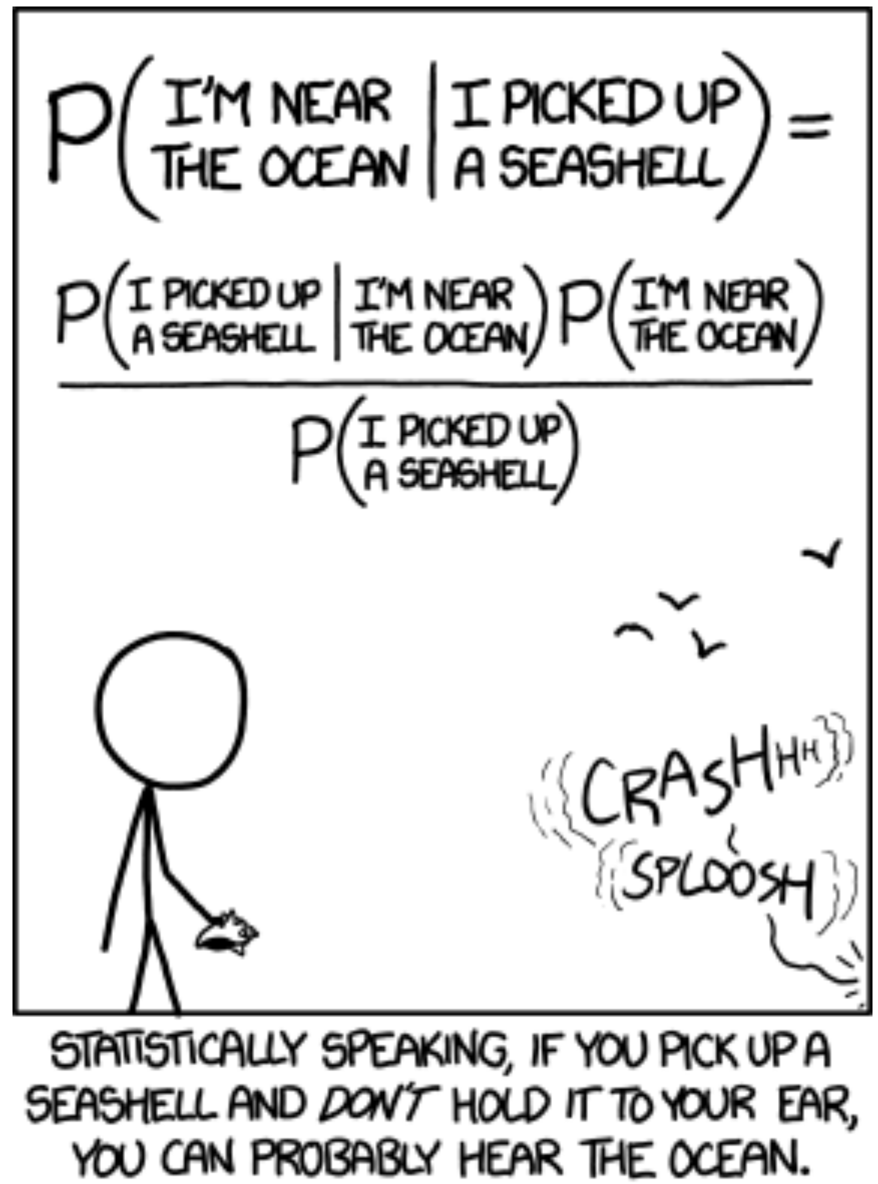
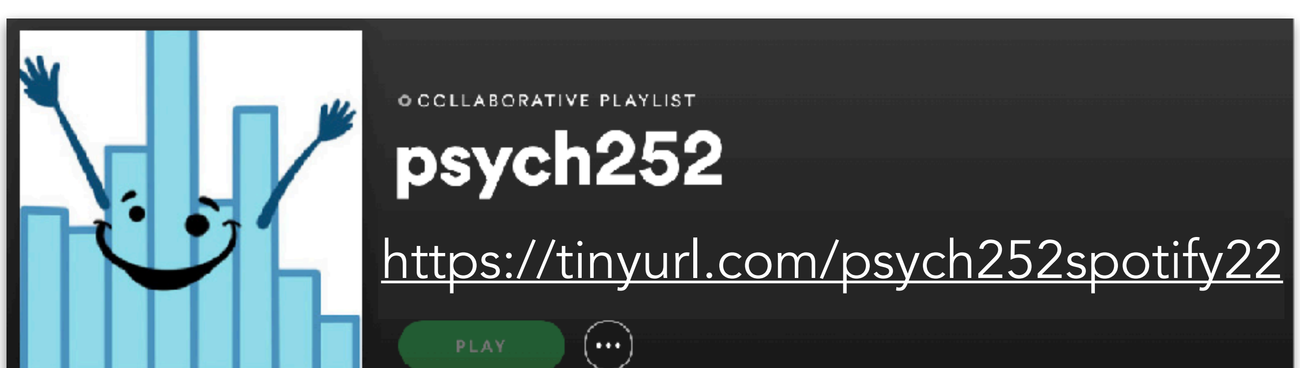
PSYCH 252: Statistical Methods for Behavioral and Social Sciences

Bayesian data analysis 4



We're listening to "Home Again" by "Michael Kiwanuka"



03/07/2022

Logistics

Homework 7 **is optional**

My name goes here

The names of the people I have worked w

2022-03-03 22:20:30

1 Instructions

This homework is due by **Thursday, March 10th, 8:00pm.**

As per usual, please upload your rendered pdf on Canvas.

Note:

- Some code chunks contain some skeleton code. The code chunk opt so that knitting the RMarkdown document doesn't throw any erro `eval=T` when you knit your homework, so that your calculations a
- Make sure to show the results of your calculations in the knitted pd function at the end of a code chunk.
- Some questions ask for a short written response as indicated by t

Good luck with the homework! If you have any questions, make sure t Thursday and/or post your questions on EdStem.

1.1 Load data

For the logistic regression question, we will use a data set which has inf transplant patients.

```
data(heart_transplant, package = "openintro")
df.heart = heart_transplant %>%
  mutate(survived = ifelse(survived == "dead", 0, 1))
```

Here is a description of the data set:

> The Stanford University Heart Transplant Study was conducted perimental heart transplant program increased lifespan. Each pat designated officially a heart transplant candidate, meaning that would most likely benefit from a new heart. Then the actual hear few weeks to several months depending on the availability of a don this waiting period show improvement and get deselected as a hear the purposes of this experiment those patients were kept in the d

1.2 Part 1: Logistic regression (3 points)

Question 1.1: (1 point)

Fit a logistic regression where you predict whether or not a person sur the model summary.

1

```
### YOUR CODE HERE ###

######################
```

Question 1.2: (1 point)

Use inverse logit to transform the log odds of the model coefficients into a probability of s interpretable, and describe these results from the model.

```
### YOUR CODE HERE ###

######################
```

Your answer:

Question 1.3: (1 point)

Visualize the results of the logistic regression model that you've fitted in Question 2. Th describe the relationship between age and the probability of survival.

```
### YOUR CODE HERE ###

######################
```

Your answer:

1.3 Part 2: Bayesian inference "by hand" (8 points)

This homework is from "PSYCH 10: Introduction to Statistical Methods" which is taugh You may find taking a look at this online chapter useful: Doing Bayesian Estimation

Question 2.1: (2 points)

Let's say that we are interested in the probability that a new drug called Bayesium w disease of frequentitis. For our purposes we are happy to estimate this variable (which we the nearest of 0.1.

Create a data frame called bayesium that includes the following variables:

- `theta`: a vector containing values of theta ranging from 0.0 to 1.0 in steps of 0.1
- `flat_prior`: a vector representing a flat, uniform prior across all possible values o the same prior value for each level of theta). This should be a probability distributio sum to one.
- `bayes_prior`: a vector containing a prior based on a binomial distribution with a reflecting the prior of a Bayesian who strongly believes that the treatment will have be obtained using the command: `dbinom(seq(0,10), 10, p = 0.8)`
- `freq_prior`: a vector containing a prior based on a binomial distribution with a reflecting the prior of a frequentist who strongly believes that the treatment will h
- `dogmatic_prior`: a vector containing a prior with all of its density on theta = 0 other values of theta, reflecting the prior of an very dogmatic Bayesian with very s

```
### YOUR CODE HERE ###

######################
```

Question 2.2: (2 points)

Let's say that we perform a clinical trial of the new drug in 20 people and we find that 10 are saved by the drug. Create a variable within the bayesium data frame called `likeli` the binomial likelihood for these data given each value of theta. (Hint: use an R function

2

function of the binomial distribution to calculate the number of 'successes' or saved lives in a certain number of trials or participants).

Then compute the posterior probabilities for each different prior and add them to the bayesium data frame within the following variables:

- `posterior_flat`: posterior given `flat_prior`
- `posterior_bayes`: posterior given `bayes_prior`
- `posterior_freq`: posterior given `freq_prior`
- `posterior_dogmatic`: posterior given `dogmatic_prior`

Each of these should be normalized so that they are probabilities (i.e. they sum to one).

```
df.bayesium = df.bayesium %>%
  ### YOUR CODE HERE ###
  mutate(likelihood =    ) %>%

  # compute posterior probabilities for each type of prior, and add to df

  ######################
```

Question 2.3: (4 points)

For each of the four priors, make a separate figure in which you plot the prior across all values of theta, and include the following: the prior with a black dotted line, the likelihood as a red dashed line, and the posterior with a solid line in blue. Add a title to the plot that includes the name of the prior. You can combine the figures into one using the "patchwork" or "cowplot" library (take a look at notes from the class `03_visualization2`). Alternatively, you can also just create one figure and make figure panels using the `facet_wrap()` or `facet_grid()` function.

```
### YOUR CODE HERE ###

# plot priors, likelihood, and posteriors

######################
```

Then, in your own words, describe how the different priors affect the respective maximum posterior estimates.

Your answer:

1.4 Part 3: Bayesian data analysis (4 points)

Question 3.1: (1 point)

Build a bayesian model from the `df.heart` data similar to the frequentist model in Question 1.1. Call this model `fit.brm`

```
### YOUR CODE HERE ###
fit.brm = brm(formula =
              family =
              data =
              file = "cache/brm",
              seed = 1)
######################
```

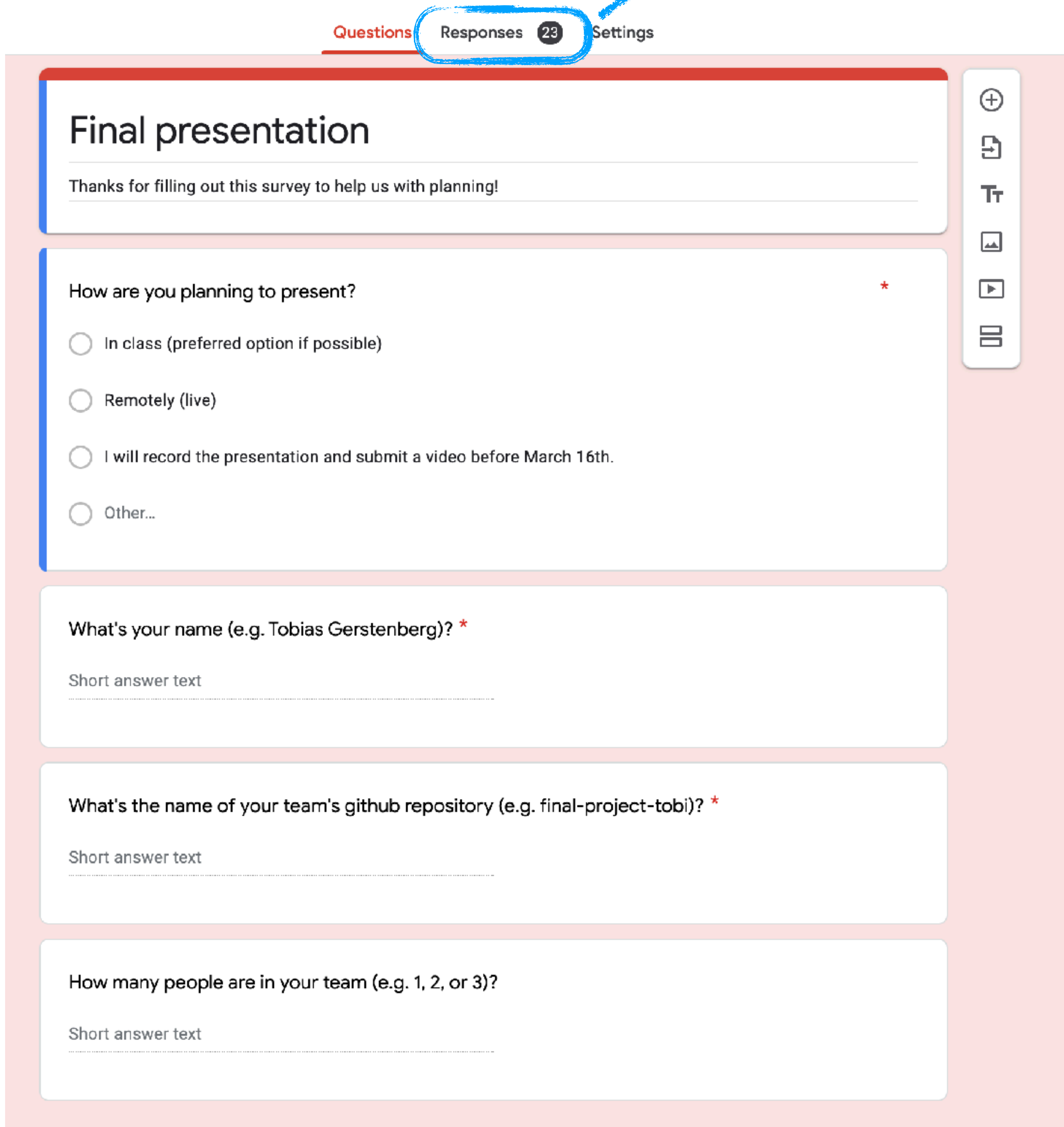
Question 3.2: (1 point)

Print the summary of this model and interpret its coefficients. Compare these to the coefficients from the frequentist model you built in Question 1.1.

3

Final presentation survey

goal: 28



https://forms.gle/DdVaig7W63JgBcPC9

Course evaluations

Course Feedback Now Open Inbox ×

Course Evaluations <course-evaluations@stanford.edu> 8:00 AM (4 minutes ago)
to Tobias

Dear Tobias,

End-term course evaluations are now open for students to provide feedback until **Monday, March 21 at 11:59PM**. Please direct your students to complete their feedback through any of the following options:

- in Canvas, students will see a pop-up notification on their Canvas dashboard page any time they log into Canvas during the evaluation period.
- direct link: http://course-evaluations.stanford.edu.
- in Axess > Student > Course and Section Evaluations > Link to the evaluation system near the top of the page.

Students who complete all of their feedback will see their grades as soon as they have been submitted. Students who do not complete all of their feedback will not be able to see their grades in Axess until the day after the grade release deadline.

Here are some tips to encourage students to complete their feedback:

- Take a few minutes of class time to allow students to complete feedback;
- Explain that you, the instructor, will only see aggregated, anonymous responses; and,
- Tell students that their feedback is very important and helps you to make improvements to the course for future students.

The student evaluation period closes on **Monday, March 21 at 11:59PM**.

Please visit our website at https://evals.stanford.edu/end-term-feedback/key-dates-end-term-feedback for a summary of key dates.

For more information about course evaluations and feedback, or if you have any questions or troubles with your course evaluations, please contact us at course-evaluations@stanford.edu

Regards,
Course Evaluations Team

Please note that all content entered into the course evaluation system by instructors and students (e.g., written responses, custom questions) is subject to Stanford's Terms of Use for Sites. For additional information on policies around the administration of course evaluations and data access and usage, please visit our Course Evaluation Data Access & Reporting page.

Course evaluations

Results / Project Response Rates / Node Response Rates / Course Response Rates

Course Response Rates Winter 2021 Course Feedback

Search

Course Code | **Course Title** | **Course Unique ID**

Search

Courses

Code	Instructor	Enrollments	Responded	Response Rate	Opted-Out	% of Enrollments Opted Out	Responded With Opted-Out Removed	Response Rate With Opted-Out Removed	View
W21-PSYCH-252 -01	Tobias Gerstenberg	30	7	23.33%	0	0%	7	23.33%	

Total **1** **Records per page** 50 **Page** 1 **of 1**

http://evaluationkit.stanford.edu/

Guest lecture: Wednesday

Patrick Bissett

I am a Research Scientist at Stanford University. I am interested in **cognitive control**, **response inhibition**, and **decision making**. I use convergent methodologies including **experimental tasks**, **computational modeling**, and **functional brain imaging** to understand these processes. I received my Ph. D in Cognition and Cognitive Neuroscience from a University in 2014 and my B. S in Brain, Behavior, and Cognitive Science from University of Michigan in 2008.

Guest lecture: Friday

Ari

Sarah

Chengxu

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 slider data
  - 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

What we've covered

visualization and data wrangling

probability, simulation, causality

linear model

power analysis

model comparison

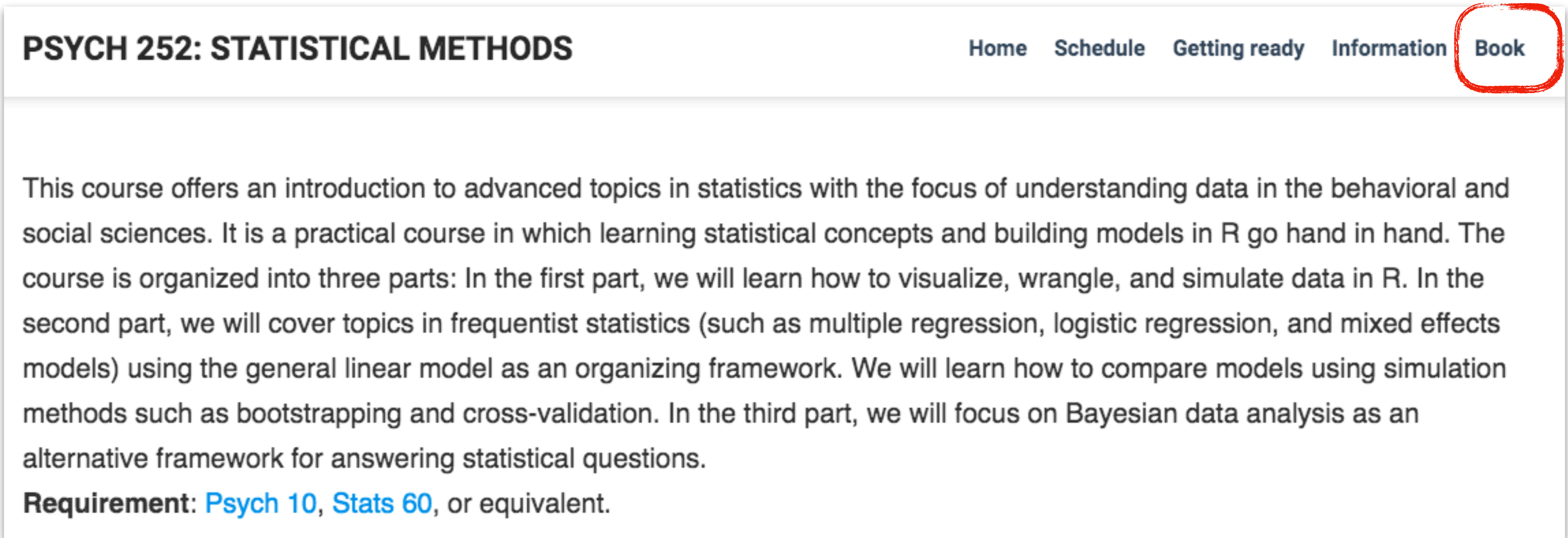
linear mixed effects models

logistic regression

Bayesian data analysis

Date	Topic
January 3rd	Introduction
January 5th	Visualization I
January 7th	Visualization II
January 10th	Data wrangling I
January 12th	Data wrangling II
January 14th	Probability
January 17th	No class (Martin Luther King Jr. Day)
January 19th	Simulation I
January 21st	Simulation II
January 24th	Modeling data
January 26th	Linear model I
January 28th	Linear model II
January 31st	Linear model III
February 2nd	Linear model IV
February 4th	Power analysis
February 7th	Model comparison
February 9th	No class (Midterm)
February 11th	Causality
	Midterm due
February 14th	Linear mixed effects models I
February 16th	Linear mixed effects models II
February 17th	Project proposal due
February 18th	Linear mixed effects models III
February 21st	No class (Presidents' Day)
February 23rd	Linear mixed effects models IV
February 25th	Generalized linear model
February 28th	Bayesian data analysis I
March 2nd	Bayesian data analysis II
March 4th	Bayesian data analysis III
March 7th	Bayesian data analysis IV
March 9th	Guest lecture
March 11th	Guest lecture
March 16th	Final project presentations
March 18th	Final project report due

I'll keep updating the course notes!

