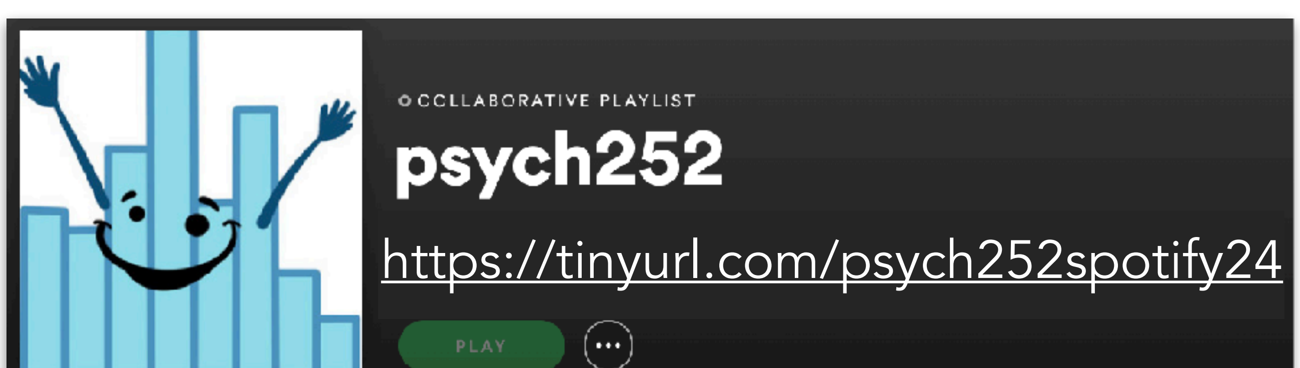
Summary and course outlook

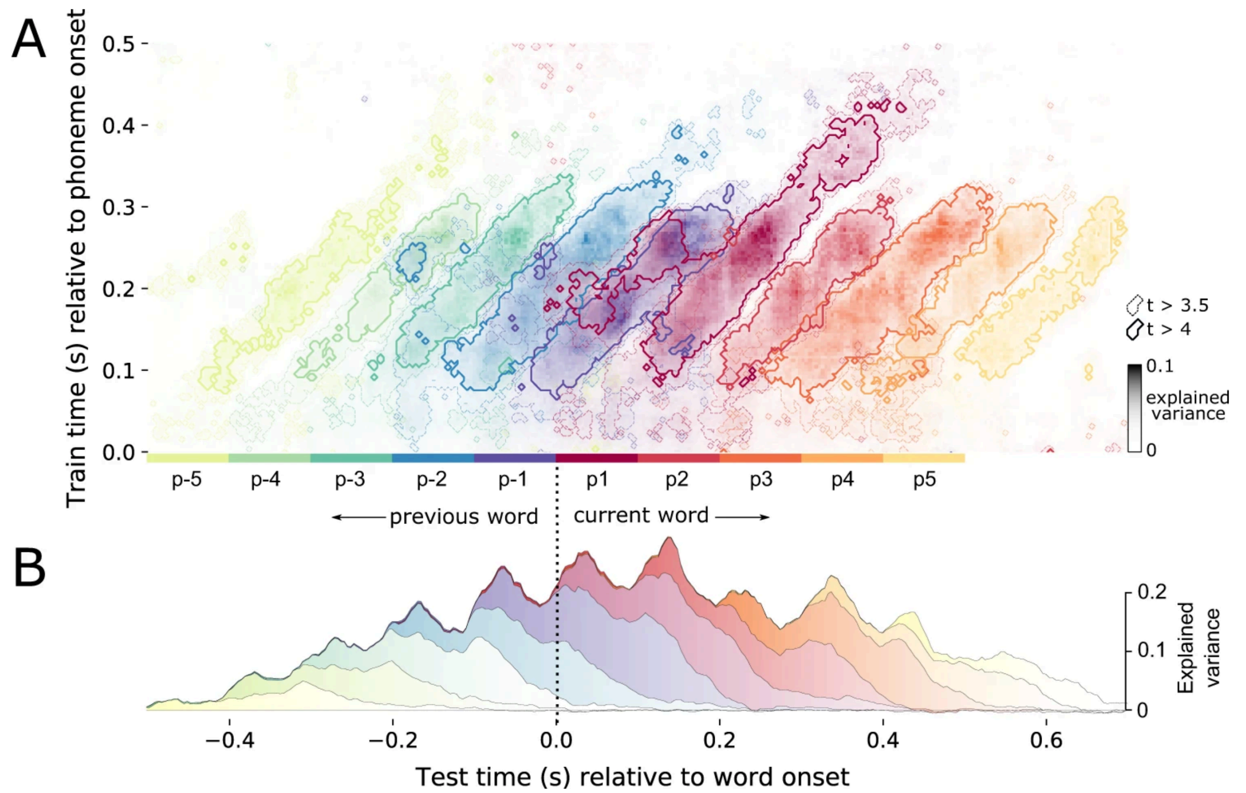




03/10/2024

Logistics

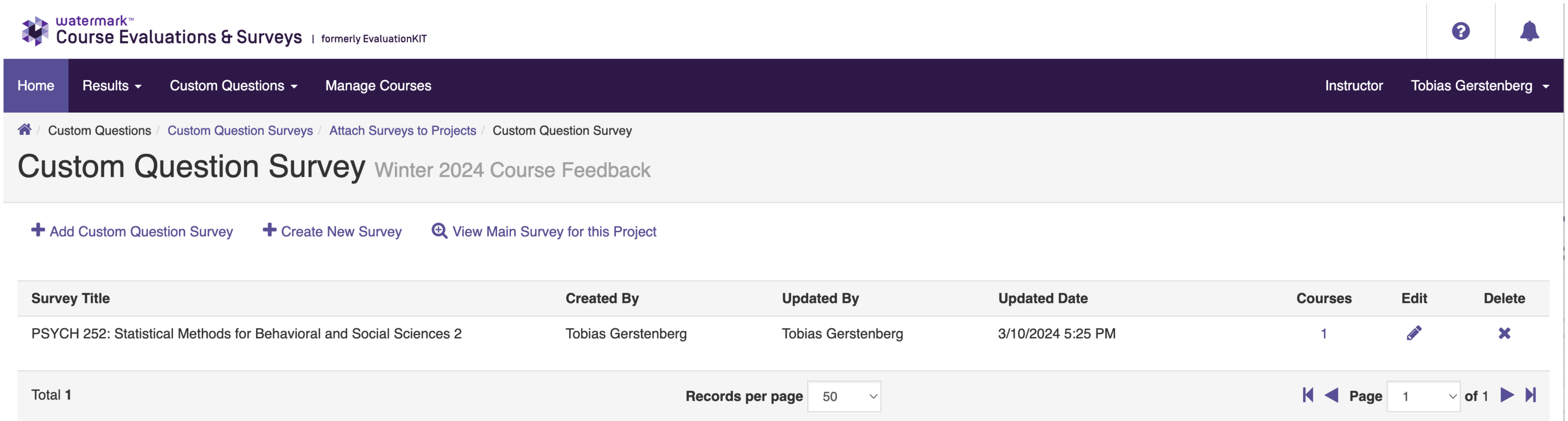
Guest lecture: Wednesday



Laura Gwilliams

Language is central to human life. The speed and accuracy with which typically-developing individuals acquire, produce and understand language is nothing short of remarkable - vastly out-performing even the most advanced artificial intelligence systems. My research aims to provide an algorithmically precise account of how the human brain achieves this feat. Such insight has the power to inform neuroscience (understanding the human brain) and engineering (building intelligent machines).

Course evaluations



Thank you for evaluating the course!!

http://evaluationkit.stanford.edu/

Guest lecture: Friday

Satchel Grant

Shawn Schwartz

Plan for today

- What we've learned
- What shall I do now?
- Quick recap
- Going beyond with Bayes
  - Evidence for the null hypothesis
  - I only want positive coefficients!
  - Dealing with unequal variance
  - Better modeling Likert scale data
- Thanks!

What we've learned

Learning goals

What you will learn

You will learn how to **use R** to …

- read, wrangle, and analyze data
- make publication-ready plots

Understand the philosophy behind null **hypothesis significance testing (NHST)** and **Bayesian statistics** through …

- running computer simulations and visualizing the results

Formulate **research questions as statistical models** and …

- determine which models work for different situations

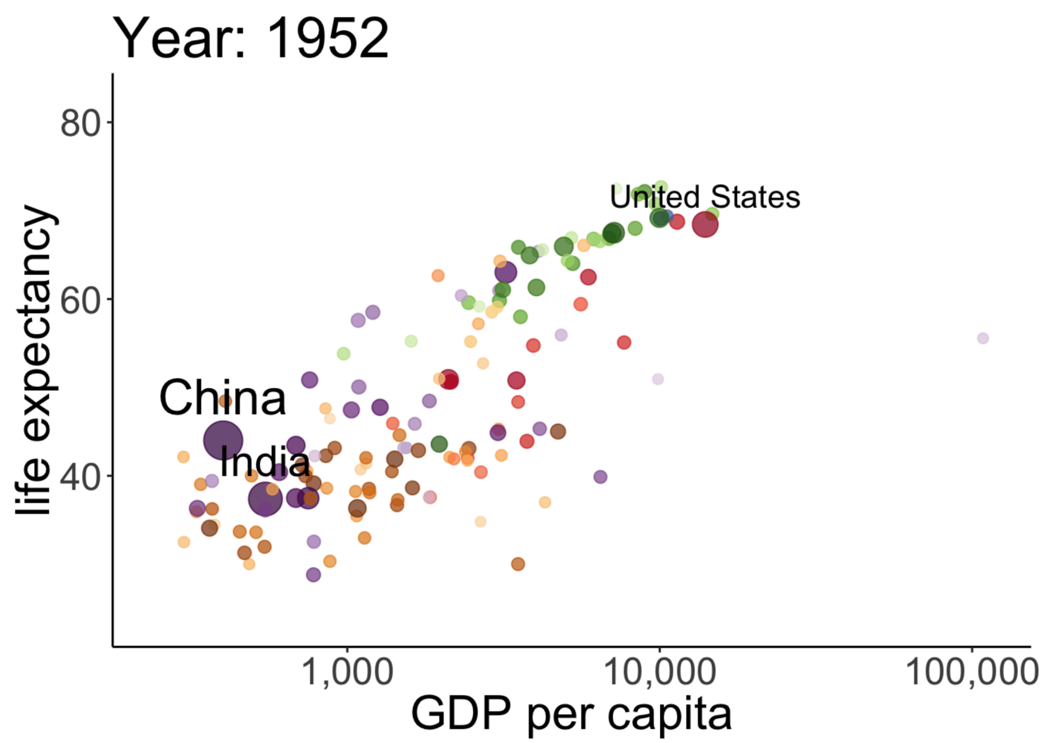
Communicate what you have learned about your data …

- in short presentations in class, showcasing your visualization and analysis
- in written reports

Contribute to open and **reproducible science** through …

- adopting good coding practices
- sharing your data and research reports online

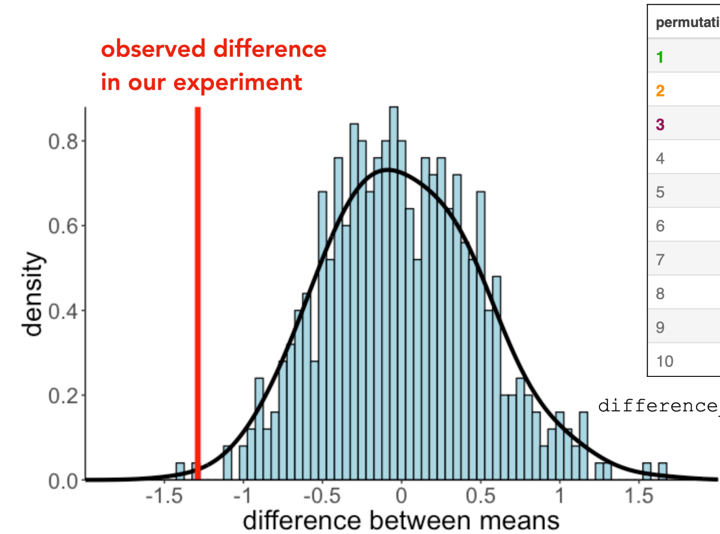
You will learn to use **R**



visualization

data wrangling

Permutation test



df.permutations

permutation	mean_difference
1	-0.88
2	-0.26
3	-0.94
4	0.47
5	-0.28
6	1.15
7	0.98
8	0.38
9	-0.08
10	0.31

difference_actual = -1.28

```
#calculate p-value of our observed result
df.permutations %>%
  summarize(p_value = sum(mean_difference <= difference_actual)/n())
```

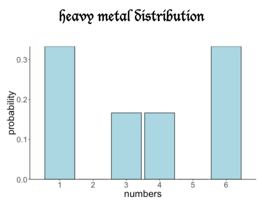
p-value = .002

64

simulation

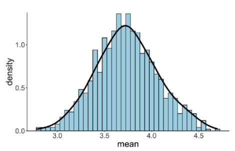
Philosophy behind **frequentist** and Bayesian stats

3 distributions in statistical inference



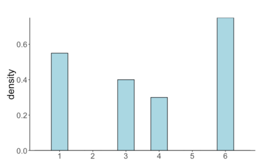
population distribution

- unknown
- our target for inference
- e.g. we might be interested in the mean of the population distribution



sampling distribution

- bridge between sample and population
- derived mathematically / computationally
- asymptotic distribution theory or resampling approaches
- shows how test statistic varies between samples



sample distribution

- our observed sample
- we compute statistics of interest (mean, variance, correlation, …)
- make an inference about the population via the sampling distribution

45

sampling distribution

Permutation test

observed data

participant	condition	performance
1	control	4.25
2	control	5.87
3	control	3.83
4	control	8.69
5	control	6.16
26	experimental	4.42
27	experimental	4.27
28	experimental	2.29
29	experimental	3.78
30	experimental	5.13

1

participant	condition	performance
1	experimental	4.25
2	control	5.87
3	control	3.83
4	experimental	8.69
5	experimental	6.16
26	control	4.42
27	experimental	4.27
28	control	2.29
29	control	3.78
30	experimental	5.13

2

participant	condition	performance
1	experimental	4.25
2	control	5.87
3	experimental	3.83
4	experimental	8.69
5	experimental	6.16
26	control	4.42
27	control	4.27
28	control	2.29
29	control	3.78
30	experimental	5.13

3

participant	condition	performance
1	control	4.25
2	experimental	5.87
3	control	3.83
4	experimental	8.69
5	control	6.16
26	control	4.42
27	experimental	4.27
28	control	2.29
29	experimental	3.78
30	experimental	5.13

⋮

permutation	mean_difference
1	-0.88
2	-0.26
3	-0.94
4	0.47
5	-0.28
6	1.15
7	0.98
8	0.38
9	-0.08
10	0.31

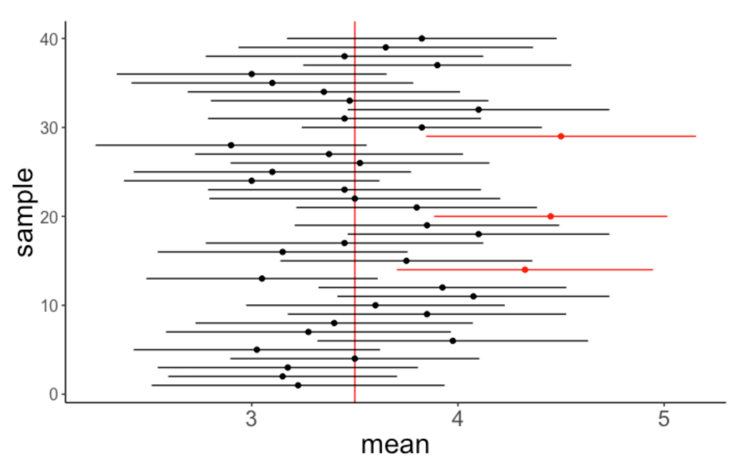
61

p-value

95% confidence interval

Definition

"If we were to repeat the experiment over and over, then 95% of the time the confidence interval contains the estimate of interest."



Hoekstra, R., Morey, R. D., Rouder, J. N., & Wagenmakers, E.-J. (2014). Robust Misinterpretation of Confidence Intervals. *Psychonomic Bulletin & Review*, 21(5), 1157-1164.

24

confidence interval

The general procedure

1. Define H_0 as Model C (compact) and H_1 as Model A (augmented)
2. Fit model parameters to the data
3. **Calculate the proportional reduction of error (PRE) in our sample**
4. Decide whether the augmented model is **worth it** by comparing the observed PRE in our sample to the sampling distribution of PRE (assuming that H_0 is true)

model comparison

Philosophy behind frequentist and **Bayesian** stats

Clue guide to probability

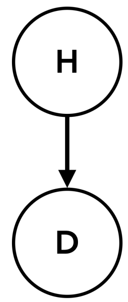
$$p(B\,|\,A) = \frac{p(A\,|\,B) \cdot p(B)}{p(A)}$$

$$\underbrace{p(H\,|\,D)}_{\text{posterior}} = \frac{\overbrace{p(D\,|\,H)}^{\text{likelihood}} \cdot \overbrace{p(H)}^{\text{prior}}}{p(D)}$$

subjective probability interpretation

H = Hypothesis

D = Data



formal framework for learning from data

updating our prior belief $p(H)$, to a posterior belief $p(H\,|\,D)$ given some data

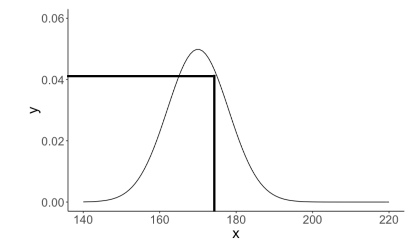
Bayes' theorem

44

Summer camp

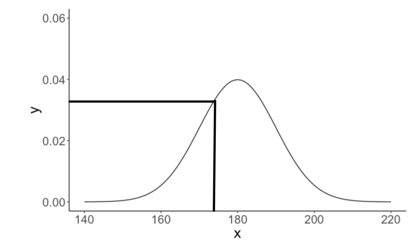
prior $p(\text{chess}) = \frac{1}{3}$ $p(\text{basketball}) = \frac{2}{3}$

likelihood



`dnorm(175, mean = 170, sd = 8)`

= 0.041



`dnorm(175, mean = 180, sd = 10)`

= 0.035

posterior

$$p(\text{basketball}\,|\,175) = \frac{p(175\,|\,\text{basketball}) \cdot p(\text{basketball})}{p(175\,|\,\text{basketball}) \cdot p(\text{basketball}) + p(175\,|\,\text{chess}) \cdot p(\text{chess})}$$

$$p(\text{basketball}\,|\,175) = \frac{0.035 \cdot 2/3}{0.035 \cdot 2/3 + 0.041 \cdot 1/3} \approx 0.63$$

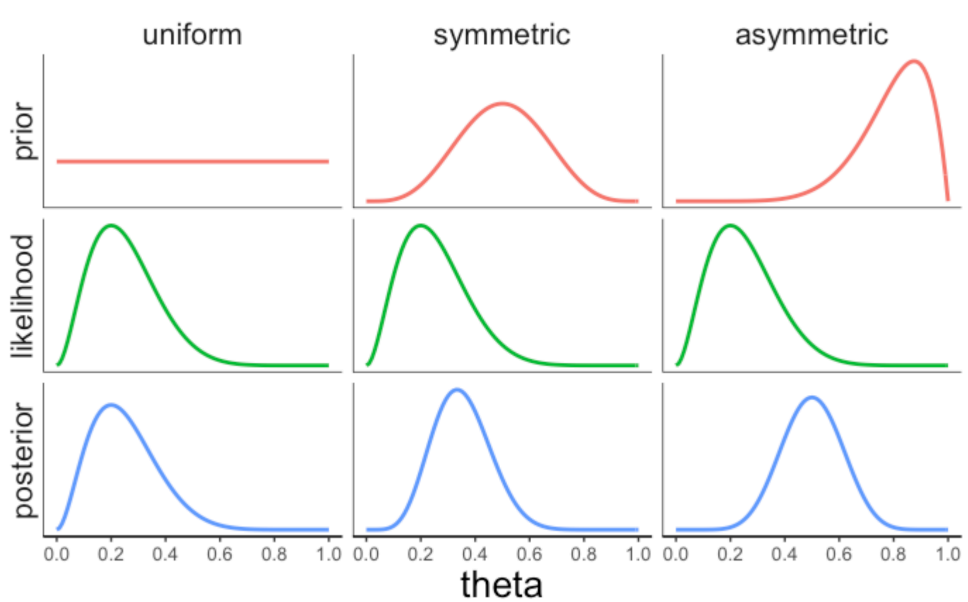
send the kid to the basketball gym!

36

Bayesian inference

The effect of the **prior**

same data, different priors



50

prior, likelihood, posterior

Recipe for Bayesian analysis with `brms`

visualization is everywhere!

