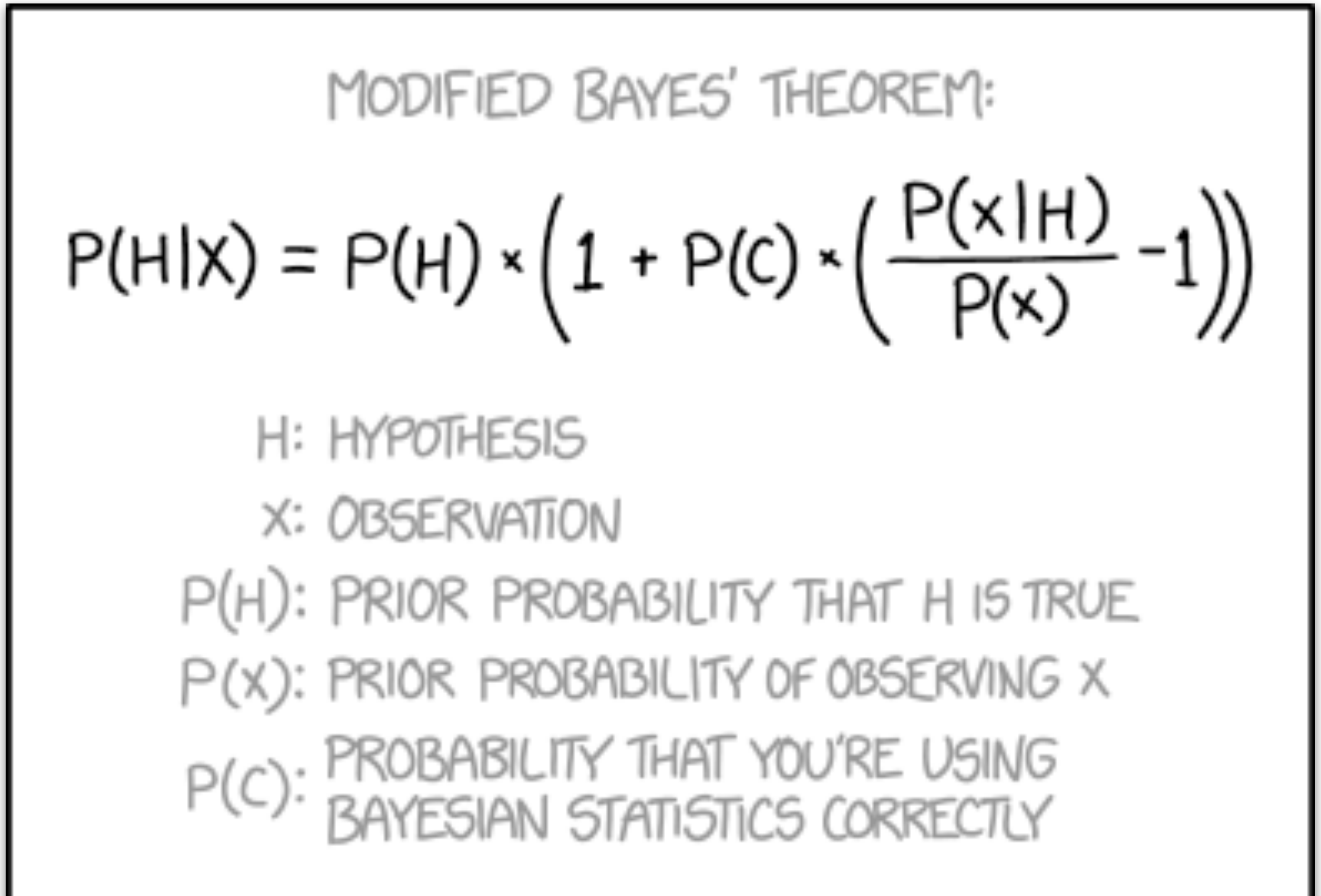
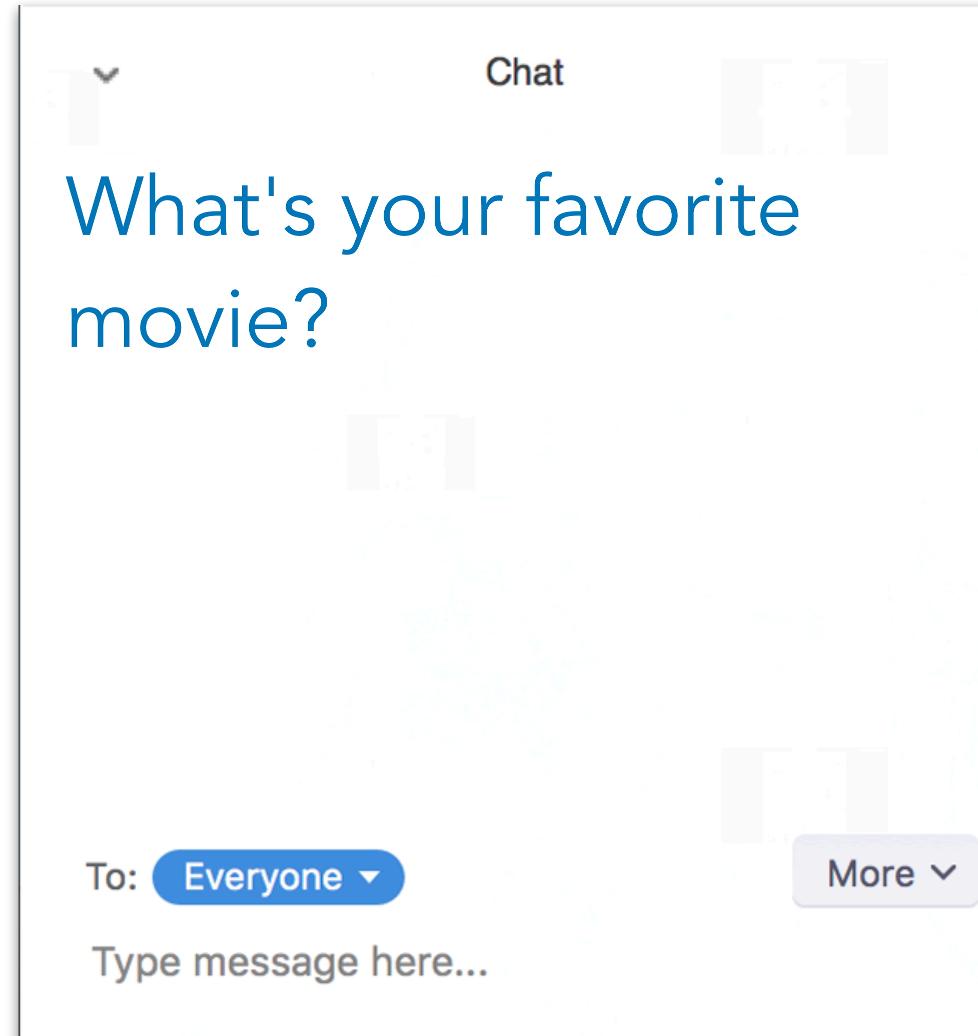
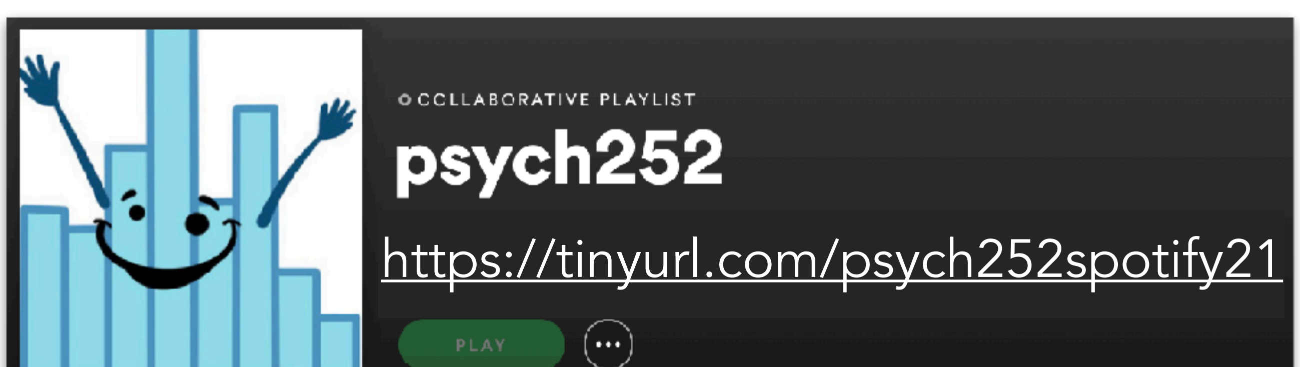
PSYCH 252: Statistical Methods for Behavioral and Social Sciences