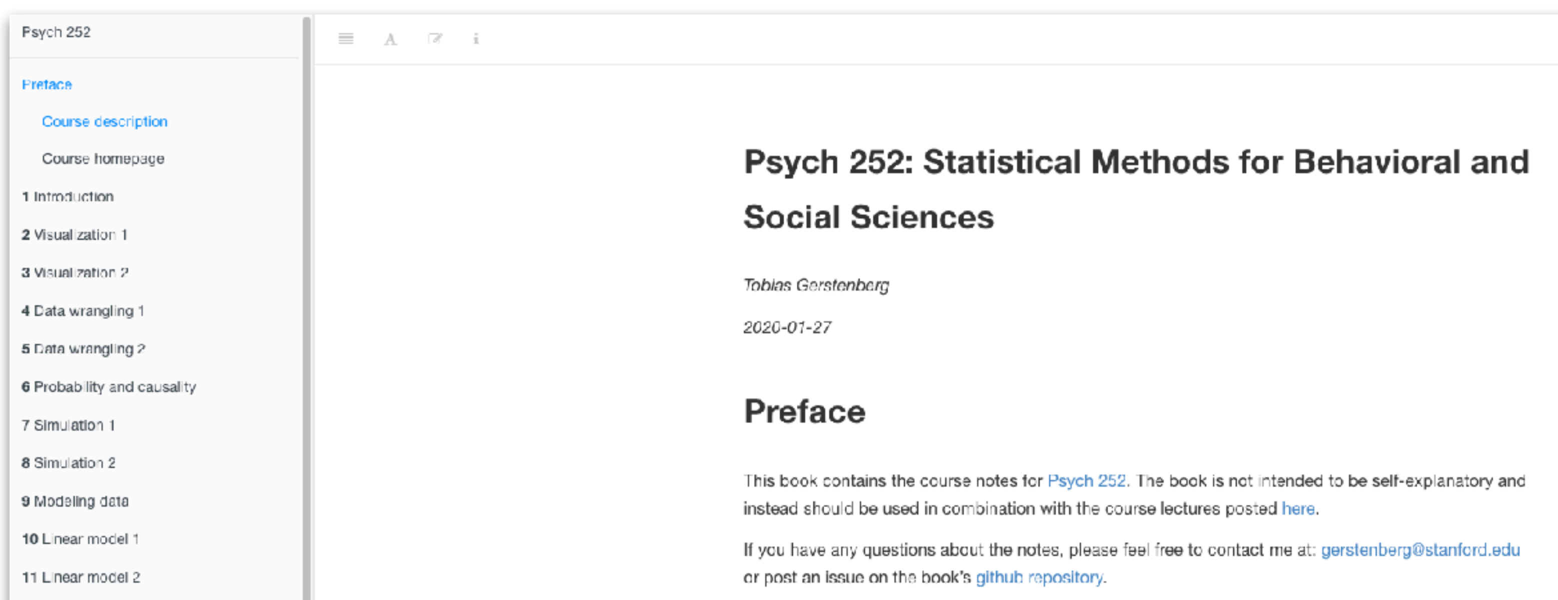


PSYCH 252: STATISTICAL METHODS

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

Preface
Course description
Course homepage
1 Introduction
2 Visualization 1
3 Visualization 2
4 Data wrangling 1
5 Data wrangling 2
6 Probability and causality
7 Simulation 1
8 Simulation 2
9 Modeling data
10 Linear model 1
11 Linear model 2

Psych 252: Statistical Methods for Behavioral and Social Sciences

Tobias Gerstenberg

2020-01-27

Preface

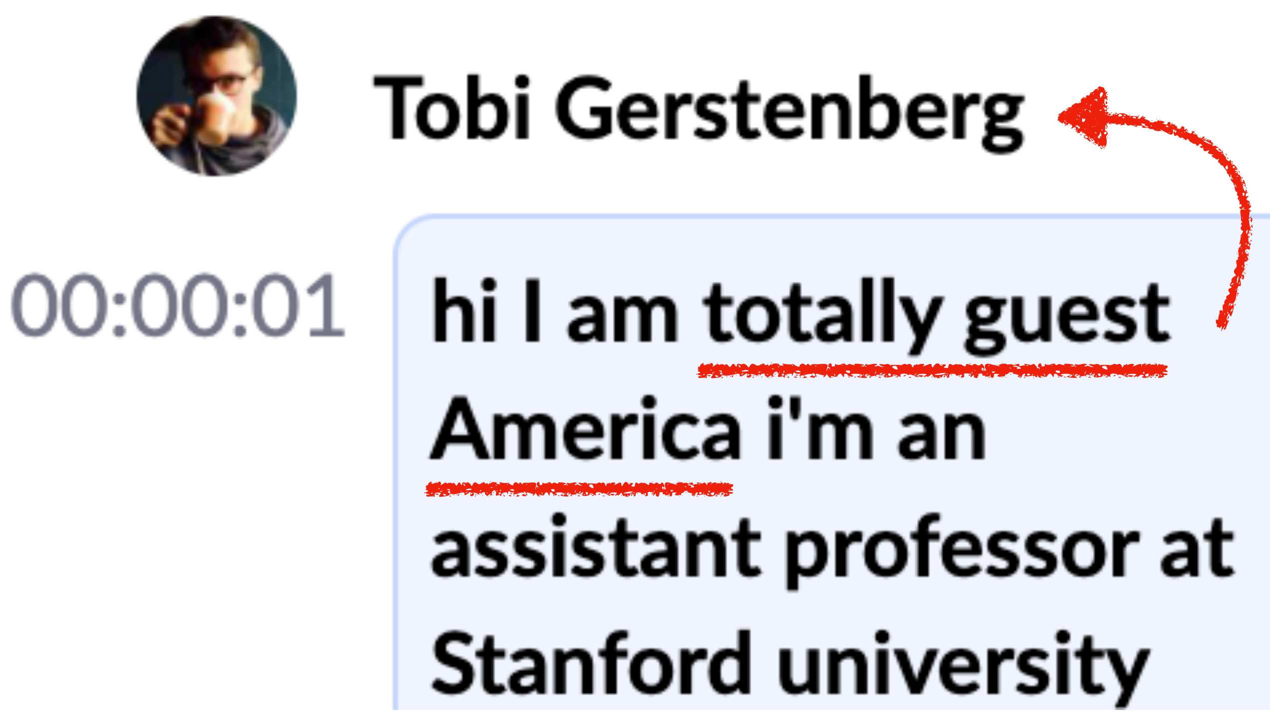
This book contains the course notes for Psych 252. The book is not intended to be self-explanatory and instead should be used in combination with the course lectures posted here.

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 have access to the lecture recordings (for at least until the summer)

including the super accurate captions!



take a hint zoom!

Datacamp (available until ~ mid june)

What shall I do now?

PSYCH 262

Nilam Ram

PSYCH 262: Measurement and the Study of Change in Social Science Research (COMM 369)

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)

Schedule for PSYCH 262

2021-2022 Spring

PSYCH 262 | 3 units | UG Reqs: None | Class # 29881 | Section 01 | Grading: Letter or Credit/No Credit | SEM | Session: 2021-2022 Spring 1 | In Person | Students enrolled: 10

03/28/2022 - 06/01/2022 Wed, Fri 8:30 AM - 10:00 AM at 120-452 with Ram, N. (PI)

Instructors: Ram, N. (PI)

Notes: Exception Granted.

PSYCH 204A

Brian Wandell

PSYCH 204A: Human Neuroimaging Methods

This course introduces the student to human neuroimaging using magnetic resonance scanners. The course is a mixture of lectures and hands-on software tutorials. The course begins by introducing basic MR principles. Then various MR measurement modalities are described, including several types of structural and functional imaging methods. Finally algorithms for analyzing and visualizing the various types of neuroimaging data are explained, including anatomical images, functional data, diffusion imaging (e.g., DTI) and magnetization transfer. Emphasis is on explaining software methods used for interpreting these types of data.

Terms: Win | Units: 3

Instructors: Wandell, B. (PI) ; Kim, I. (TA)

Schedule for PSYCH 204A

2021-2022 Winter

PSYCH 204A | 3 units | UG Reqs: None | Class # 14950 | Section 01 | Grading: Letter or Credit/No Credit | SEM | Session: 2021-2022 Winter 1 | In Person | Students enrolled: 16 / 30

01/03/2022 - 03/11/2022 Tue, Thu 1:30 PM - 3:00 PM at 300-300 with Wandell, B. (PI); Kim, I. (TA)

Exam Date/Time: 2022-03-15 3:30pm - 6:30pm (Exam Schedule)

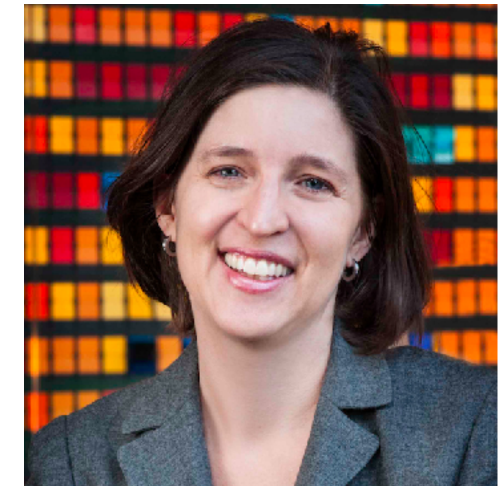
Instructors: Wandell, B. (PI); Kim, I. (TA)

Additional Resources: Syllabus

https://syllabus.stanford.edu/syllabus/#/viewSyllabus/W22-PSYCH-204A-01/W22-PSYCH-204A-01/fromExploreCourse

ECON 293: Machine Learning and Causal Inference

Susan Athey



ECON 293: Machine Learning and Causal Inference

This course will cover statistical methods based on the machine learning literature that can be used for causal inference. In economics and the social sciences more broadly, empirical analyses typically estimate the effects of counterfactual policies, such as the effect of implementing a government policy, changing a price, showing advertisements, or introducing new products. This course will review when and how machine learning methods can be used for causal inference, and it will also review recent modifications and extensions to standard methods to adapt them to causal inference and provide statistical theory for hypothesis testing. We consider causal inference methods based on randomized experiments as well as observational studies, including methods such as instrumental variables and those based on longitudinal data. We consider the estimation of average treatment effects as well as personalized policies. Lectures will focus on theoretical developments, while classwork will consis more »

Terms: Spr | Units: 3

Instructors: Athey, S. (PI) ; Wager, S. (PI) ; Wager, S. (SI)

Schedule for ECON 293

2020-2021 Spring

ECON 293 | 3 units | Class # 20461 | Section 01 | Grading: Letter or Credit/No Credit Exception | CAS | Remote: Asynchronous | Students enrolled: 26

03/29/2021 - 06/02/2021 - at Remote with Athey, S. (PI); Wager, S. (SI)

Instructors: Athey, S. (PI); Wager, S. (SI)

ECON 293: Machine Learning and Causal Inference

Susan Athey

ANNUAL REVIEWS

Annual Review of Economics

Machine Learning Methods That Economists Should Know About

Susan Athey[1,2,3] and Guido W. Imbens[1,2,3,4]

[1]Graduate School of Business, Stanford University, Stanford, California 94305, USA; email: athey@stanford.edu, imbens@stanford.edu

[2]Stanford Institute for Economic Policy Research, Stanford University, Stanford, California 94305, USA

[3]National Bureau of Economic Research, Cambridge, Massachusetts 02138, USA

[4]Department of Economics, Stanford University, Stanford, California 94305, USA

Annu. Rev. Econ. 2019. 11:685–725

First published as a Review in Advance on June 10, 2019

The *Annual Review of Economics* is online at economics.annualreviews.org

https://doi.org/10.1146/annurev-economics-080217-053433

JEL code: C30

Keywords

machine learning, causal inference, econometrics

Abstract

We discuss the relevance of the recent machine learning (ML) literature for economics and econometrics. First we discuss the differences in goals, methods, and settings between the ML literature and the traditional econometrics and statistics literatures. Then we discuss some specific methods from the ML literature that we view as important for empirical researchers in economics. These include supervised learning methods for regression and classification, unsupervised learning methods, and matrix completion methods. Finally, we highlight newly developed methods at the intersection of ML and econometrics that typically perform better than either off-the-shelf ML or more traditional econometric methods when applied to particular classes of problems, including causal inference for average treatment effects, optimal policy estimation, and estimation of the counterfactual effect of price changes in consumer choice models.

CS109: Probability for Computer Scientists

CS109: Probability for Computer Scientists 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. **Read more here to learn what CS109 is all about.** This is going to be a great quarter and we are looking forward to the chance to teach you.

Teaching Team

Lecturer: Chris Piech
piech @ cs
OH Online
M 2:00pm-3:00pm

Lecturer: Jerry Cain
jerry @ cs
OH Online
TBD

- 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 253

Basics / Course Schedule & Syllabus / Software Tools

offered again next year

Advanced Statistical Modeling
Spring 2021, Stanford University

Introduction to high-dimensional data analysis and machine learning methods for use in the behavioral and neurosciences, including: supervised methods such as linear regression and classification, linear mixed-effects and hierarchical models, structural equation modeling, and regularization techniques; statistical methods such as bootstrapping, signal detection, and reliability theory; metrics for model/data comparison such as representational similarity analysis; and unsupervised methods such as clustering, PCA, and exploratory factor analysis. Students will learn how to both use existing statistical data analysis packages (such as scikit-learn) as well to build and estimate simple custom models in Python. Requirement: Psych 251, and familiarity with Python programming and introductory linear algebra.

Time: Mondays & Wednesdays 13:00 - 14:20 PDT.

- focus on high-dimensional data
- classification
- representational similarity analysis
- clustering ...
- learn Python!

Dan Yamins

Russ Poldrack

https://web.stanford.edu/class/psych253/

PSYCH 204

offered again next year

Noah Goodman

PSYCH 204: Computation and Cognition: The Probabilistic Approach (CS 428)

This course will introduce the probabilistic approach to cognitive science, in which learning and reasoning are understood as inference in complex probabilistic models. Examples will be drawn from areas including concept learning, causal reasoning, social cognition, and language understanding. Formal modeling ideas and techniques will be discussed in concert with relevant empirical phenomena.

Terms: Spr | Units: 3

Instructors: Goodman, N. (PI) ; Nam, A. (TA)

Schedule for PSYCH 204

2020-2021 Spring

PSYCH 204 | 3 units | Class # 31428 | Section 01 | Grading: Letter or Credit/No Credit Exception | LEC | Remote: Synchronous | Students enrolled: 14

03/29/2021 - 06/04/2021 Tue, Thu 12:30 PM - 1:50 PM at Remote with Goodman, N. (PI); Nam, A. (TA)

Instructors: Goodman, N. (PI); Nam, A. (TA)

Advanced regression analysis

Sanne Smith

last offered 2021

EDUC 326: Advanced Regression Analysis

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)

Schedule for EDUC 326

2020-2021 Spring

EDUC 326 | 3-5 units | Class # 27329 | Section 01 | Grading: Letter or Credit/No Credit Exception | LEC | Remote: Synchronous | Students enrolled: 5

03/29/2021 - 06/04/2021 Fri 1:00 PM - 3:50 PM at Remote with Smith, S. (PI)

Instructors: Smith, S. (PI)

Additional Resources: (Login to view additional resources)

STATS 271: Applied Bayesian Statistics

Scott Linderman

last offered 2021

STATS 271: Applied Bayesian Statistics (STATS 371)

This course is a modern treatment of applied Bayesian statistics with a focus on high-dimensional problems. We will study a collection of canonical methods that see heavy use in applications, including high-dimensional linear and generalized linear models, hierarchical/random effects models, Gaussian processes, variable-dimension and Dirichlet process mixtures, graphical models, and methods used in Bayesian inverse problems. Each method will be accompanied by one or more motivating datasets. Through these examples the course will cover: (1) Bayesian hypothesis testing, multiplicity correction, selection, shrinkage, and model averaging; (2) prior choice; (3) Frequentist properties of Bayesian procedures in high dimensions; and (4) computation by Markov chain Monte Carlo, including constructing efficient Gibbs, Metropolis, and more exotic samplers, empirical convergence analysis, strategies for scaling computation to high dimensions (approximations, divide-and-conquer, minibatching, et cetera), and the theory of convergence rates.

Terms: Spr | Units: 3

Instructors: Linderman, S. (PI)

Schedule for STATS 271

2020-2021 Spring

STATS 271 | 3 units | Class # 19123 | Section 01 | Grading: Letter or Credit/No Credit Exception | LEC | Remote: Asynchronous | Students enrolled: 14 / 20

03/29/2021 - 06/04/2021 - at Remote with Linderman, S. (PI)

Instructors: Linderman, S. (PI)

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 Tobi, Ari, Sarah, Chengxu 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

Dawn Finzi

dfinzi@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	4.09	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

⋮

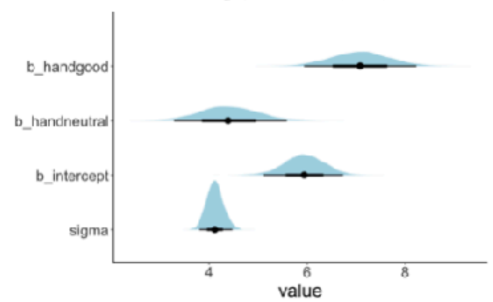
summary of posterior

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

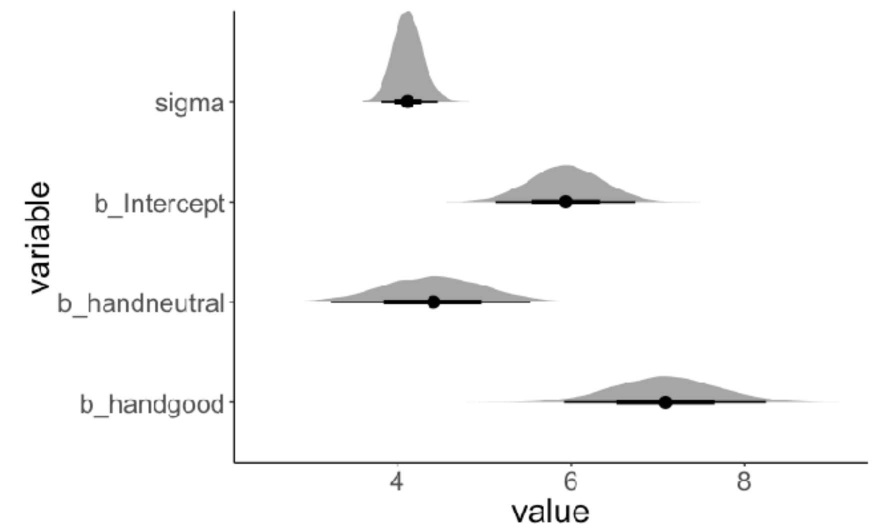
visualization



12

Asking questions based on the posterior

Do good hands make twice as much as bad hands?