1. Visualize the data
2. Specify and fit the model
3. Model evaluation
   a) Did the inference work?
   b) Visualize model predictions
4. Interpret the model parameters
5. Test specific hypotheses
6. Report results

39

Bayesian data analysis

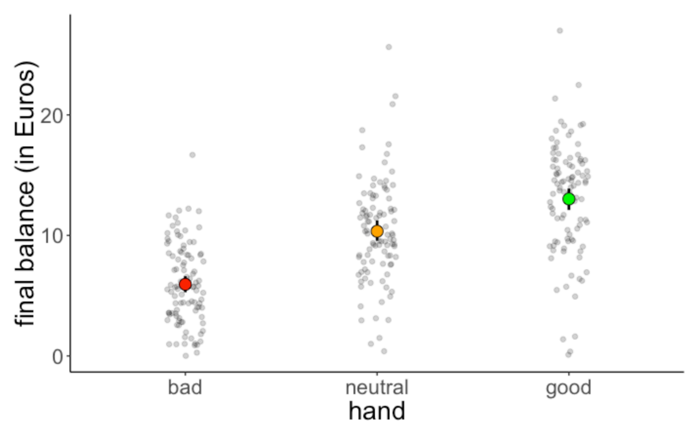
Formulate research questions as statistical models

Do better hands win more money?

participant	skill	hand	limit	balance
1	expert	bad	fixed	4.00
2	expert	bad	fixed	5.55
26	expert	bad	none	5.52
27	expert	bad	none	8.28
51	expert	neutral	fixed	11.74
52	expert	neutral	fixed	10.04
76	expert	neutral	none	21.55
77	expert	neutral	none	3.12
101	expert	good	fixed	10.86
102	expert	good	fixed	8.68

hand = {bad, neutral, good}

Visualize the data first

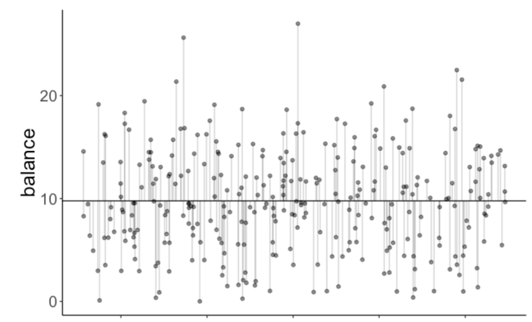


H_0: Card quality does not affect the final balance.

Model C

$balance_i = \beta_0 + \epsilon_i$

Model prediction



SSE(C) = 7580

Fitted model

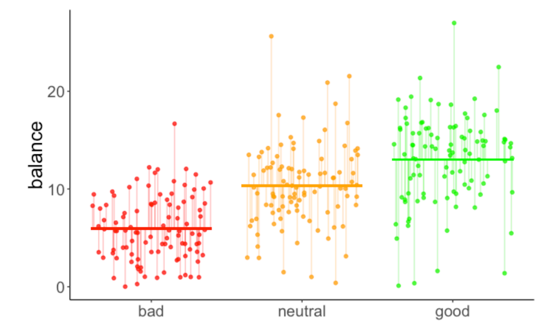
$\widehat{balance}_i = 9.77$

H_1: Card quality affects the final balance.

Model A

$balance_i = \beta_0 + \beta_1 hand\_neutral_i + \beta_2 hand\_good_i + \epsilon$

Model prediction



SSE(A) = 5021

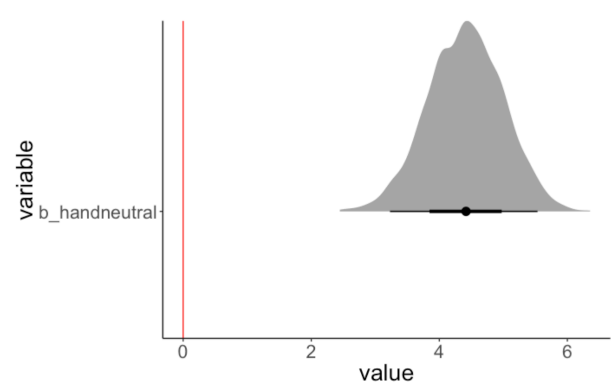
Fitted model

$\widehat{balance}_i = 5.94 + 4.41 \cdot hand\_neutral_i + 7.08 \cdot hand\_good_i$

31

Asking questions based on the posterior

Do neutral hands earn more money than bad hands?



What's the probability that handneutral is less than 0?

```
hypothesis(fit.brm,
           hypothesis = "handneutral < 0")
```

p = 0

52

Communicate what you've learned about your data

2.1: Regressions (1.5 point)

Run three linear regression models that estimate the relationships between `quarter_of_birth` and `log_weekly_wage`, `education` and `log_weekly_wage`, and `quarter_of_birth` and `education`. Then run a multiple regression model that predicts `log_weekly_wage` based on both `quarter_of_birth` and `education`. Print the summaries of all four regressions and comment on their significances.

```
### YOUR CODE HERE ###
lm1 = lm(formula = log_weekly_wage ~ quarter_of_birth,
         data = df.qob)
lm2 = lm(formula = log_weekly_wage ~ education,
         data = df.qob)
lm3 = lm(formula = education ~ quarter_of_birth,
         data = df.qob)

lm_multi = lm(formula = log_weekly_wage ~ quarter_of_birth + education,
              data = df.qob)

summary(lm1)
```

```
Call:
lm(formula = log_weekly_wage ~ quarter_of_birth, data = df.qob)

Residuals:
    Min      1Q  Median      3Q     Max
-8.2511 -0.2629  0.0605  0.3720  4.2099

Coefficients:
                 Estimate Std. Error t value Pr(>|t|)
(Intercept)      5.887183   0.007114 827.584   <2e-16 ***
quarter_of_birth 0.007366   0.002913   2.529   0.0115 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6908 on 49998 degrees of freedom
Multiple R-squared:  0.0001279, Adjusted R-squared:  0.0001079
F-statistic: 6.394 on 1 and 49998 DF,  p-value: 0.01145
```

homework/midterm

A catchy project title goes here

My team's name goes here

The team members' names go here

2021-03-19 12:44:19

The final project is due on **Friday, March 20th at 8pm**.

Here are some guidelines:

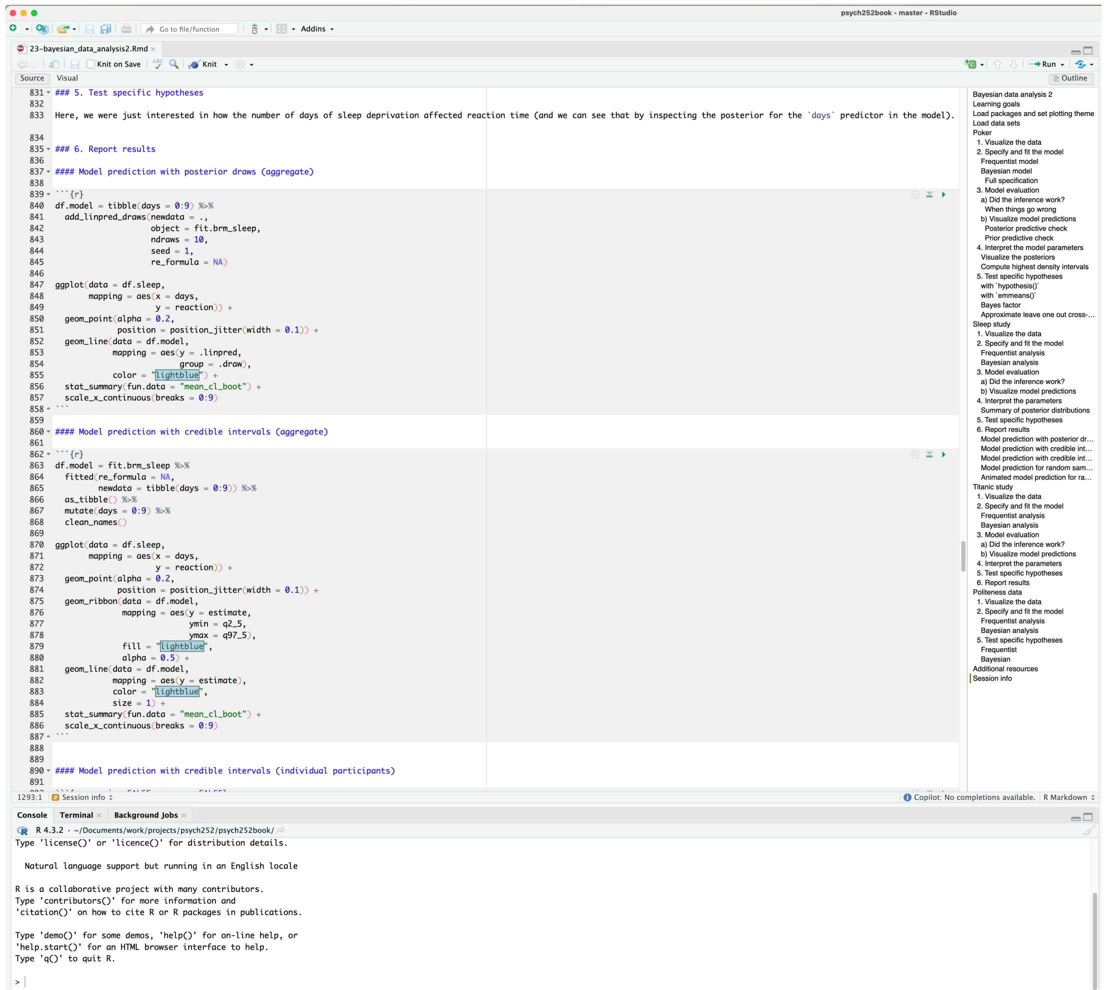
- The length of the final report should be around 2000 words per person in the group.
- All the code should be contained in this RMarkdown file (from reading in the messy data file, to making beautiful plots).
- Feel free to make the final report look like an actual paper. So you can hide all the code chunks that do data wrangling etc. from the output (by setting the code chunk option to `echo=F`), and only show the figures and tables that you need to explain your work.
- Show us what you've learned :) We're excited to read it!

For more information on how to do stuff in RMarkdown, check out:

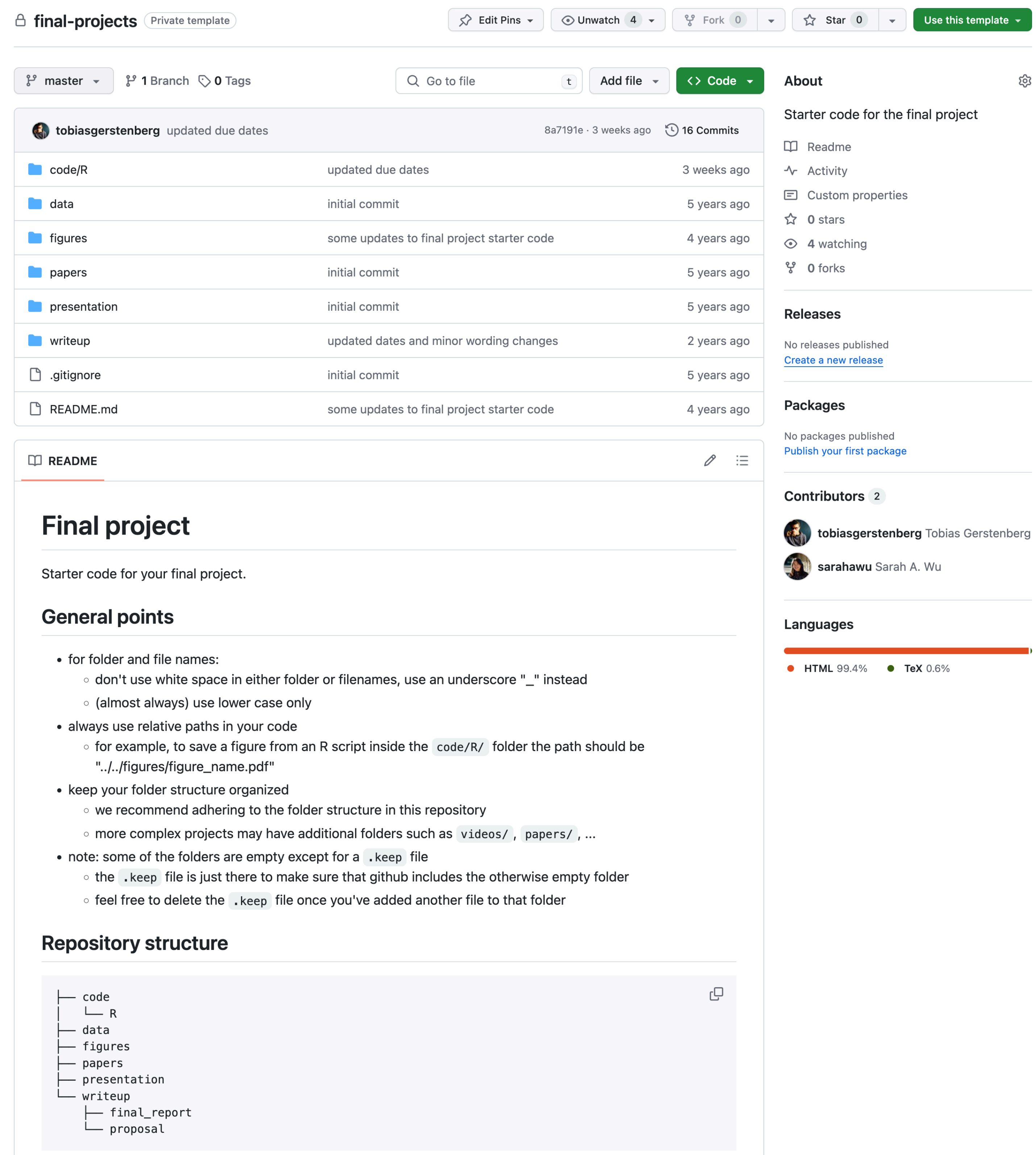
- bookdown documentation
- RMarkdown code chunks
- RMarkdown cheat sheet
- Citations

final project (proposal, presentation, report)

Contribute to open and reproducible science



RStudio



github

What we've covered

visualization and data wrangling

probability, simulation, causality

linear model

power analysis

model comparison

linear mixed effects models

logistic regression

Bayesian data analysis

Date	Topic
January 8th	Introduction
January 10th	Visualization 1
January 12th	Visualization 2
January 15th	Martin Luther King Jr. Day
January 17th	Data wrangling 1
January 19th	Data wrangling 2
January 22nd	Probability
January 24th	Simulation 1
January 26th	Simulation 2
January 29th	Modeling data
January 31st	Linear model 1
February 2nd	Linear model 2
February 5th	Linear model 3
February 7th	Linear model 4
February 9th	Power analysis
February 12th	Model comparison
February 14th	No class (due to Midterm)
February 16th	Causation
February 19th	President's Day
February 21st	Linear mixed effects models 1
February 23rd	Linear mixed effects models 2
February 26th	Linear mixed effects models 3
February 28th	Linear mixed effects models 4
March 1st	Generalized linear model
March 4th	Bayesian data analysis 1
March 6th	Bayesian data analysis 2
March 8th	Bayesian data analysis 3
March 11th	Summary and course outlook
March 13th	Guest lecture: Laura Gwilliams
March 15th	Guest lecture: Satchel Grant & Shawn Schwartz

I'll keep updating the course notes!

PSYCH 252: STATISTICAL METHODS Home Schedule Getting ready Information Book

This course offers an introduction to advanced topics in statistics with the focus of understanding data in the behavioral and social sciences. It is a practical course in which learning statistical concepts and building models in R go hand in hand. The course is organized into three parts: In the first part, we will learn how to visualize, wrangle, and simulate data in R. In the second part, we will cover topics in frequentist statistics (such as multiple regression, logistic regression, and mixed effects models) using the general linear model as an organizing framework. We will learn how to compare models using simulation methods such as bootstrapping and cross-validation. In the third part, we will focus on Bayesian data analysis as an alternative framework for answering statistical questions.

Requirement: Psych 10, Stats 60, or equivalent.

Psych 252: Statistical Methods for Behavioral and Social Sciences

Tobias Gerstenberg

2024-01-06

Preface

This book contains the course notes for Psych 252. The book is not intended to be self-explanatory and instead should be used in combination with the course lectures posted here.

If you have any questions about the notes, please feel free to contact me at: gerstenberg@stanford.edu or post an issue on the book's github repository.

https://psych252.github.io/

you'll still have access to the lecture recordings from 2022

including the super accurate captions!

Tobi Gerstenberg

take a hint zoom!

00:00:01

hi I am totally guest America i'm an assistant professor at Stanford university

Datacamp (available until ~ mid june)

datacamp

What shall I do now?

PSYCH 253

Nilam Ram

COMM 369: Measurement and the Study of Change in Social Science Research (PSYCH 253)

This course is a survey of methodological issues associated with the measurement of psychological constructs and processes of change. General areas to be covered include use of latent variable models (structural equation modeling), classical test theory, generalizability theory, principal component analysis, factor analysis, item response theory and how these models facilitate and/or constrain the study of change processes. Students will work through application/implementation of the models through hands-on analysis of simulated and empirical data, acquire experiences in the formulation of research questions and study designs that are appropriately tethered to the different theoretical perspectives invoked by the different models.

Terms: Spr | Units: 3

Instructors: Ram, N. (PI) ; Mehta, A. (TA) ; Tan, A. (TA)

Schedule for COMM 369

CS109: Probability for Computer Scientists

Schedule

The class starts by providing a fundamental grounding in combinatorics, and then quickly moves into the basics of probability theory. We will then cover many essential concepts in probability theory, including particular probability distributions, properties of probabilities, and mathematical tools for analyzing probabilities. Finally, the last third of the class will focus on data analysis and Machine Learning as a means for seeing direct applications of probability in this exciting and quickly growing subfield of computer science.

Overview of Topics

Counting Theory | Core Probability | Random Variables | Probabilistic Models | Uncertainty Theory | Machine Learning

- learn more about probability theory through programming
- gain a deeper understanding of the fundamental underlying concepts

http://web.stanford.edu/class/cs109/schedule.html

PSYCH 204B: Computational Neuroimaging: Data Analyses and Experimental Designs

This course provides an in-depth survey and understanding of modern computational approaches to design and analyses of neuroimaging data. The course is a mixture of lectures and projects geared to give the student an understanding of the possibilities as well as limitations of different computational approaches. Topics include: signal and noise in MRI; general linear modeling; fMRI-adaptation; multivoxel pattern analyses; decoding and encoding algorithms; modeling population receptive fields. Required: Psych 204A; Recommended: Cognitive Neuroscience, Stats

Terms: Spr | Units: 1-3

Instructors: Grill-Spector, K. (PI)

Schedule for PSYCH 204B

Advanced regression analysis

Sanne Smith

EDUC 326: Advanced Regression Analysis (SOC 384)

Social science researchers often deal with complex data and research questions that traditional statistics models like linear regression cannot adequately address. This course offers the opportunity to understand and apply two widely used types of advanced regression analysis that allow the examination of 1) multilevel data structures (multilevel models) and 2) multivariate research questions (structural equation models).

Terms: Spr | Units: 3-5

Instructors: Smith, S. (PI) ; Bhat, K. (TA)

Schedule for EDUC 326

2023-2024 Spring

EDUC 326 | 3-5 units | UG Reqs: None | Class # 2200 | Section 01 | Grading: Letter or Credit/No Credit | LEC | Session: 2023-2024 Spring 1 | In Person | Students enrolled: 9

04/01/2024 - 06/05/2024 Thu 1:30 PM - 4:20 PM at Ceras 108 with Smith, S. (PI); Bhat, K. (TA)

Instructors: Smith, S. (PI); Bhat, K. (TA)

Additional Resources: (Login to view additional resources)

and many more …

- **EDUC 423B**: Introduction to Data Science II: Machine learning (SOC 302B) (overview of machine learning techniques)
- **EDUC 430A**: Experimental Research Design and Analysis (learn how to do field experiments and causal inference)
- **EDUC 430B**: Quasi-Experimental Research Design & Analysis (SOC 258B) ((seeking to) get causal inference without doing experiments)
- **MS&E 226**: Fundamentals of Data Science: Prediction, Inference, Causality (a bit redundant with this class but great if you want to reinforce this knowledge and get an intro to ML)
- **MS&E 231**: Introduction to Computational Social Science (SOC 278) (I heard this was very good. it hasn't been offered for a couple years though)
- **STATS 209A**: Topics in Causal Inference (MS&E 327) (haven't taken but seems like a good intro to causal inference)
- **STATS 216**: Introduction to Statistical Learning
- **CS 228**: Probabilistic Graphical Models: Principles and Techniques

What shall I not do?

Email the psych252 teaching team for stats questions I have in the future

We'd love to hear from you, but we can't help with stats questions.