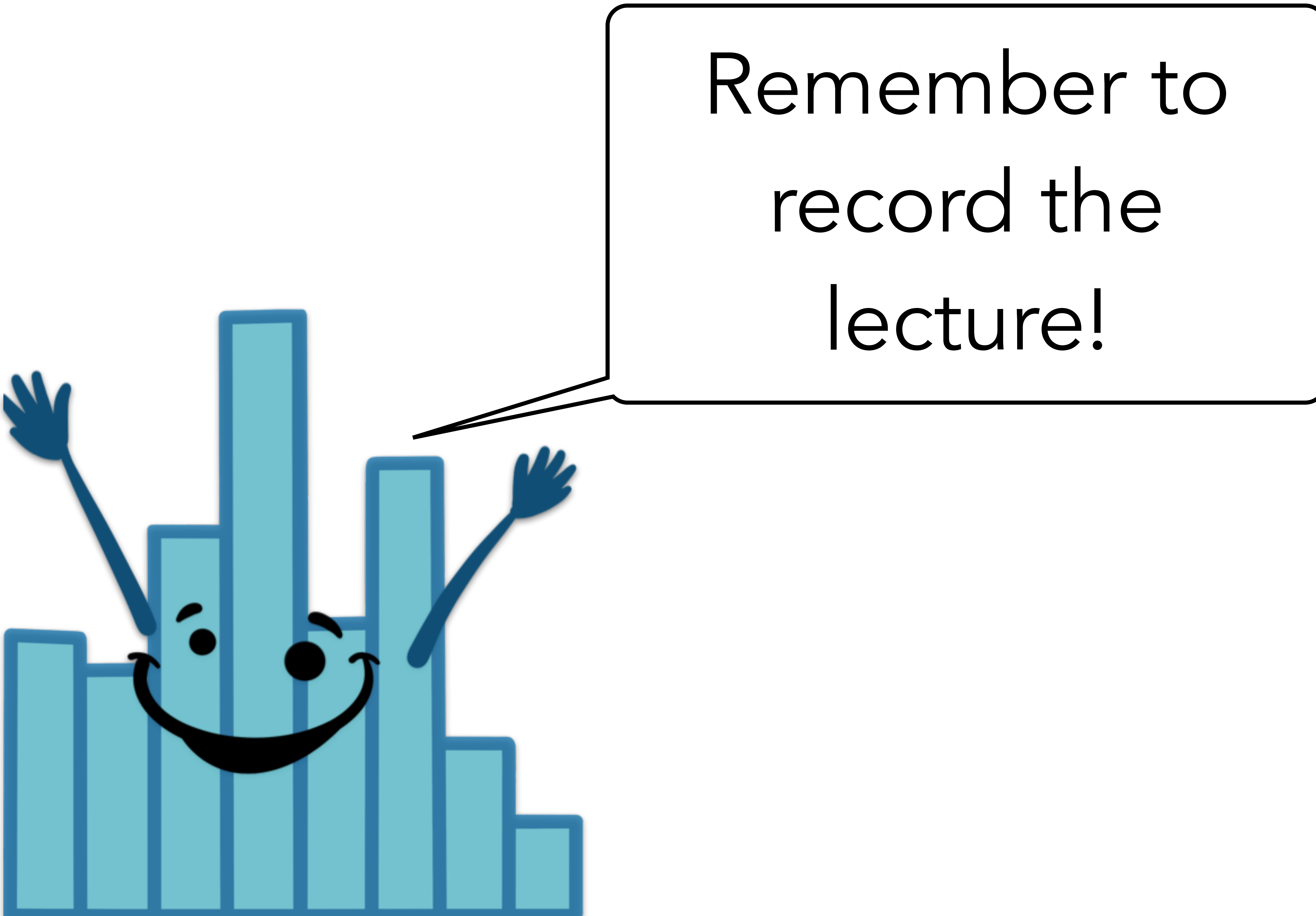
Bayesian data analysis 3





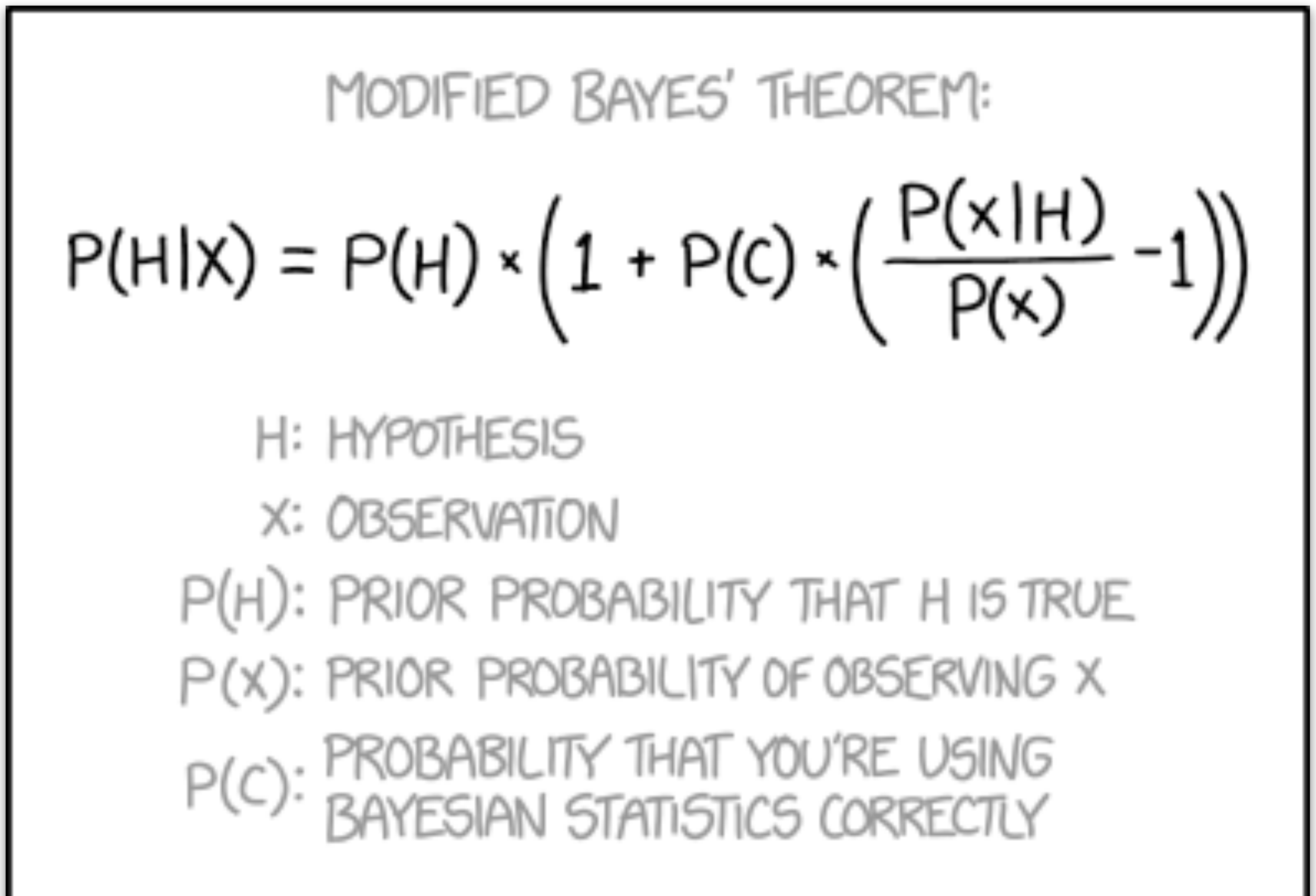


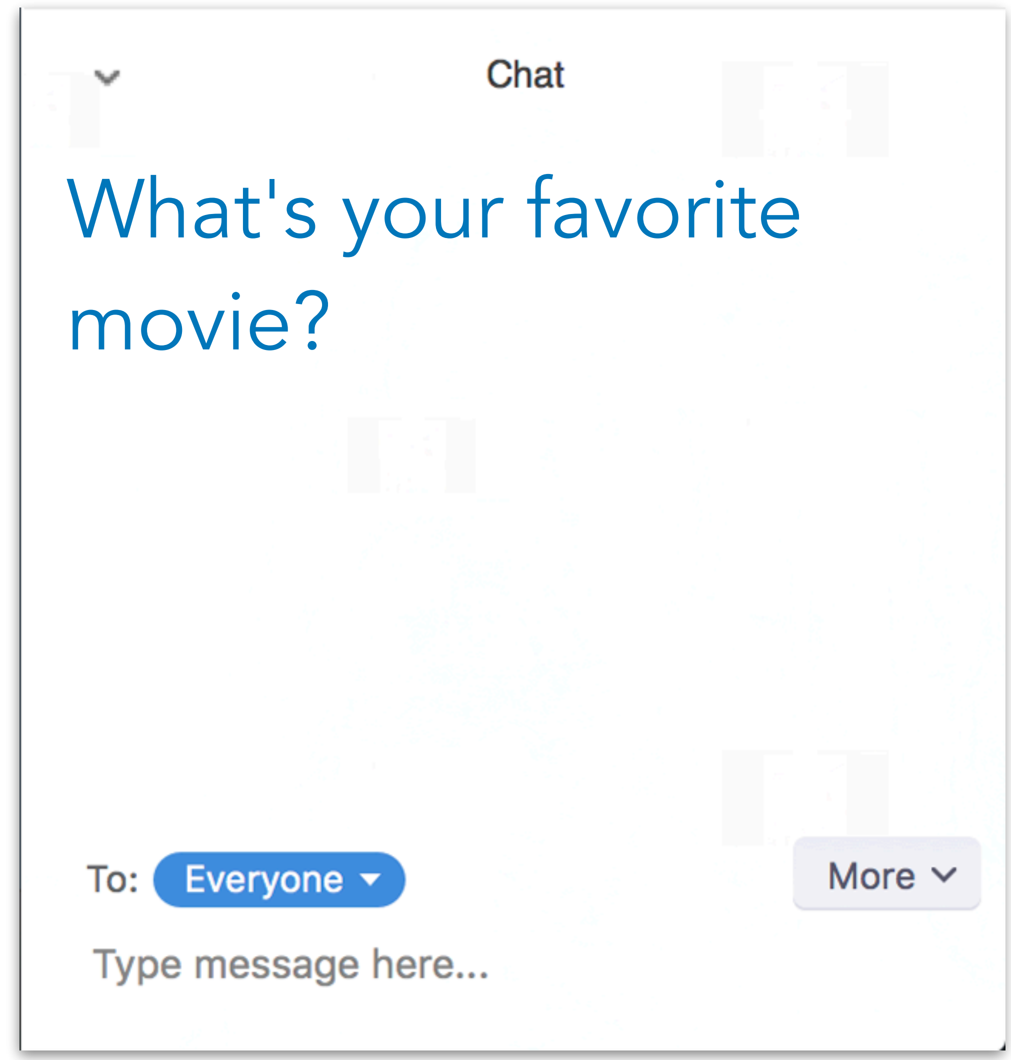
03/10/2021

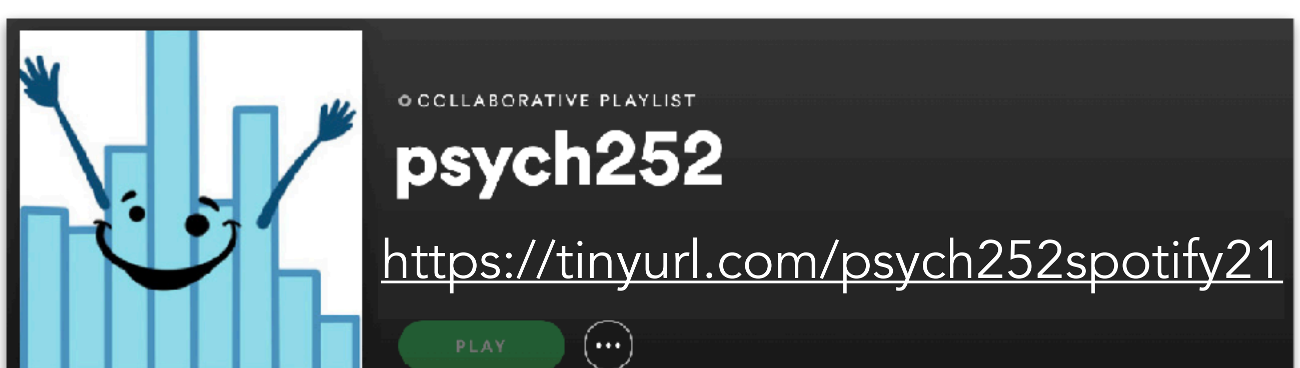

Remember to record the lecture!