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

p = 0.89

16

Testing hypothesis

```
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	6.18	10.63	8.94	FALSE
5.86	4.46	7.68	10.32	9.70	FALSE

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

p = 0.04

18

The "emmeans" package is your friend!

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

estimated mean for each group

contrasts

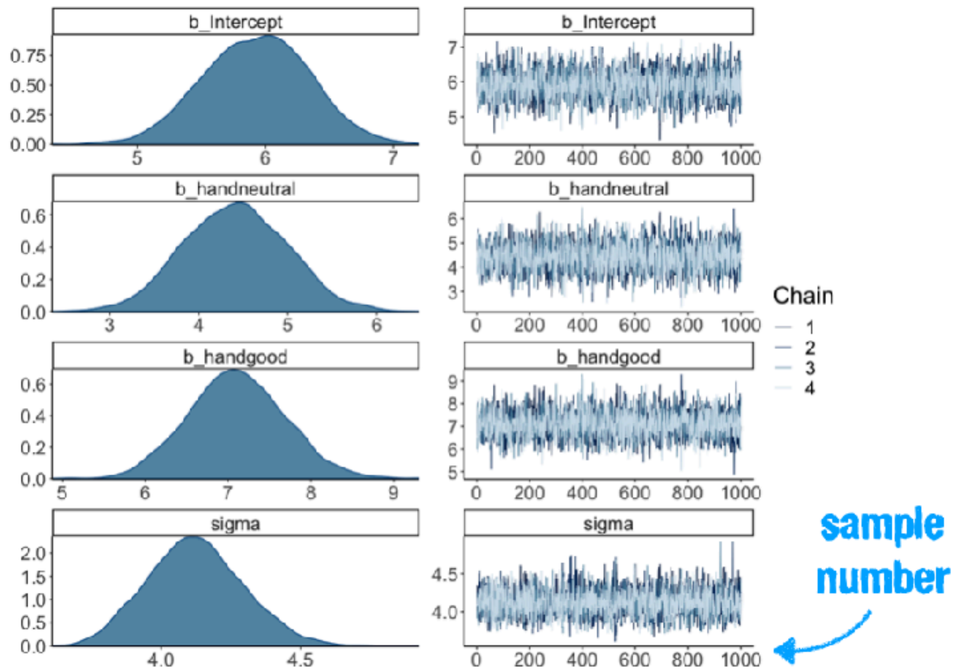
```
$emmeans
 hand     emmean lower.HPD upper.HPD
 bad        5.94      5.16      6.78
 neutral   10.34      9.55     11.15
 good      13.02     12.22     13.82

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)
```

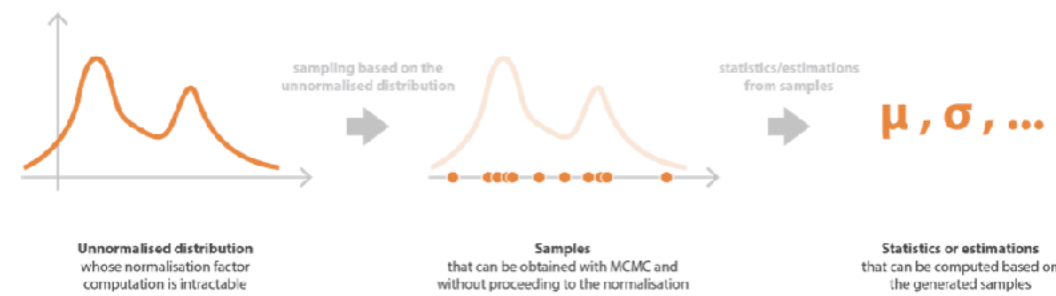


https://chi-feng.github.io/mcmc-demo/

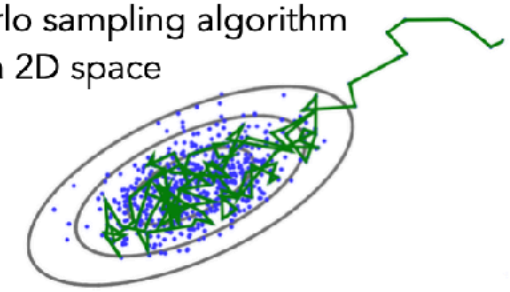
24

Can we trust the model results?

Inference via Markov Chain Monte Carlo (MCMC)



Markov Chain Monte Carlo sampling algorithm in a 2D space

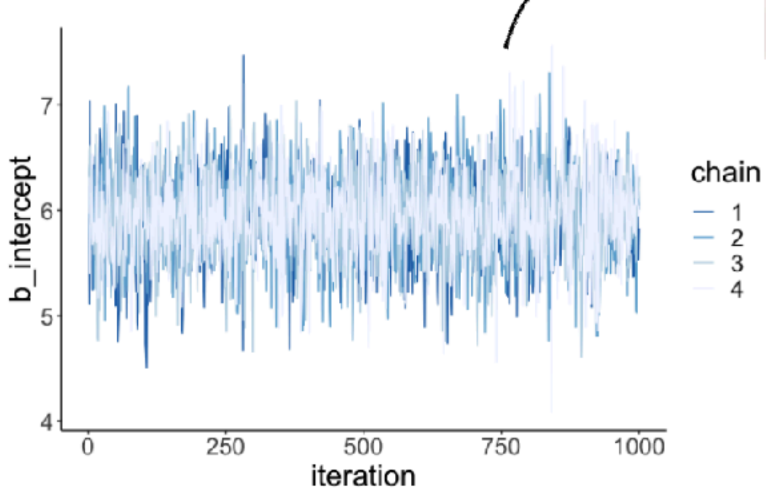


goal: draw **independent** samples from the posterior distribution

25

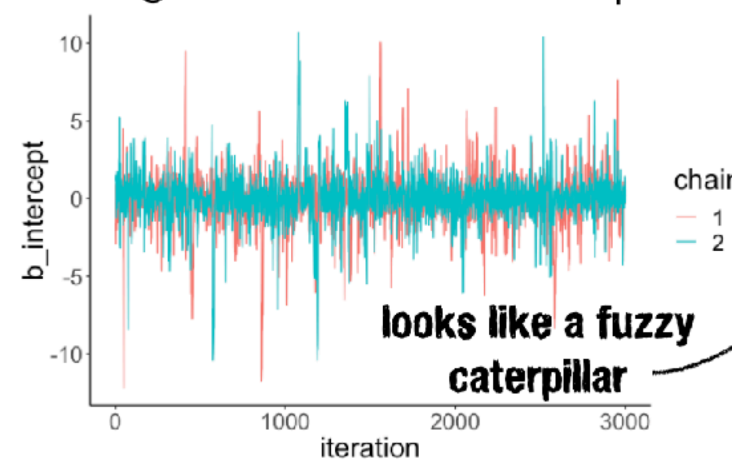
Can we trust the model results?

looks like a fuzzy caterpillar



26

Having somewhat informative priors fixes things



if things go wrong:

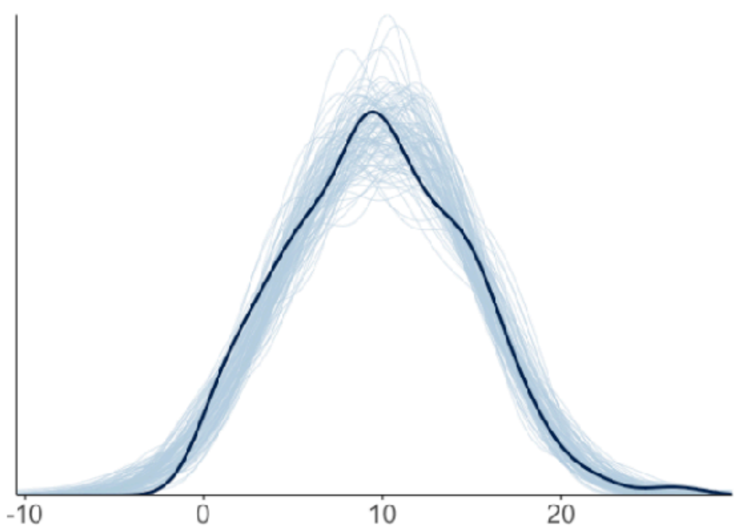
- set more informative priors
- run more warm-up samples
- adjust the sampling algorithm as suggested via the `control` argument

32

Quick recap: Posterior predictive check & reporting

Posterior predictive check

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

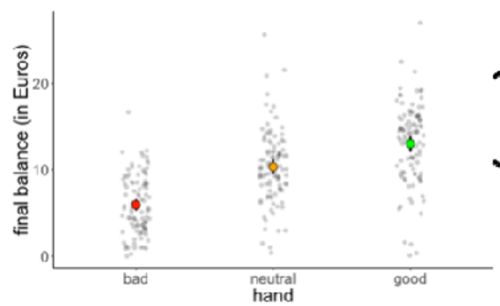


The model accurately captures the distribution of the response variable

34

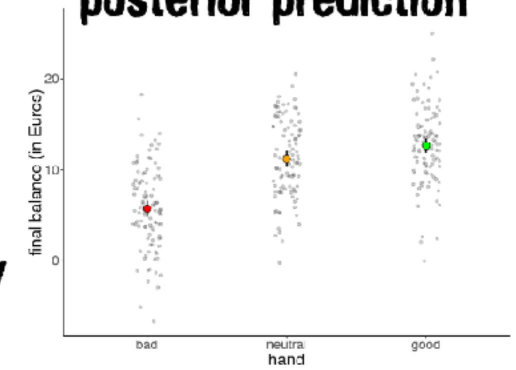
Posterior predictive check

original data

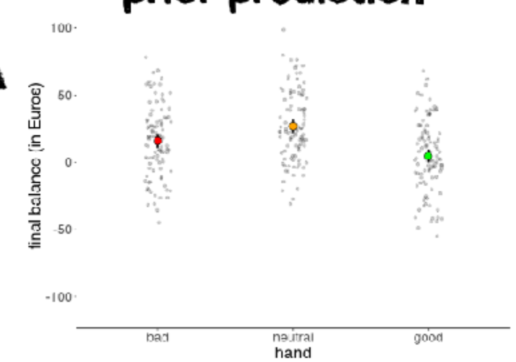


take a look at course notes for how to make these

posterior prediction



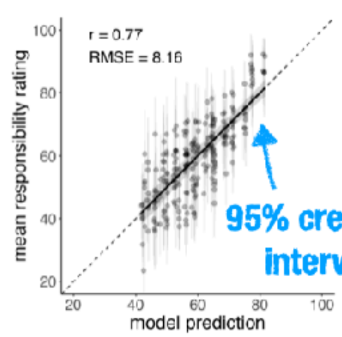
prior prediction



35

Reporting results

Plots



95% credible interval

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	59.94	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

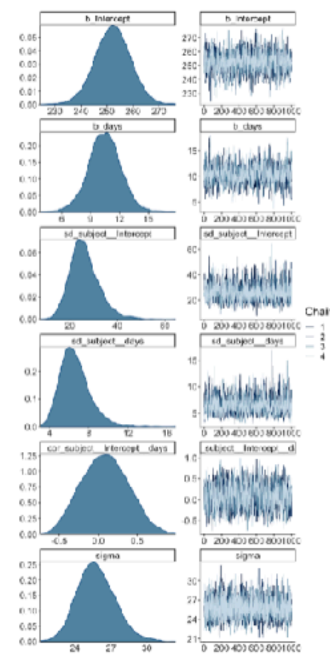
```r
fit.brm_sleep = brm(formula = reaction ~ 1 + days + (1 + days | subject),
                    data = df.sleep,
                    seed = 1,
                    file = "cache/brm_sleep")
```

43

a) Did the inference work?

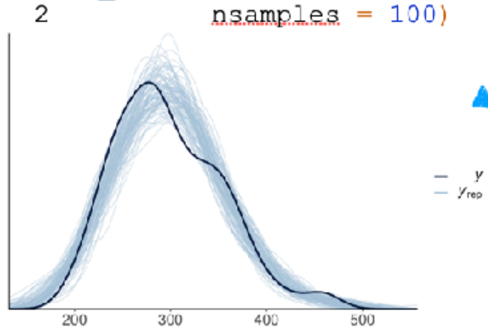
```r
fit.brm_sleep %>%
  plot(N = 6)
```

these look good!



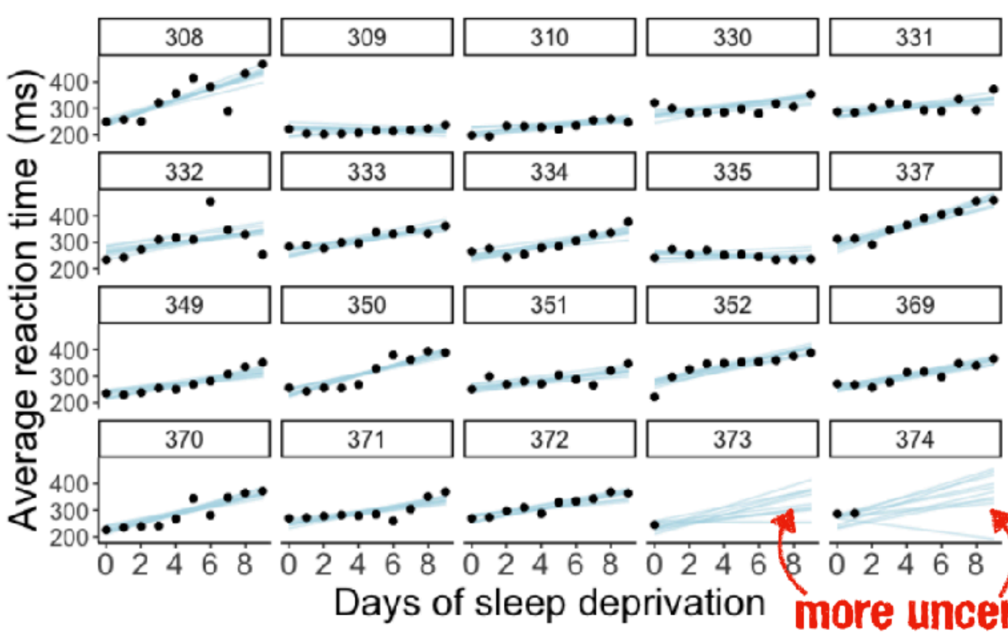
also looks good!

```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)            25.18      6.25    15.65    40.54 1.00     1879     2463
sd(days)                  6.59      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

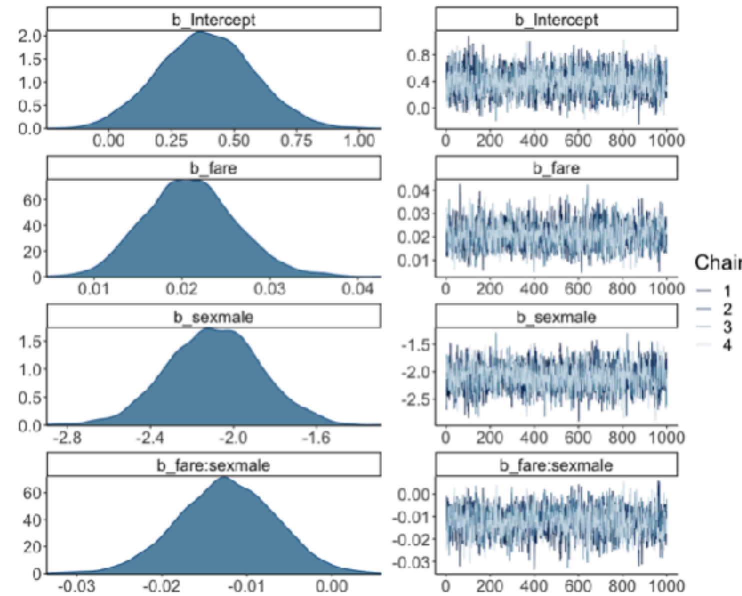
```r
fit.brm_titanic = brm(formula = survived ~ 1 + fare * sex,
                      family = "bernoulli",
                      data = df.titanic,
                      file = "cache/brm_titanic",
                      seed = 1)
```

just need to change the family

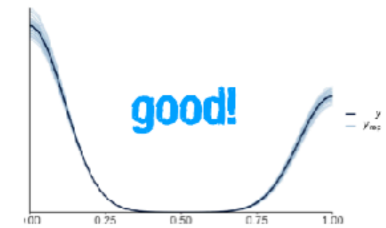
63

a) Did the inference work?

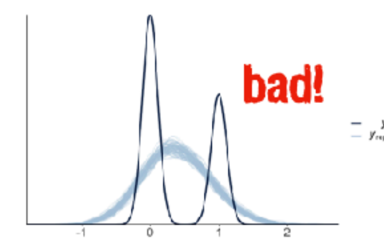
```r
fit.brm_titanic %>%
  plot()
```




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



model with Gaussian family

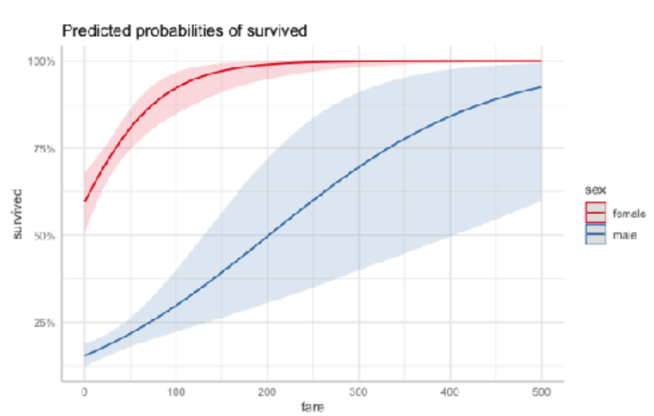


66

4. Interpret the model parameters

```r
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?

```r
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. PMCID: PMC4114196

Published online 2014 Jul 29. doi: 10.3389/fpsyg.2014.00781 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 …

☆ 99 Cited by 966 Related articles All 14 versions Web of Science: 583 Import into BibTeX

1257 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

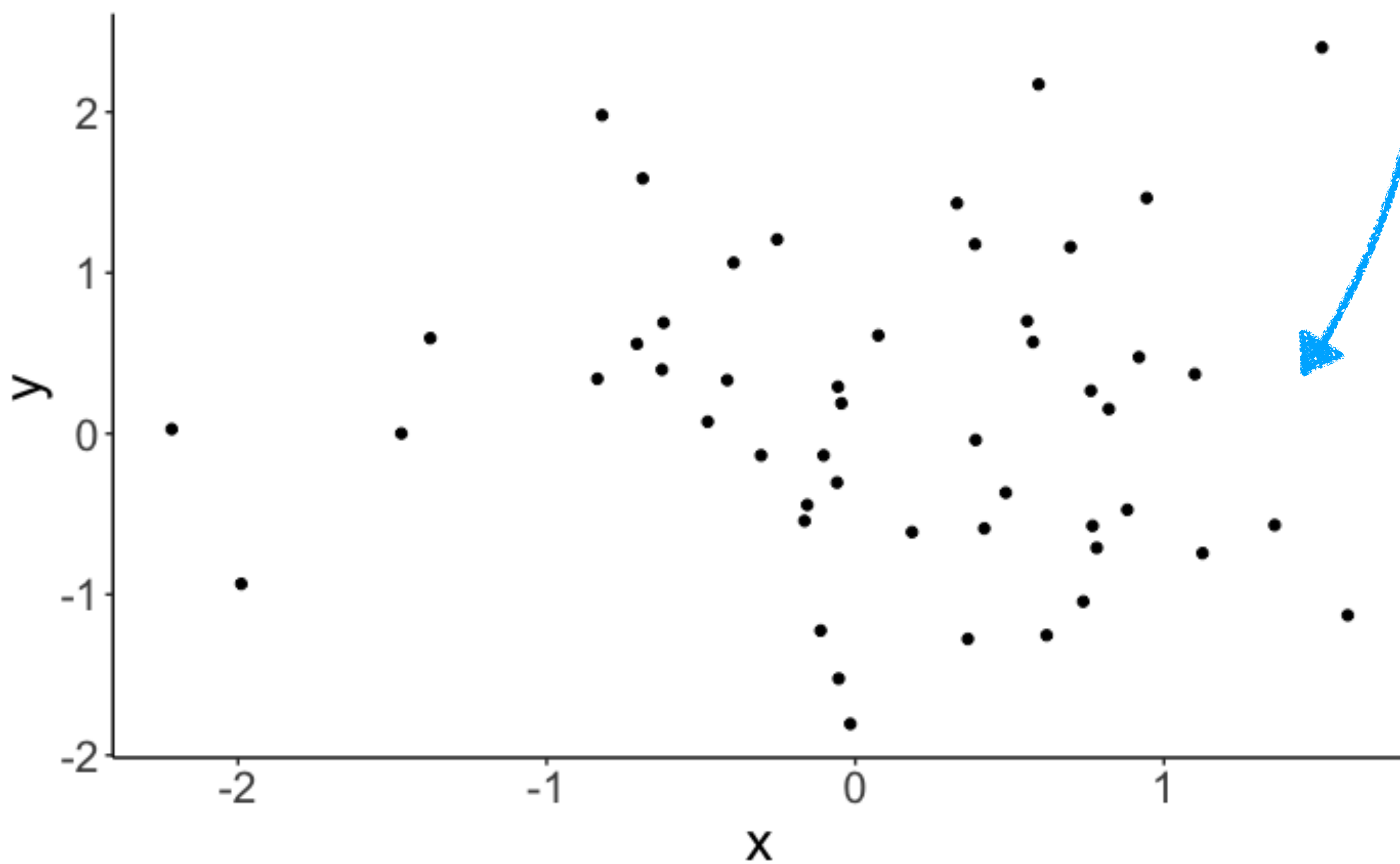
https://vuorre.netlify.com/post/2017/03/21/bayes-factors-with-brms/

Approximate LOO

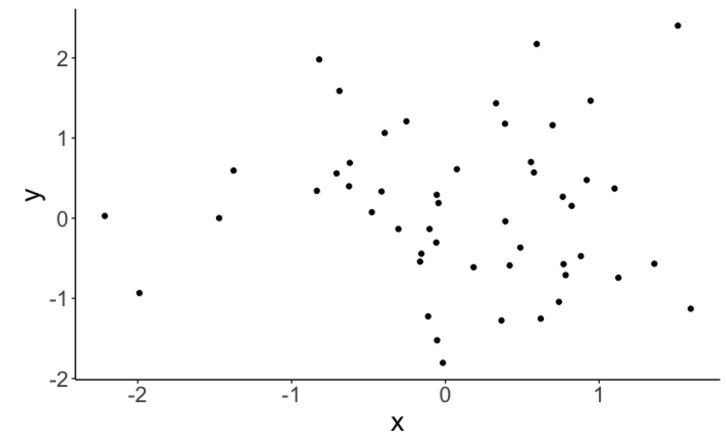
Evidence for the null hypothesis

```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

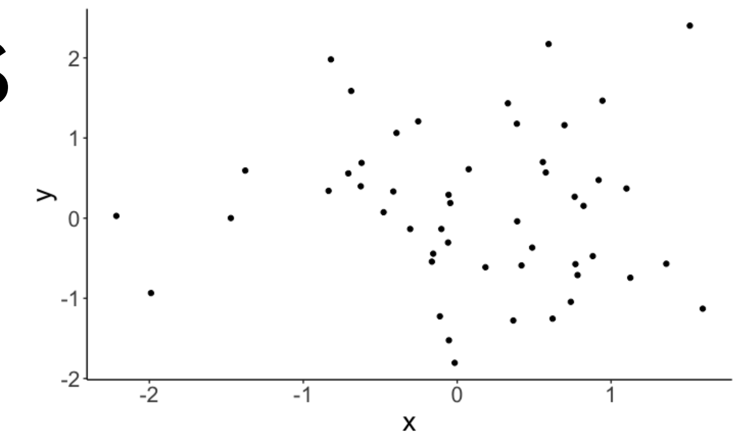



```r
fit.lm_loo = lm(formula = y ~ 1 + x,
                data = df.loo)

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!