Getting help with stats

for anyone

Consulting Services

The Department of Statistics offers a free online consulting service to members of the broader research community during each Stanford academic quarter.

Under the supervision of a senior faculty member, Statistics graduate students arrange Zoom meetings with clients to help with statistical research questions in areas such as:

- Experimental design and data acquisition
- Data exploration, analysis, and interpretation
- Modeling data and model fitting
- Statistical inference for estimation, testing, and prediction

Students taking statistics courses should understand that **this is not a tutoring service**.

https://statistics.stanford.edu/resources/consulting

for psych grads

Shawn Schwartz
stschwartz@stanford.edu

Ari Beller
abeller@stanford.edu

Andrew Nam
ajhnam@stanford.edu

Quick recap

Quick recap: Testing hypotheses

Results

posterior samples

b_Intercept	b_handneutral	b_handgood	sigma
5.97	4.27	7.48	3.94
5.11	5.25	7.40	3.91
7.03	3.78	5.80	4.48
5.72	4.18	7.25	4.00
6.01	4.44	6.15	4.57
5.94	4.69	6.72	4.36
6.39	3.84	6.40	3.92
5.24	5.15	7.69	4.16
6.12	4.51	7.20	4.14
6.43	3.71	6.37	4.13
5.85	5.01	7.32	4.00
6.51	3.58	6.62	3.95
5.85	4.45	7.62	4.17
5.80	5.45	6.36	4.10
5.48	5.51	7.22	3.99

⋮

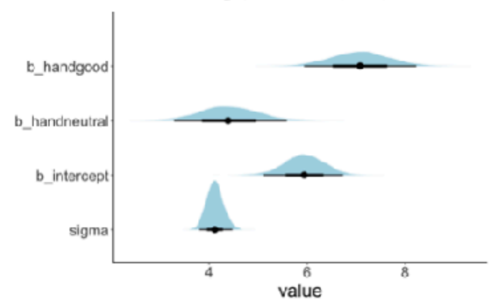
summary of posterior

parameter	lower	mode	upper
b_handgood	5.97	7.07	8.27
b_handneutral	3.21	4.43	5.51
b_intercept	5.17	5.95	6.77
sigma	3.81	4.12	4.47

maximum a posteriori

MAP estimate and 95% highest density interval

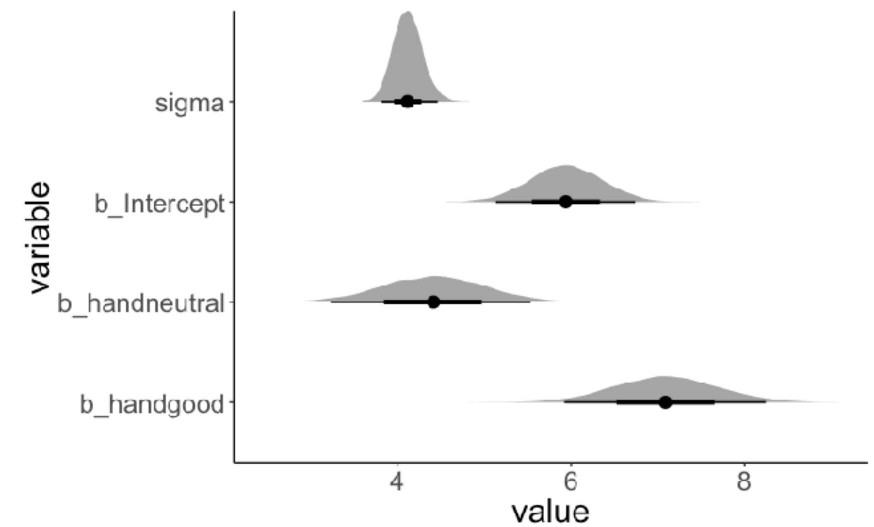
visualization



12

Asking questions based on the posterior

Do good hands make twice as much as bad hands?




```r
hypothesis(fit.brm,
           hypothesis = "handgood + Intercept > 2 * Intercept")
```

p = 0.89

16

Testing hypothesis

```r
df.hypothesis = fit.brm %>%
  posterior_samples() %>%
  clean_names() %>%
  select(starts_with("b_")) %>%
  mutate(neutral = b_intercept + b_handneutral,
         bad_good_average = (b_intercept + b_intercept + b_handgood)/2,
         hypothesis = neutral < bad_good_average)
```

samples from the posterior

b_intercept	b_handneutral	b_handgood	neutral	bad_good_average	hypothesis
6.07	4.10	7.20	10.17	9.67	FALSE
6.06	4.44	6.95	10.49	9.53	FALSE
5.88	5.00	6.73	10.87	9.24	FALSE
5.85	4.78	6.16	10.63	8.94	FALSE
5.86	4.46	7.68	10.32	9.70	FALSE

```r
df.hypothesis %>%
  summarize(p = sum(hypothesis)/n())
```

p = 0.04

18

The "emmeans" package is your friend!

```r
fit.brm_poker %>%
  emmeans(specs = consec ~ hand)
```

estimated mean for each group

contrasts

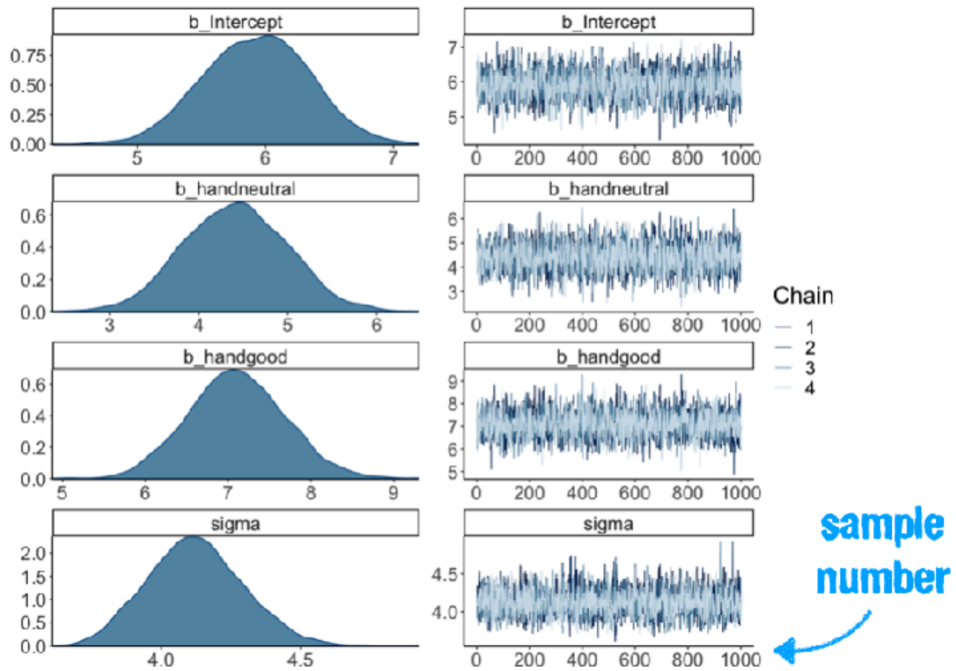
```
$emmeans
 hand    emmean lower.HPD upper.HPD
 bad       5.94      5.16      6.78
 neutral  10.34      9.55     11.15
 good     13.02     12.22     13.82

Point estimate displayed: median
HPD interval probability: 0.95

$contrasts
 contrast        estimate lower.HPD upper.HPD
 neutral - bad       4.38      3.24      5.52
 good - neutral      2.69      1.51      3.78

Point estimate displayed: median
HPD interval probability: 0.95
```

20

Quick recap: Model evaluation

Can we trust the model results?

```
plot(fit.brm_poker)
```

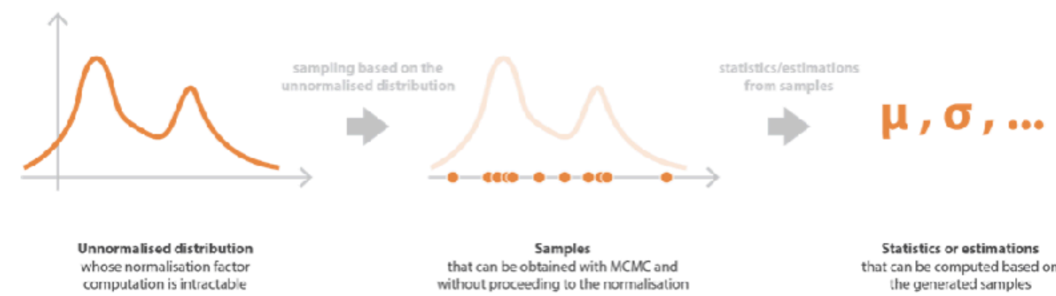


https://chi-feng.github.io/mcmc-demo/

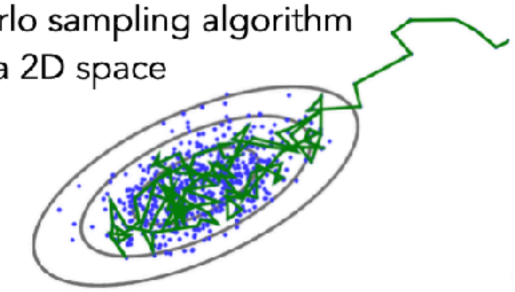
24

Can we trust the model results?

Inference via Markov Chain Monte Carlo (MCMC)



Markov Chain Monte Carlo sampling algorithm in a 2D space

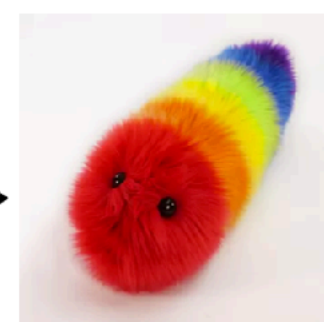


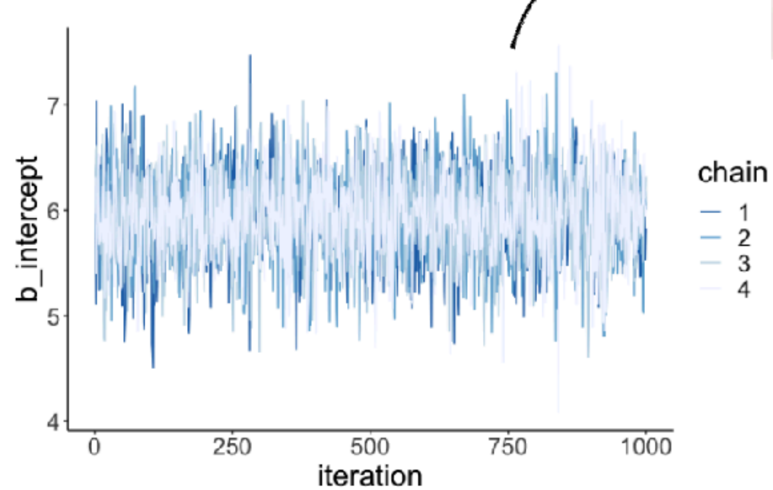
goal: draw **independent** samples from the posterior distribution

25

Can we trust the model results?

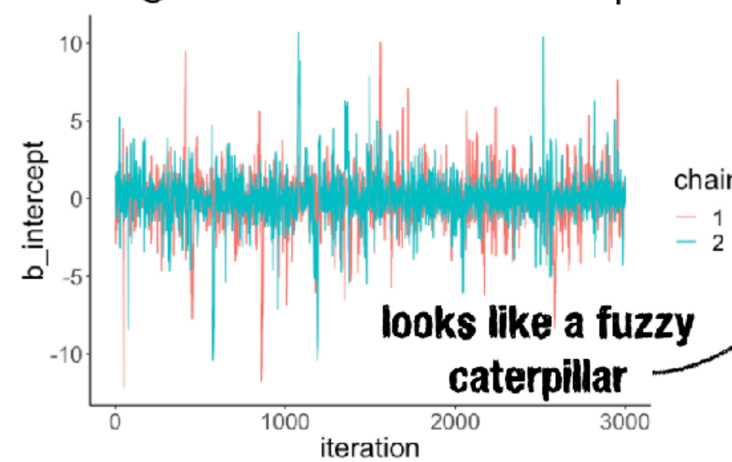
looks like a fuzzy caterpillar





26

Having somewhat informative priors fixes things



looks like a fuzzy caterpillar

if things go wrong:

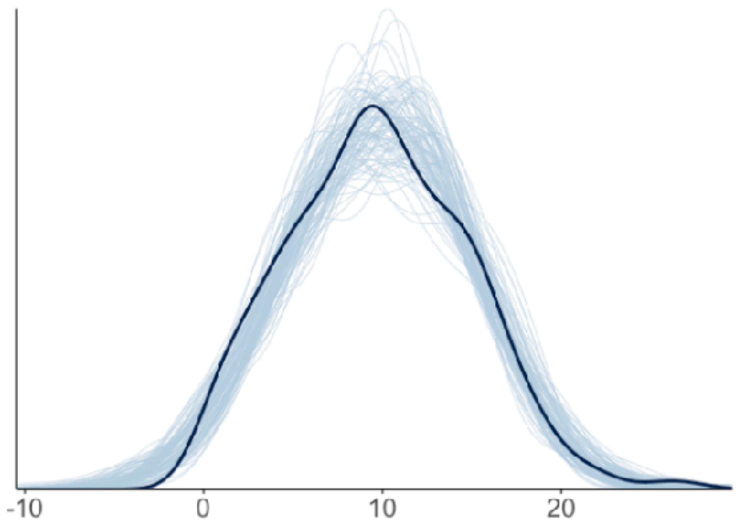
- set more informative priors
- run more warm-up samples
- adjust the sampling algorithm as suggested via the `control` argument

32

Quick recap: Posterior predictive check & reporting

Posterior predictive check

```
pp_check(fit.brm, nsamples = 100)
```

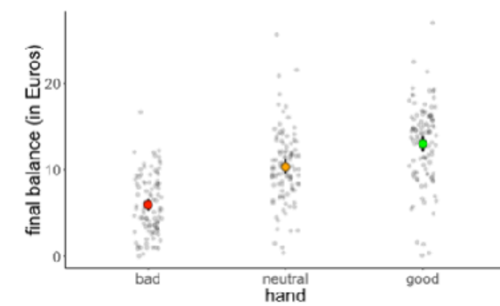


The model accurately captures the distribution of the response variable

34

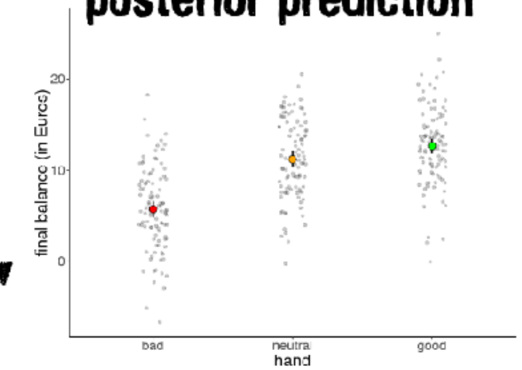
Posterior predictive check

original data

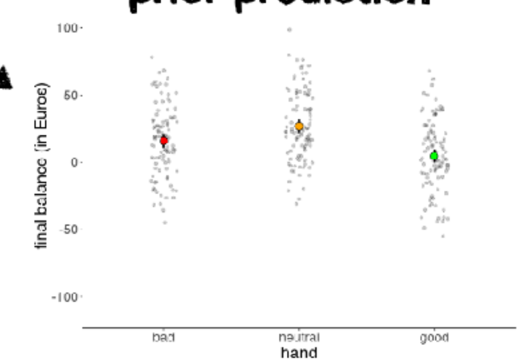


take a look at course notes for how to make these

posterior prediction



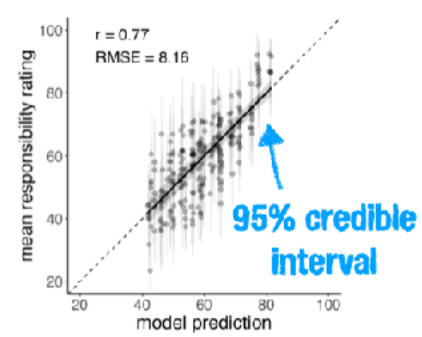
prior prediction



35

Reporting results

Plots



Tables

Table 1

Estimates of the mean, standard error, and 95% HDIs of the different predictors in the Bayesian mixed effects model. Note: `n_causes` *= number of causes.*

`responsibility ~ 1 + surprise + pivotality + n_causes + (1 + surprise + pivotality + n_causes | participant)`

term	estimate	std.error	lower 95% HDI	upper 95% HDI
intercept	58.94	3.25	54.70	65.22
surprise	21.68	4.57	14.17	29.23
pivotality	13.52	1.82	10.47	16.53
n_causes	-5.72	0.50	-6.55	-4.90

model formula

parameter estimates

Text

We computed a Bayesian mixed effects model with random intercepts and slopes to predict participants' responsibility judgments (see Table 1). Figure 6b shows a scatter plot of the model predictions and participants' responsibility judgments for the full set of 170 scenarios (with 250 judgments). Overall, the model predicts participants' responsibility judgments well with r = .77 and RMSE = 8.16. Table 1 shows the estimates of the different predictors. As can be seen, none of the predictors' 95% HDIs overlap with 0.[1]

[1]For any statistical claim, we report the mean of the posterior distribution together with the 95% highest-density interval (HDI). All Bayesian models were written in Stan (Carpenter et al., 2017) and accessed with the brms package (Bürkner, 2017) in R (R Core Team, 2019).

37

Quick recap: Sleep data

1. Specify and fit the model

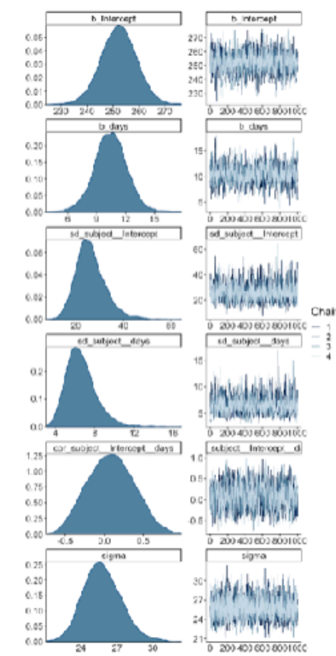
```r
fit.brm_sleep = brm(formula = reaction ~ 1 + days + (1 + days | subject),
                    data = df.sleep,
                    seed = 1,
                    file = "cache/brm_sleep")
```

43

a) Did the inference work?

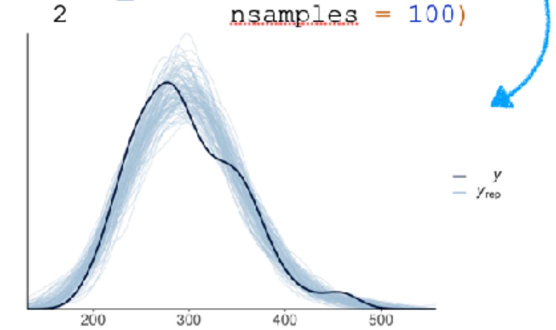
```r
fit.brm_sleep %>%
  plot(N = 6)
```

these look good!