PSYCH 252: Statistical Methods for Behavioral and Social Sciences

Bayesian data analysis 3



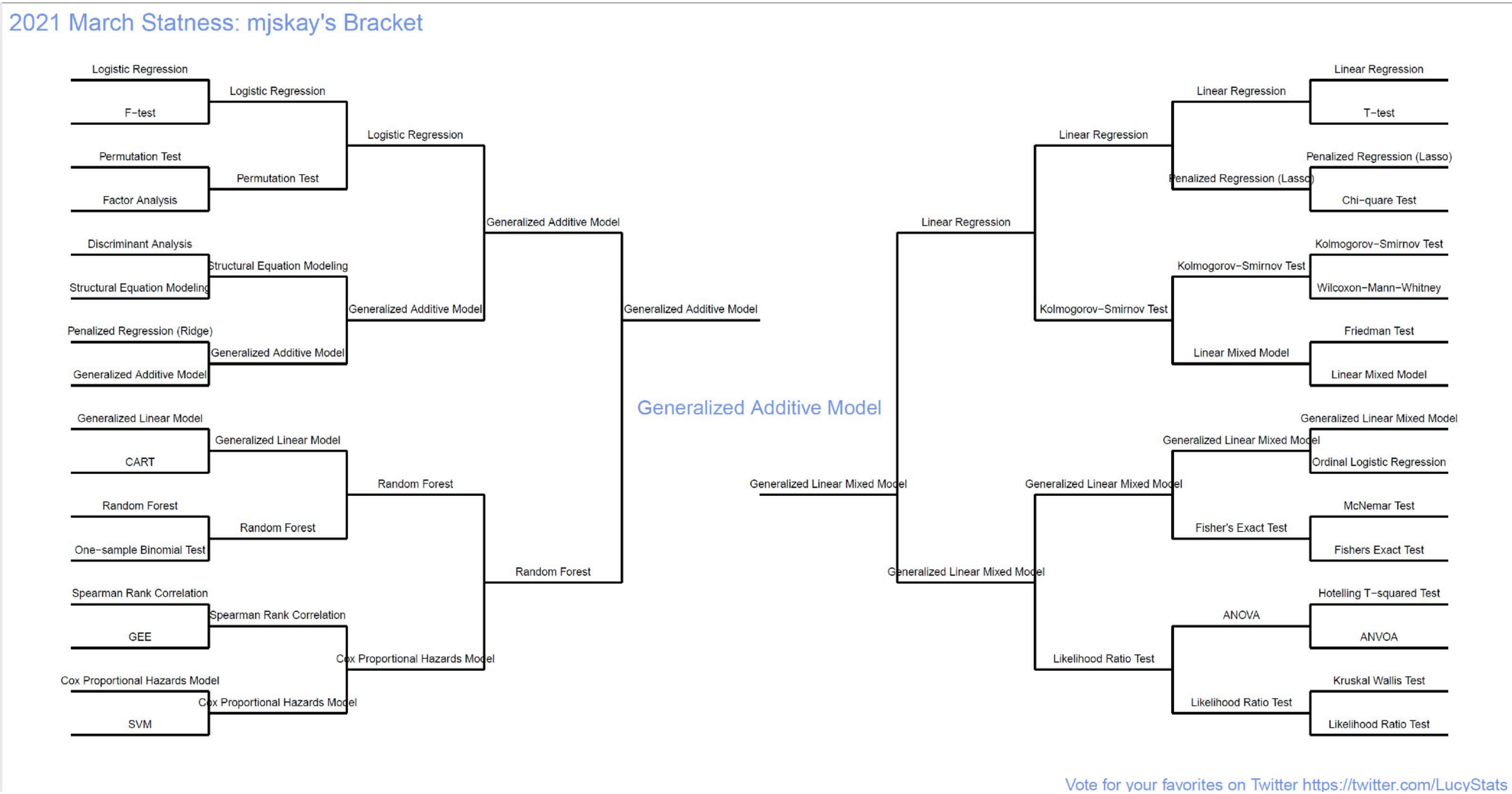




03/10/2021

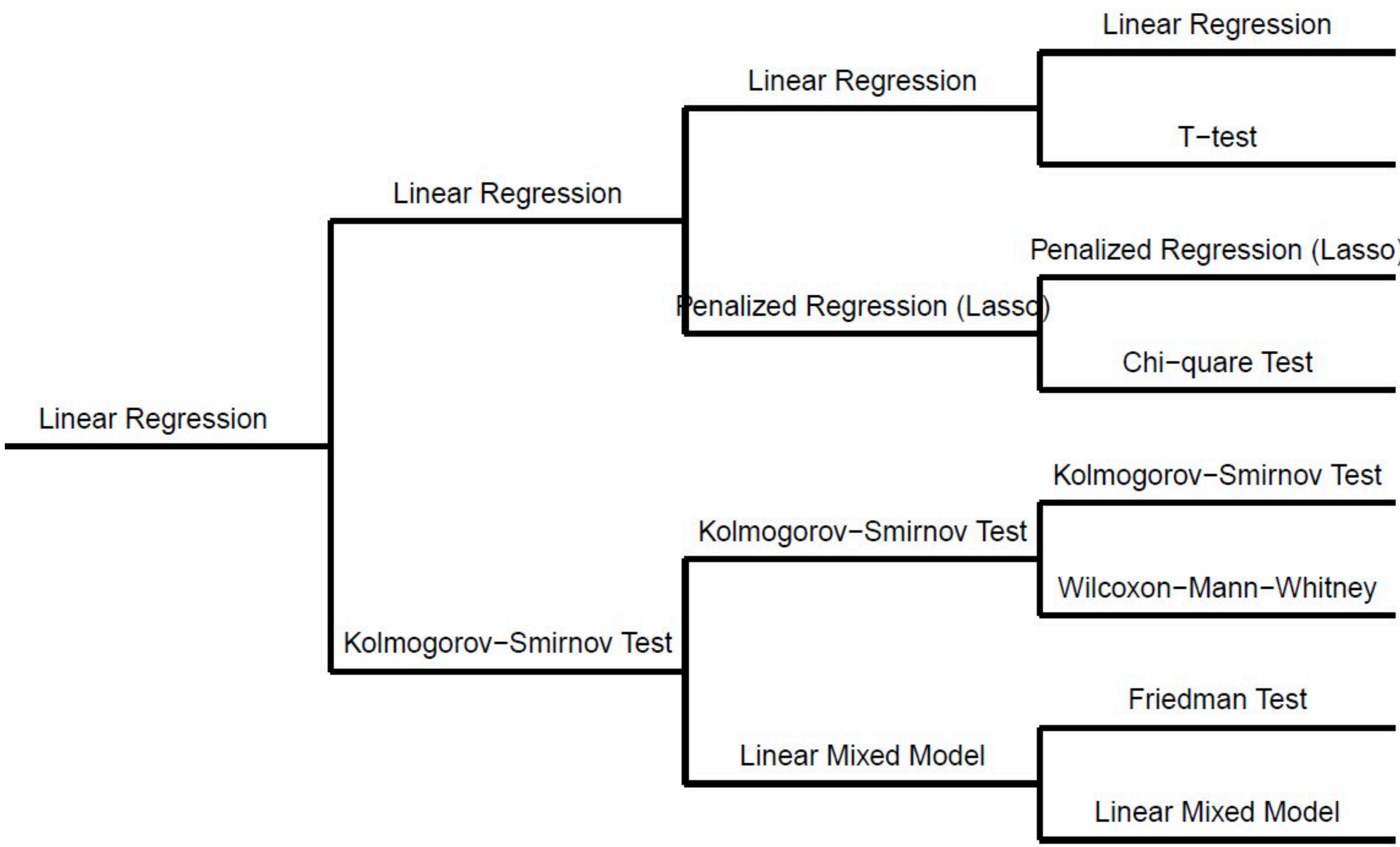
Things that came up

2021 March Statness Bracket



https://lucy.shinyapps.io/march-statness/

2021 March Statness Bracket



https://lucy.shinyapps.io/march-statness/

Adjusting for multiple comparisons Bayesian style

ESSAY | Open Access | Published: 13 May 2019

Frequentist versus Bayesian approaches to multiple testing

Arvid Sjölander & Stijn Vansteelandt

European Journal of Epidemiology 34, 809–821(2019) | Cite this article

https://link.springer.com/article/10.1007/s10654-019-00517-2

Focus Article

Bayesian multiple comparisons and model selection

Andrew A. Neath,[1] Javier E. Flores[2] and Joseph E. Cavanaugh[2*]