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

coefficients in the model

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

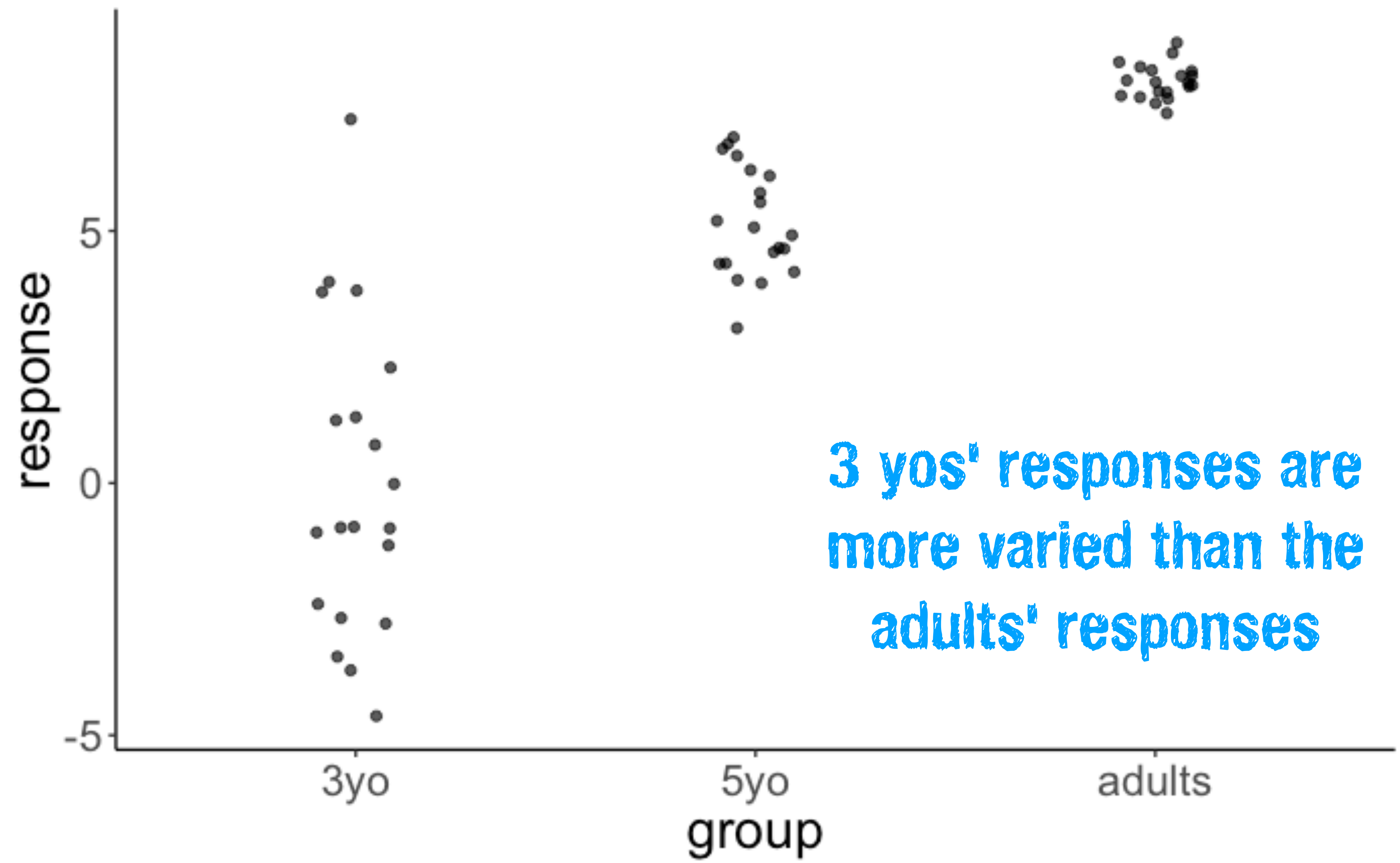
lower bound = 0

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

Dealing with unequal variance

Unequal variance aka heteroscedasticity

```
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)))
```



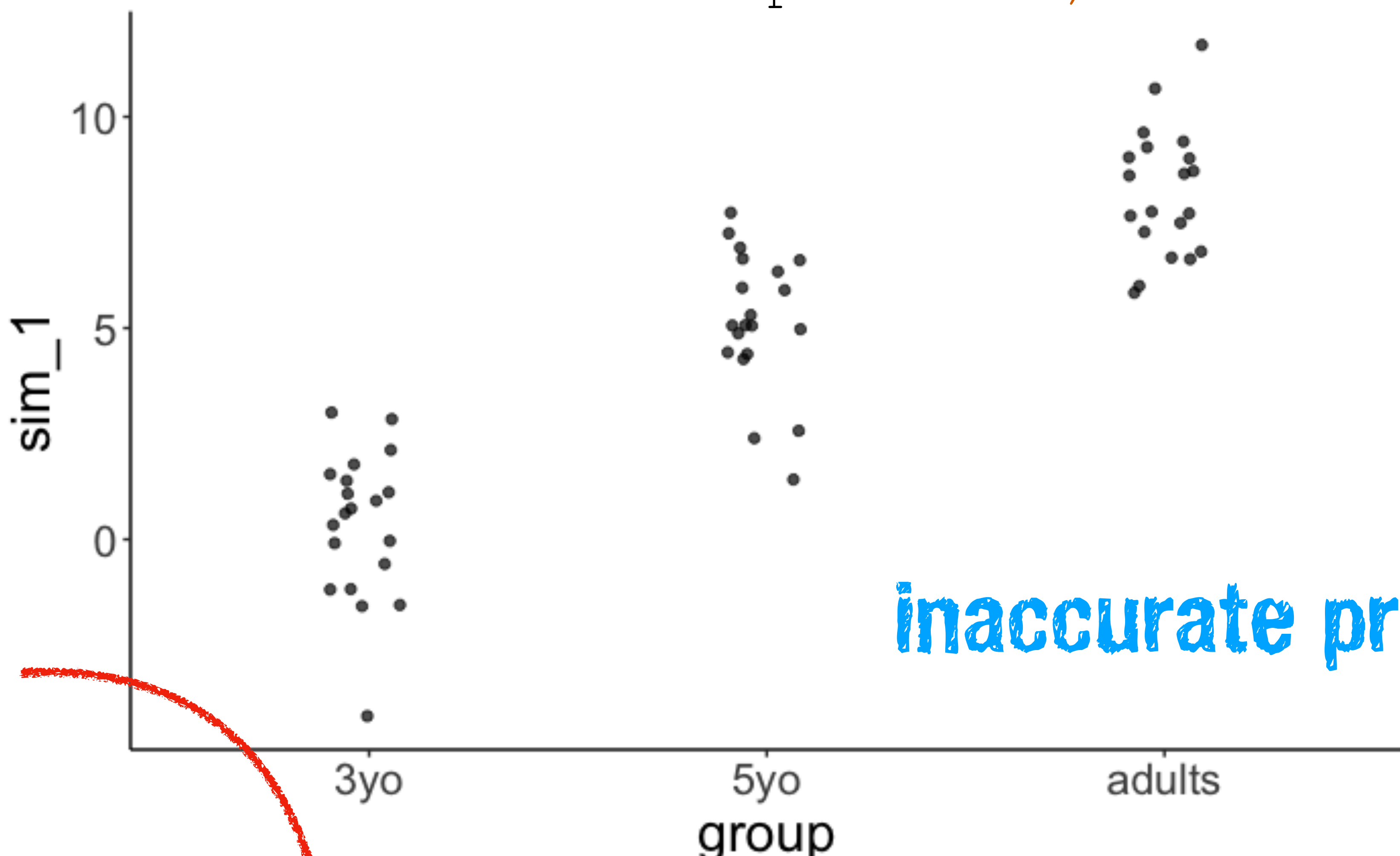
Unequal variance aka heteroscedasticity

```r
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!

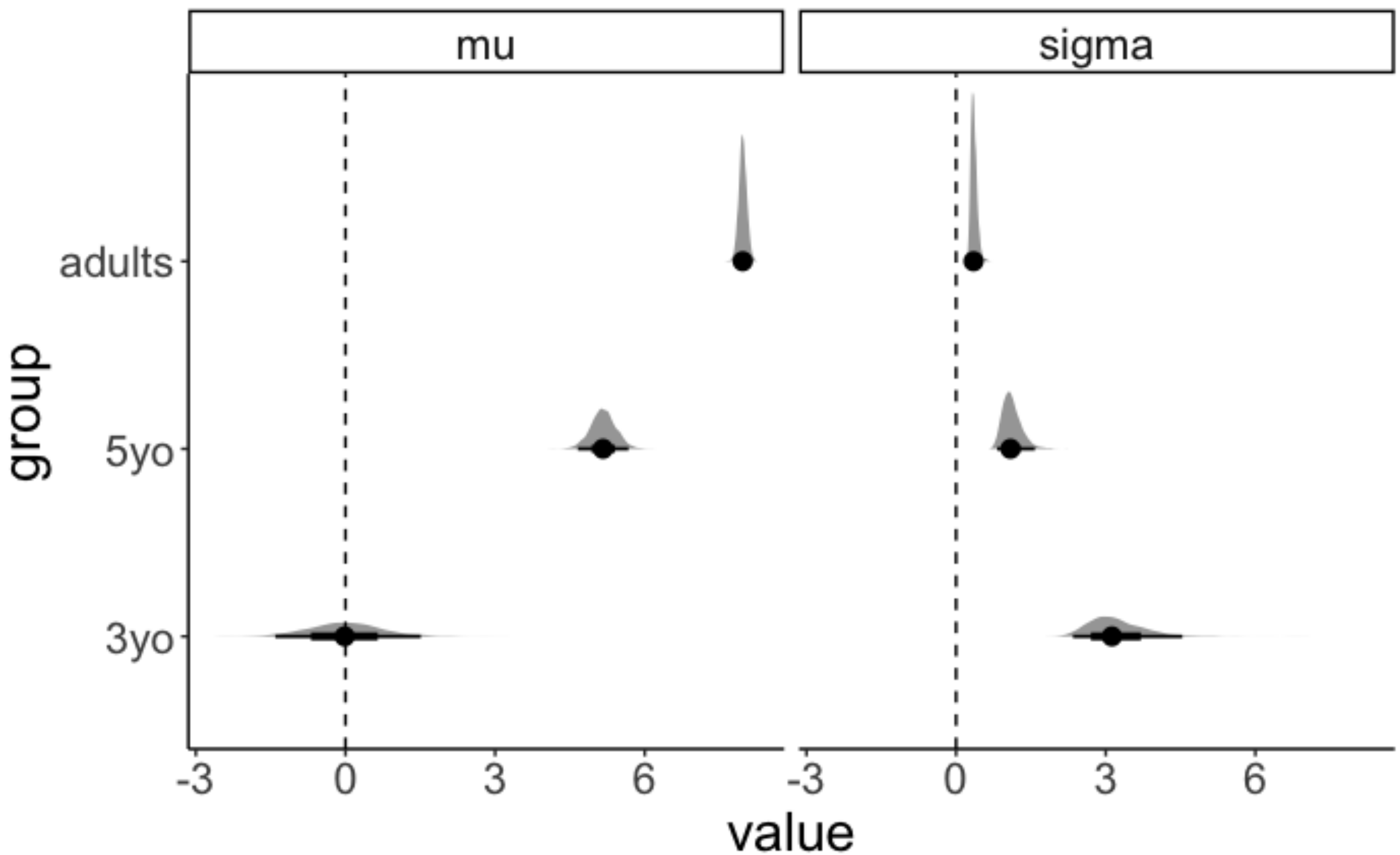
3 year olds $e^{1.15} = 3.16$

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

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

Unequal variance aka heteroscedasticity

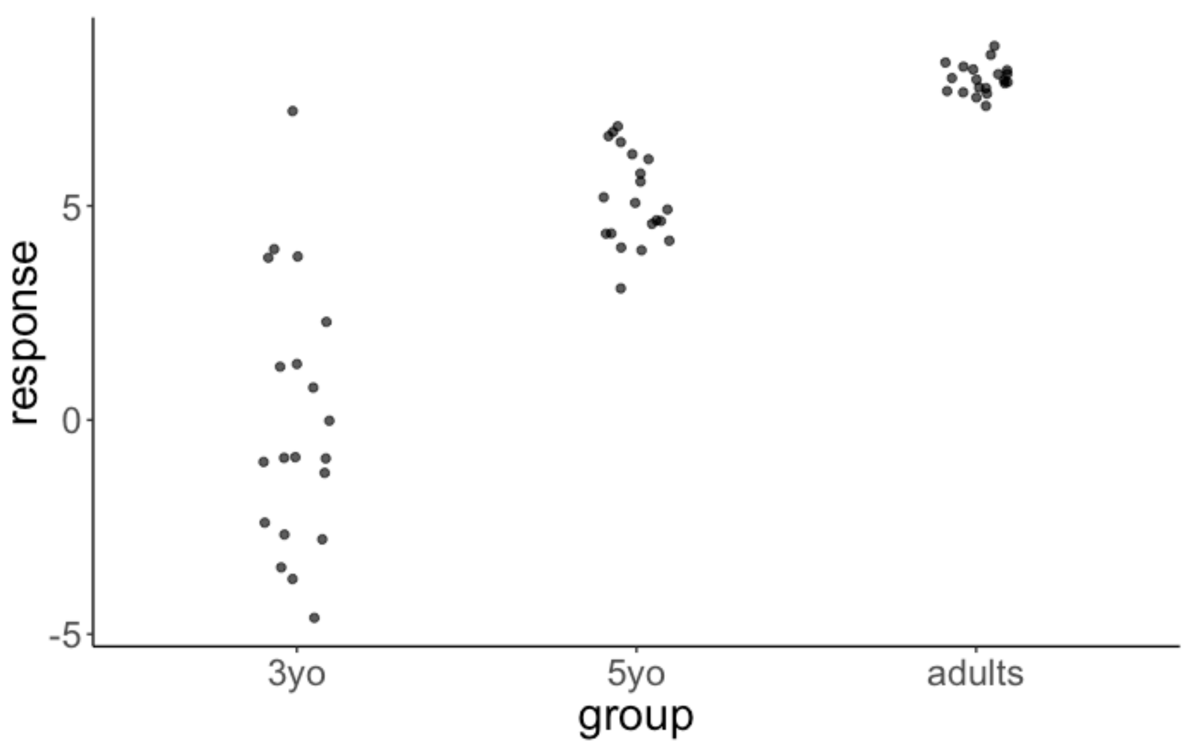
```
df.variance %>%
  expand(group) %>%
  add_fitted_draws(fit.brm1, dpar = TRUE) %>%
  select(group, .row, .draw, posterior = .value, mu, sigma) %>%
  pivot_longer(cols = c(mu, sigma),
               names_to = "index",
               values_to = "value") %>%
  ggplot(aes(x = value, y = group)) +
 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