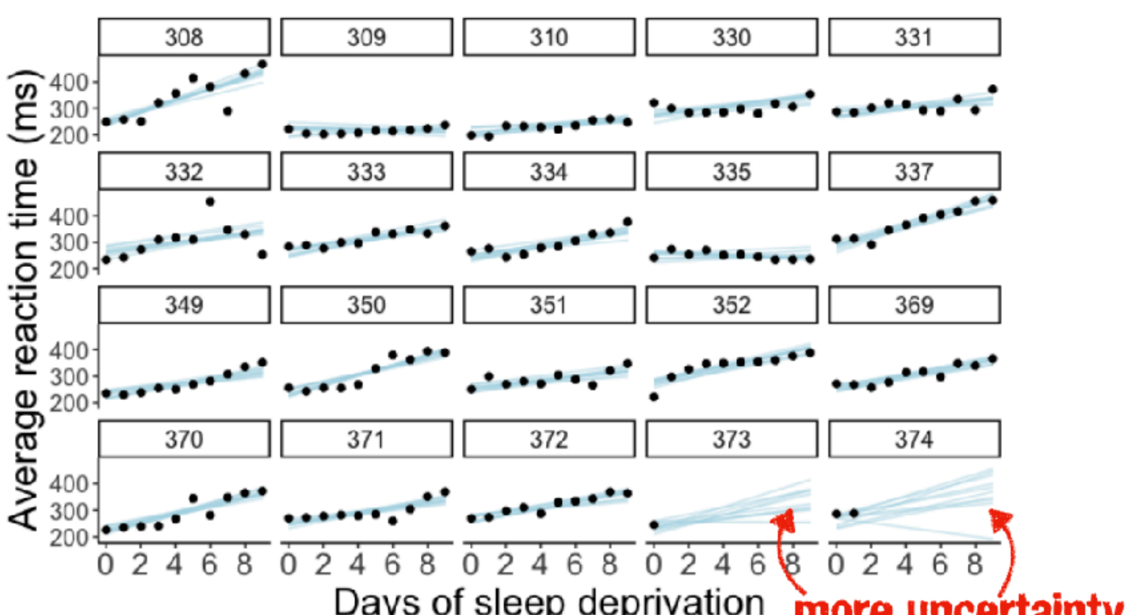
also looks good!

```r
pp_check(fit.brm_sleep,
         nsamples = 100)
```



46

b) Visualize the model predictions



10 random samples from the posterior distribution

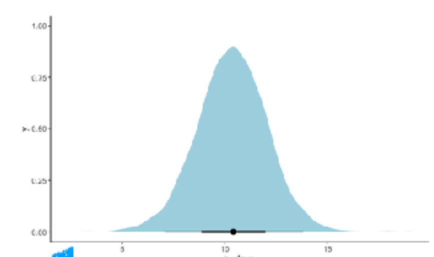
48

5. Test specific hypotheses

Did reaction times increase with the number of days of sleep deprivation?

```r
fit.brm_sleep %>%
  summary()
```

```
Family: gaussian
  Links: mu = identity; sigma = identity
Formula: reaction ~ 1 + days + (1 + days | subject)
   Data: df.sleep (Number of observations: 183)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Group-Level Effects:
~subject (Number of levels: 20)
                    Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)          26.18      6.25    15.65    40.54 1.00     1879     2463
sd(days)                6.39      1.53     4.14    10.13 1.00     1145     1625
cor(Intercept,days)     0.09      0.29    -0.46     0.67 1.00      993     1526

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept   252.18      6.86   238.47   265.42 1.00     1826     2766
days         10.46      1.69     7.13    13.78 1.00     1203     1782

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma    25.77      1.57    22.93    29.14 1.00     3864     2773

Samples were drawn using sampling(NUTS). For each parameter, Eff.Sample
is a crude measure of effective sample size, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```



54

Quick recap: Titanic data

1. Specify and fit the model

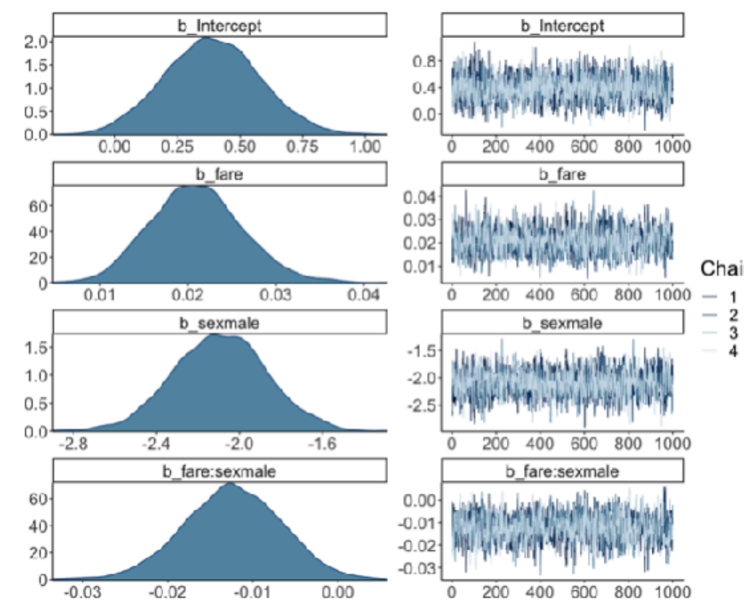
```
fit.brm_titanic = brm(formula = survived ~ 1 + fare * sex,
                      family = "bernoulli",
                      data = df.titanic,
                      file = "cache/brm_titanic",
                      seed = 1)
```

just need to change the family

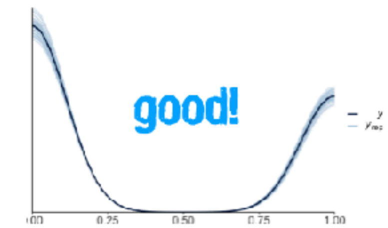
63

a) Did the inference work?

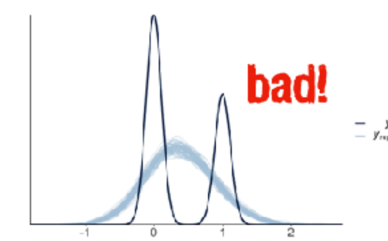
```
fit.brm_titanic %>%
  plot()
```




```
pp_check(fit.brm_titanic,
         nsamples = 100)
```



model with Gaussian family

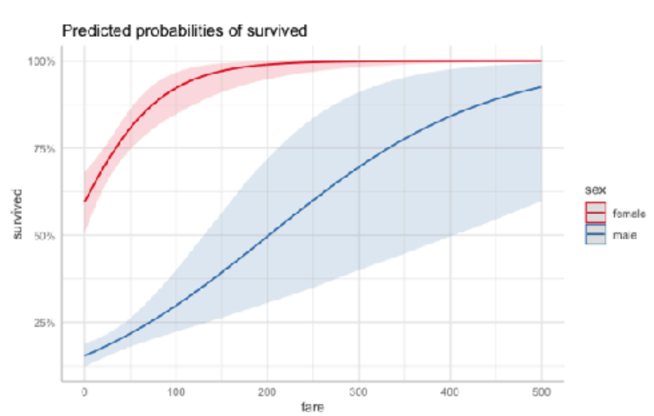


66

4. Interpret the model parameters

```
fit.brm_titanic %>%
  ggpredict(terms = c("fare [0:500]", "sex"))
```

```
# Predicted probabilities of survived
# x = fare

# sex = female

  x | Predicted |        95% CI
-----------------------------------
  0 |      0.60 | [0.51, 0.68]
 83 |      0.89 | [0.82, 0.95]
167 |      0.98 | [0.93, 1.00]
250 |      1.00 | [0.97, 1.00]
333 |      1.00 | [0.99, 1.00]
500 |      1.00 | [1.00, 1.00]

# sex = male

  x | Predicted |        95% CI
-----------------------------------
  0 |      0.15 | [0.12, 0.19]
 83 |      0.27 | [0.21, 0.35]
167 |      0.43 | [0.28, 0.62]
250 |      0.60 | [0.35, 0.84]
333 |      0.75 | [0.43, 0.94]
500 |      0.93 | [0.60, 0.99]
```



70

5. Test specific hypotheses

Was the effect of fare on survival different for men vs women?

```
fit.brm_titanic %>%
  emtrends(specs = pairwise ~ sex,
           var = "fare")
```

```
$emtrends
 sex    fare.trend lower.HPD upper.HPD
 female    0.02083   0.01129    0.0316
 male      0.00845   0.00385    0.0135

Point estimate displayed: median
HPD interval probability: 0.95

$contrasts
 contrast      estimate lower.HPD upper.HPD
 female - male   0.0124  0.000884    0.0232

Point estimate displayed: median
HPD interval probability: 0.95
```

the chance of survival increased more with fare for female than male passengers

73

Going beyond

Evidence for the null hypothesis

Evidence for the null hypothesis

frontiers in Psychology

Front Psychol. 2014; 5: 781.
Published online 2014 Jul 29. doi: 10.3389/fpsyg.2014.00781

PMCID: PMC4114196
PMID: 25120503

Using Bayes to get the most out of non-significant results

Zoltan Dienes*

‣ Author information ‣ Article notes ‣ Copyright and License information Disclaimer

[HTML] Using Bayes to get the most out of non-significant results
Z Dienes - Frontiers in psychology, 2014 - frontiersin.org
No scientific conclusion follows automatically from a statistically non-significant result, yet people routinely use non-significant results to guide conclusions about the status of theories (or the effectiveness of practices). To know whether a non-significant result counts against a theory, or if it just indicates data insensitivity, researchers must use one of: power, intervals (such as confidence or credibility intervals), or else an indicator of the relative evidence for one theory over another, such as a Bayes factor. I argue Bayes factors allow theory to be …
Cited by 966 Related articles All 14 versions Web of Science: 583 Import into BibTeX

2070 now

- There is nothing special about H_0 compared to H_1 in Bayesian inference
- We can get evidence of H_0 over H_1 (e.g. using the Bayes factor approach)

Rolling the dice

Four sided

Six sided

both dice are equally likely to be picked

p() = p() = 0.5

both dice are equal sided (uniform probability over the different numbers)

Which die do you think was rolled?

4 p( | data) = ?

4, 2, 1 p( | data) = 0.77

4, 2, 1, 3, 1 p( | data) = 0.88

4, 2, 1, 3, 1, 5 p( | data) = 0

Bayes factor

probability of the data
given H_0

$$BF_{01} = \frac{p(D \mid H_0)}{p(D \mid H_1)}$$

probability of the data
given H_1

check this out

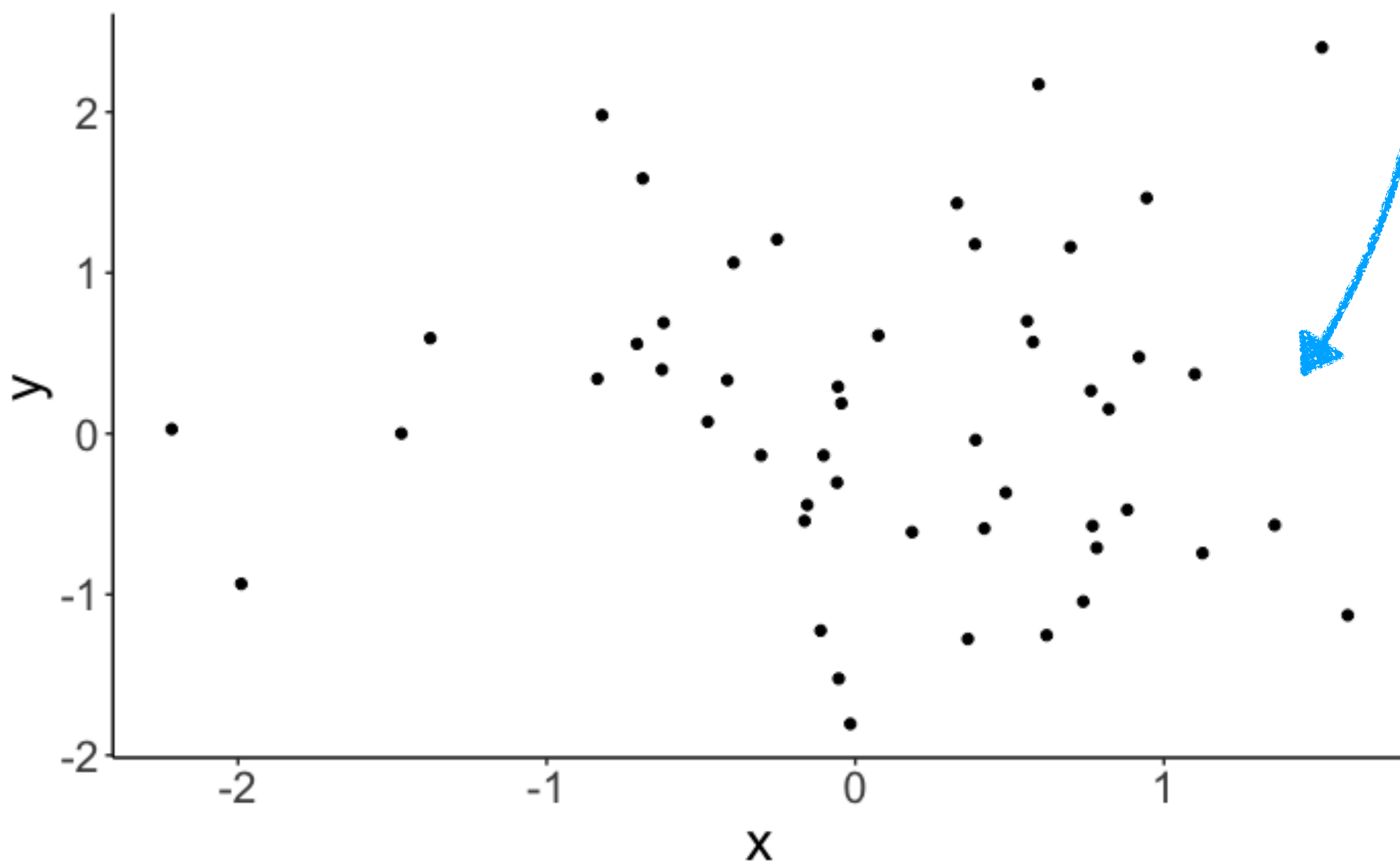
https://vuorre.netlify.com/post/2017/03/21/bayes-factors-with-brms/

Approximate LOO

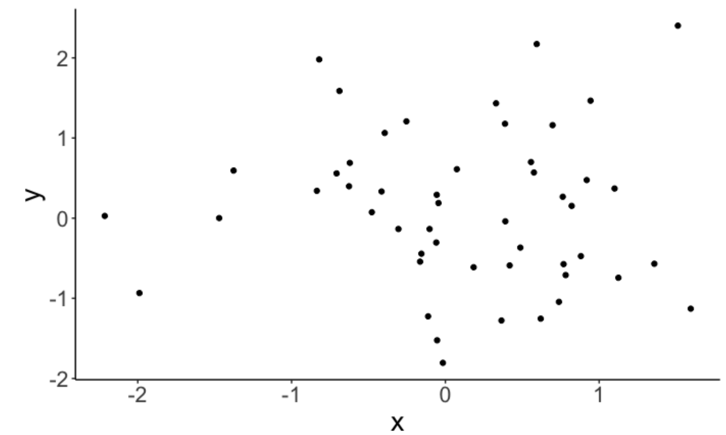
Evidence for the null hypothesis

```r
set.seed(1)
df.loo = tibble(x = rnorm(n = 50),
                y = rnorm(n = 50))

ggplot(data = df.loo,
       mapping = aes(x = x,
                     y = y)) +
  geom_point()
```

no relationship between x and y



Evidence for the null hypothesis

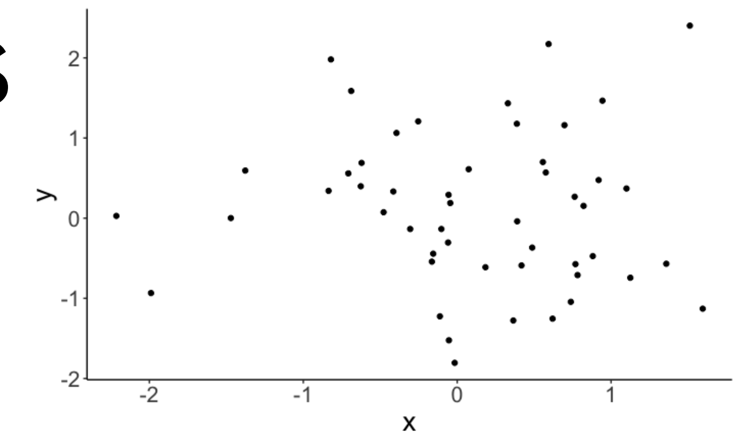



```r
fit.lm_loo = lm(formula = y ~ 1 + x,
                data = df.loo)

fit.lm_loo %>%
  summary()
```

```
Call:
lm(formula = y ~ 1 + x, data = df.loo)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.2185 -0.6735  0.0018  0.6734  4.2428 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.0006437  0.0031639   0.203    0.839
x           -0.0019184  0.0031541  -0.608    0.543

Residual standard error: 1.001 on 99998 degrees of freedom
Multiple R-squared:  3.7e-06,  Adjusted R-squared:  -6.301e-06 
F-statistic:  0.37 on 1 and 99998 DF,  p-value: 0.543
```

cannot reject the H_0 that the reduction in error due to x is what one would have expected by chance

Evidence for the null hypothesis




```r
fit.brm_loo1 = brm(formula = y ~ 1, data = df.loo)

fit.brm_loo2 = brm(formula = y ~ 1 + x, data = df.loo)

fit.brm_loo1 = add_criterion(fit.brm_loo1, criterion = "loo")

fit.brm_loo2 = add_criterion(fit.brm_loo2, criterion = "loo")
```

approximate
leave-one-out
cross-validation

```r
loo_compare(fit.brm_loo1, fit.brm_loo2)
```

```
             elpd_diff se_diff
fit.brm_loo1  0.0       0.0
fit.brm_loo2 -1.1       0.5
```

```r
model_weights(fit.brm_loo1, fit.brm_loo2)
```

```
fit.brm_loo1 fit.brm_loo2
99.99999     0.00001
```

I want only positive coefficients!

I only want positive coefficients!

```
brm(formula = how_much_i_love_stats ~ 1 + tobi + ari + sarah + chengxu,
    data = df.stats_love)
```

coefficients in the model

```
# priors
priors = c(set_prior("normal(0,10)", class = "b", lb = 0))
```

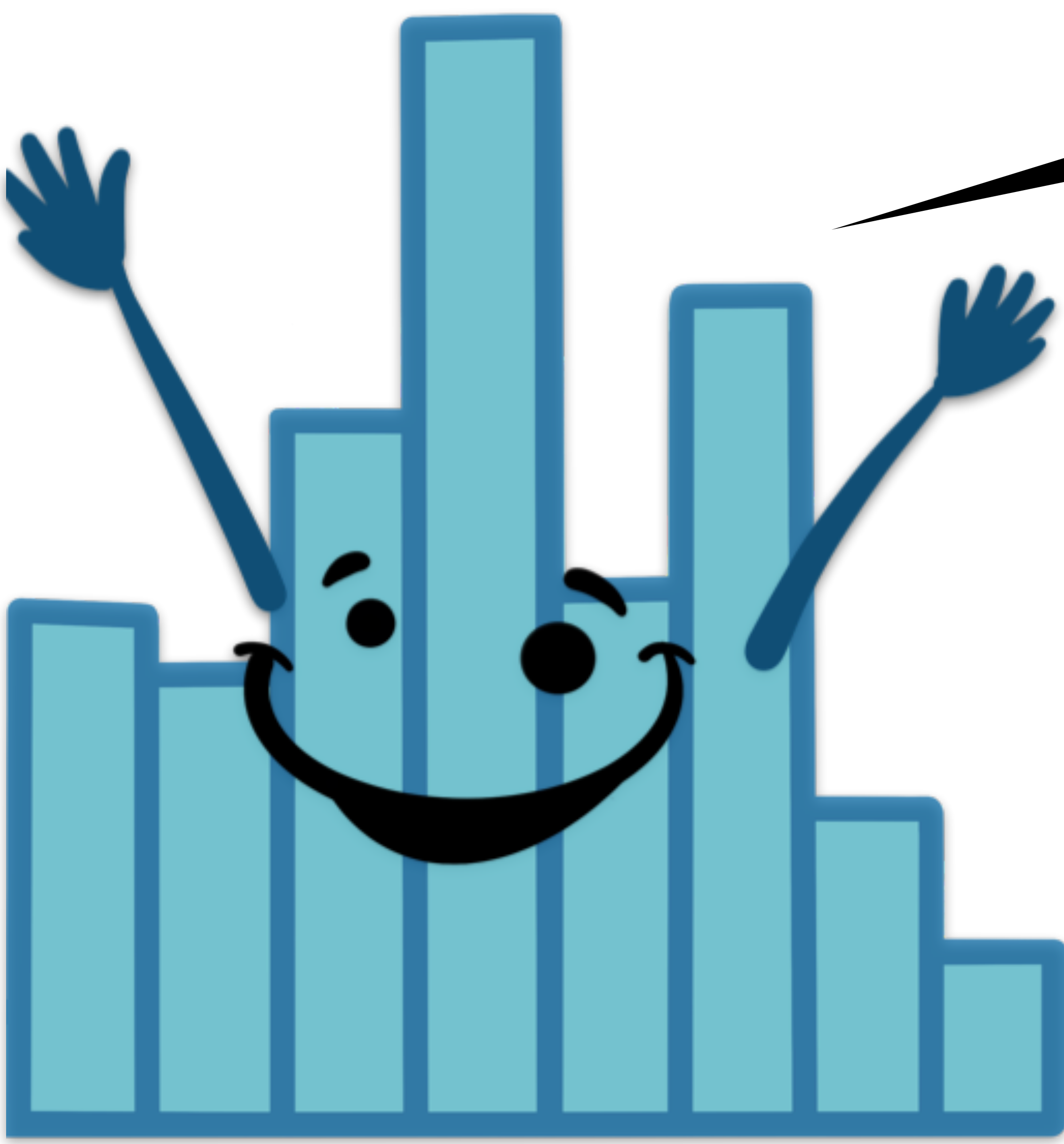
lower bound = 0