https://onlinelibrary.wiley.com/doi/pdf/10.1002/wics.1420

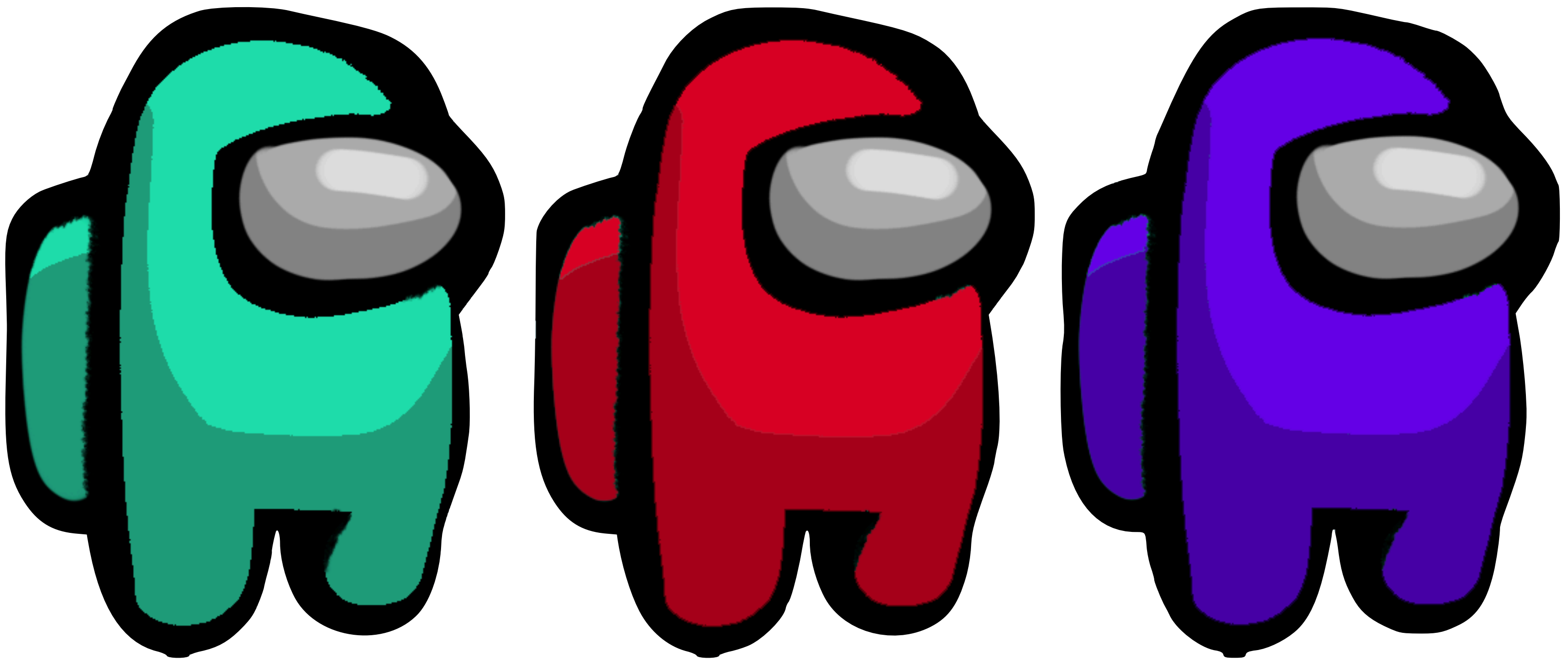
good blog post

http://eointravers.com/post/hypothesis/#bayesian-multiple-comparisons

Logistics

Psych252 Social via Zoom

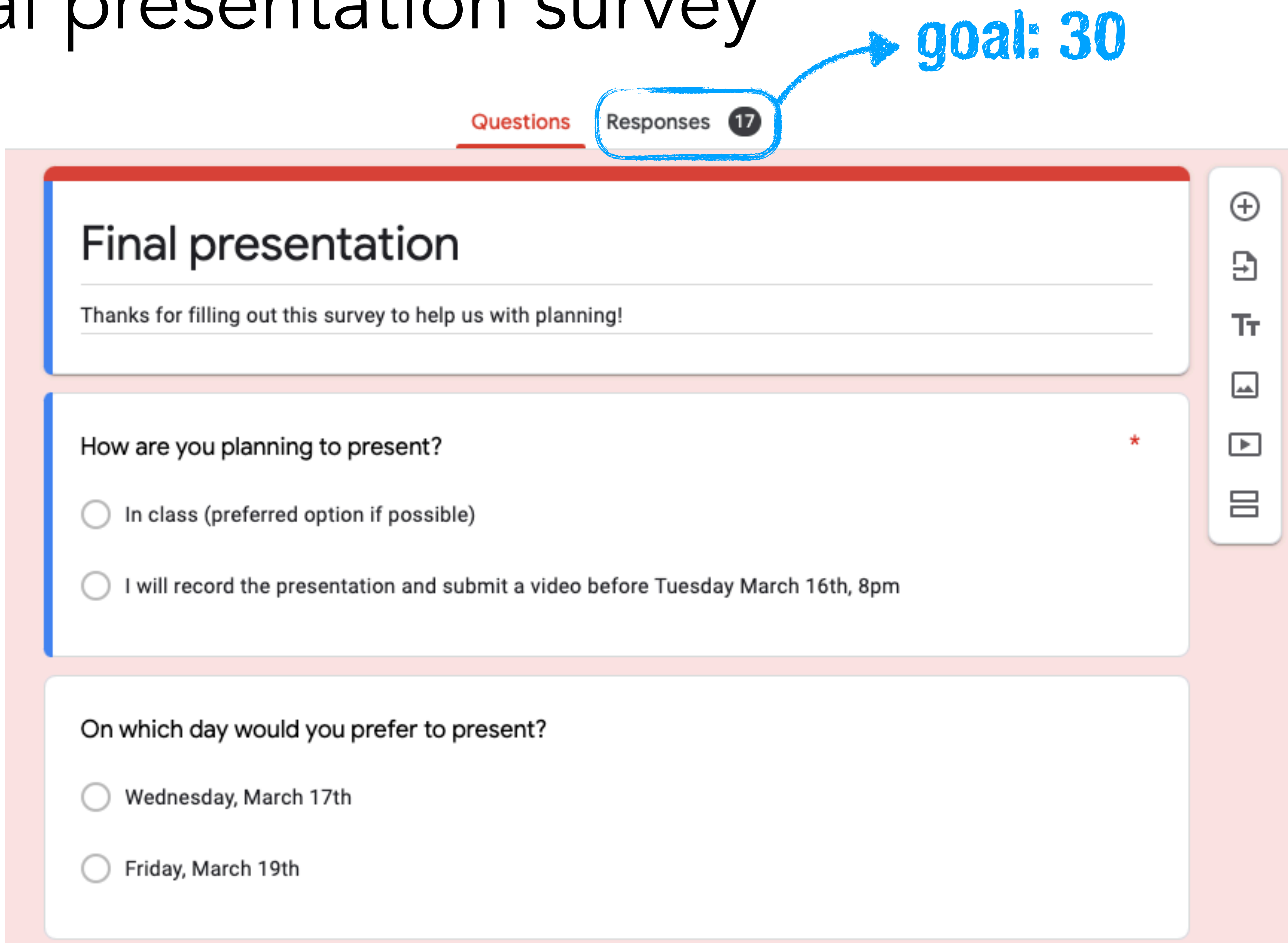
Among us and/or other games



Friday, 12th at 5pm PT

zoom link

Final presentation survey



https://forms.gle/moacbtGBKKexGWUk8

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

For our guest lectures

it would be great if you could turn on your videos,
so they feel welcome :)

Plan for today

- What we've learned
- What shall I do now?
- Quick recap of the Bayesian analysis recipe
- Going beyond
    - 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

data wrangling and visualization

probability, causality, simulation

linear model

power analysis

model comparison

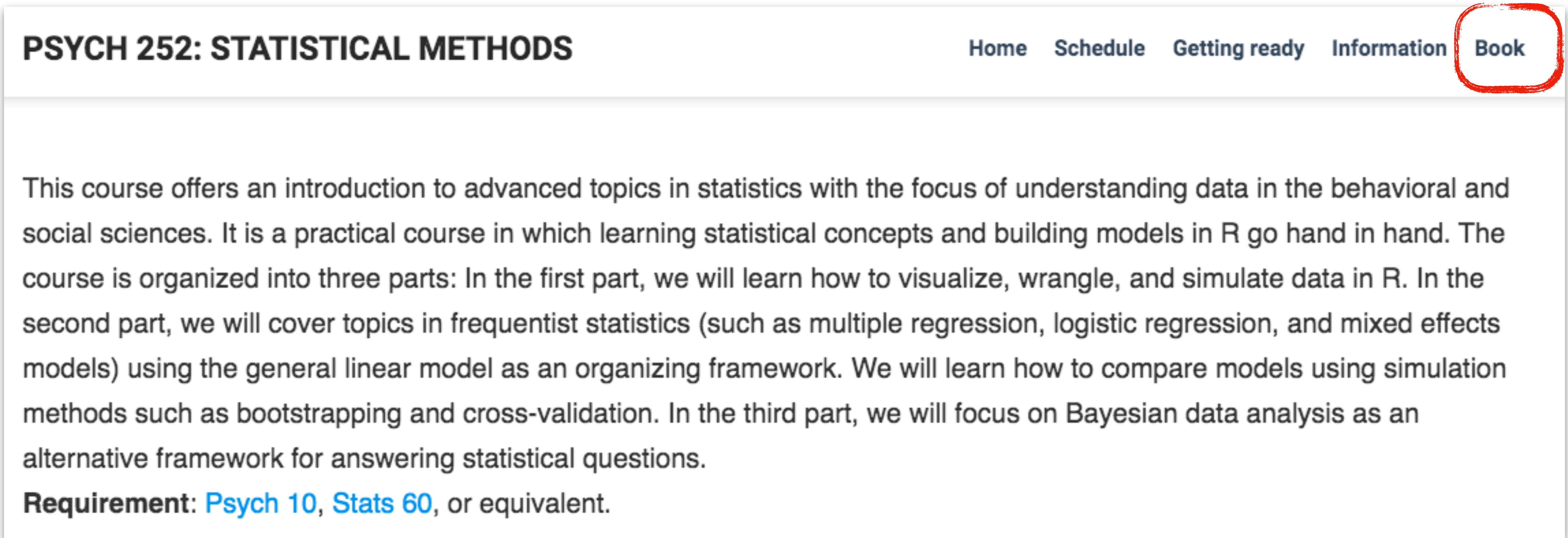
linear mixed effects models

logistic regression

Bayesian data analysis

Date	Topic
January 11th	Introduction
January 13th	Visualization I
January 15th	Visualization II
January 18th	No class (Martin Luther King Jr. Day)
January 20th	Data wrangling I
January 22nd	Data wrangling II
January 25th	Probability
January 27th	Simulation I
January 29th	Simulation II
February 1st	Modeling data
February 3rd	Linear model I
February 5th	Linear model II
February 8th	Linear model III
February 10th	Linear model IV
February 12th	Power analysis
February 15th	No class (Presidents' Day)
February 17th	No class (time to work on midterm)
February 18th	Midterm due
February 19th	Mediation and moderation
February 22nd	Model comparison
February 24th	Linear mixed effects models I
February 25th	Project proposal due
February 26th	Linear mixed effects models II
March 1st	Linear mixed effects models III
March 3rd	Generalized linear model
March 5th	Bayesian data analysis I
March 8th	Bayesian data analysis II
March 10th	Bayesian data analysis III
March 12th	Guest lecture: Matthew Kay
March 15th	Guest lecture: Nilam Ram
March 17th	Final project presentations
March 19th	Final project presentations
	Final project report due

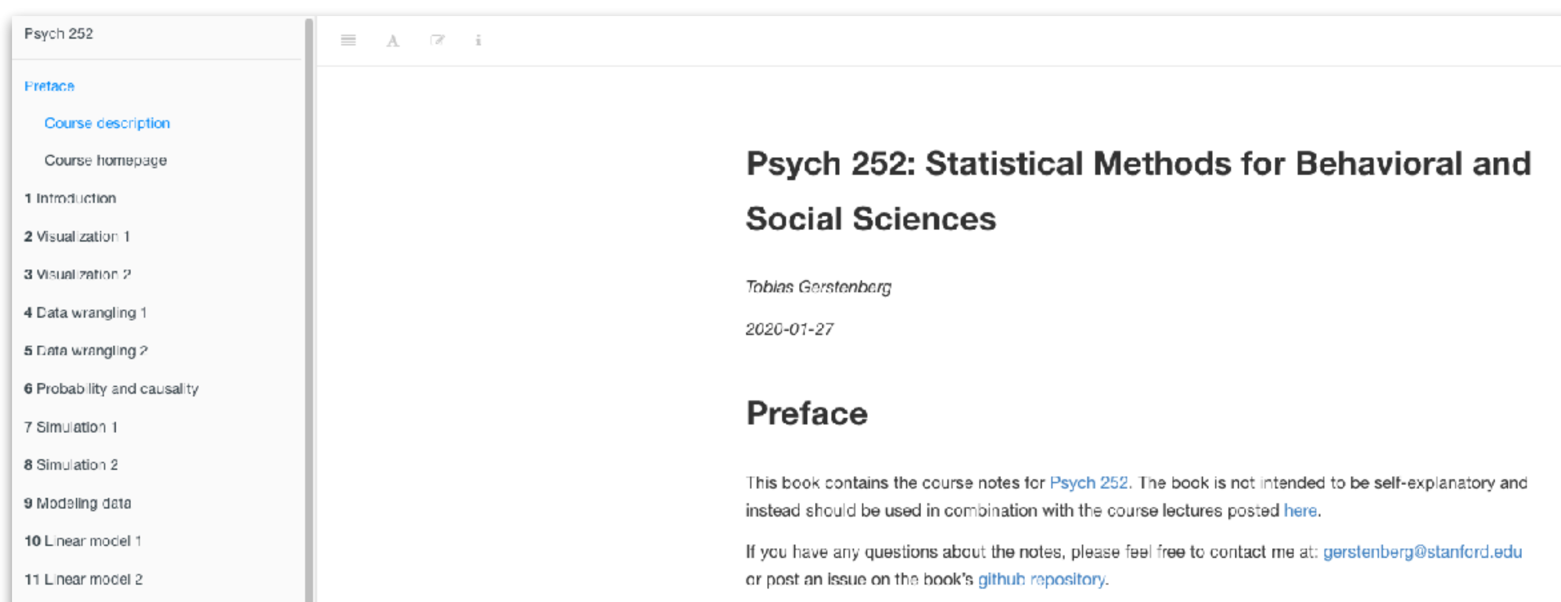
I'll keep updating the course notes!


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/

I'll make sure that you'll have access to the video recordings

and 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

What shall I do now?

Psych 253

Basics / Course Schedule & Syllabus / Software Tools

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 262

we'll get a teaser on Monday, yay!