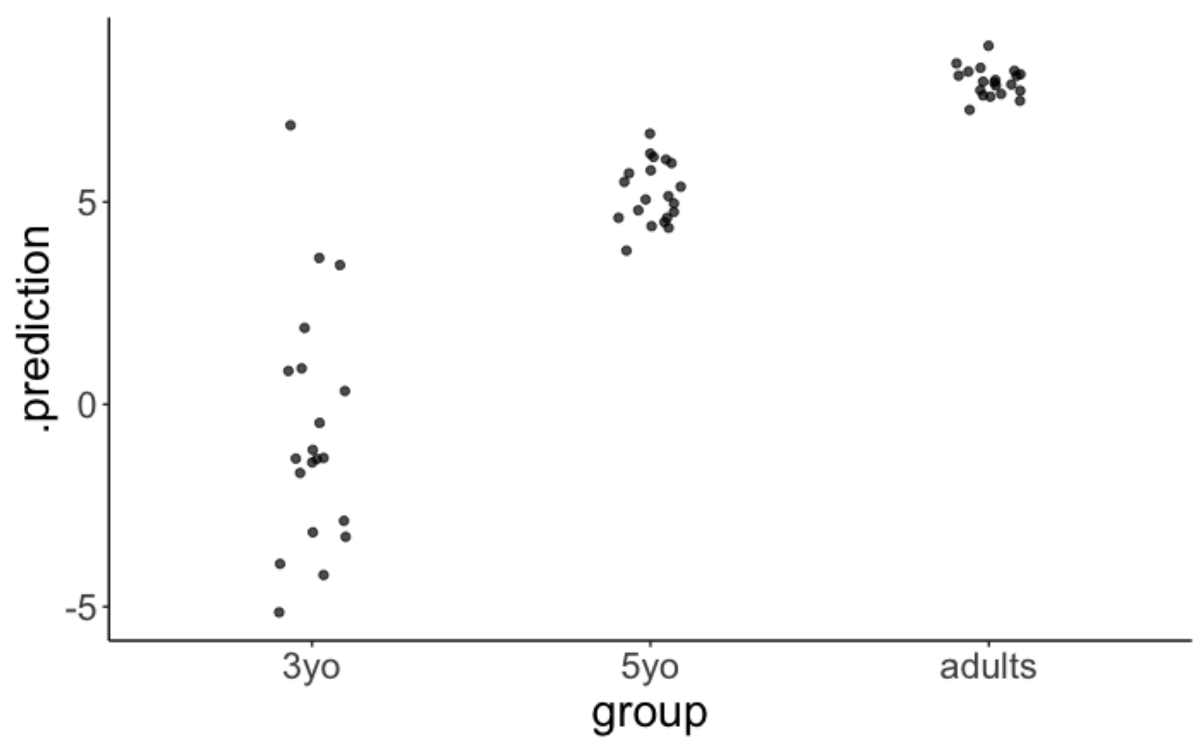
Unequal variance aka heteroscedasticity

```
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

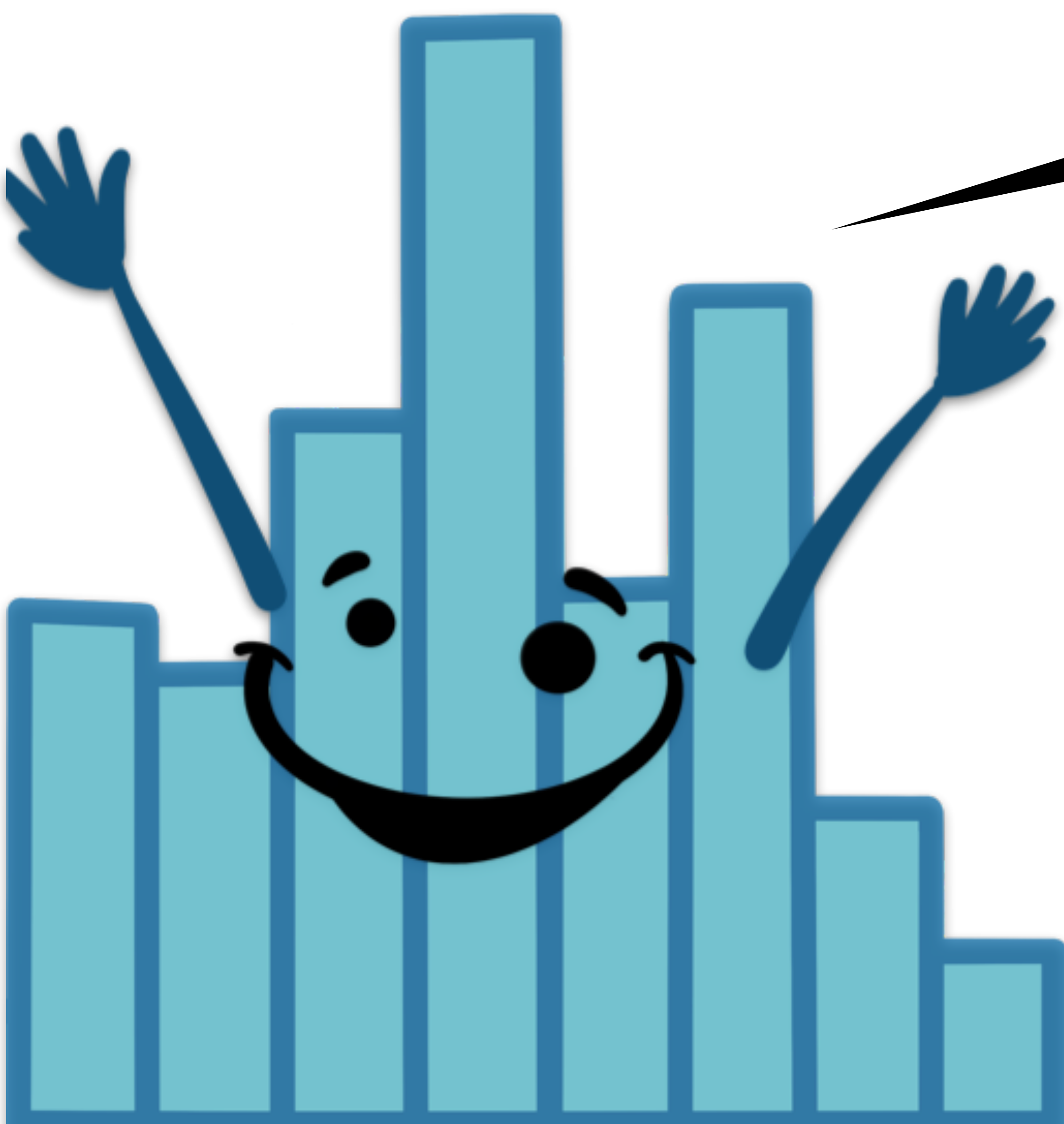


these predictions look good!

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

02:00

stretch break!



Better modeling slider data

Example taken from …

How to analyze visual analog (slider) scale data?

A reasonable choice might be the zero-one-inflated beta model

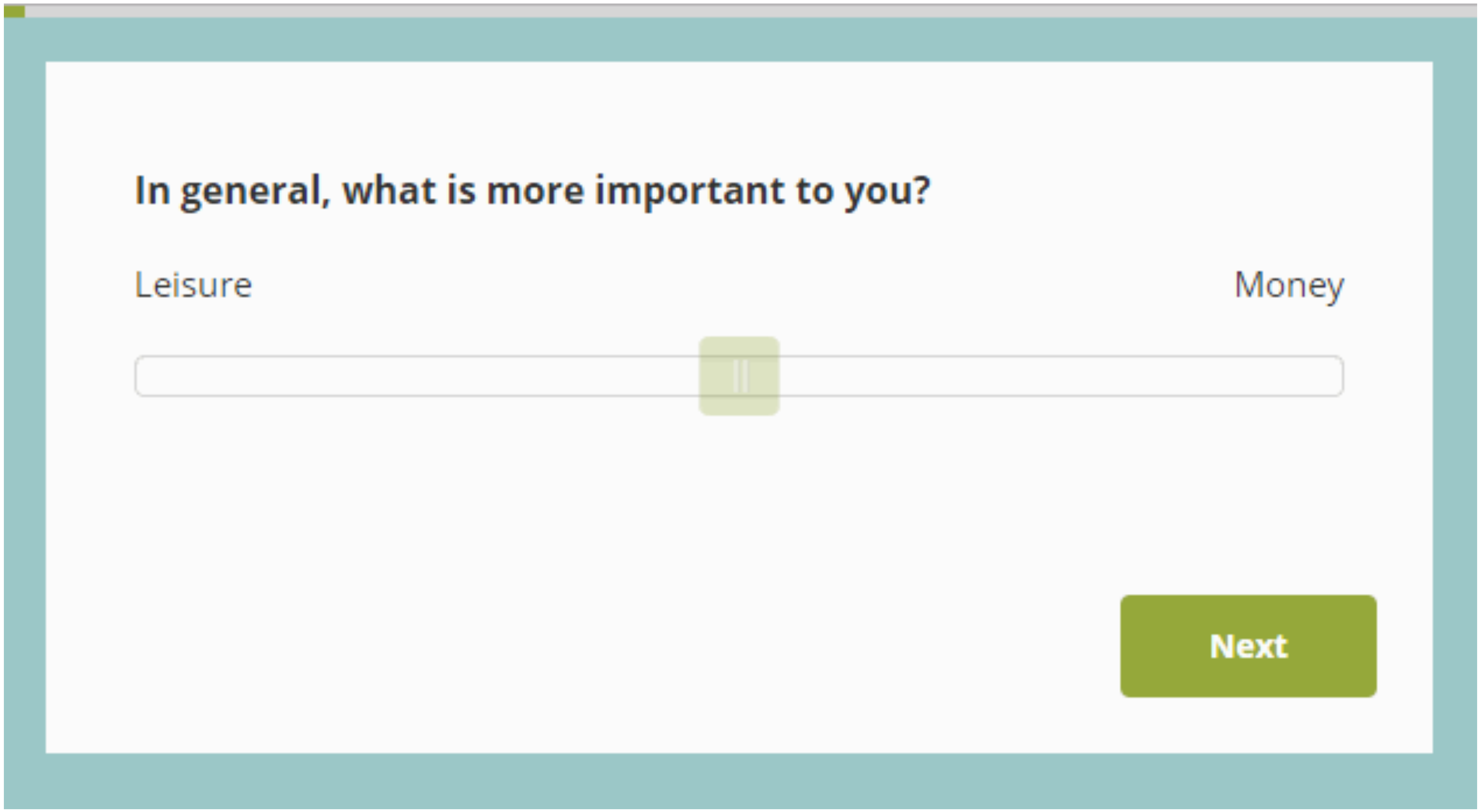
Feb 18, 2019 · 25 min read · psychology, statistics

Introduction

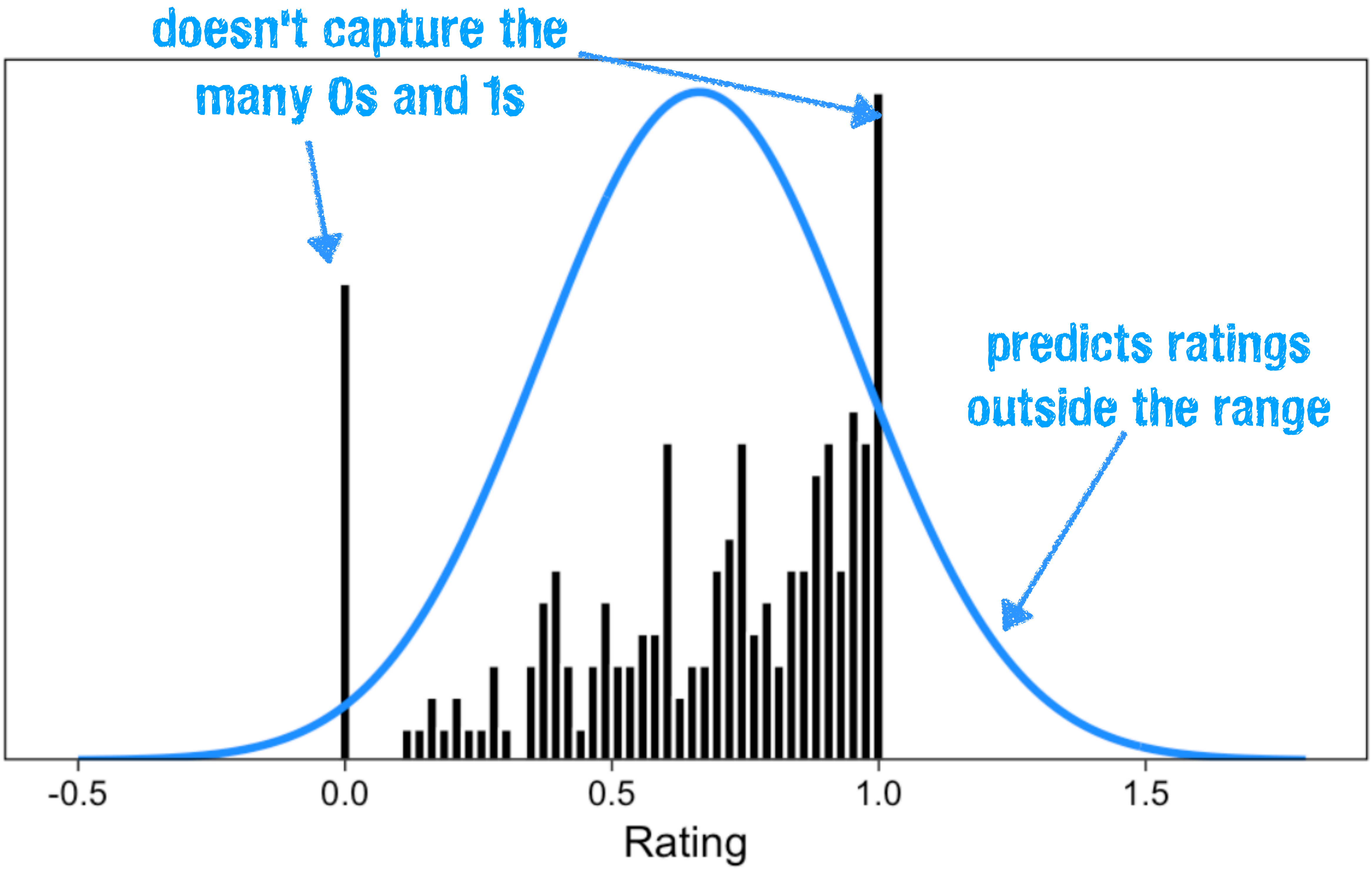
In psychological experiments, subjective responses are often collected using two types of response scales: ordinal and visual analog scales. These scales are unlikely to provide normally distributed data. However, researchers often analyze responses from these scales with models that assume normality of the data.[1]