```
brm(formula = how_much_i_love_stats ~ 1 + tobi + ari + sarah + chengxu,
    prior = priors,
    data = df.stats_love)
```

We're listening to "Big Bad Trumpet Player" by "Kormac"

02:00

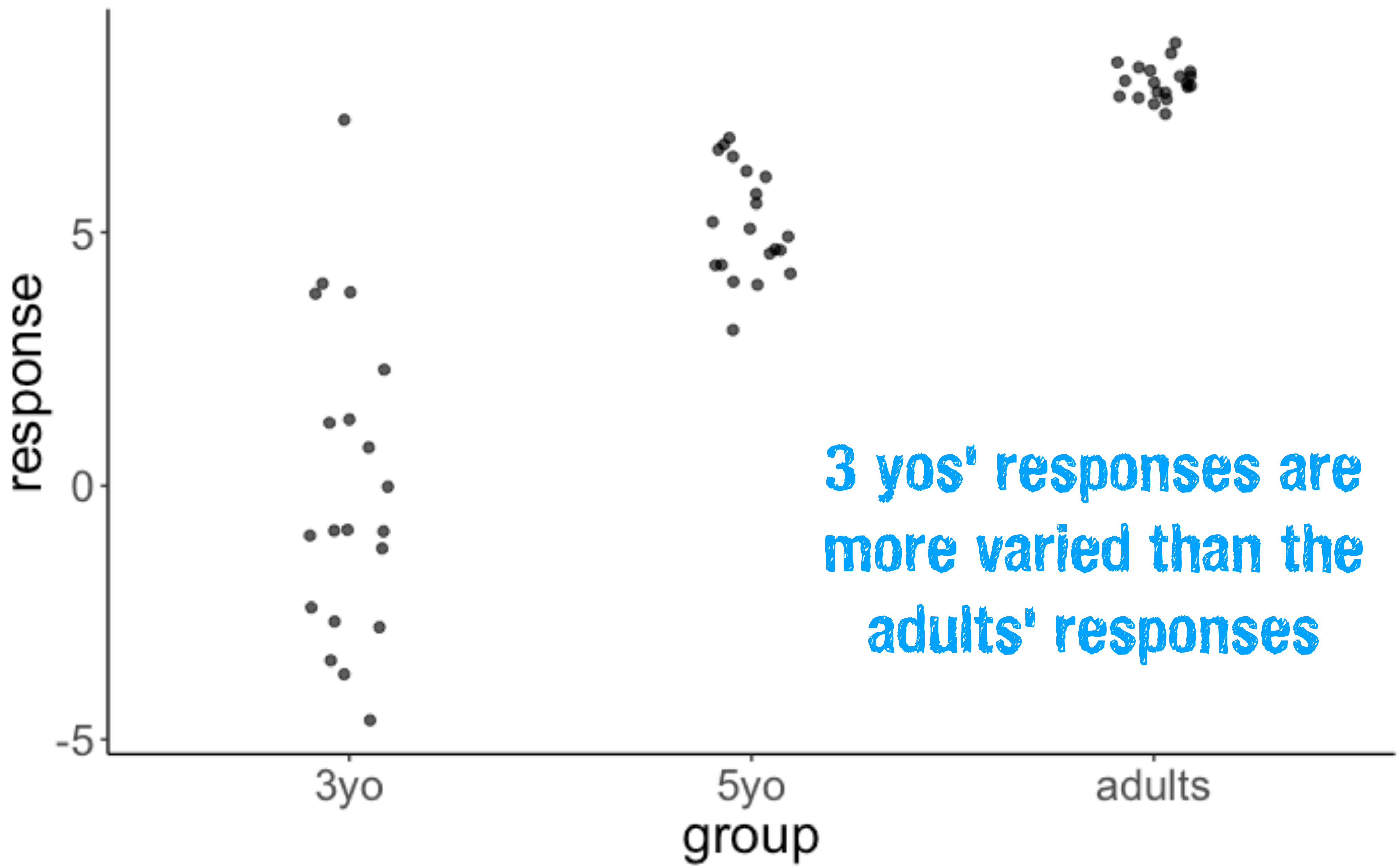
stretch break!



Dealing with unequal variance

Unequal variance aka heteroscedasticity

```r
df.variance = tibble(group = rep(c("3yo", "5yo", "adults"), each = 20),
                     response = rnorm(n = 60,
                                      mean = rep(c(0, 5, 8), each = 20),
                                      sd = rep(c(3, 1.5, 0.3), each = 20)))
```



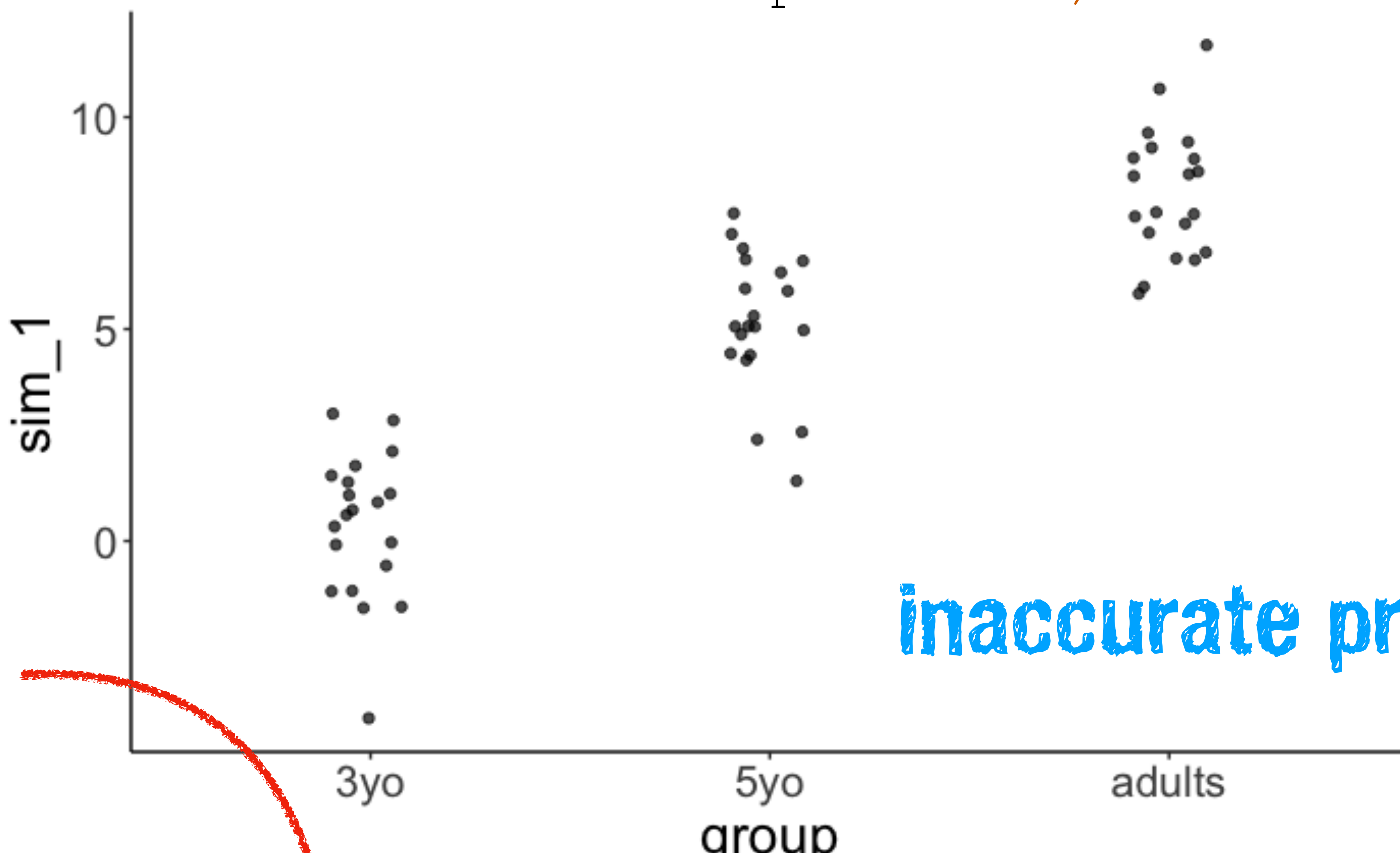
Unequal variance aka heteroscedasticity

```
fit.lm1 = lm(formula = response ~ 1 + group,
             data = df.variance)

fit.lm1 %>%
  summary()
```

```
Call:
lm(formula = response ~ 1 + group, data = df.variance)

Residuals:
    Min      1Q  Median      3Q     Max 
-4.6145 -0.8288 -0.0879  0.6315  7.2193 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) -0.005336   0.421618  -0.013     0.99    
group5yo     5.172810   0.596258   8.675 5.25e-12 ***
groupadults  7.970655   0.596258  13.368  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 1.886 on 57 degrees of freedom
Multiple R-squared:  0.7635, Adjusted R-squared:  0.7552 
F-statistic: 91.99 on 2 and 57 DF,  p-value: < 2.2e-16
```

Unequal variance aka heteroscedasticity

```r
fit.lm1 %>%
  simulate() %>%
  bind_cols(df.variance) %>%
  ggplot(aes(x = group, y = sim_1)) +
  geom_jitter(height = 0,
              width = 0.1,
              alpha = 0.7)
```



only one standard deviation fitted

inaccurate predictions ...

r.squared	adj.r.squared	sigma	statistic	p.value	df	logLik	AIC	BIC	deviance	df.residual
0.76	0.76	1.89	91.99	0	3	-121.65	251.3	259.68	202.65	57

Unequal variance aka heteroscedasticity

```r
fit.brm1 = brm(formula = bf(response ~ group,
                            sigma ~ group),
               data = df.variance,
               file = "cache/brm1",
               seed = 1)
```

modeling both the means and variances

```
 Family: gaussian
  Links: mu = identity; sigma = log
Formula: response ~ group
         sigma ~ group
   Data: df.variance (Number of observations: 60)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
                    Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept              -0.01      0.73    -1.41     1.51 1.01     1107     1072
sigma_Intercept         1.15      0.17     0.85     1.51 1.00     1991     1922
group5yo                5.18      0.77     3.60     6.65 1.00     1252     1327
groupadults             7.98      0.74     6.47     9.37 1.01     1110     1079
sigma_group5yo         -1.05      0.24    -1.51    -0.57 1.00     2249     2420
sigma_groupadults      -2.19      0.24    -2.66    -1.74 1.00     2171     2427

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Unequal variance aka heteroscedasticity

```
 Family: gaussian
  Links: mu = identity; sigma = log
Formula: response ~ group
         sigma ~ group
   Data: df.variance (Number of observations: 60)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
                  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept            -0.01      0.73    -1.41     1.51 1.01     1107     1072
sigma_Intercept       1.15      0.17     0.85     1.51 1.00     1991     1922
group5yo              5.18      0.77     3.60     6.65 1.00     1252     1327
groupadults           7.98      0.74     6.47     9.37 1.01     1110     1079
sigma_group5yo       -1.05      0.24    -1.51    -0.57 1.00     2249     2420
sigma_groupadults    -2.19      0.24    -2.66    -1.74 1.00     2171     2427

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

on a log scale!

```
mean = c(0, 5, 8)
sd = c(3, 1.5, 0.3)
```

pretty good!

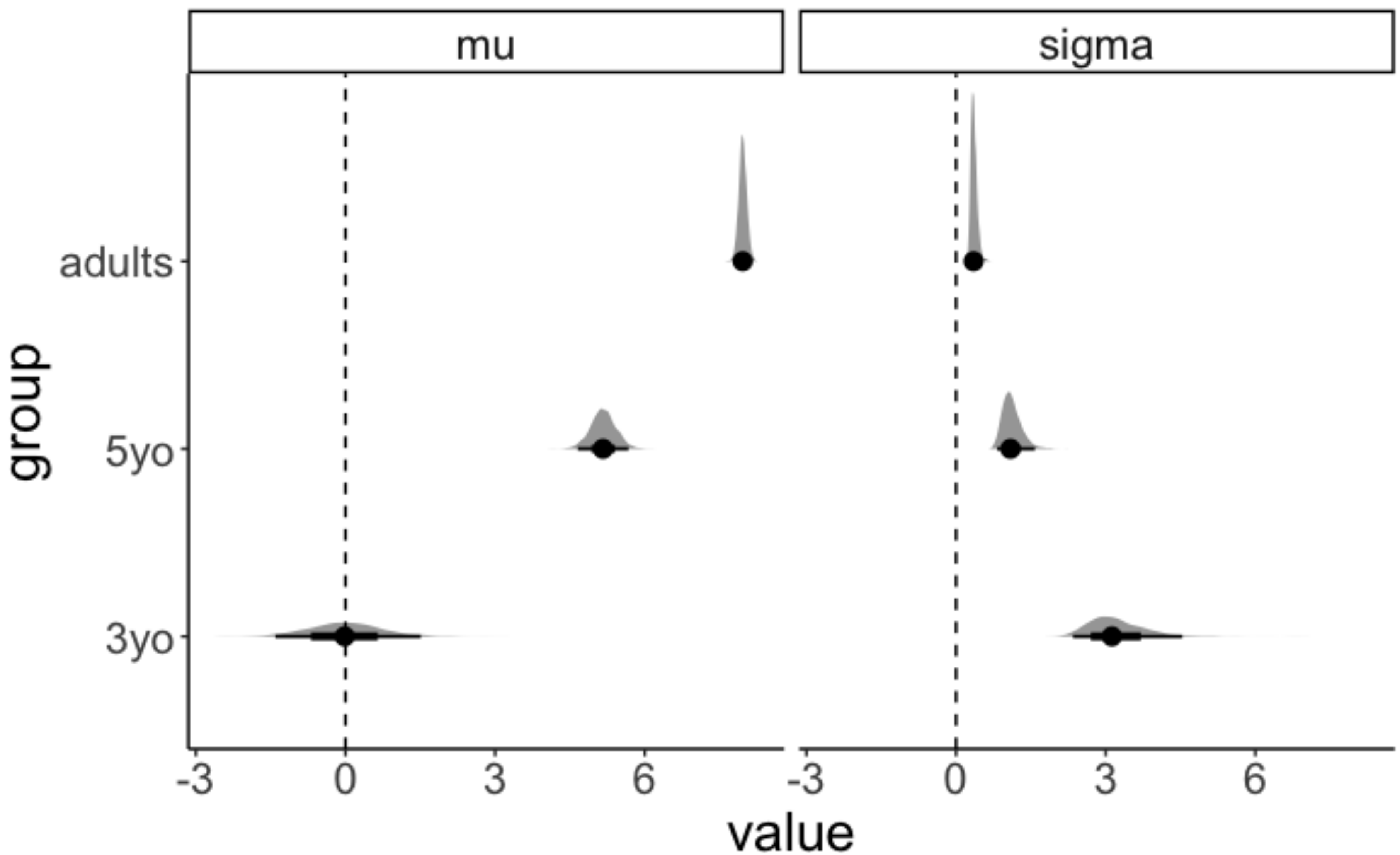
3 year olds $e^{1.15} = 3.16$

5 year olds $e^{1.15+(-1.05)} = 1.10$

adults $e^{1.15+(-2.19)} = 0.35$

Unequal variance aka heteroscedasticity

```
df.variance %>%
  expand(group) %>%
  add_fitted_draws(fit.brm1, dpar = TRUE) %>%
  select(group, .row, .draw, posterior = .value, mu, sigma) %>%
  pivot_longer(cols = c(mu, sigma),
               names_to = "index",
               values_to = "value") %>%
  ggplot(aes(x = value, y = group)) +
  geom_halfeyeh() +
  geom_vline(xintercept = 0, linetype = "dashed") +
  facet_grid(cols = vars(index))
```

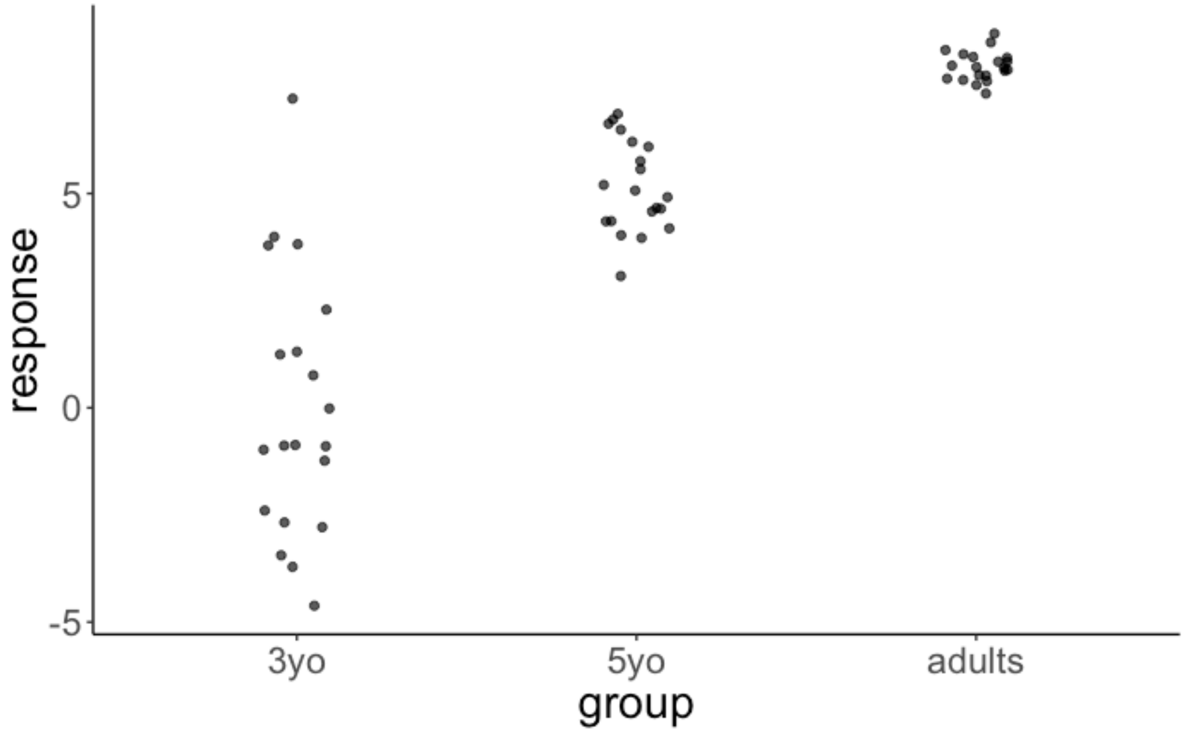


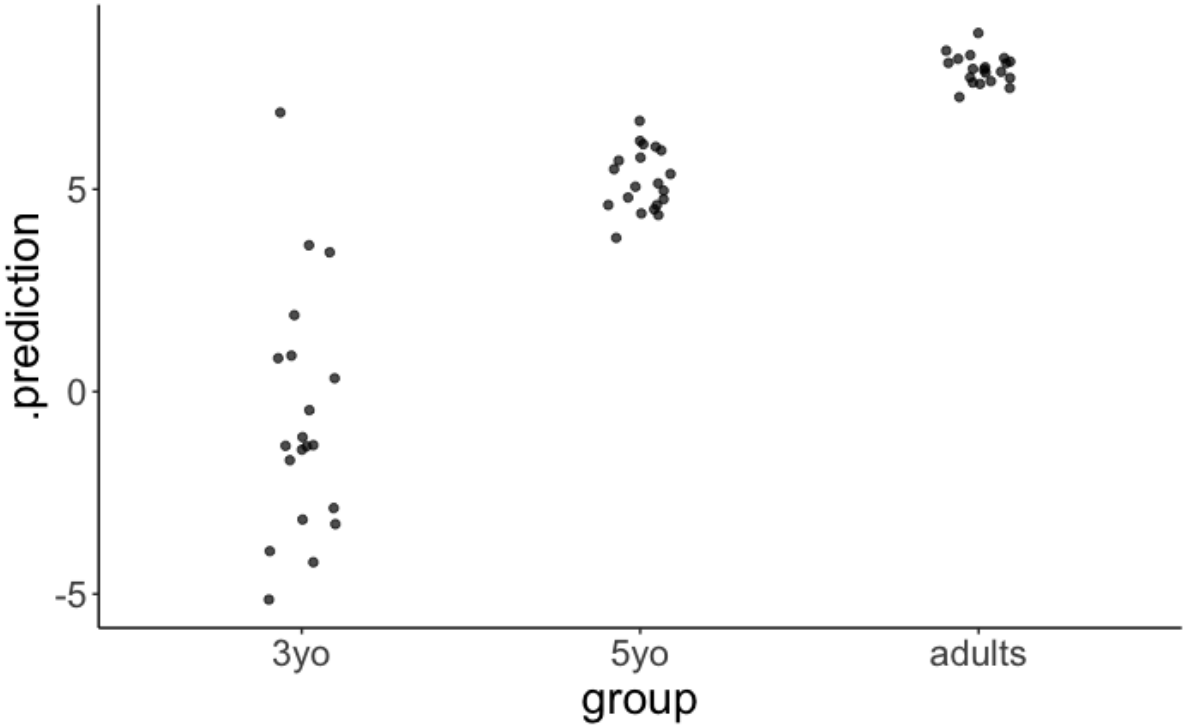
Unequal variance aka heteroscedasticity