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: 1-5 | Repeatable 5 times (up to 25 units total)

Instructors: Ram, N. (PI)

Schedule for PSYCH 262

2020-2021 Spring

PSYCH 262 | 1-5 units | Class # 31926 | Section 01 | Grading: Letter or Credit/No Credit Exception | SEM | Remote: Synchronous | Students enrolled: 4

03/29/2021 - 06/04/2021 Mon, Wed 8:30 AM - 9:50 AM at Remote with Ram, N. (PI)

Instructors: Ram, N. (PI)

Notes: Enrollment is by instructor consent. Please contact faculty for permission.

PSYCH 204

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

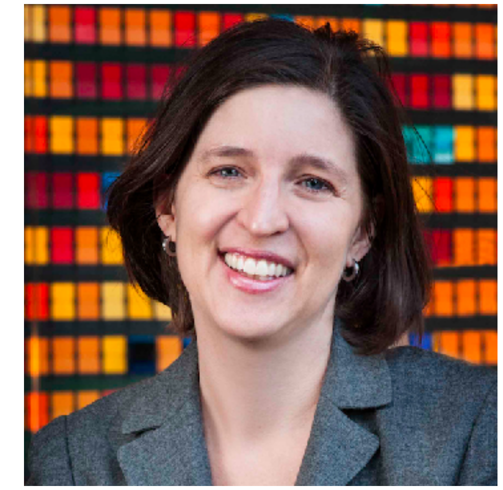
PSYCH 204 | 3 units | Class # 31428 | Section 01 | Grading: Letter or Credit/No Credit Exception | LEC | Remote: Synchronous | Students enrolled: 14

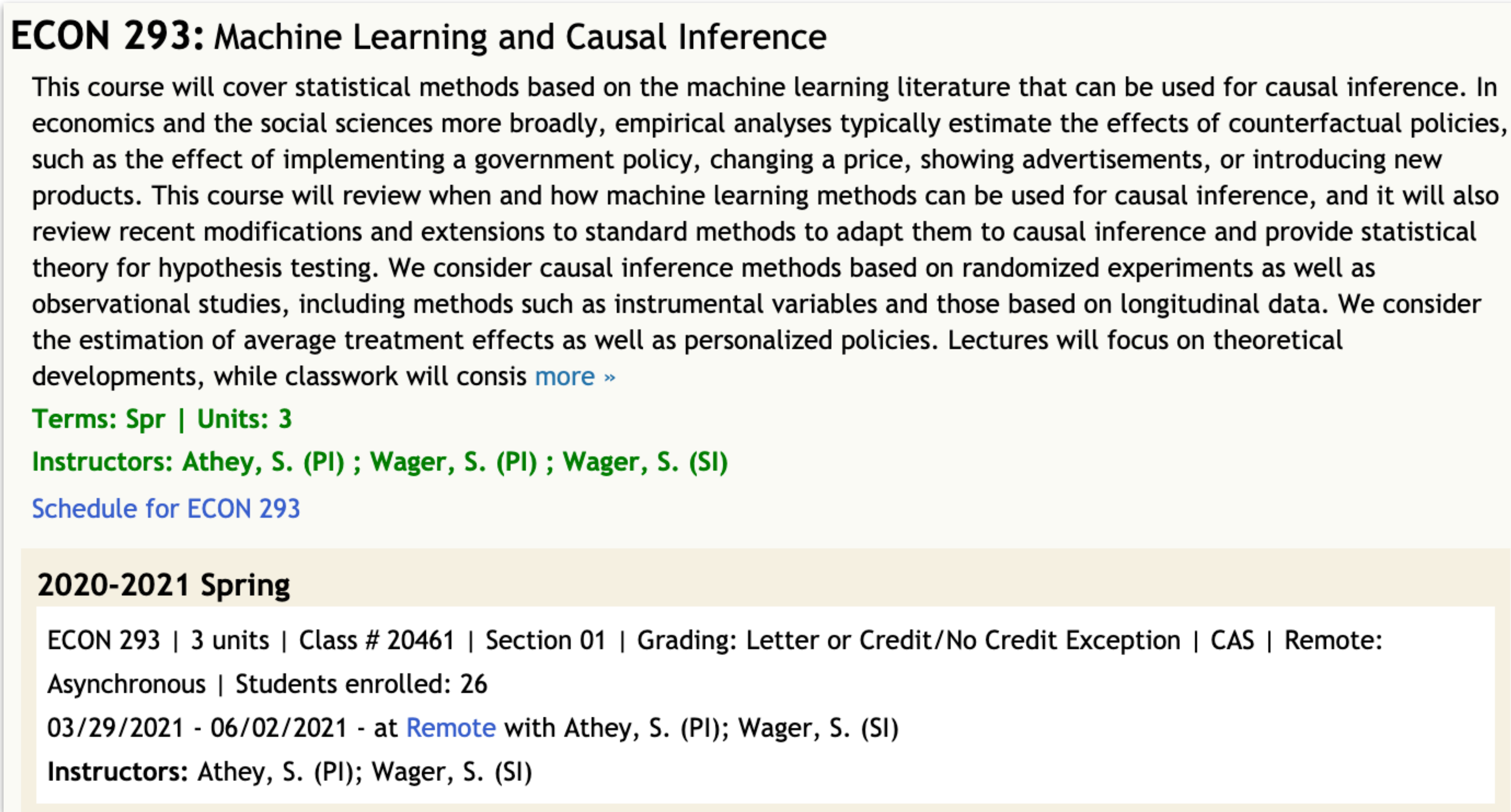
03/29/2021 - 06/04/2021 Tue, Thu 12:30 PM - 1:50 PM at Remote with Goodman, N. (PI); Nam, A. (TA)

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

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)

Advanced regression analysis

Sanne Smith

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

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)

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

Data Challenge Lab

Where students develop their data skills by solving a progression of increasingly difficult challenges

ENGR 150: Data Challenge Lab
Terms: Win, Spr | Units: 5
Instructors: Bill Behrman, Hadley Wickham

Prof. Tidyverse

https://datalab.stanford.edu/challenge-lab

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 hasnt 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 for any stats questions I have in the future

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

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

Back to Bayes 1

Recipe for Bayesian analysis with `brms`

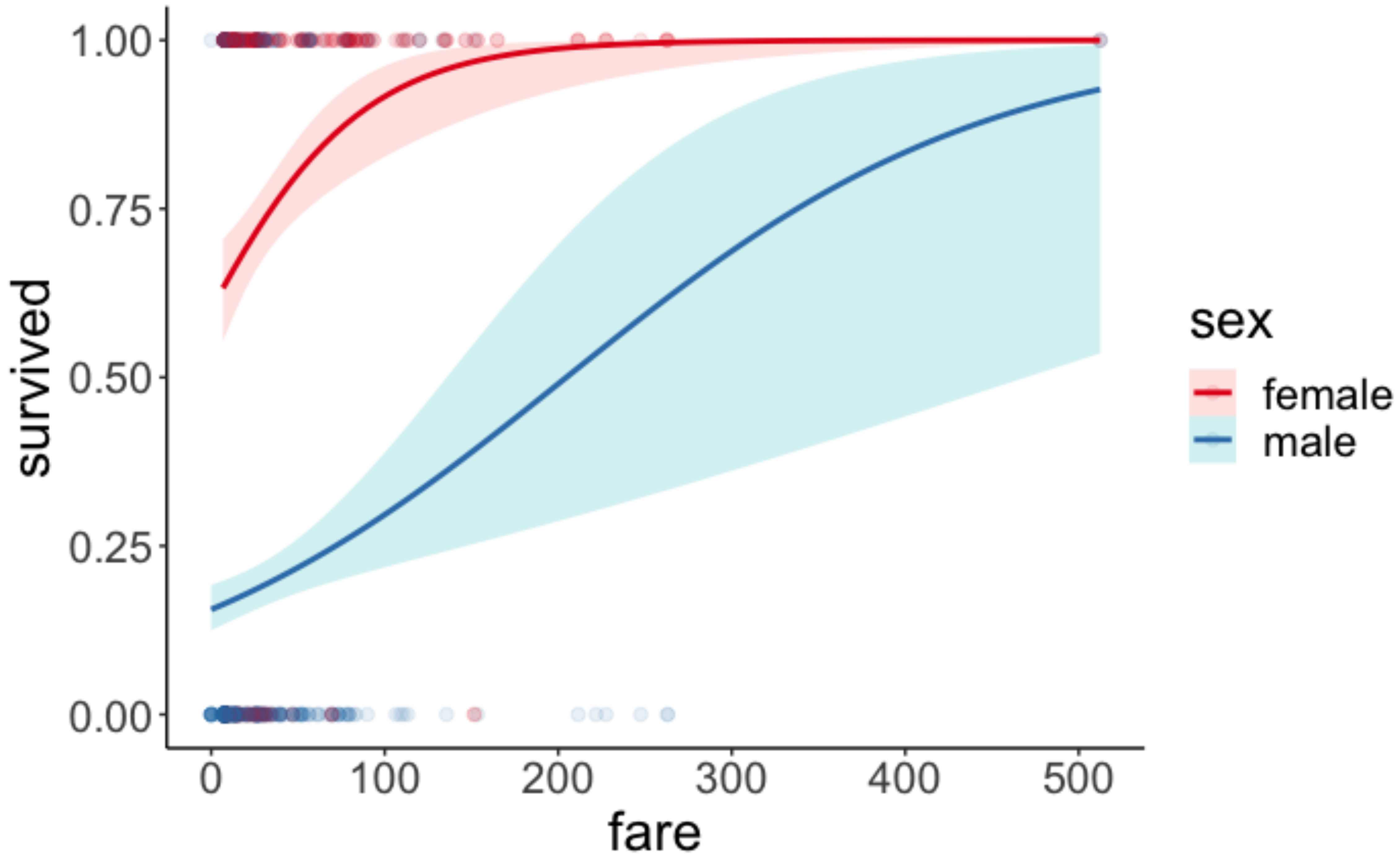
visualization is everywhere!

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

Titanic data

1. Visualize the data

Feeling cold?



2. Specify and fit the model

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

3. Model evaluation

a) Did the inference work?