Ordinal scales, of which binary ratings are a special case, provide ordinal data and are thus better analyzed using ordinal models (Bürkner and Vuorre 2018; Liddell and Kruschke 2018).

https://vuorre.netlify.com/post/2019/02/18/analyze-analog-scale-ratings-with-zero-one-inflated-beta-models/#zoib-regression


In general, what is more important to you?
Leisure
Money
Next

Normality assumption is (almost always) violated



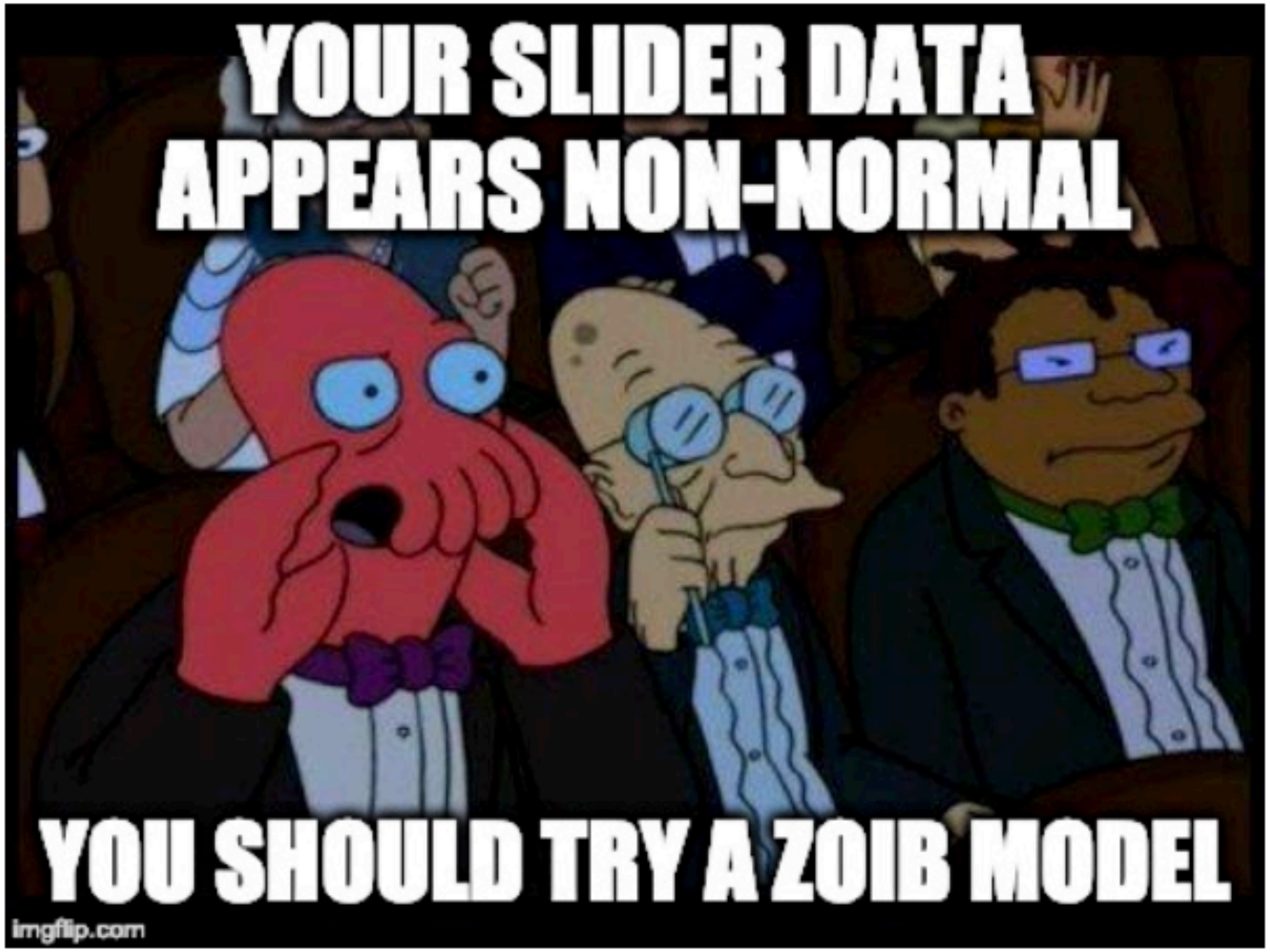
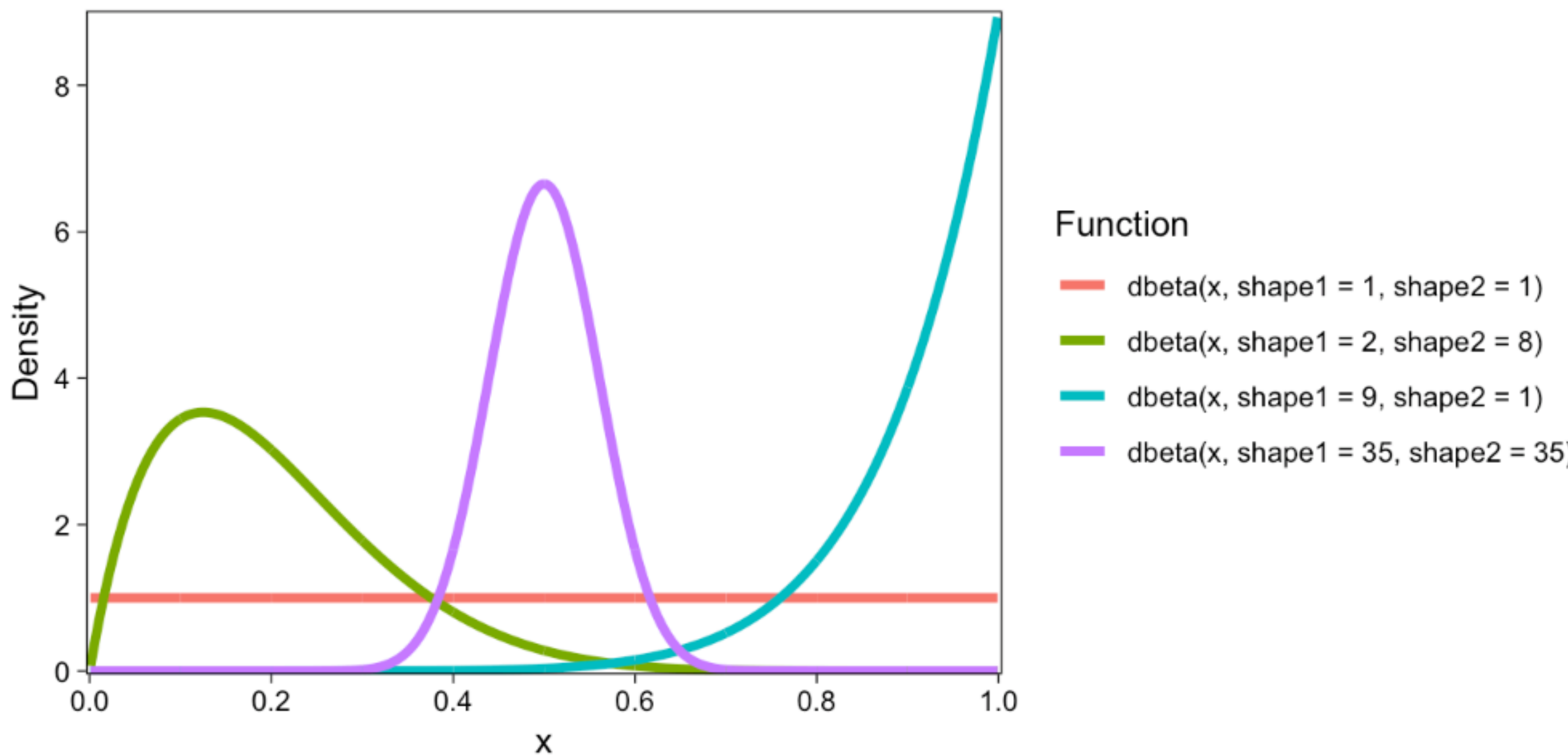


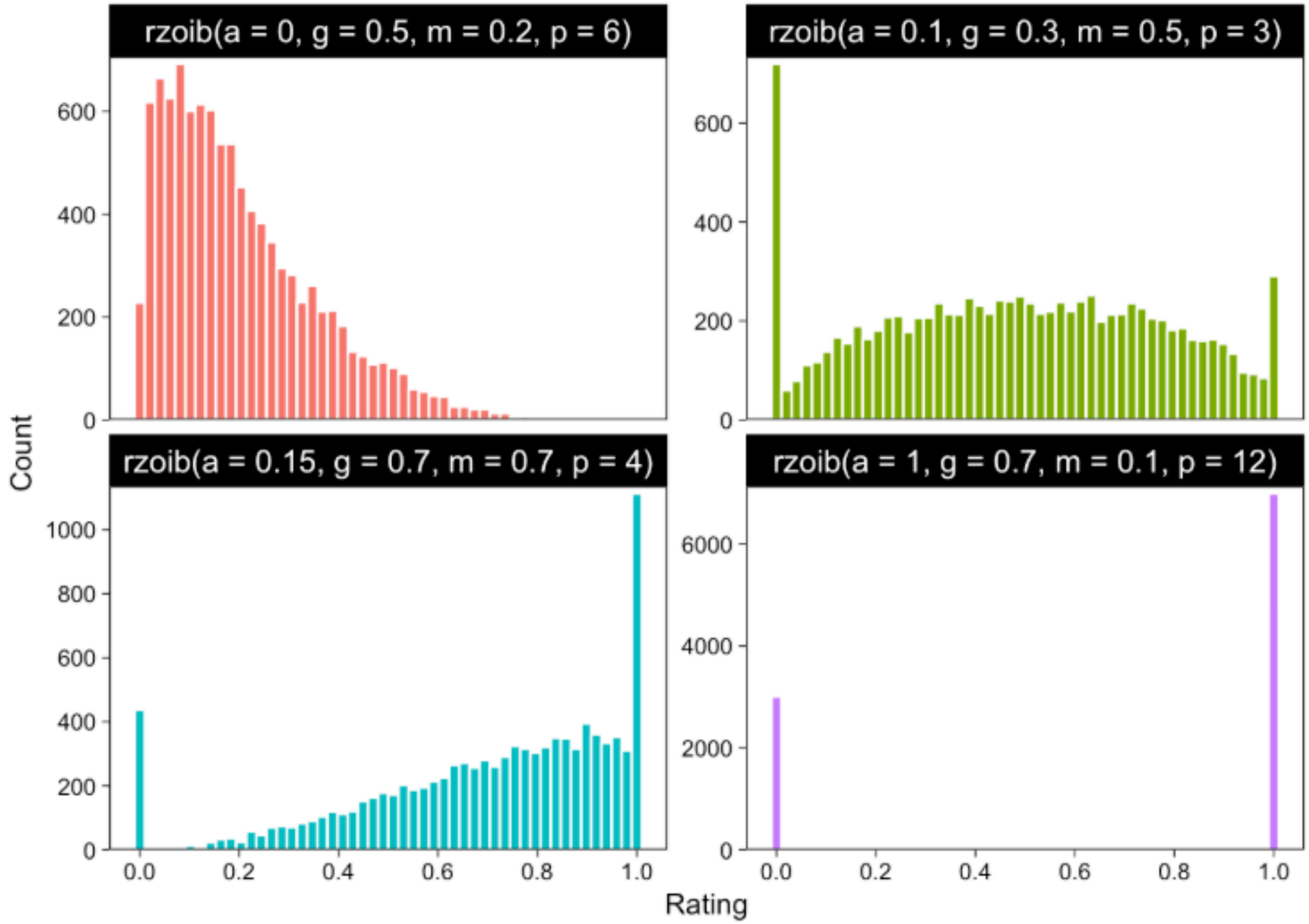
Figure 3: Dr. John A. Zoidberg thinks you should try a ZOIB model on your slider scale data.

Beta distribution



Zero-one inflated **beta** binomial model

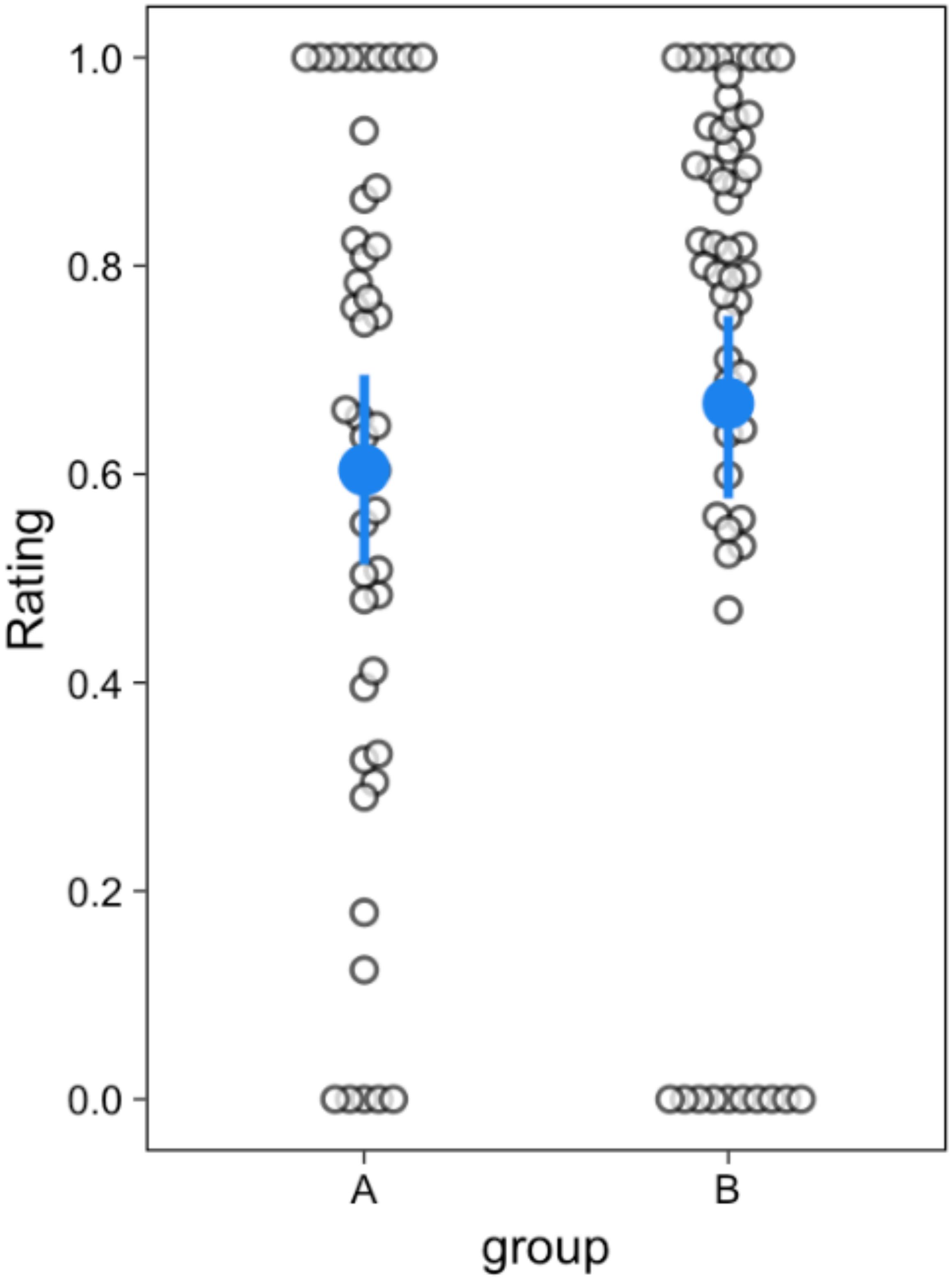
Zero-one inflated beta binomial model



Generative process

Some chance the a person will pick a 0 or 1, if not then response is determined by the beta distribution.

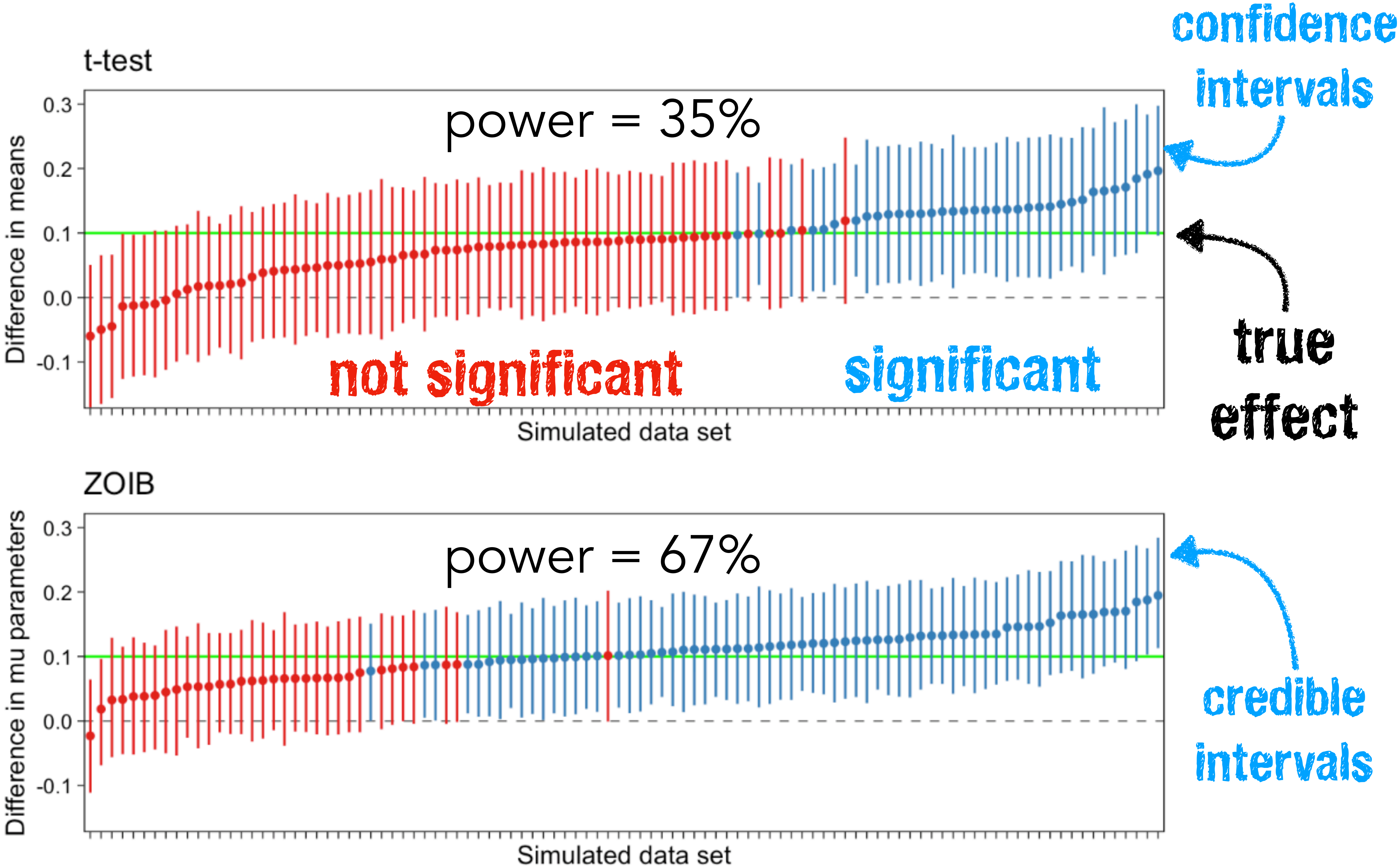
Example data



Fit the ZOIB