```
df.variance %>%
  add_predicted_draws(model = fit.brm1,
                      n = 1) %>%
  ggplot(aes(x = group, y = .prediction)) +
  geom_jitter(height = 0,
              width = 0.1,
              alpha = 0.7)
```

original data

predicted data





these predictions look good!

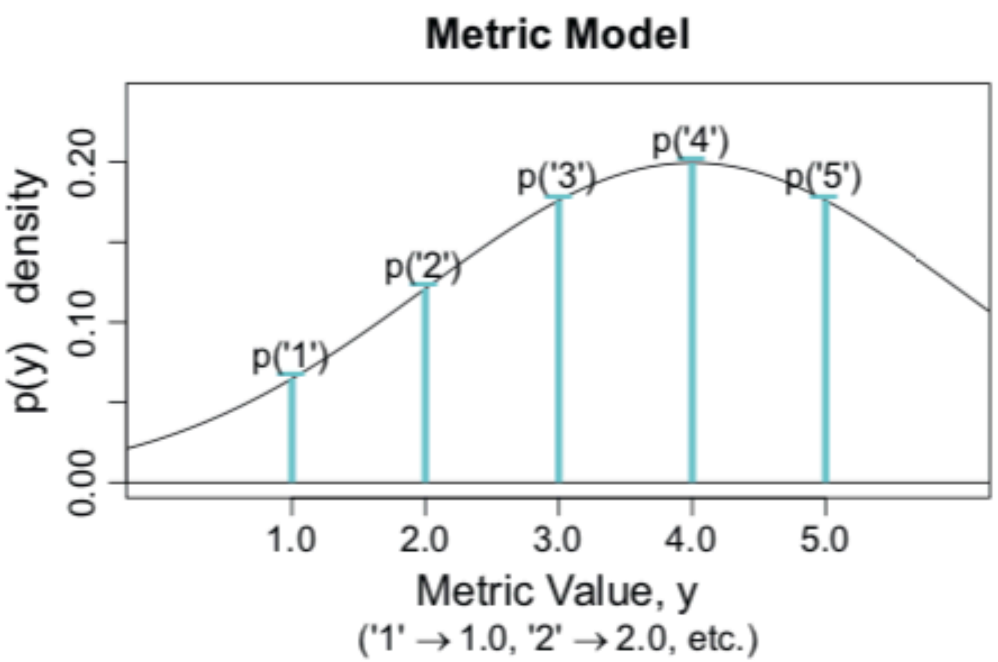
Better modeling Likert scale

We surveyed all articles in the *Journal of Personality and Social Psychology (JPSP)*, *Psychological Science (PS)*, and the *Journal of Experimental Psychology: General (JEP:G)* that mentioned the term "Likert," and found that **100% of the articles** that analyzed ordinal data did so using a metric model.

great paper!

Liddell & Kruschke (2018) Analyzing ordinal data with metric models: What could possibly go wrong?. *Journal of Experimental Social Psychology*

Ordinal regression



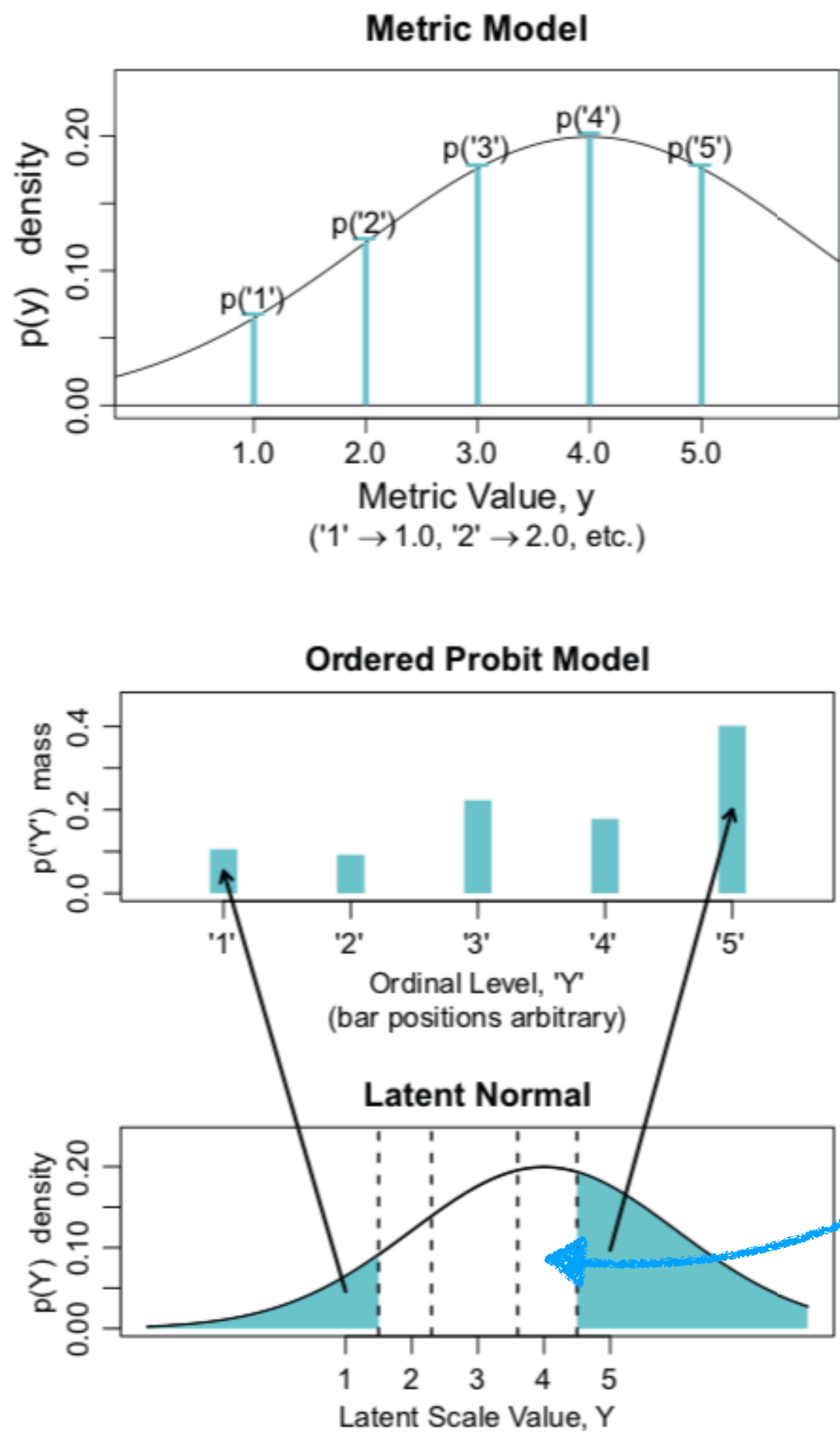
metric model

- fits a Gaussian distribution with mean and standard deviation
- makes the assumption that categories are **equidistant**

I looooove Bayesian statistics!

- ○ completely disagree (1)
- ○ moderately disagree (2)
- ○ neither disagree nor agree (3)
- ○ moderately agree (4)
- ◉ completely agree (5)

Ordinal regression



ordered probit model

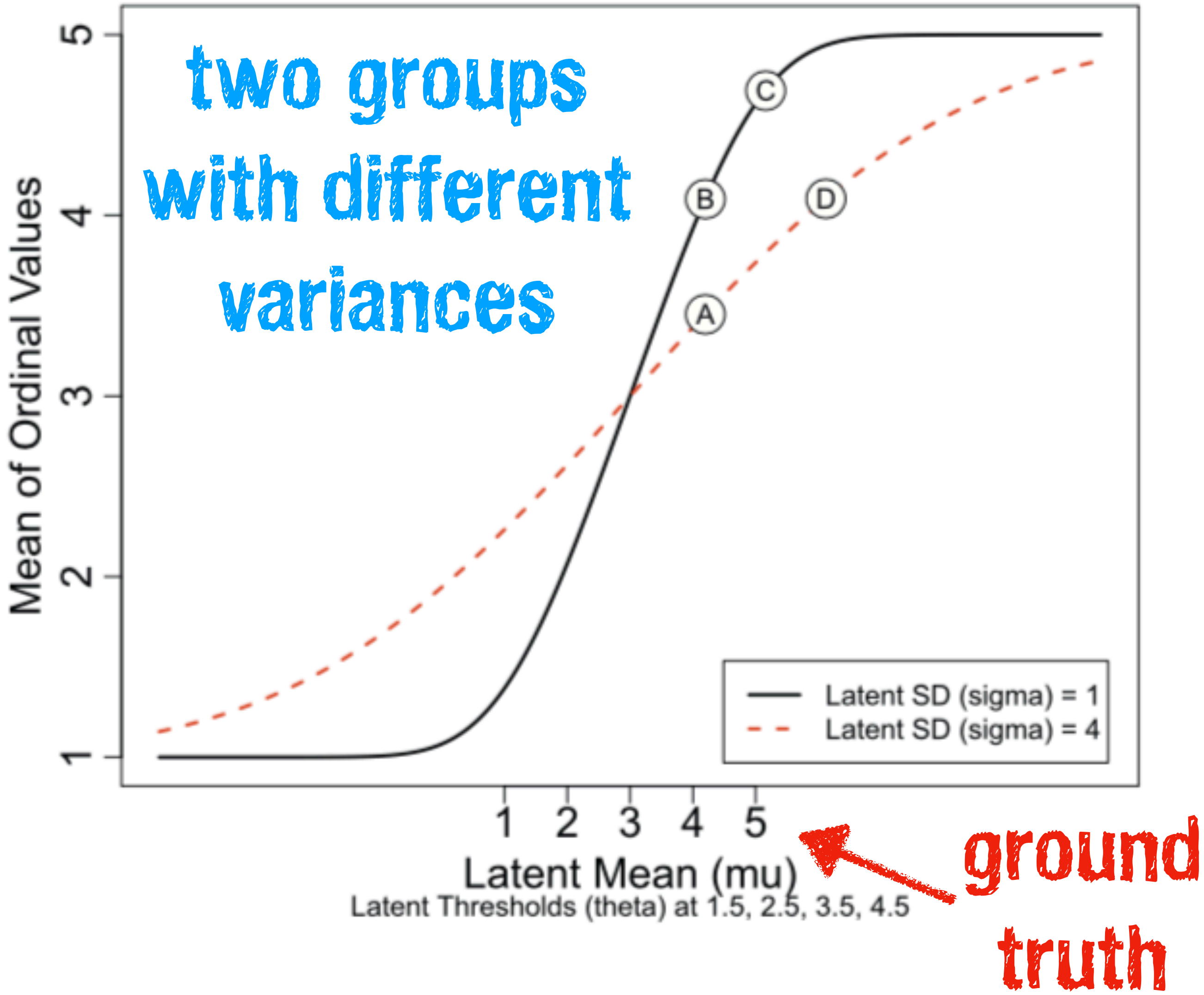
- assumes an ordering of the categories
- assumes a normal distribution in latent space
- finds a good mapping from that normal distribution in latent space to the ordered responses

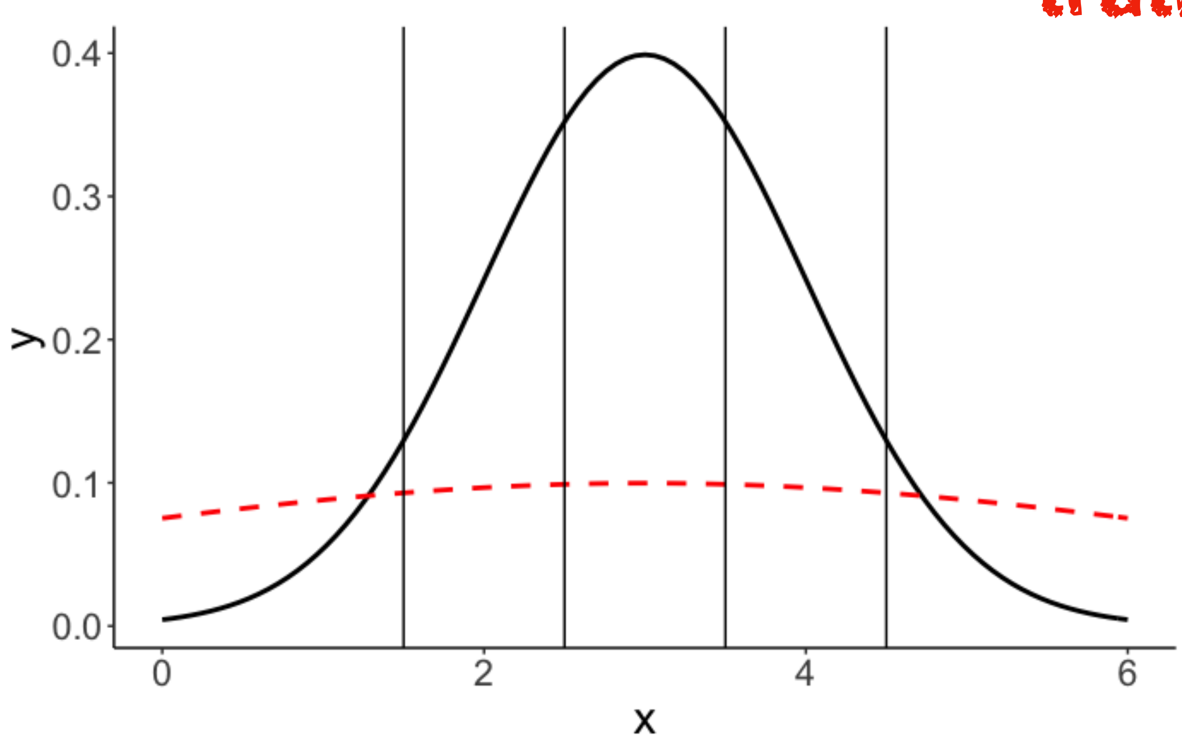
finds thresholds that separate the categories

Does it matter?

- treating ordinal data as metric can lead to:
  - low rates of correct detection (Type II error)
  - distorted effect size estimates
  - inflated false alarms (Type I error)
  - inversions of differences between groups
- main reasons for why this happens:
  - response categories may not be equidistant
  - response distribution may be non-normal
  - variances of unobserved variables may differ between groups, conditions, time points, ...

What could possibly go wrong?





- **A** vs. **B**: false positive (Type I error)

  no difference in latent space

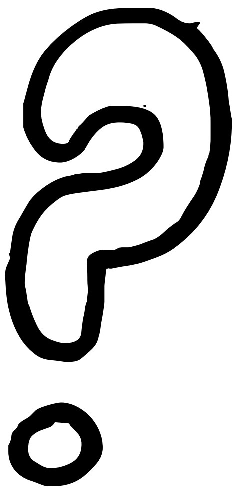
- **B** vs. **D**: false negative (Type II error)

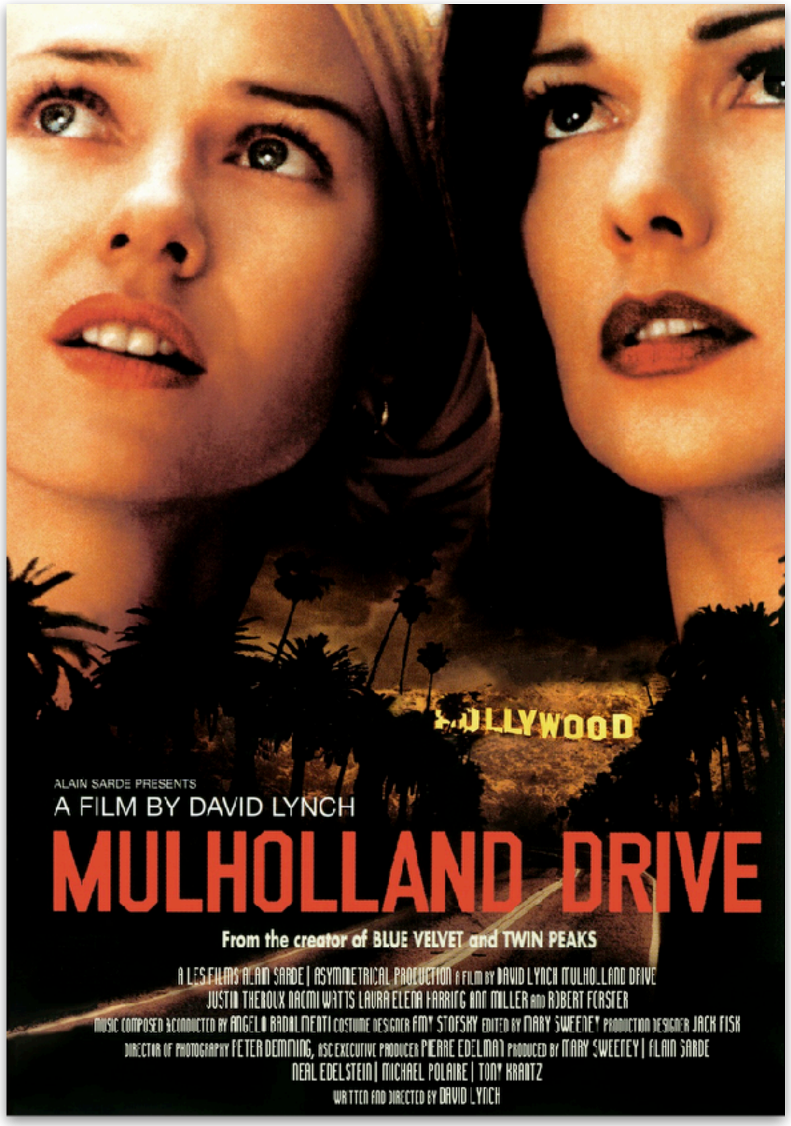
  difference in latent space

- **C** vs. **D**: reversal

  C is greater than D in metric space, but D is greater than C in latent space

Which movie shall I watch?





Customer reviews

4.5 out of 5

2,029 customer ratings

5 star	78%
4 star	10%
3 star	5%
2 star	3%
1 star	5%

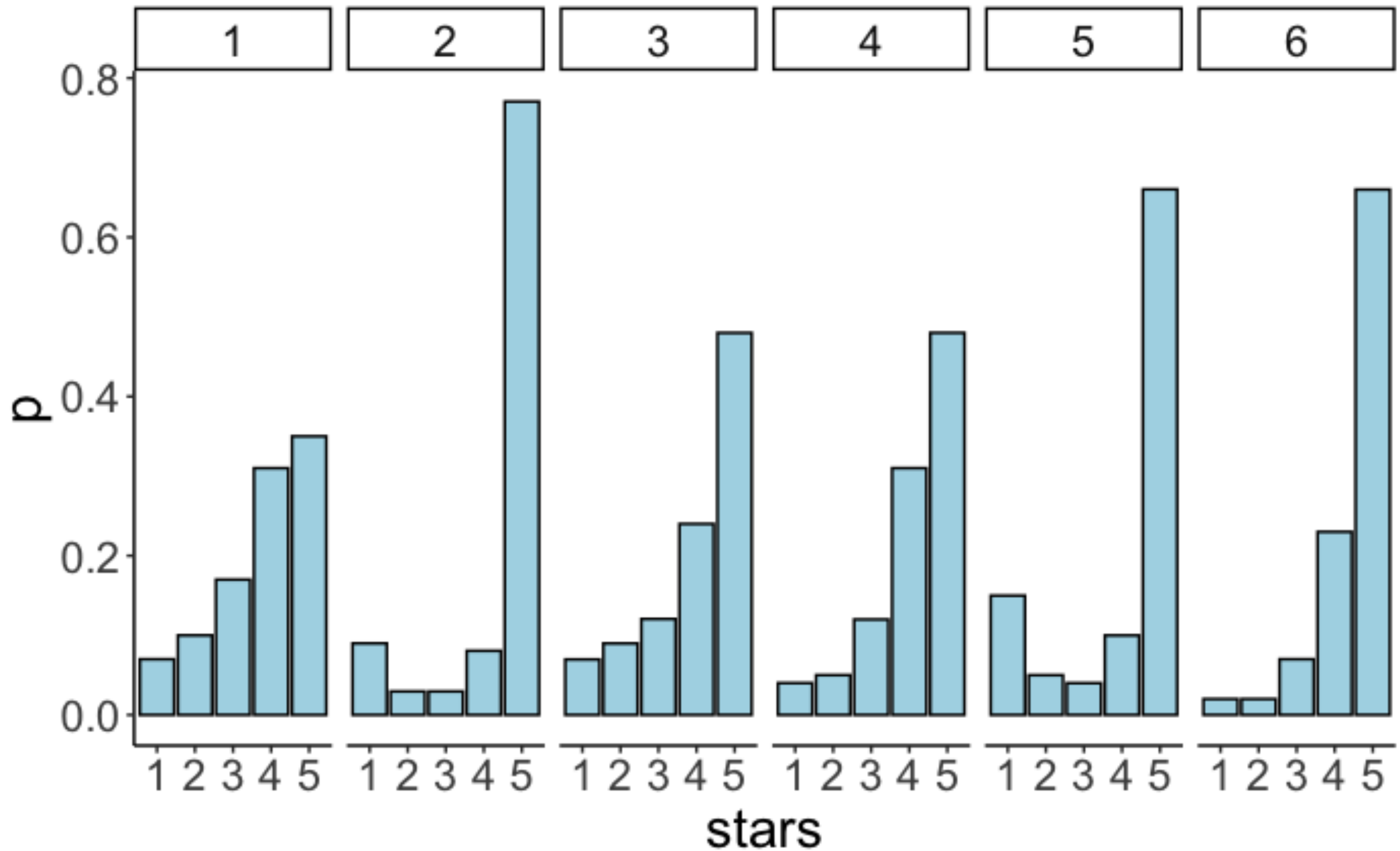
Customer reviews

3.8 out of 5

1,558 customer ratings

5 star	53%
4 star	13%
3 star	8%
2 star	8%
1 star	17%

Amazon movie ratings



Fit the ordinal regression model