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

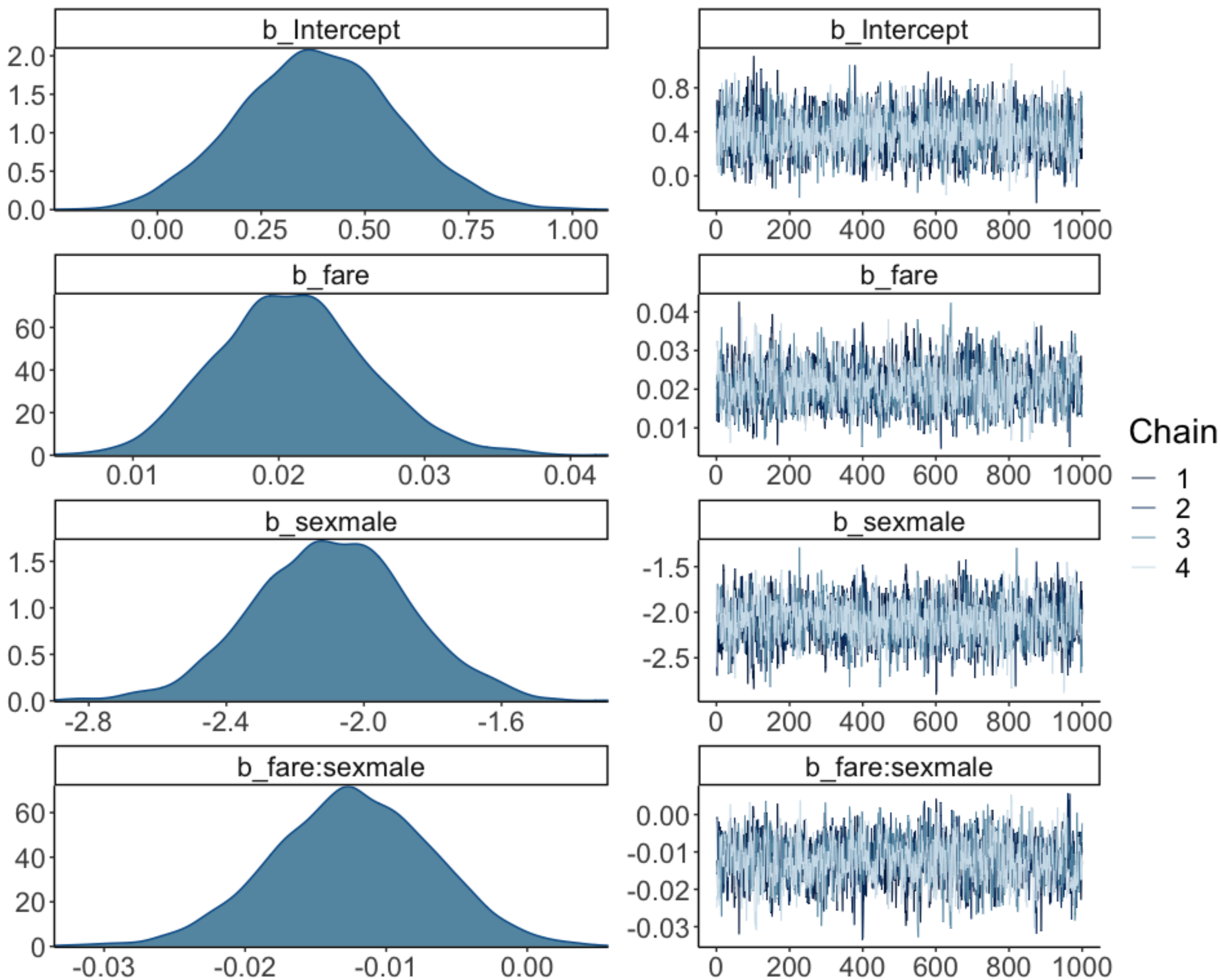
```
 Family: bernoulli
  Links: mu = logit
Formula: survived ~ 1 + fare * sex
   Data: df.titanic (Number of observations: 891)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept        0.39      0.19     0.03     0.76 1.00     2010     2625
fare             0.02      0.01     0.01     0.03 1.00     1545     2124
sexmale         -2.09      0.23    -2.54    -1.65 1.00     1754     1984
fare:sexmale    -0.01      0.01    -0.02    -0.00 1.00     1479     2041

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).
```

looks good

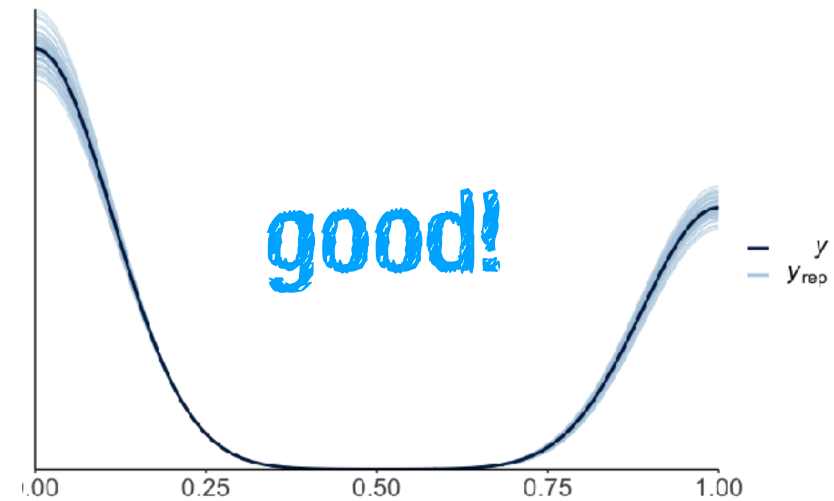
https://mc-stan.org/misc/warnings.html

a) Did the inference work?

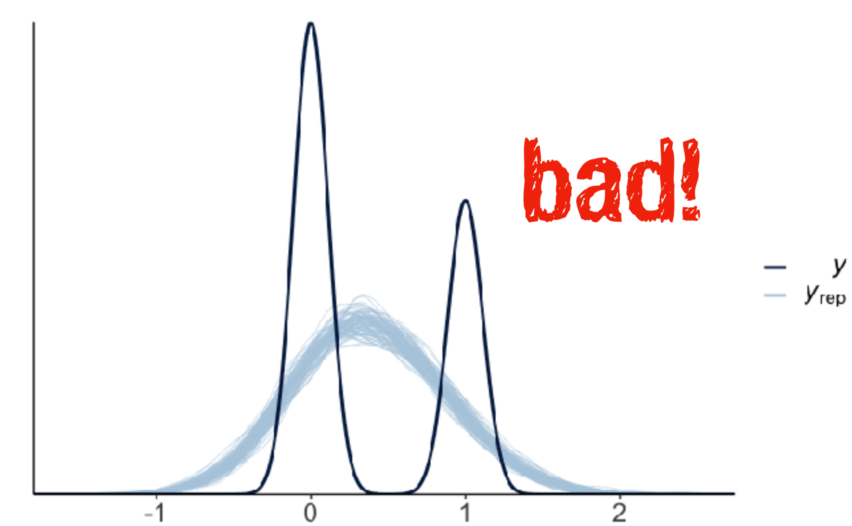
```
fit.brm_titanic %>%
  plot()
```



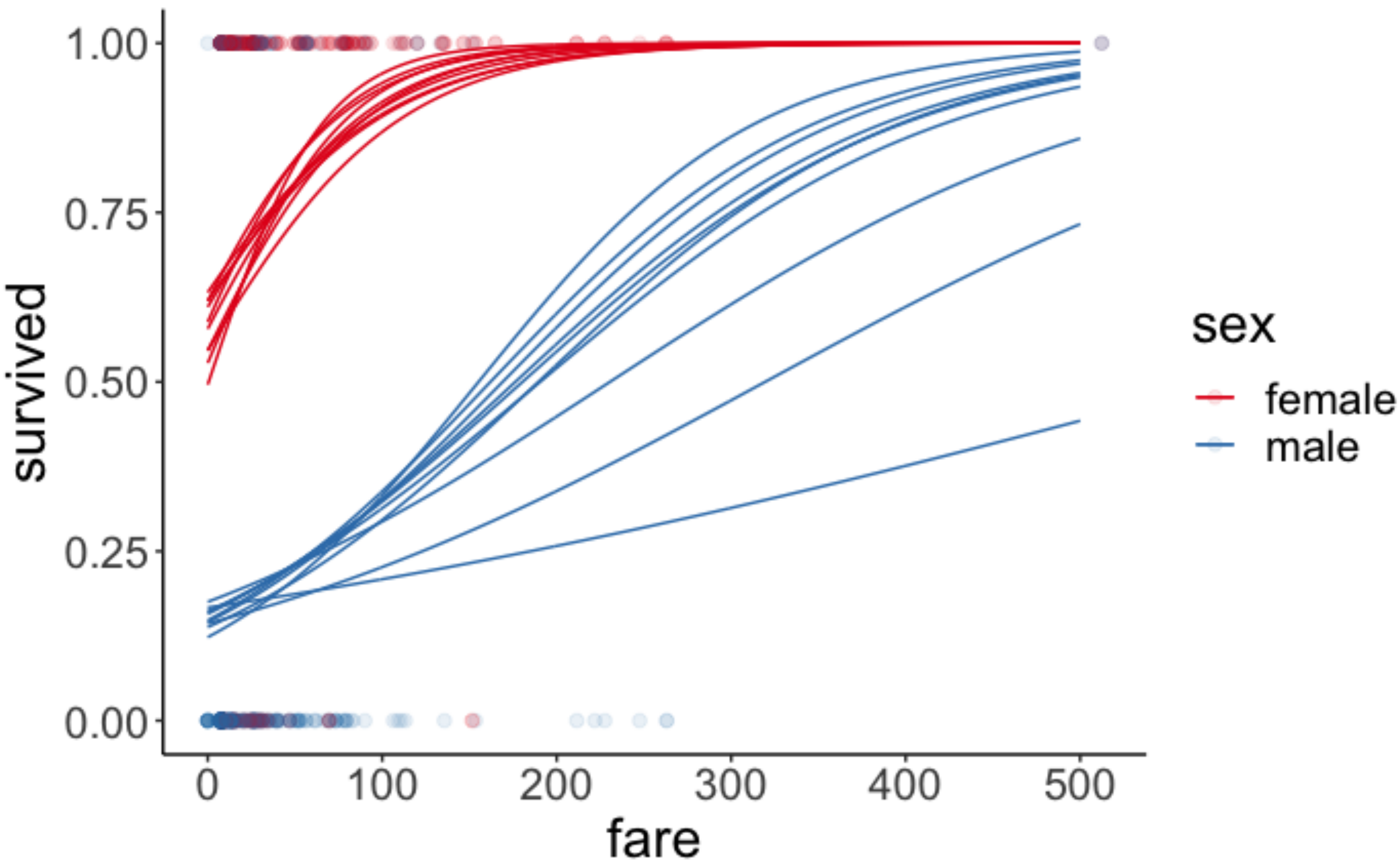

```
pp_check(fit.brm_titanic,
         nsamples = 100)
```