```
zoib_model = bf(
  Rating ~ group,
  phi ~ group,
  zoi ~ group,
  coi ~ group,
  family = zero_one_inflated_beta()
)

fit = brm(
  formula = zoib_model,
  data = dat
)
```

Capturing the data-generating process gives you power



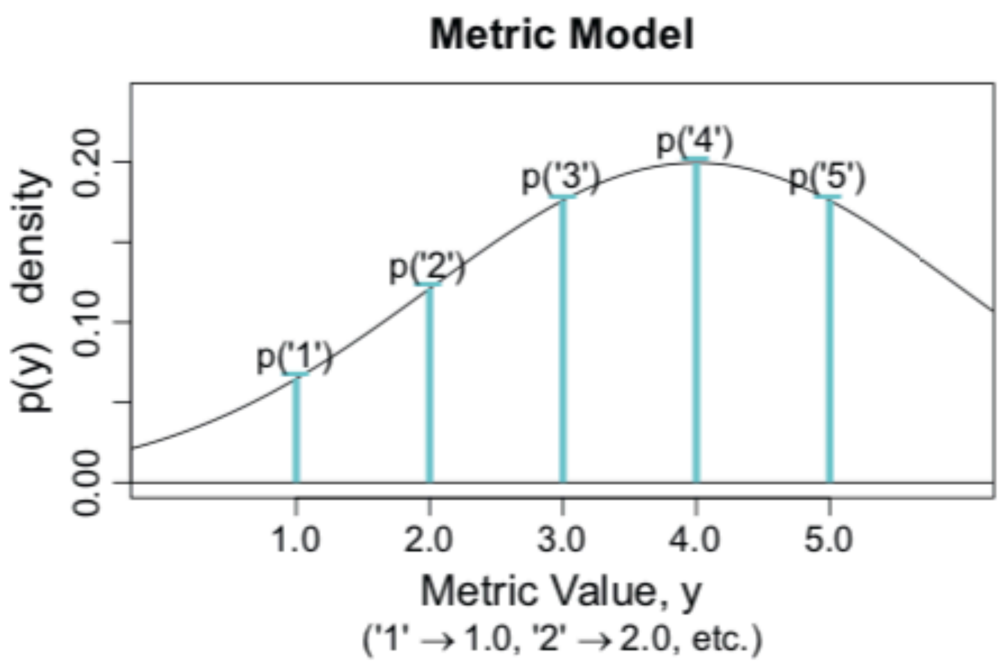
Better modeling Likert scale

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



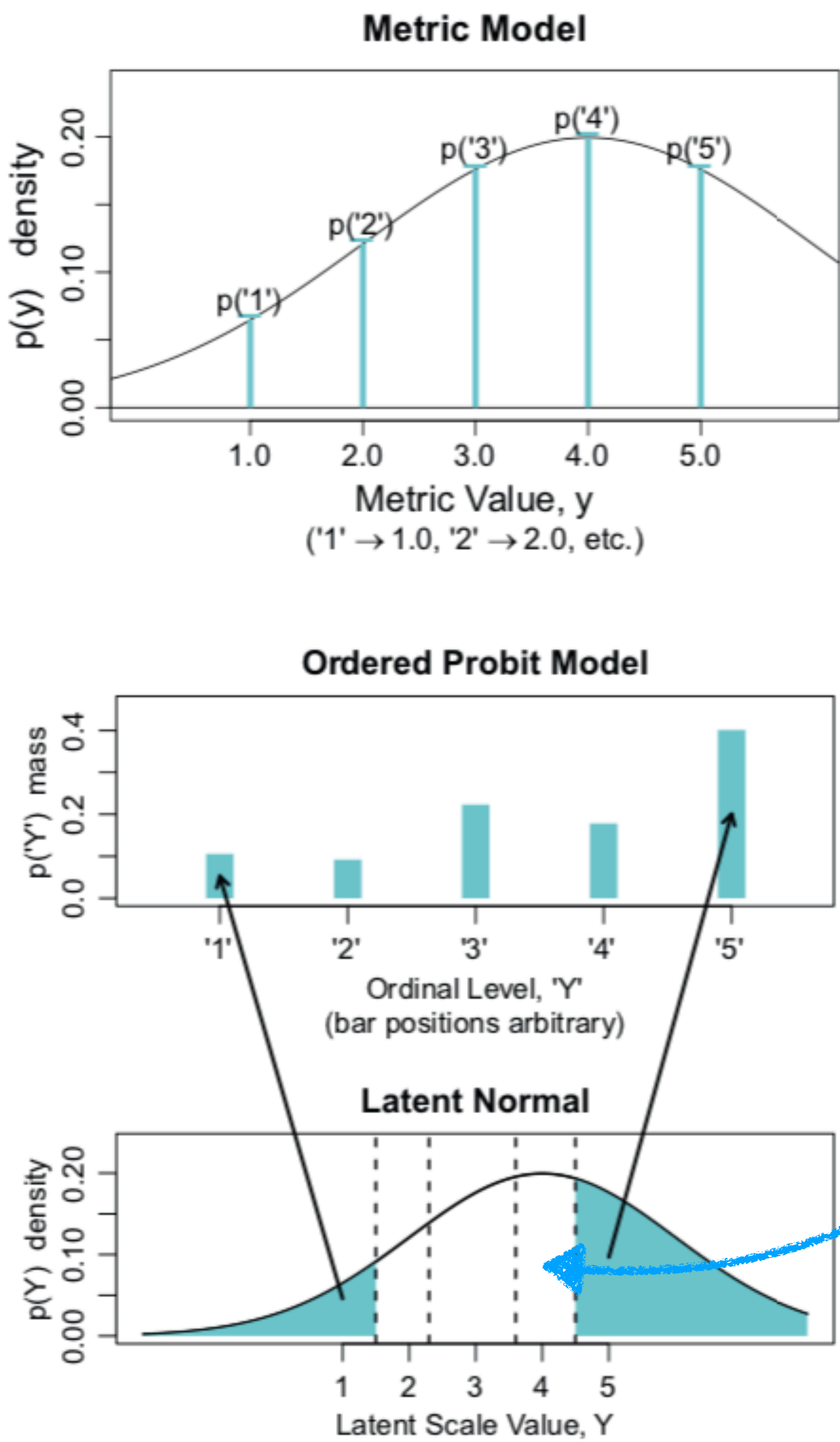
metric model

- 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

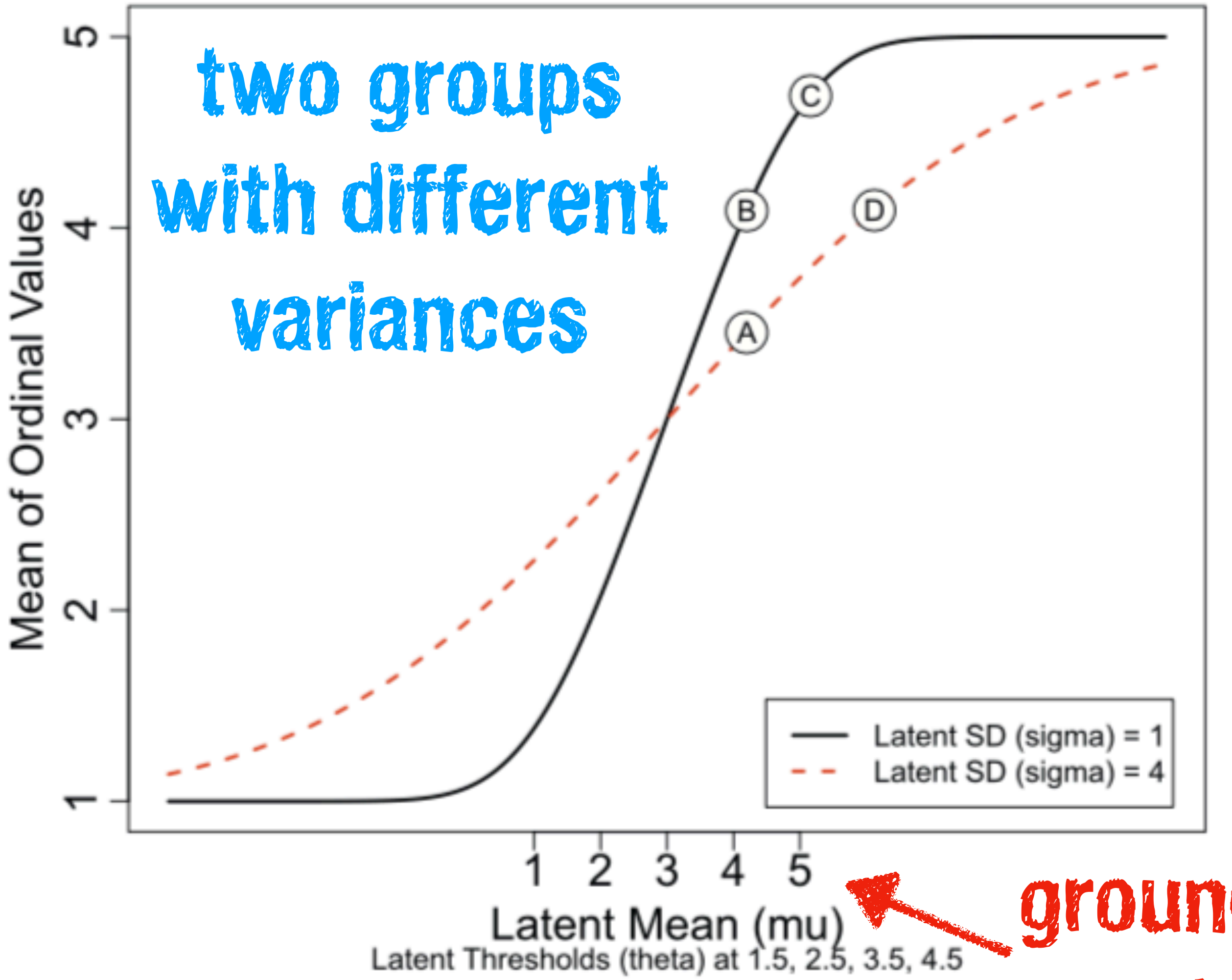
- assumes an ordering of the categories
- assumes a normal distribution in latent space
- finds a good mapping from that normal distribution in latent space to the ordered responses

finds thresholds that separate the categories

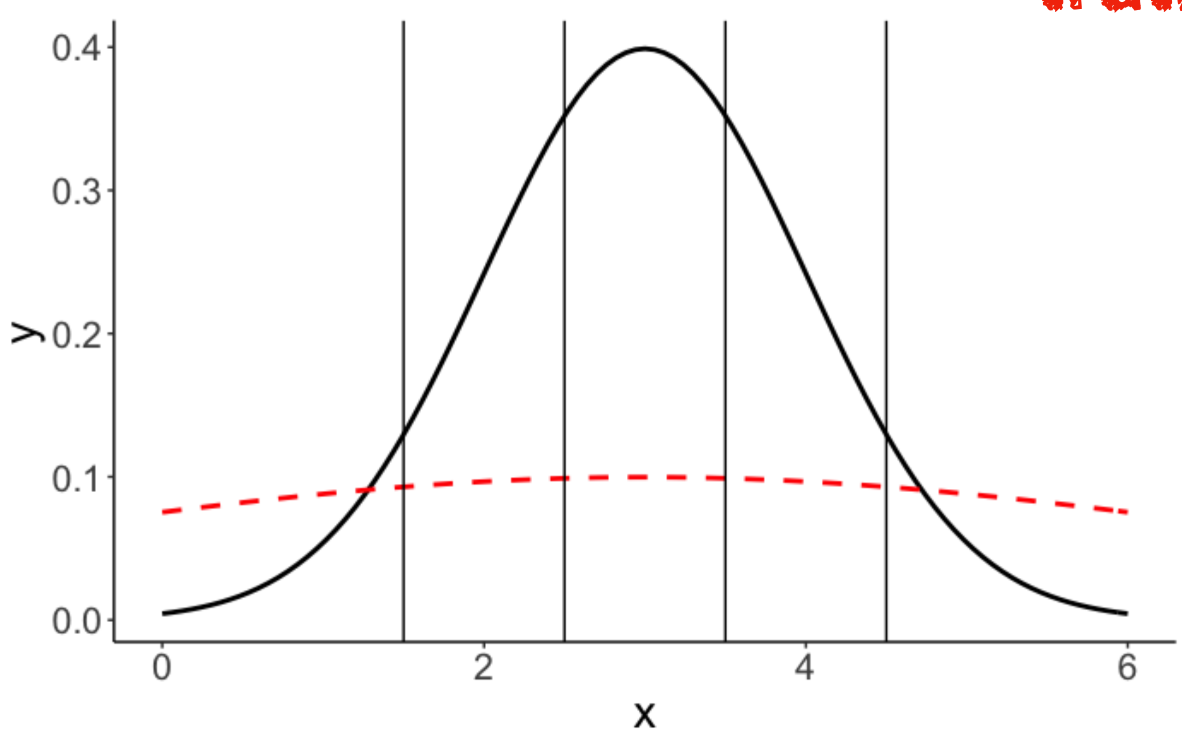
Does it matter?

- 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?



ground truth



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

  no difference in latent space

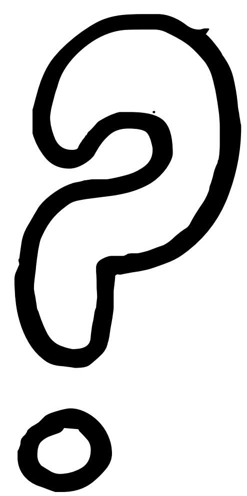
- **B** vs. **D**: false negative (Type II error)

  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%

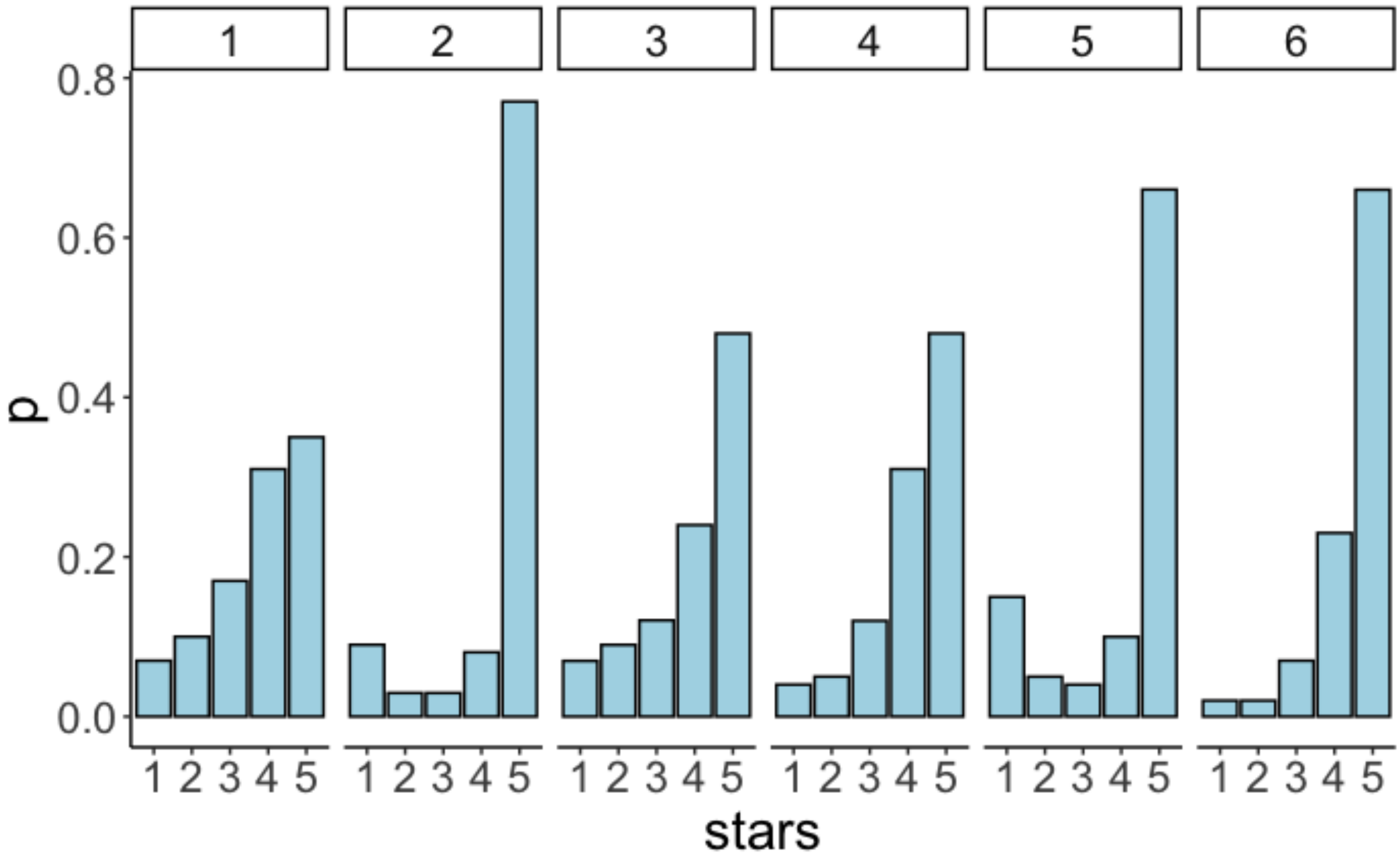
Customer reviews

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



these aren't normally distributed ...

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

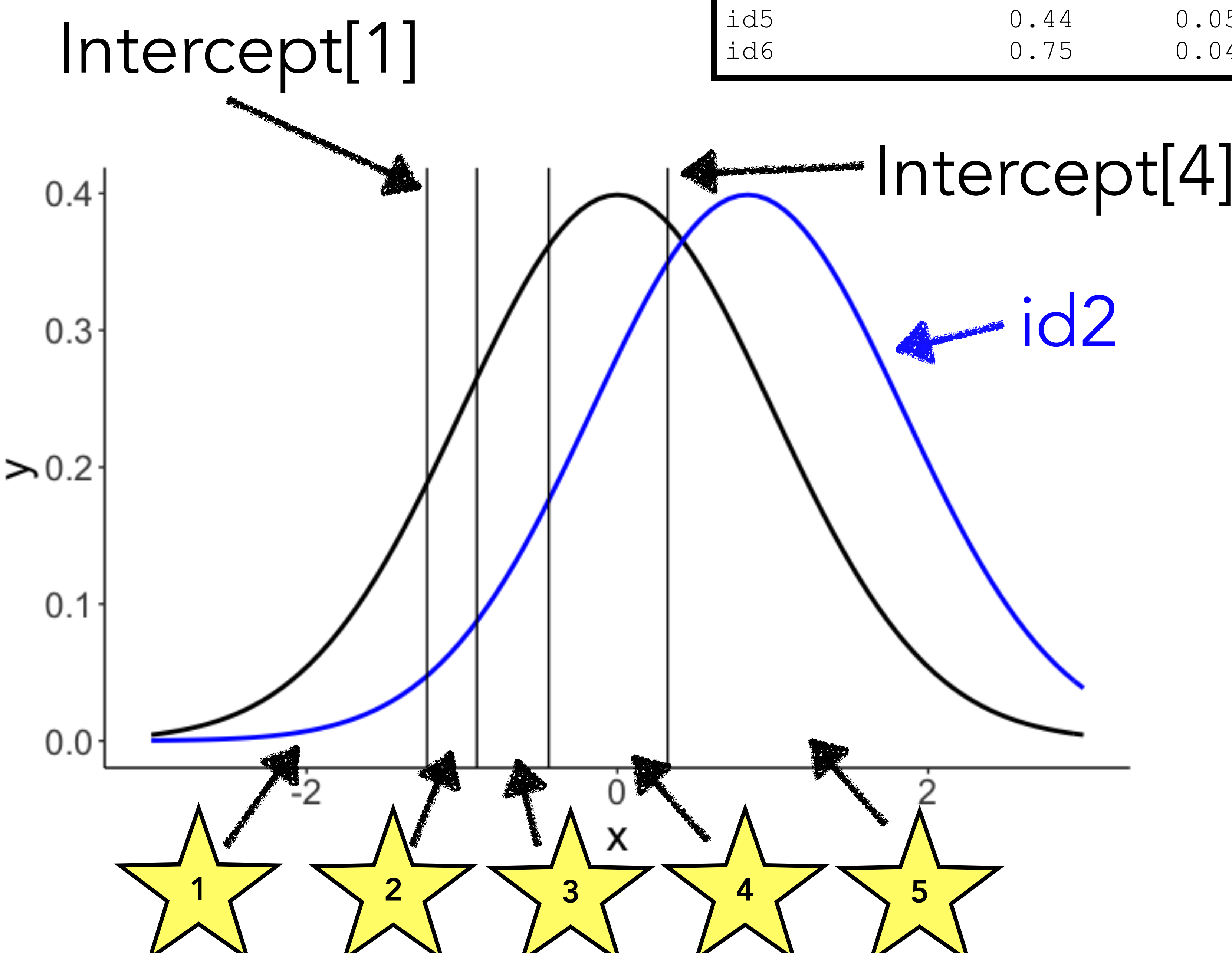
```
Family: cumulative
  Links: mu = probit; disc = identity
Formula: stars ~ 1 + id
   Data: df.movies (Number of observations: 21708)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

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

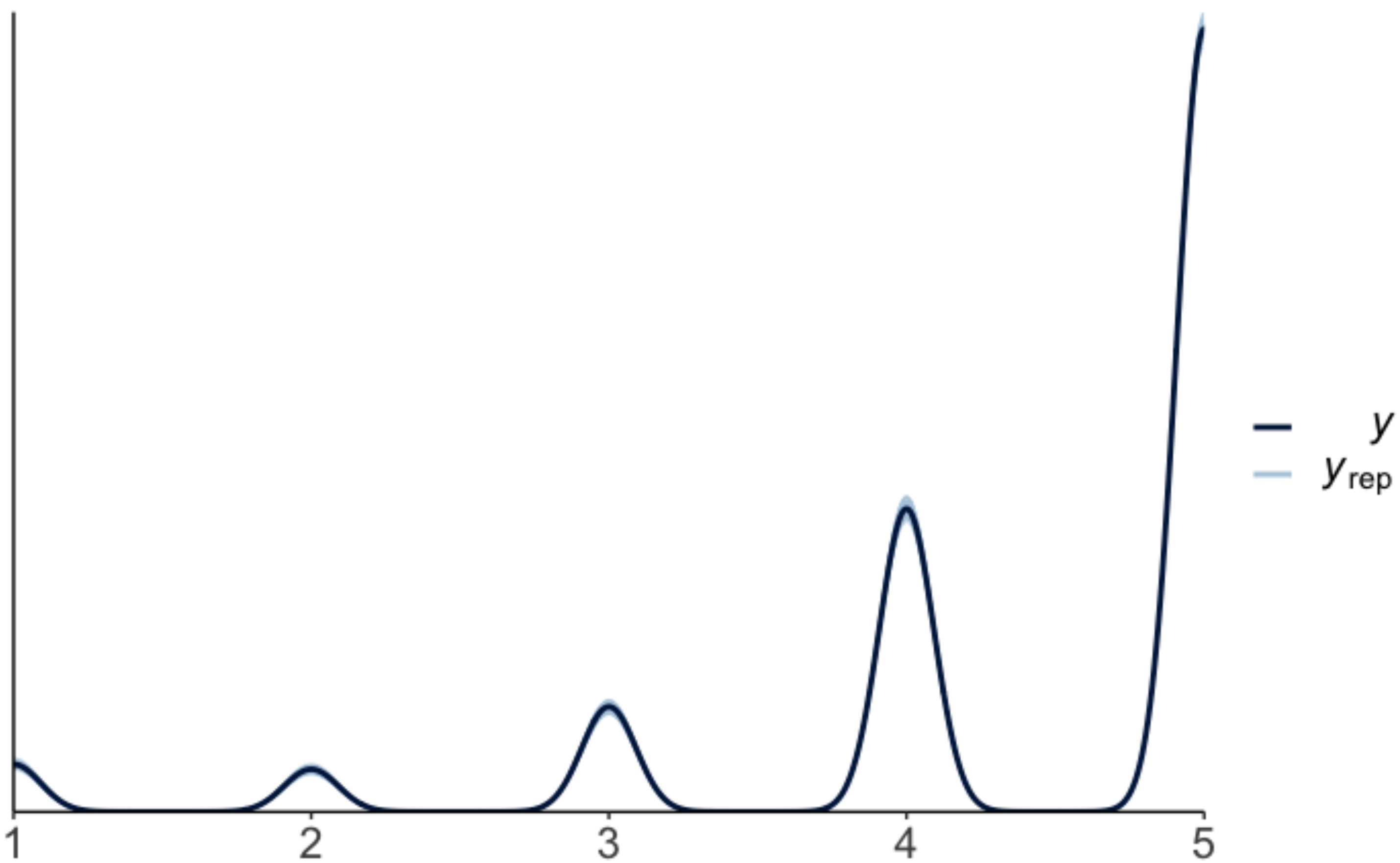
```
Family: cumulative 
  Links: mu = probit; disc = identity 
Formula: stars ~ 1 + id 
   Data: df.movies (Number of observations: 21708) 
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

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
```

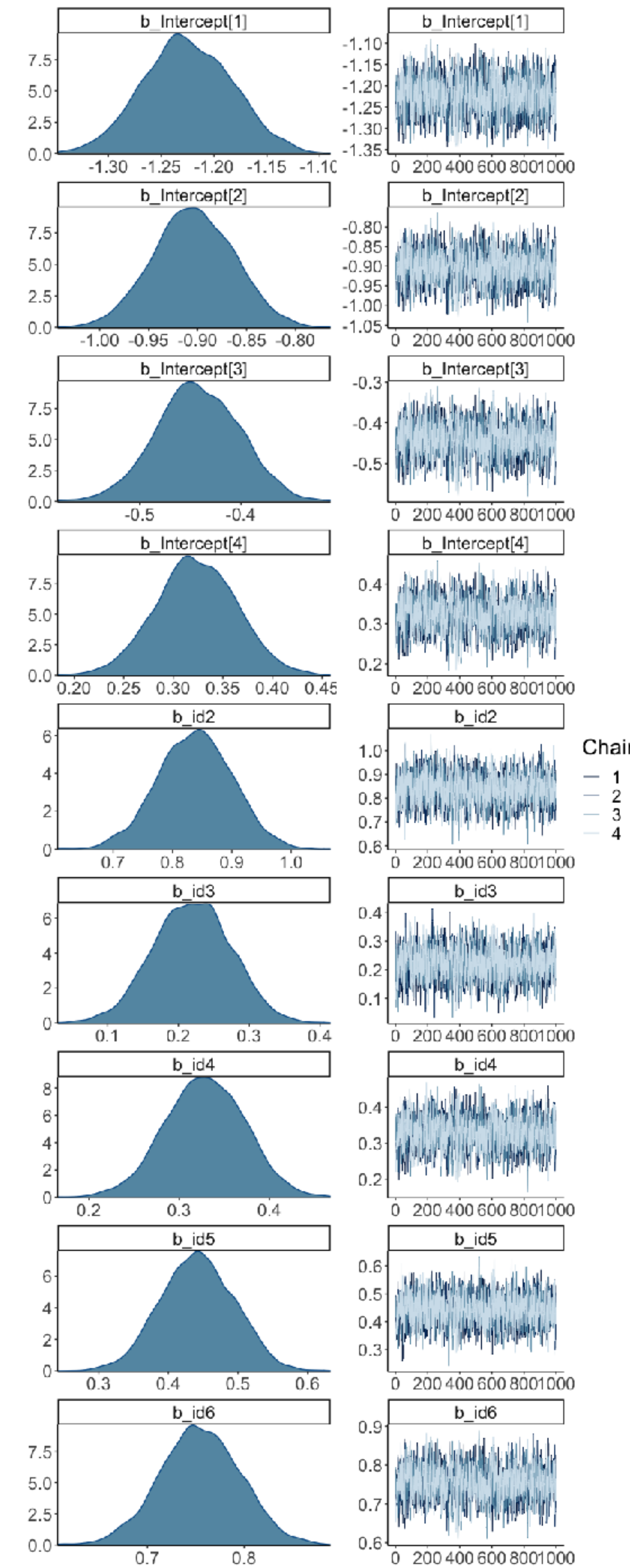


Check the model

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

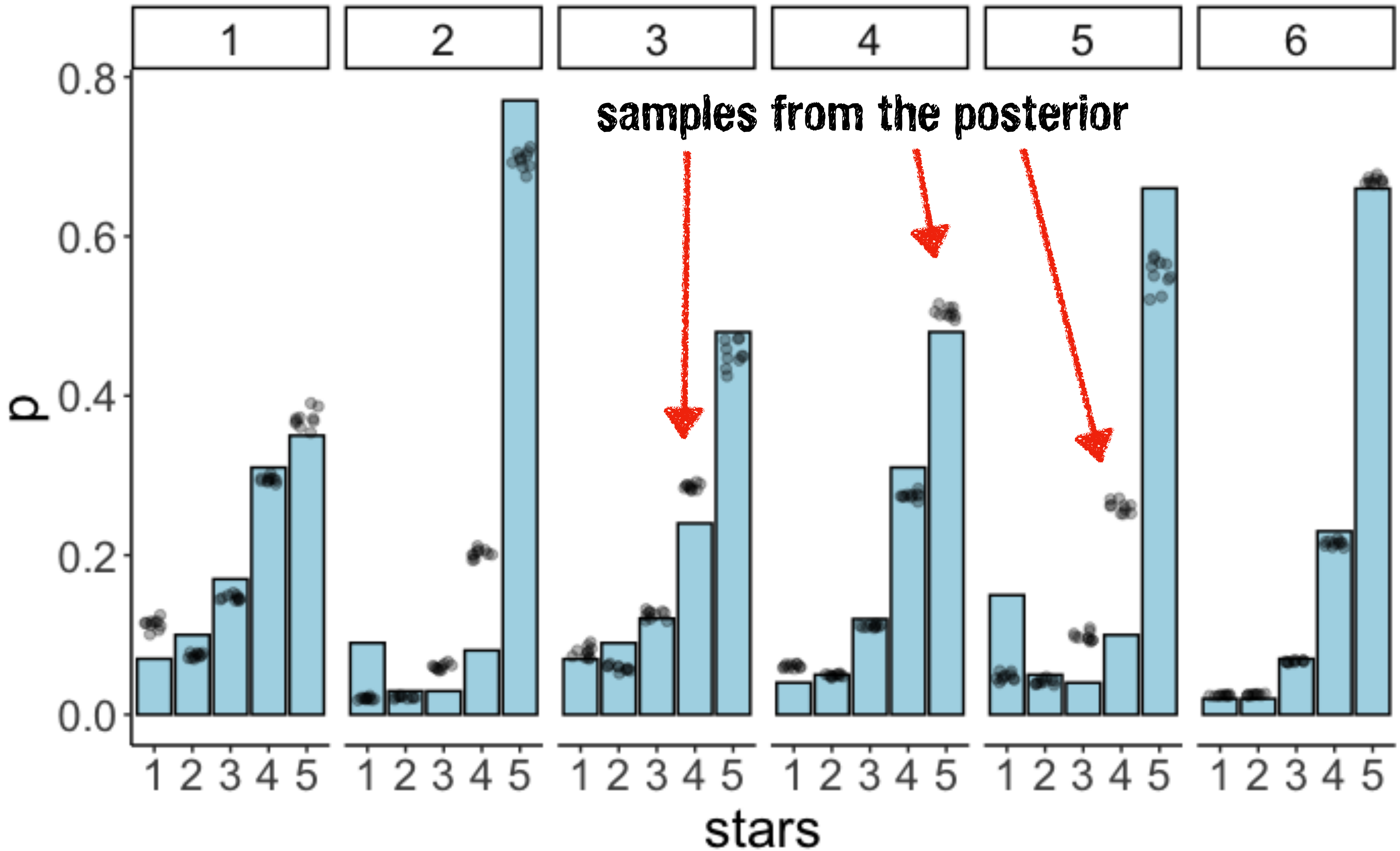
```
fit.brm5 %>%
  pp_check(nsamples = 100)
```





looking good!