```r
fit.brm5 = brm(formula = stars ~ 1 + id,
               family = cumulative(link = "probit"),
               data = df.movies,
               file = "cache/brm5",
               seed = 1)
```

linking function

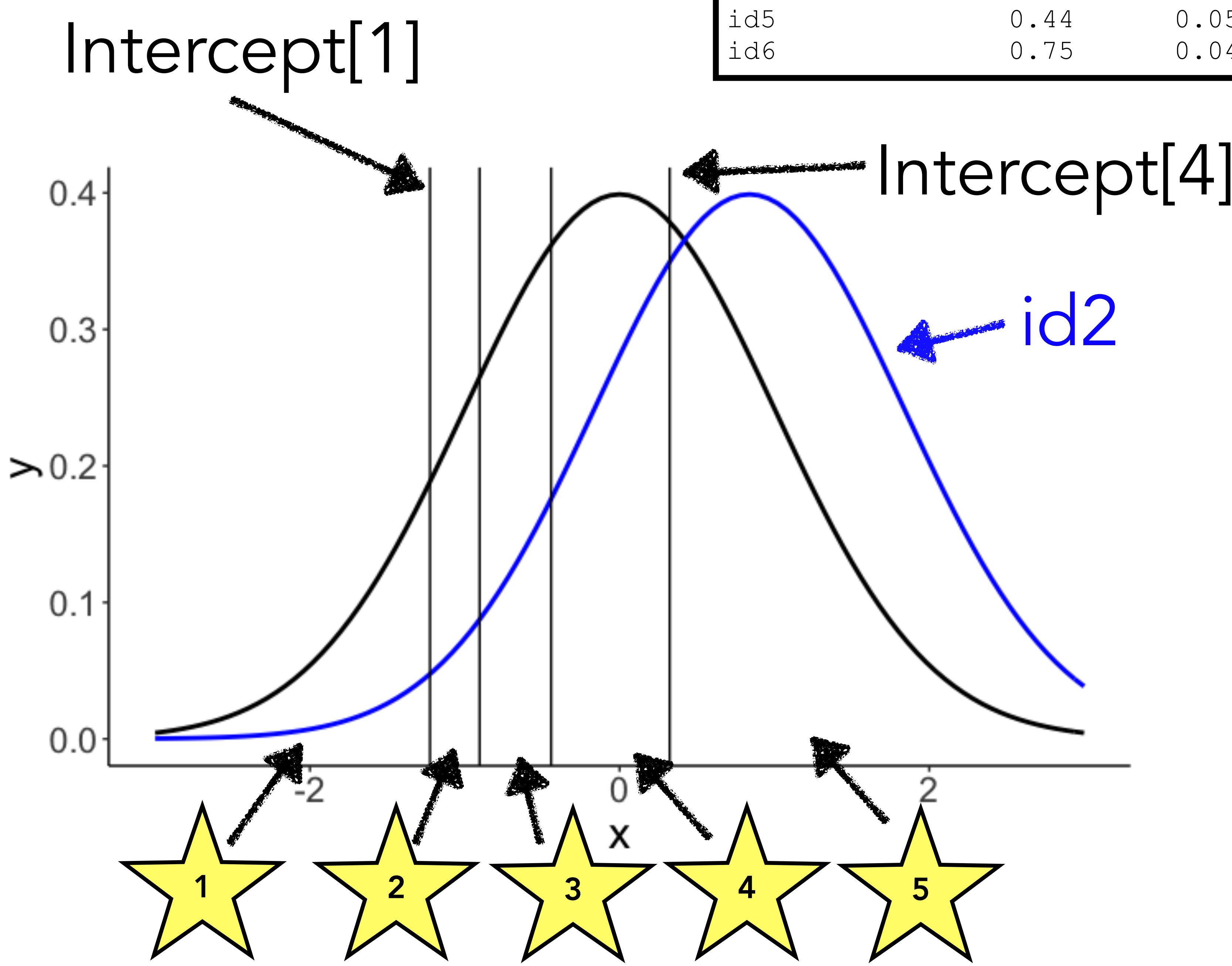
```
Family: cumulative
  Links: mu = probit; disc = identity
Formula: stars ~ 1 + id
   Data: df.movies (Number of observations: 21708)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -1.22      0.04    -1.31    -1.14 1.00     1877     2488
Intercept[2]    -0.90      0.04    -0.98    -0.82 1.00     1787     2419
Intercept[3]    -0.44      0.04    -0.52    -0.36 1.00     1692     2185
Intercept[4]     0.32      0.04     0.24     0.40 1.00     1634     2101
id2              0.84      0.06     0.71     0.96 1.00     2354     2553
id3              0.22      0.05     0.11     0.32 1.00     2146     2516
id4              0.33      0.04     0.24     0.41 1.00     1647     2315
id5              0.44      0.05     0.34     0.54 1.00     1982     2608
id6              0.75      0.04     0.67     0.83 1.00     1659     2158

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

thresholds

difference in mean to reference category

Interpretation

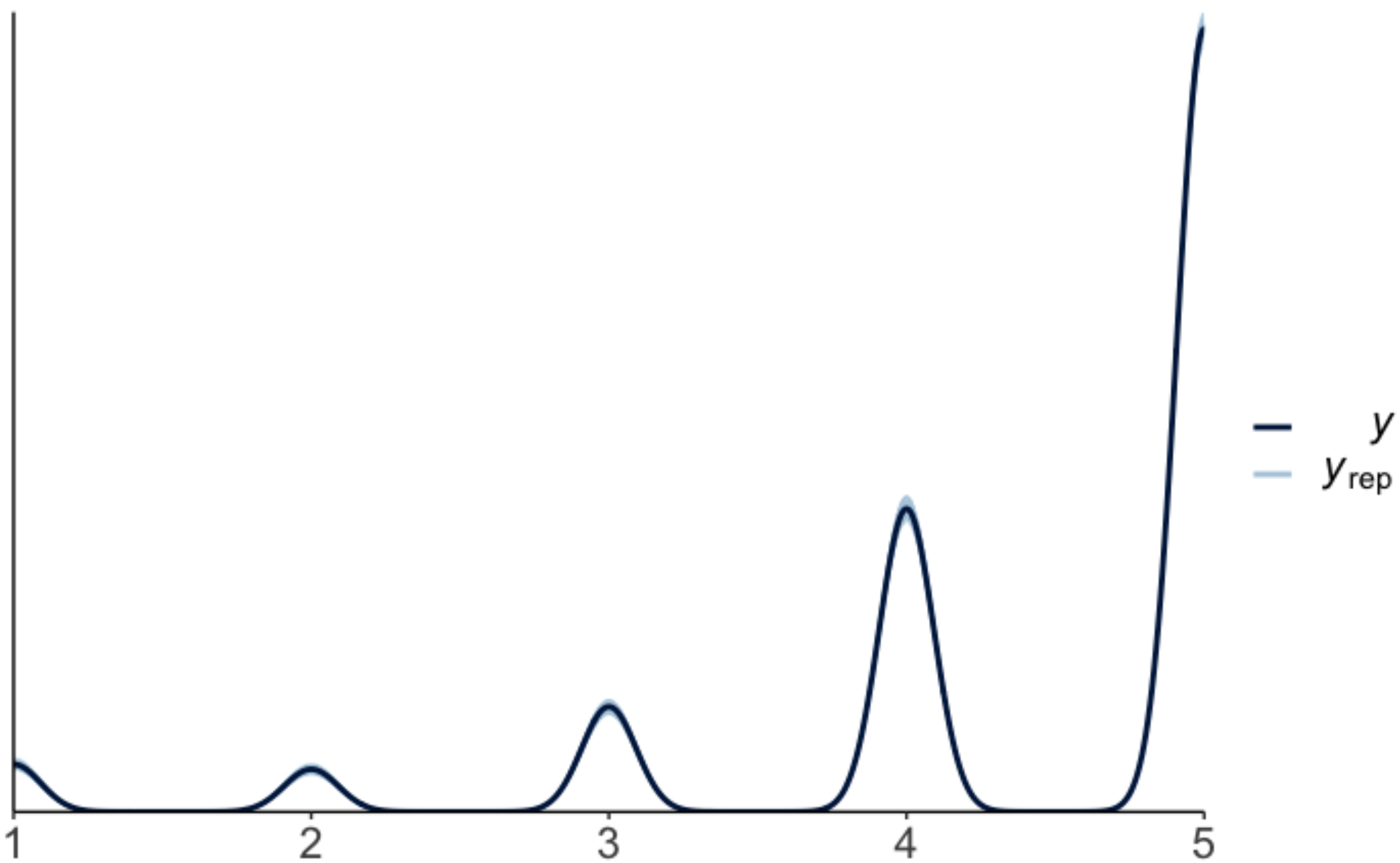
```
Family: cumulative
  Links: mu = probit; disc = identity
Formula: stars ~ 1 + id
   Data: df.movies (Number of observations: 21708)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -1.22      0.04    -1.31    -1.14 1.00     1877     2488
Intercept[2]    -0.90      0.04    -0.98    -0.82 1.00     1787     2419
Intercept[3]    -0.44      0.04    -0.52    -0.36 1.00     1692     2185
Intercept[4]     0.32      0.04     0.24     0.40 1.00     1634     2101
id2              0.84      0.06     0.71     0.96 1.00     2354     2553
id3              0.22      0.05     0.11     0.32 1.00     2146     2516
id4              0.33      0.04     0.24     0.41 1.00     1647     2315
id5              0.44      0.05     0.34     0.54 1.00     1982     2608
id6              0.75      0.04     0.67     0.83 1.00     1659     2158
```

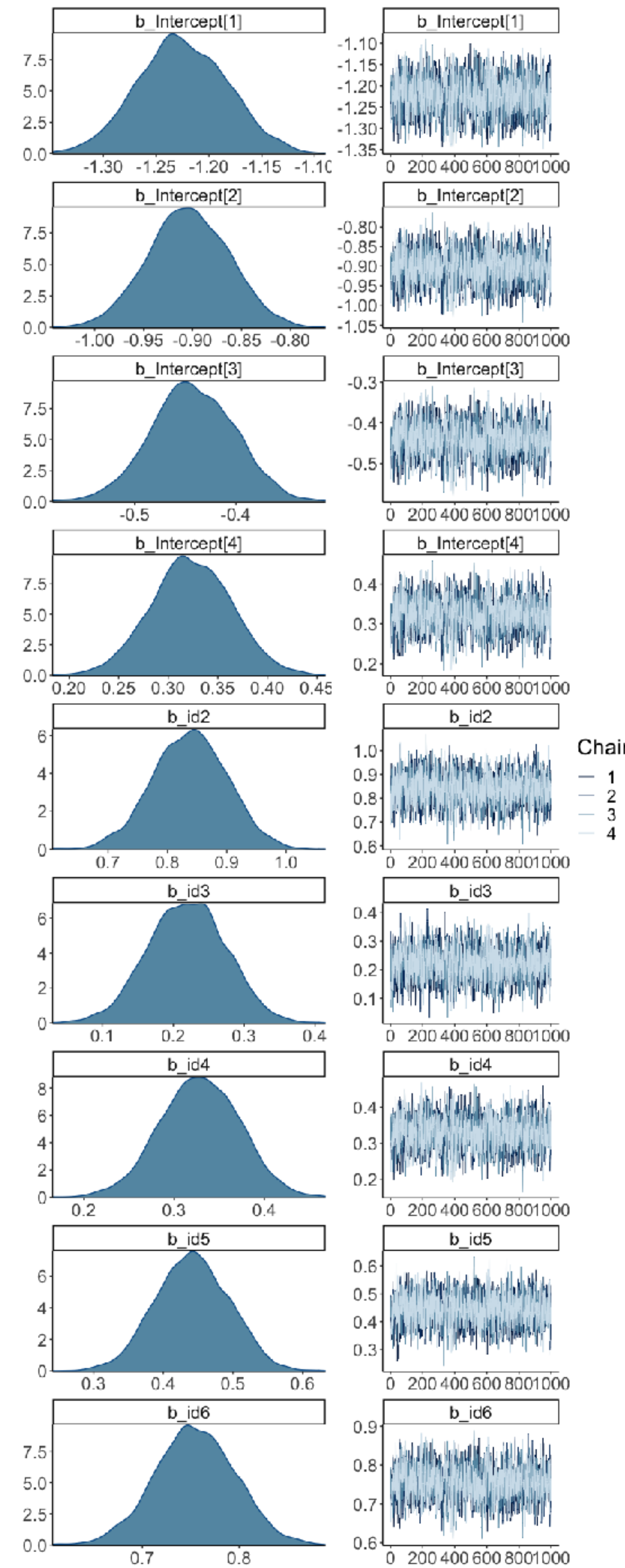


Check the model

```
fit.brm5 %>%
  plot(N = 9)
```

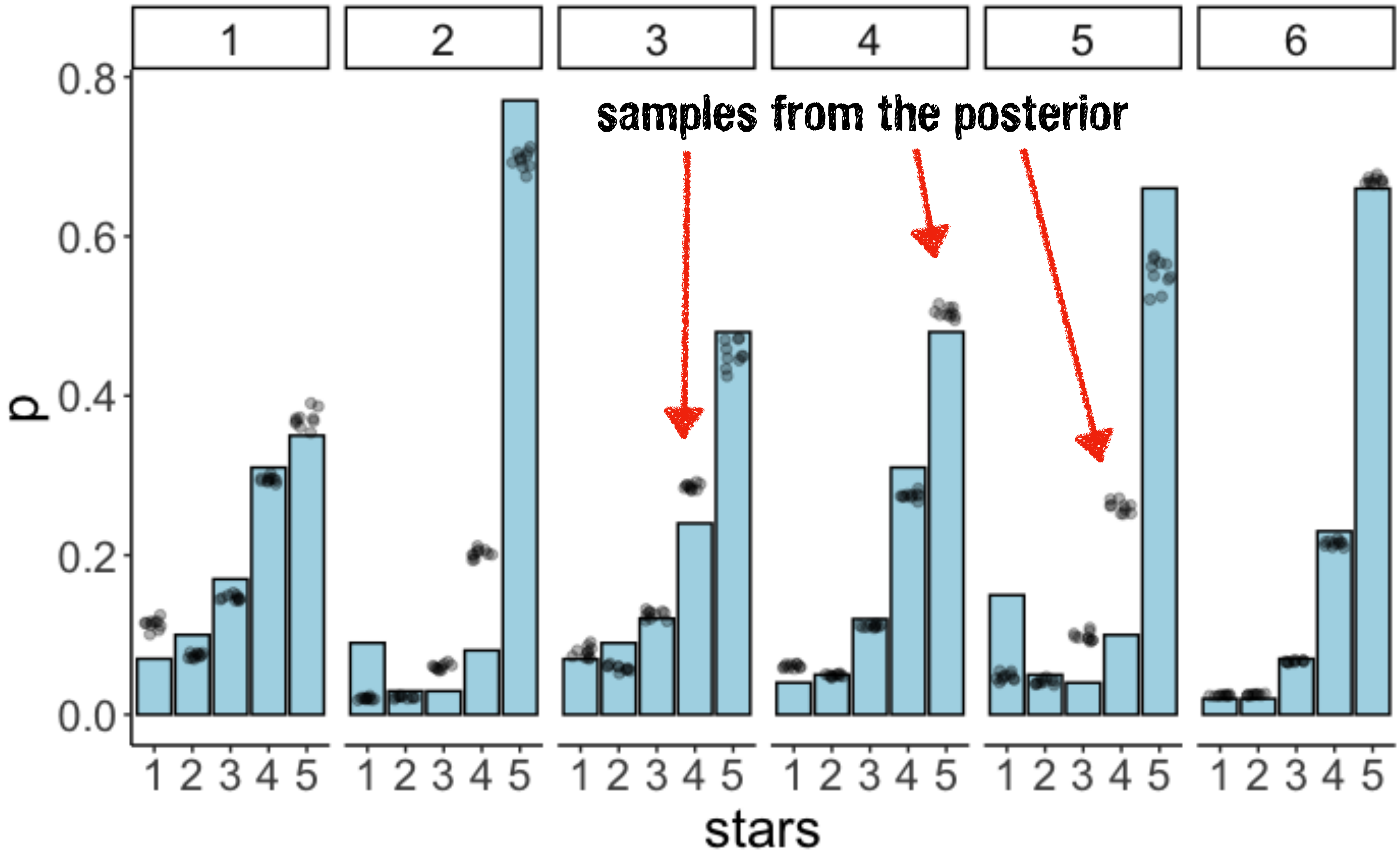
```
fit.brm5 %>%
  pp_check(nsamples = 100)
```





looking good!

Illustrate the predictions



predictions look pretty good but
maybe we can do better?

Before doing better, let's do worse!

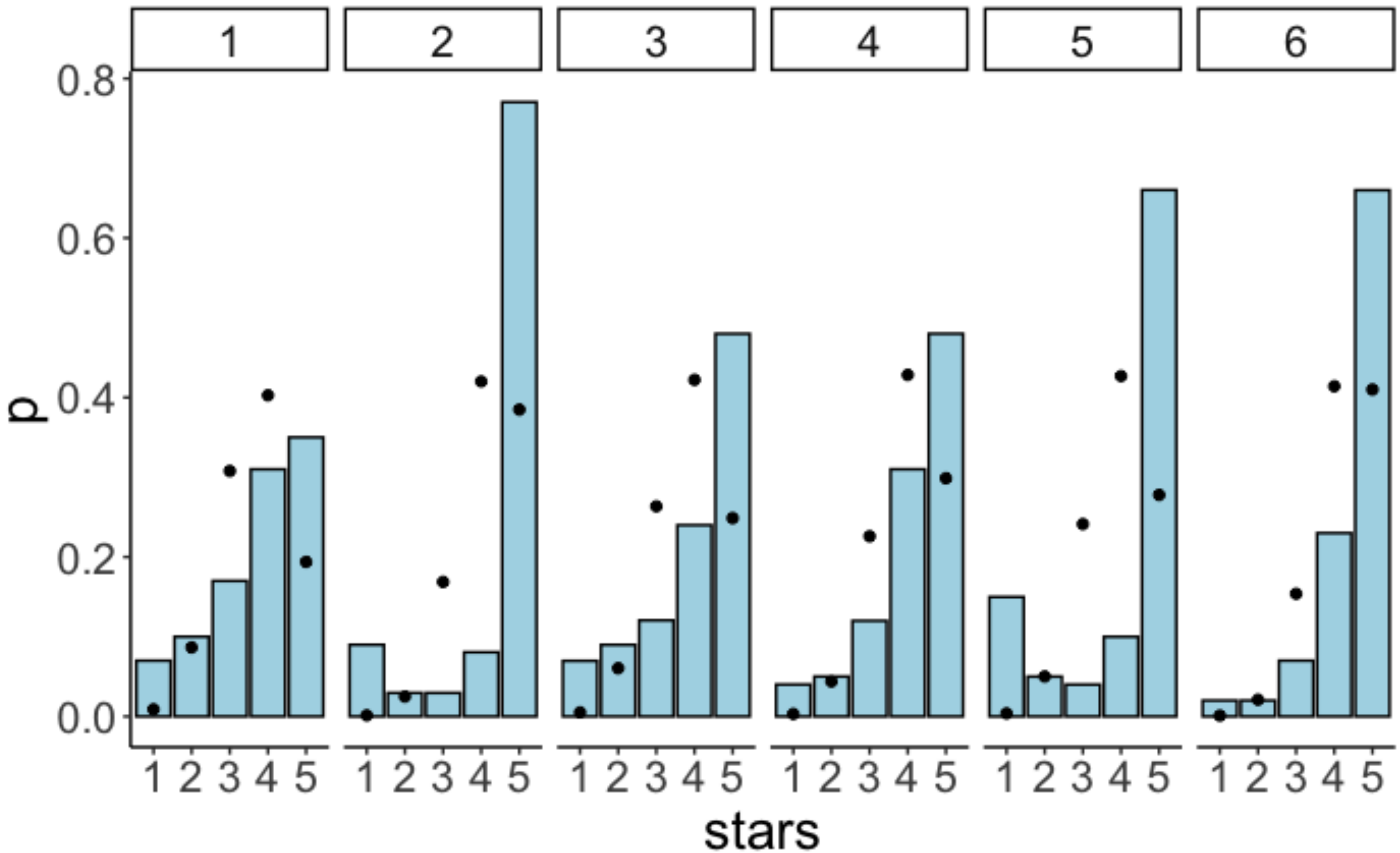
```r
fit.brm6 = brm(formula = stars ~ 1 + id,
               data = df.movies,
               file = "cache/brm6",
               seed = 1)
```