model with Gaussian family



b) Visualize the model predictions



4. Interpret the model parameters

4. Interpret the model parameters

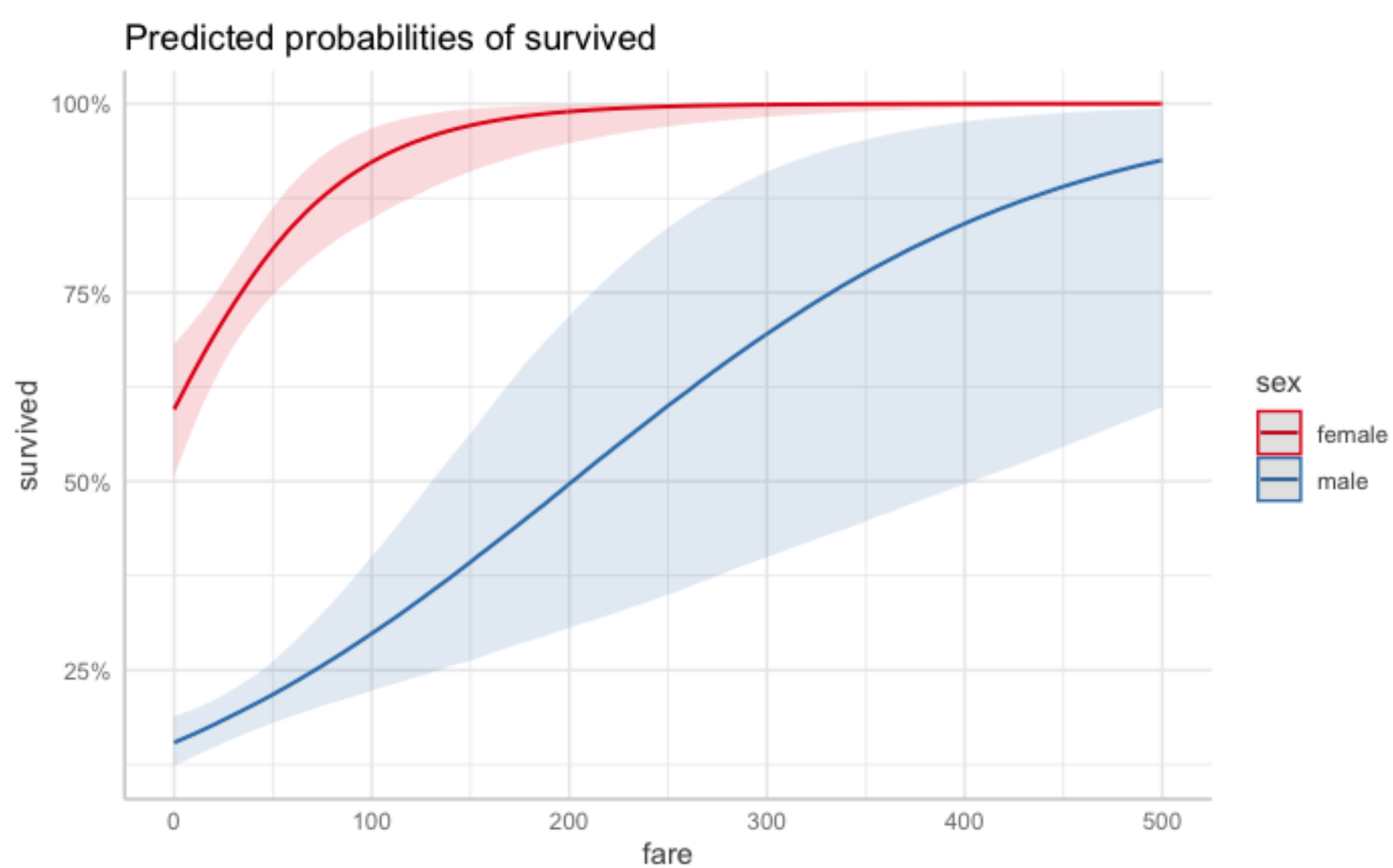
```
 Family: bernoulli
  Links: mu = logit
Formula: survived ~ 1 + fare * sex
   Data: df.titanic (Number of observations: 891)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

Population-Level Effects:
             Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept        0.39      0.19     0.03     0.76 1.00     2010     2625
fare             0.02      0.01     0.01     0.03 1.00     1545     2124
sexmale         -2.09      0.23    -2.54    -1.65 1.00     1754     1984
fare:sexmale    -0.01      0.01    -0.02    -0.00 1.00     1479     2041

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).
```

log odds

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]
```



5. Test specific hypotheses

5. Test specific hypotheses

Were women more likely to survive than men?

```
fit.brm_titanic %>%
  emmeans(specs = pairwise ~ sex,
          type = "response")
```

$$\frac{\left(\frac{p_f}{1-p_f}\right)}{\left(\frac{p_m}{1-p_m}\right)}$$

```
NOTE: Results may be misleading due to involvement in interactions
$emmeans
 sex    response lower.HPD upper.HPD
 female    0.743      0.69     0.795
 male      0.194      0.16     0.225

Point estimate displayed: median
Results are back-transformed from the logit scale
HPD interval probability: 0.95

$contrasts
 contrast      odds.ratio lower.HPD upper.HPD
 female / male       12.1      8.39      16.6

Point estimate displayed: median
Results are back-transformed from the log odds ratio scale
HPD interval probability: 0.95
```

5. Test specific hypotheses

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

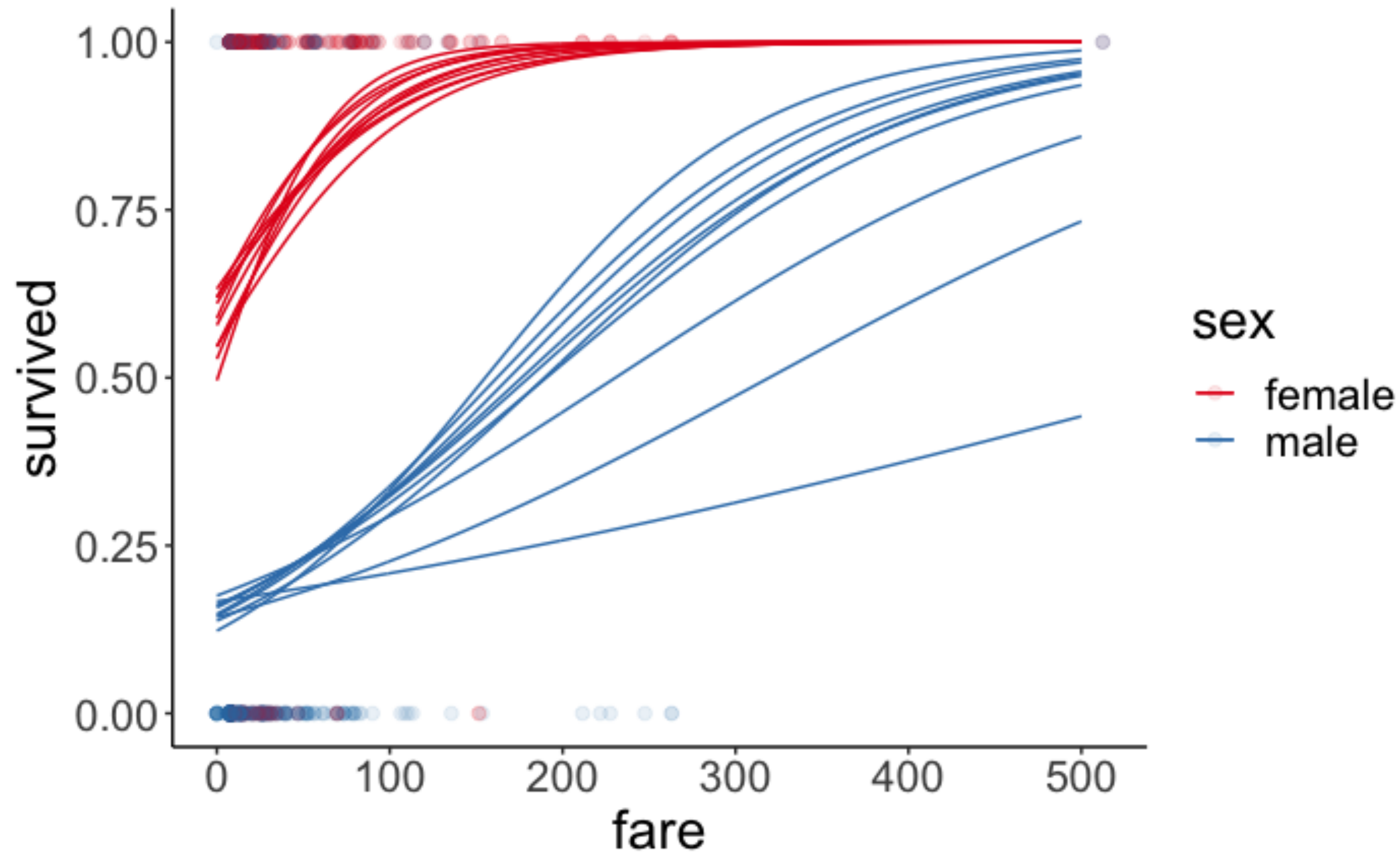
```
fit.brm_titanic %>%
  emtrends(specs = pairwise ~ sex,
           var = "fare")
```

```
$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

6. Report results

6. Report results



Female passengers were more likely to survive (74.3%) than male passengers (19.4%). The estimated odds ratio of survival for female vs. male passengers was 12.1 [8.39, 16.6].

The chance of survival increased more with fare for female compared to male passengers. The difference in slopes on the log odds scale was 0.0124 [0.000884, 0.0232].

Going beyond

Plan for today

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

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)

Bayes factor

check this out

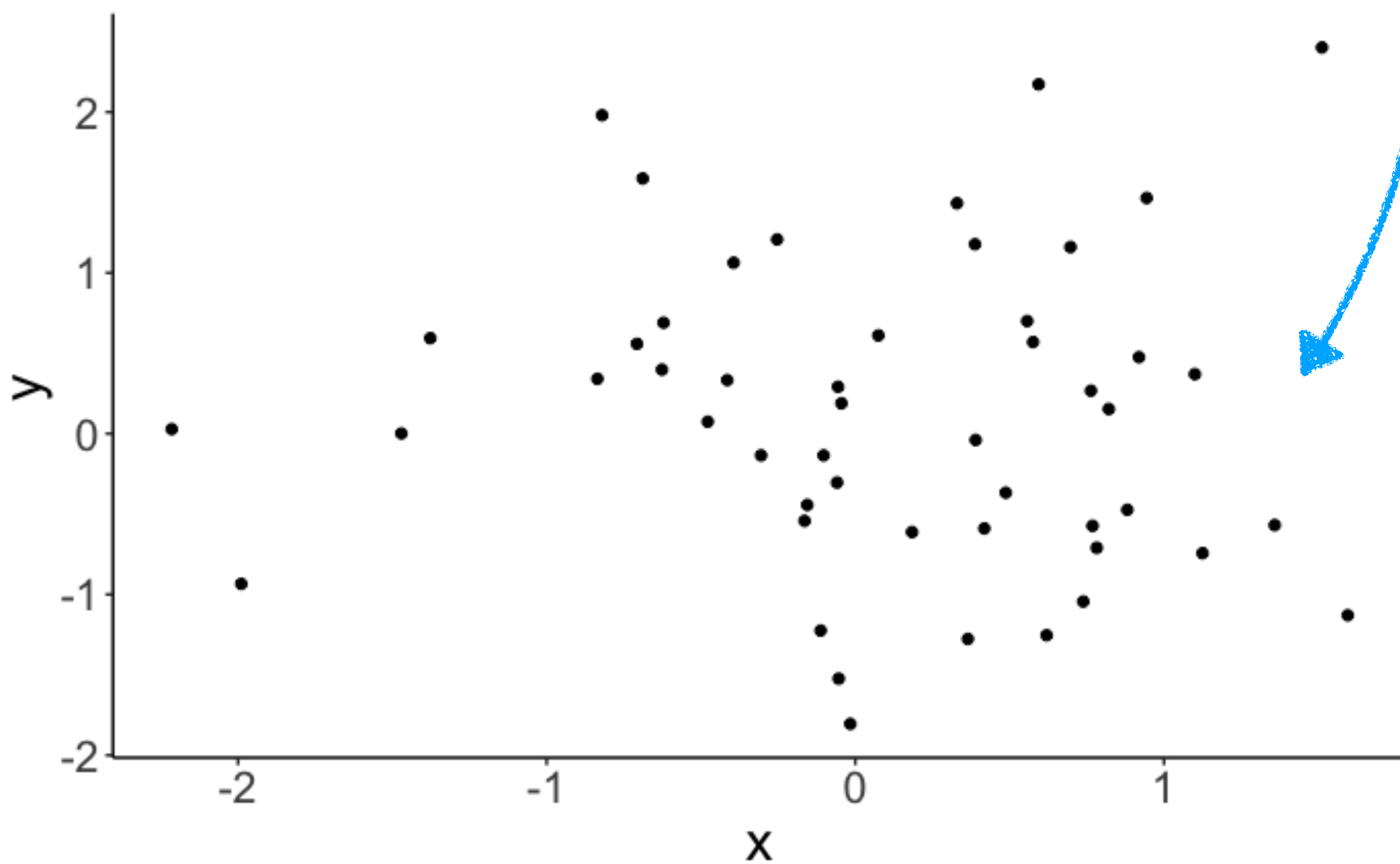
https://vuorre.netlify.com/post/2017/03/21/bayes-factors-with-brms/