Illustrate the predictions



predictions look pretty good but
maybe we can do better?

Before doing better, let's do worse!

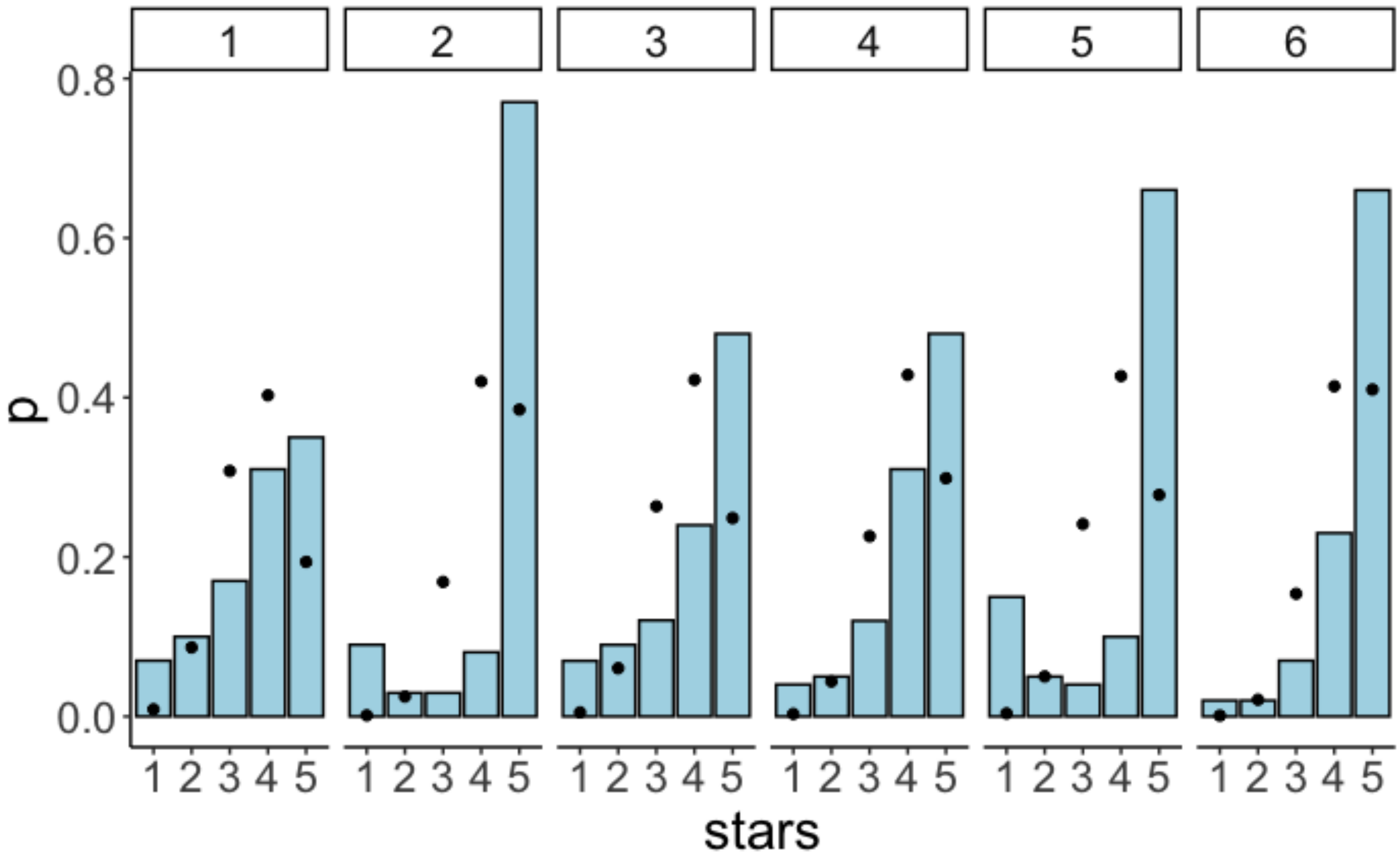
```
fit.brm6 = brm(formula = stars ~ 1 + id,
               data = df.movies,
               file = "cache/brm6",
               seed = 1)
```

mean for the reference category

difference in mean to reference category

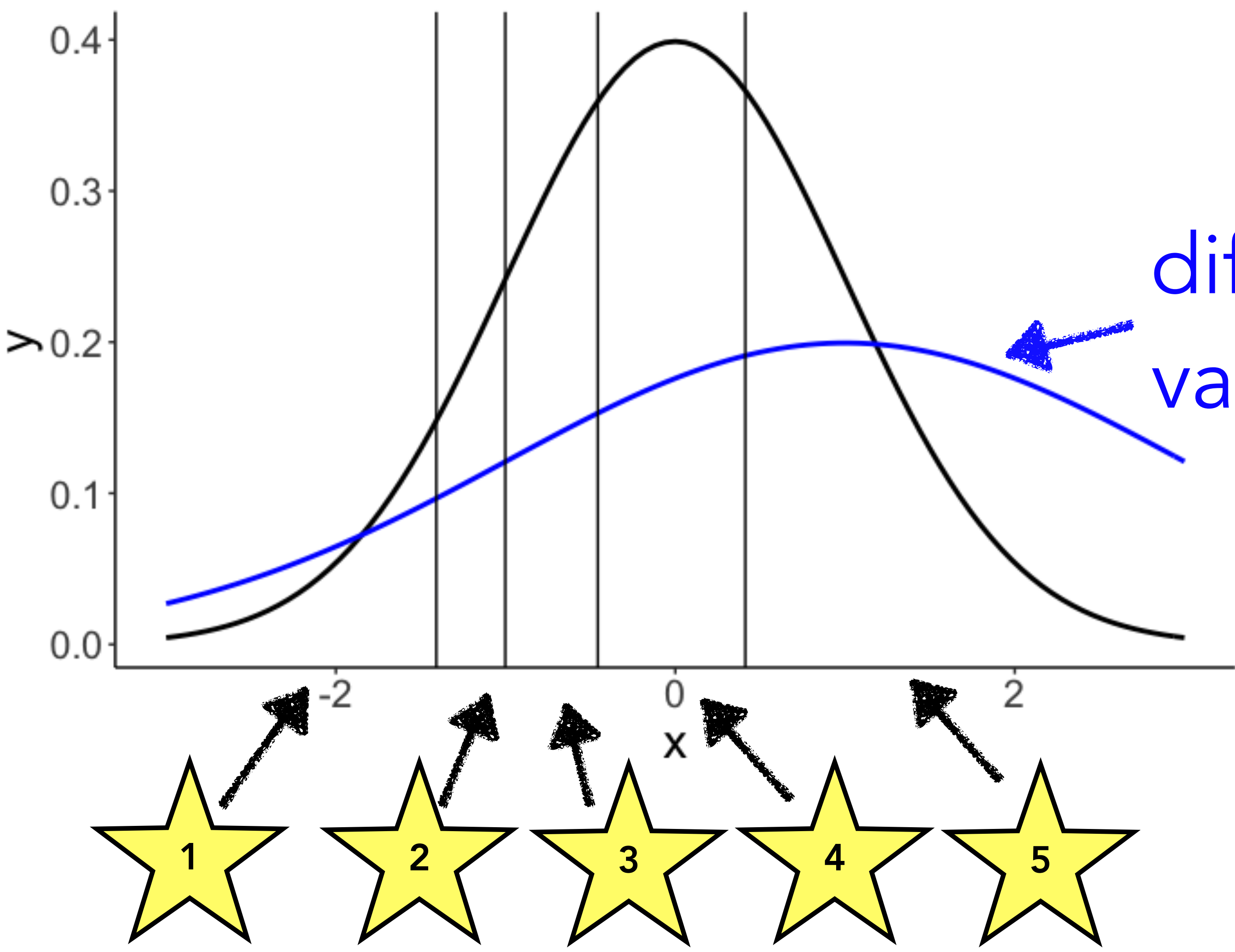
```
 Family: gaussian 
  Links: mu = identity; sigma = identity 
Formula: stars ~ 1 + id 
   Data: df.movies (Number of observations: 21708) 
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

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

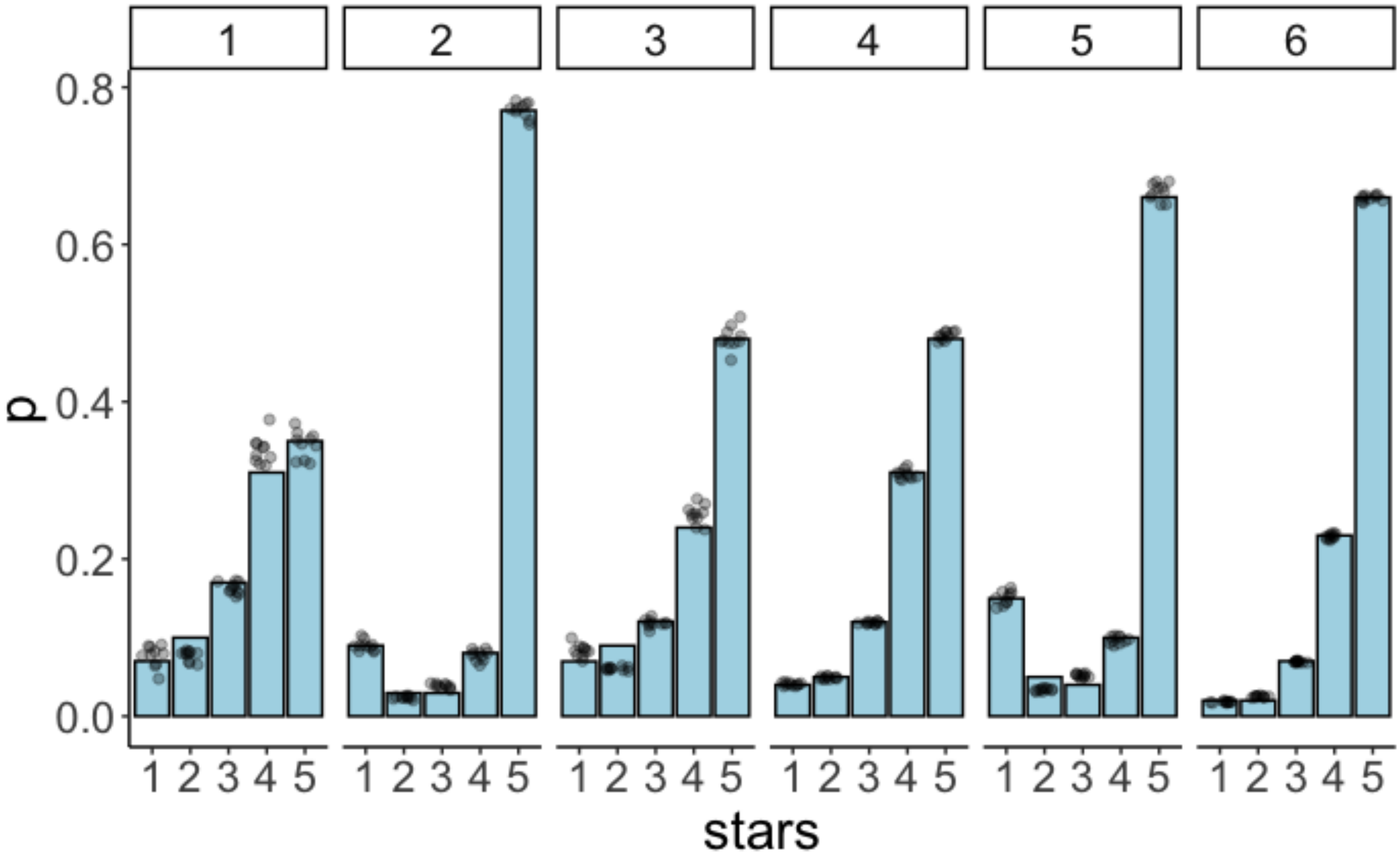
difference in mean

difference in variance

see for details

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

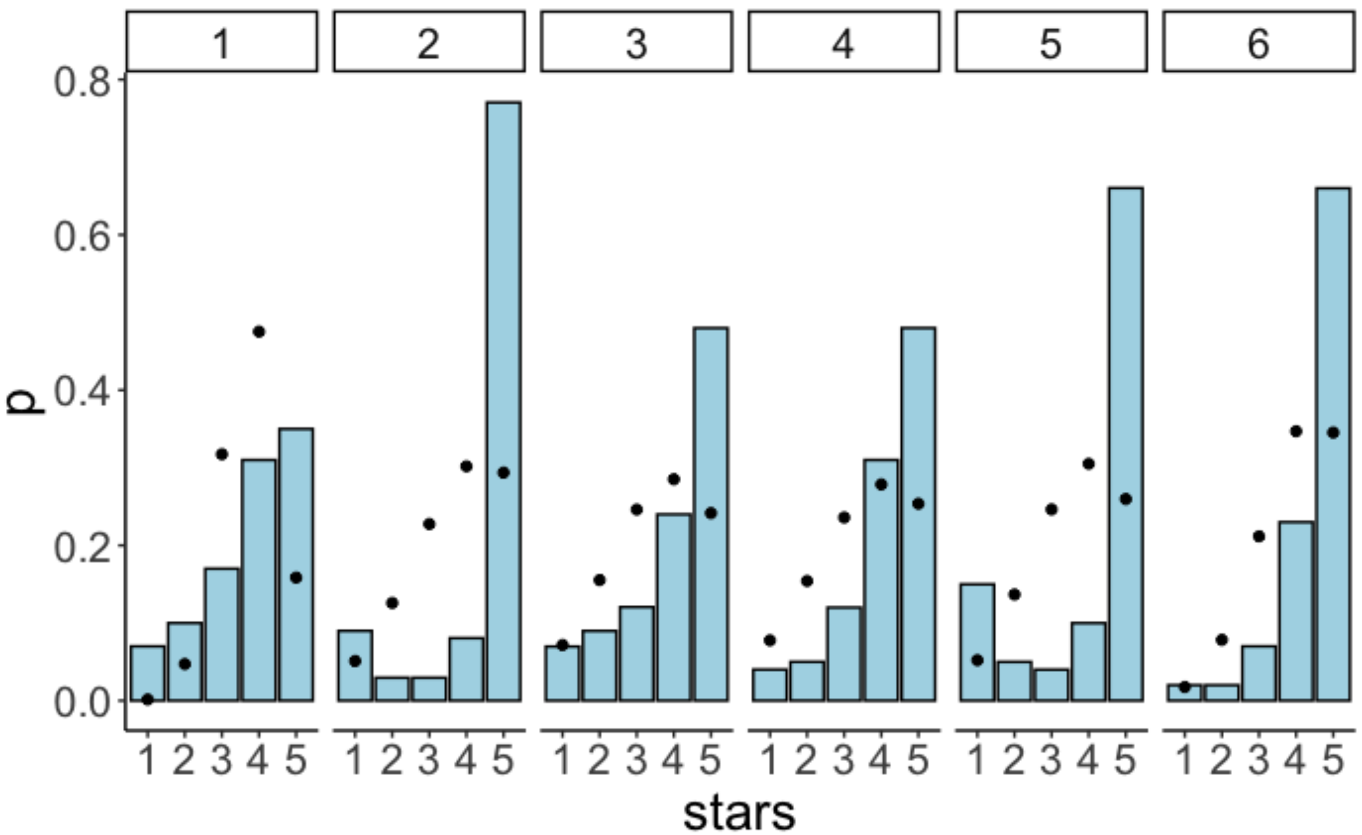
Illustrate the predictions



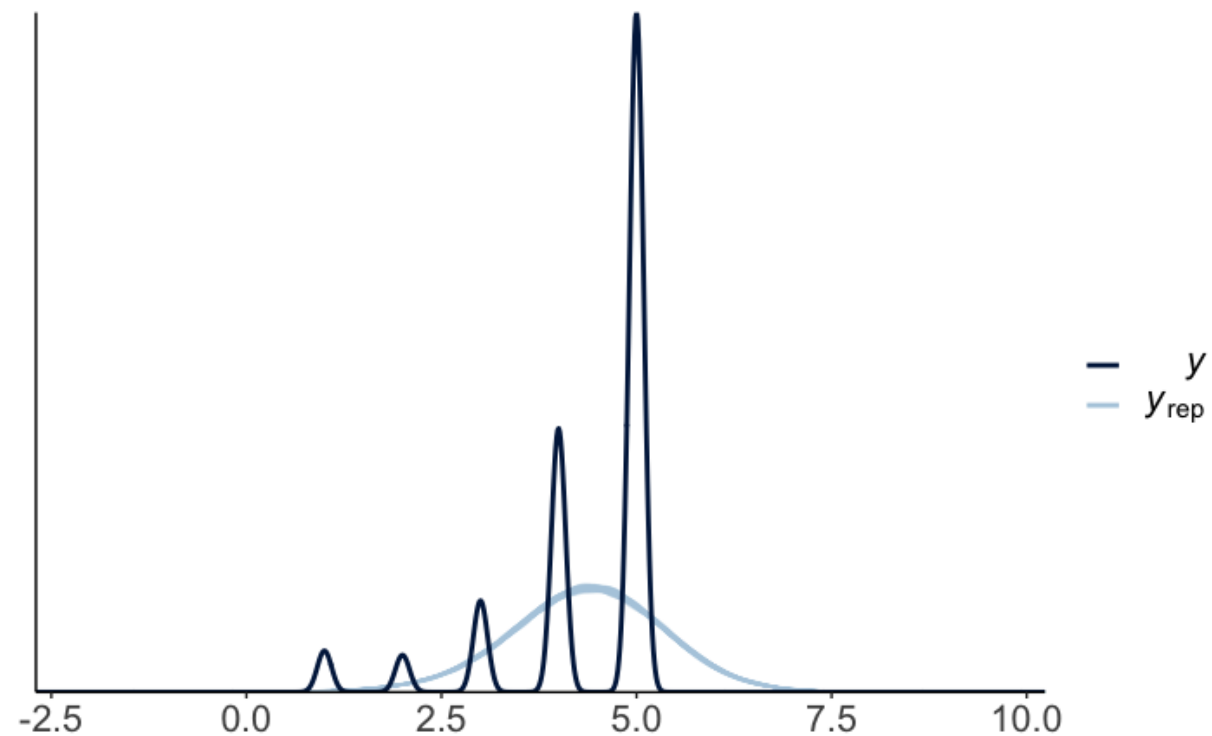
this looks excellent!

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

Sarah

Chengxu

All of you!

Abigail Rose Bergman
Alan Cheng
Cyan DeVeaux
Evelyn Rocio Fernandez-Lizarraga
Lynde Folsom
Nava Haghighi
Alexia Maria Hernandez
Steve Gutierrez Juarez
Lukas Kisunas
Michelle Lam
Alexander Peter Landry
Yoonji Lee
Wanjing Anya Ma
Dean Manko
Michelle Ng
Javier Omar
Jasmin Elena Palmer
Penny Pan
Jaylen Pittman
Ben Prystawski
Parker Ruth
Kimia Saadatian
Shawn Tyler Schwartz
Megumi Emily Takada
Josephine Chow Ying Tan
Emma Velterop
Josh Wilson
Xi Jia Zhou
Peter Guandi Zhu

Thanks!