mean for the reference category

difference in mean to reference category

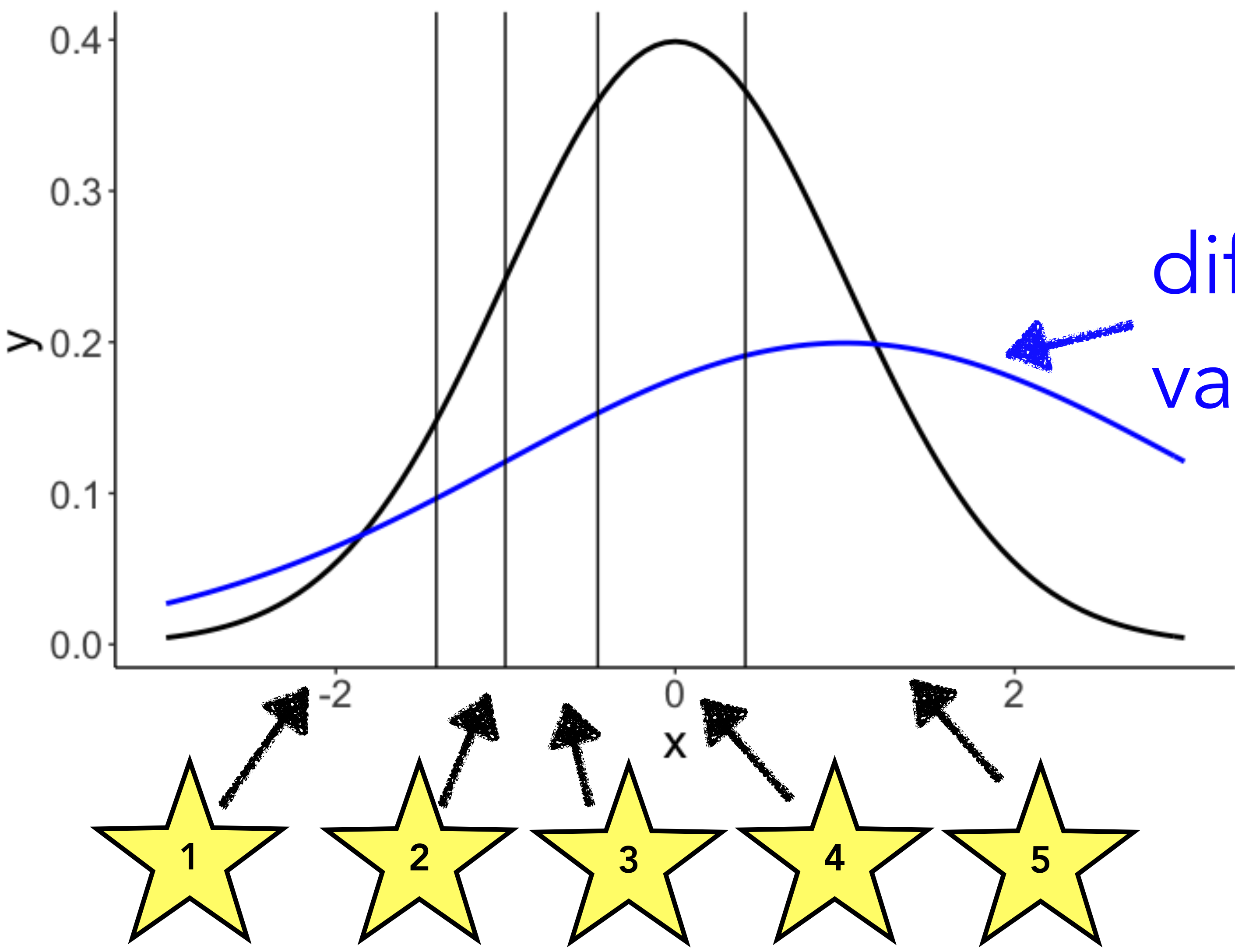
```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: stars ~ 1 + id
   Data: df.movies (Number of observations: 21708)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
          Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept     3.77      0.04     3.70     3.84 1.00     1203     1621
id2           0.64      0.05     0.54     0.75 1.00     1605     2335
id3           0.20      0.05     0.10     0.30 1.00     1558     2147
id4           0.37      0.04     0.29     0.45 1.00     1267     1862
id5           0.30      0.05     0.21     0.40 1.00     1441     2154
id6           0.72      0.04     0.65     0.79 1.00     1205     1720

Family Specific Parameters:
      Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sigma     1.00      0.00     0.99     1.01 1.00     3300     2723

Samples were drawn using sampling(NUTS). For each parameter, Bulk_ESS
and Tail_ESS are effective sample size measures, and Rhat is the potential
scale reduction factor on split chains (at convergence, Rhat = 1).
```

Predictions from the metric model



that's not looking good ...

Let's relax the assumption of equal variances



different mean and variance

Fit a model that doesn't assume equal variance

```r
fit.brm7 = brm(formula = bf(stars ~ 1 + id) + lf(disc ~ 0 + id, cmc = FALSE),
               family = cumulative(link = "probit"),
               data = df.movies,
               file = "cache/brm7",
               seed = 1)
```

tricky formula

```
 Family: cumulative 
  Links: mu = probit; disc = log 
Formula: stars ~ 1 + id 
         disc ~ 0 + id
   Data: df.movies (Number of observations: 21708) 
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects: 
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept[1]    -1.41      0.06    -1.53    -1.29 1.00     1484     2421
Intercept[2]    -1.00      0.05    -1.10    -0.90 1.00     1852     2561
Intercept[3]    -0.46      0.04    -0.54    -0.37 1.00     2405     2684
Intercept[4]     0.41      0.05     0.32     0.51 1.00     1336     2161
id2              2.71      0.33     2.14     3.44 1.00     1681     1865
id3              0.33      0.07     0.20     0.47 1.00     1961     2618
id4              0.36      0.05     0.26     0.46 1.00     1525     2753
id5              1.65      0.17     1.34     2.00 1.00     1929     2281
id6              0.86      0.06     0.74     0.98 1.00     1112     1769
disc_id2        -1.12      0.10    -1.33    -0.94 1.00     1672     1943
disc_id3        -0.23      0.06    -0.34    -0.11 1.00     1342     1955
disc_id4        -0.01      0.04    -0.09     0.07 1.00     1043     1747
disc_id5        -1.09      0.07    -1.23    -0.95 1.00     1681     1996
disc_id6        -0.08      0.04    -0.15     0.00 1.00      941     1489
```

thresholds

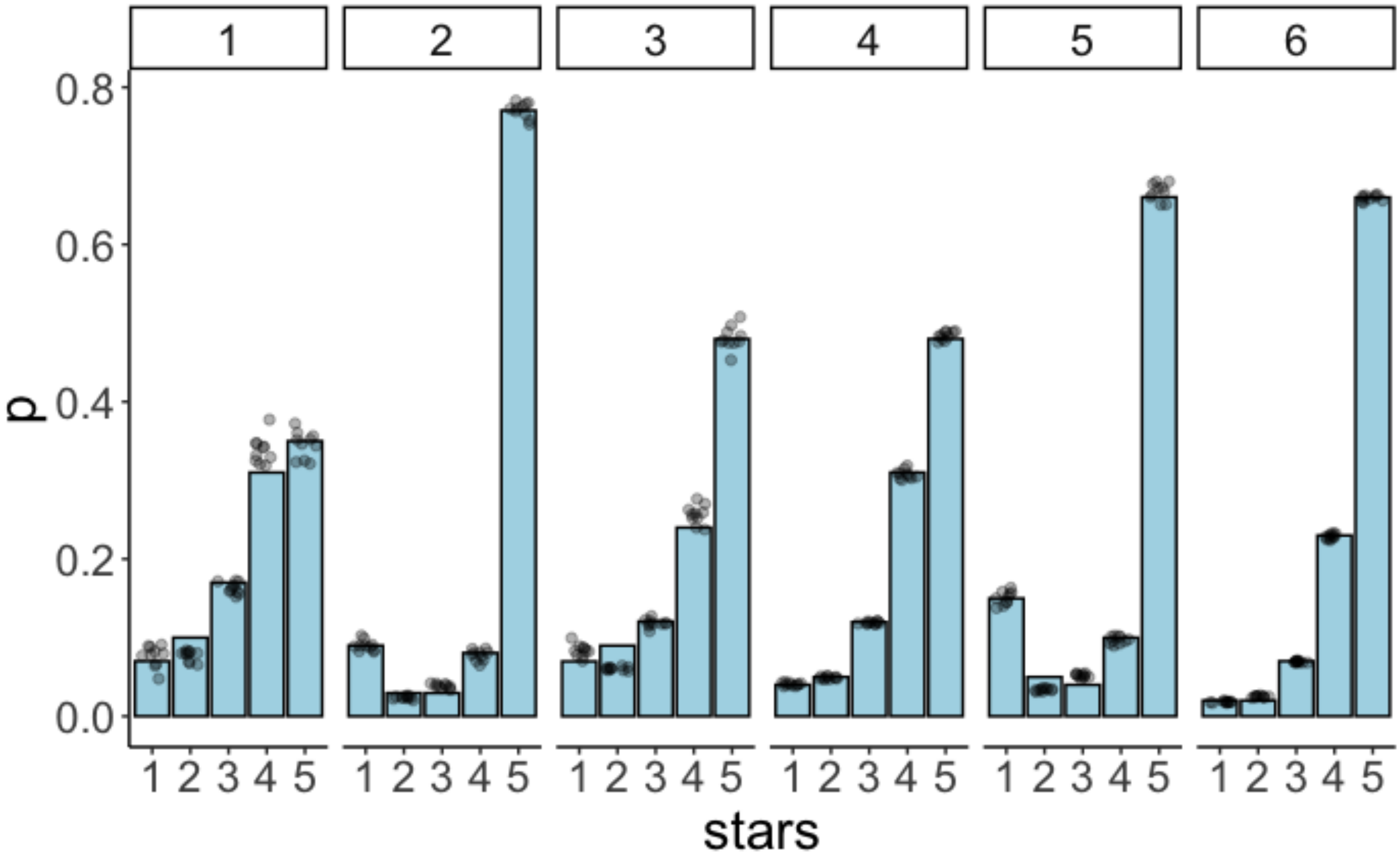
difference in mean

difference in variance

see for details

Bürkner & Vuorre (2019) Ordinal Regression Models in Psychology: A Tutorial. Advances in Methods and Practices in Psychological Science

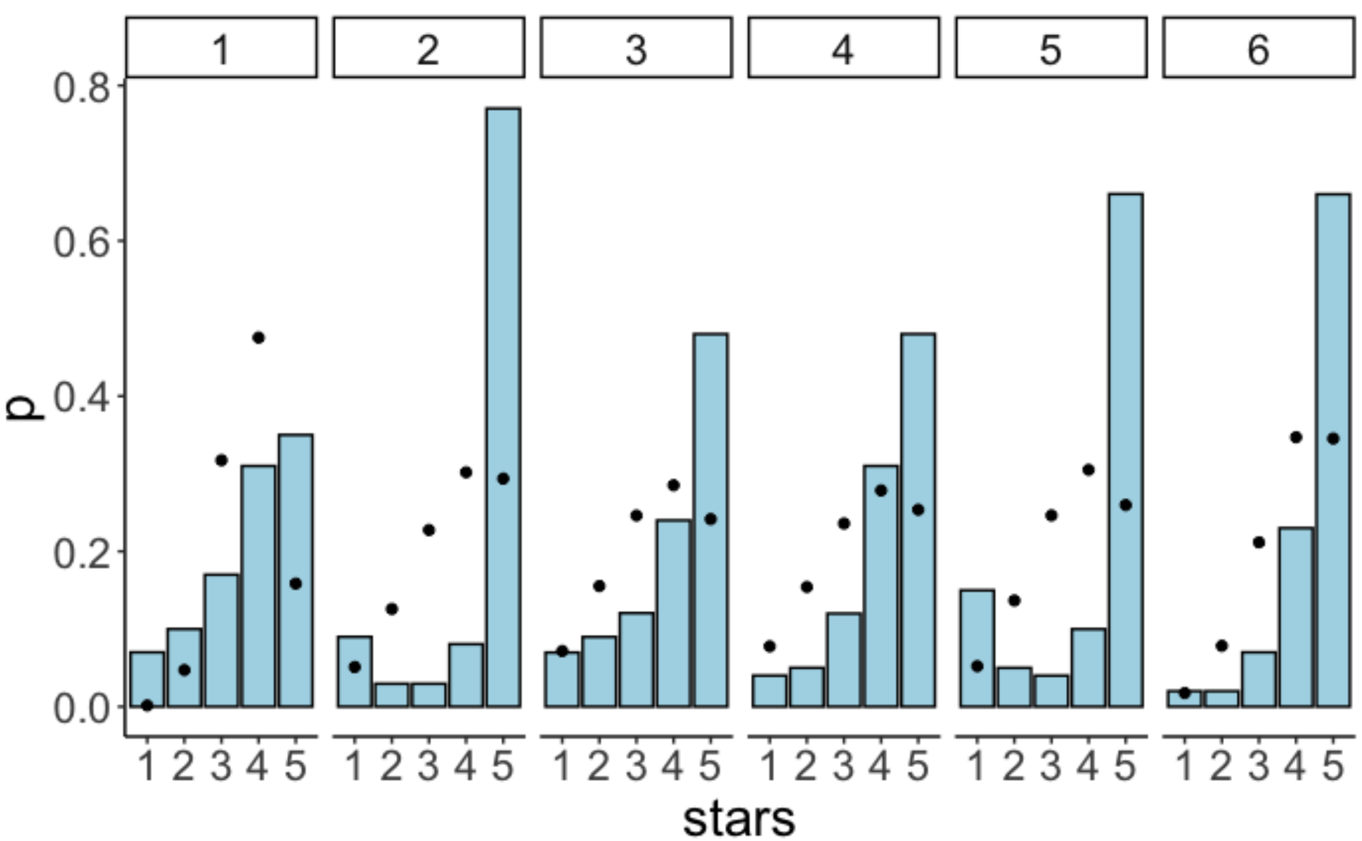
Illustrate the predictions



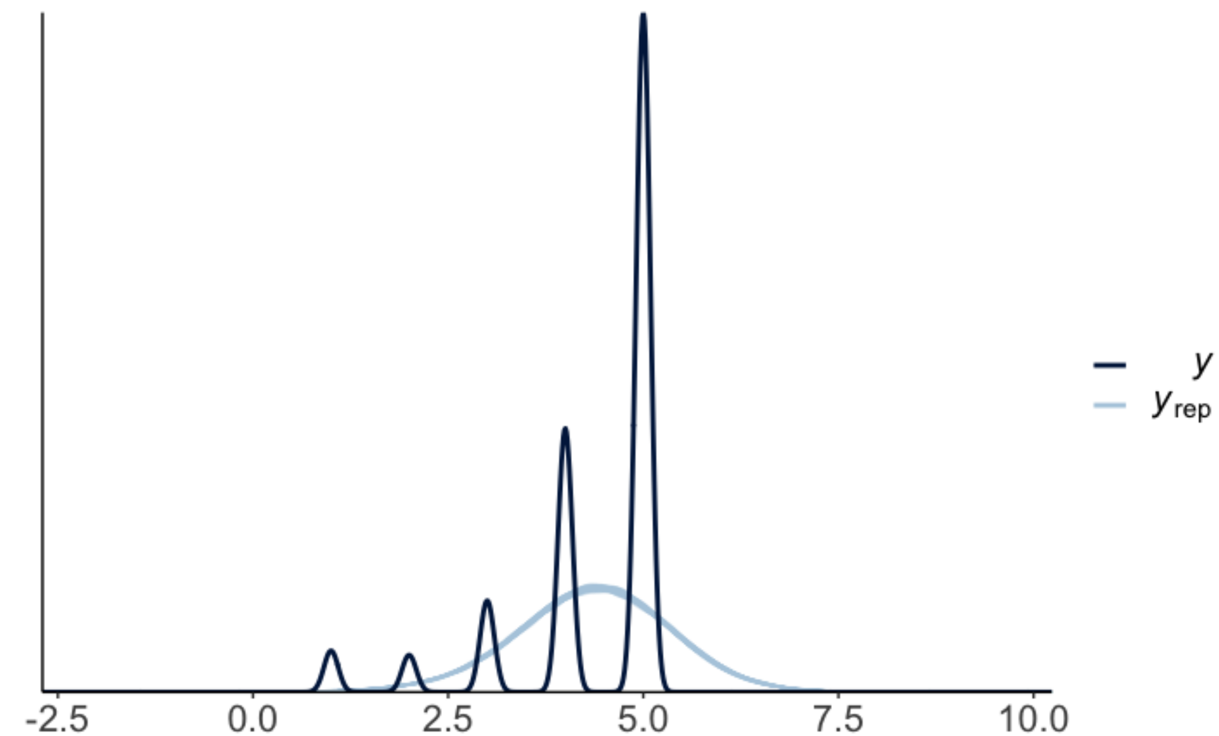
this looks excellent!

Metric model with different variances still doesn't work well ...

```
fit.brm8 = brm(formula = bf(stars ~ 1 + id,
                            sigma ~ 1 + id),
               data = df.movies,
               file = "cache/brm8",
               seed = 1)
```



posterior predictive check



still no good ...

Model comparison

```
fit.brm5 = add_criterion(fit.brm5,
                         criterion = "loo")

fit.brm6 = add_criterion(fit.brm6,
                         criterion = "loo")

loo_compare(fit.brm5, fit.brm6)
```

	elpd_diff	se_diff
fit.brm5	0.0	0.0
fit.brm6	-7657.1	109.2

the ordinal regression model is muuuuuuch better!

Which movie shall I watch?

run an ordinal regression to find out!

?

Customer reviews

4.5 out of 5

2,029 customer ratings

5 star	78%
4 star	10%
3 star	5%
2 star	3%
1 star	5%

Customer reviews

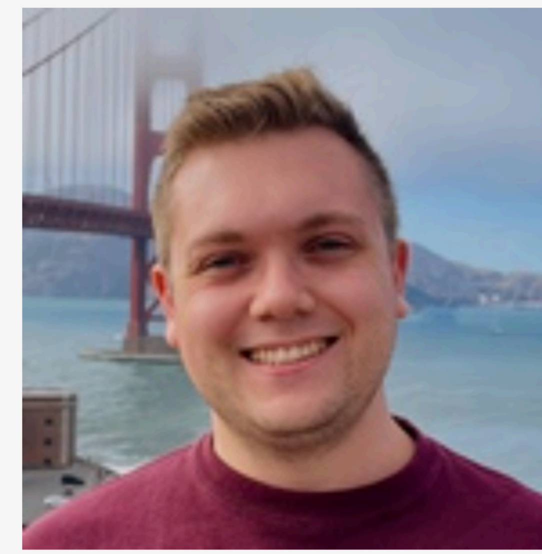
3.8 out of 5

1,558 customer ratings

5 star	53%
4 star	13%
3 star	8%
2 star	8%
1 star	17%

Thanks

Psych 252 Team

Ari Beller	Beth Rispoli	Satchel Grant	Shawn Schwartz
			
Teaching assistant	Teaching assistant	Teaching assistant	Teaching assistant
they/them	she/her	he/him	he/him

All of you!

Abutto, Adani Bennett
Anderson, Sean Paul
Blakey, Will William
Chen, Emily M
Chiu, Howard
Decatur, Cid Halsey-Steve
Dhingra, Monisha
Ergin, Irmak
Fang, David
Fendrich, Sarah
Fischer, Stephanie
Goodwin, Emily
Gupta, Anmol
Hu, Sherry
Johnson, Katherine Abigail
Kemmann, Bendix
Klevak, Nastasia Ruth
Lai, Wilson Cyrus Shi Hao
Lepp, Haley
Li, Qianqian
Morante Mazzaferro, Mateus
Myoung, Eunjung
Park, Joon
Portillo, Sofia Analise
Rajagopalan, Neha
Tao, Yujie
Tufts, L'Nard Evan Travis, II
Tung, Sarah Shi
Uricher, Raphael
Zamora, Tatiana Iza