Approximate LOO

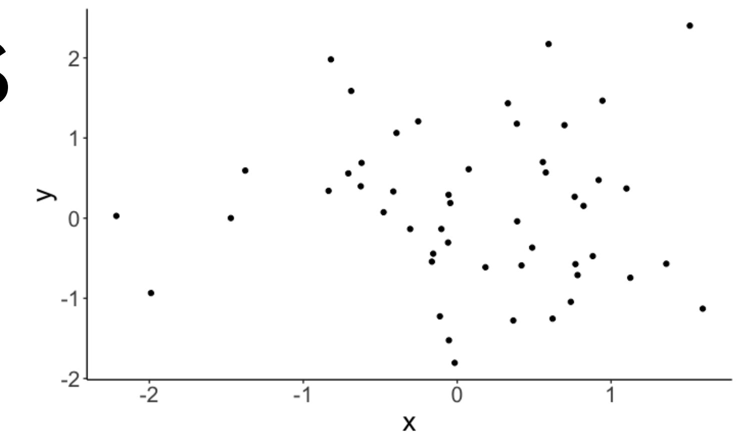
Evidence for the null hypothesis

```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

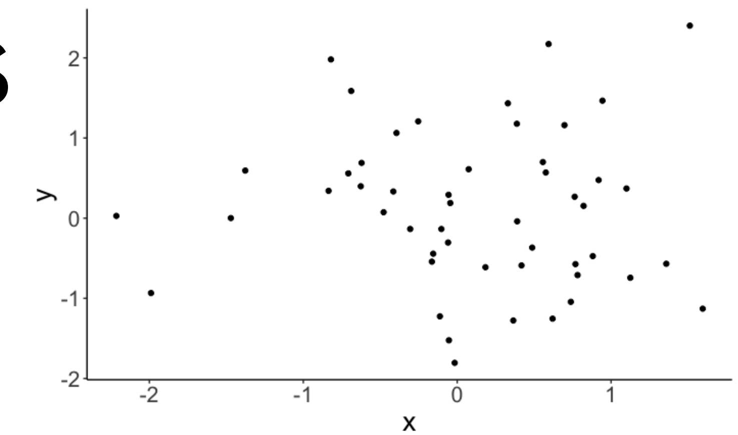



```
fit.lm_loo = lm(formula = y ~ 1 + x,
                data = df.loo)

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




```
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

```
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
```

```
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 + andrew + catherine + jon + dan,
    data = df.stats_love)
```

coefficients in the model

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

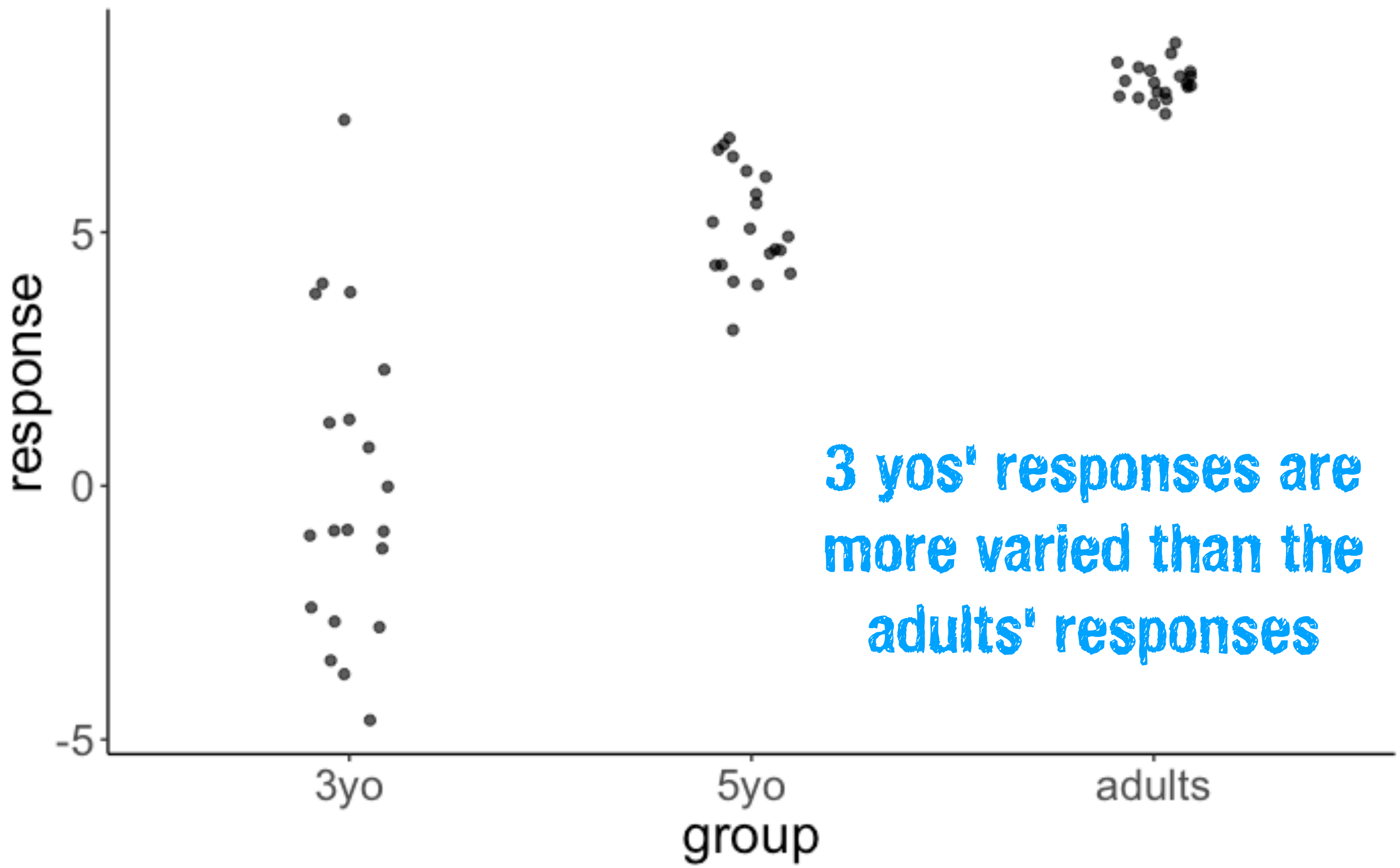
lower bound = 0

```
brm(formula = how_much_i_love_stats ~ 1 + tobi + andrew + catherine + jon + dan,
    prior = priors,
    data = df.stats_love)
```

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)))
```



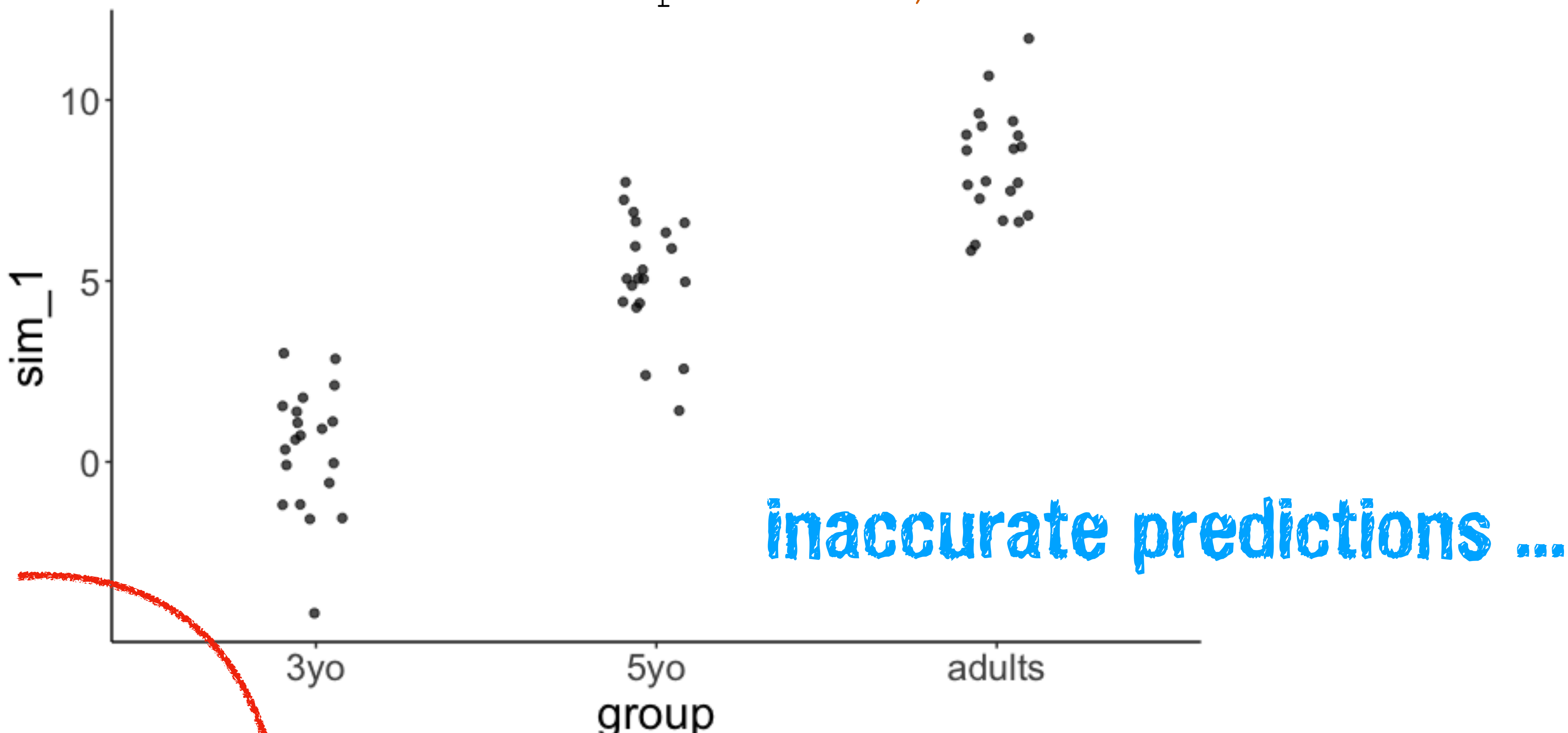
Unequal variance aka heteroscedasticity

```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

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

```
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!

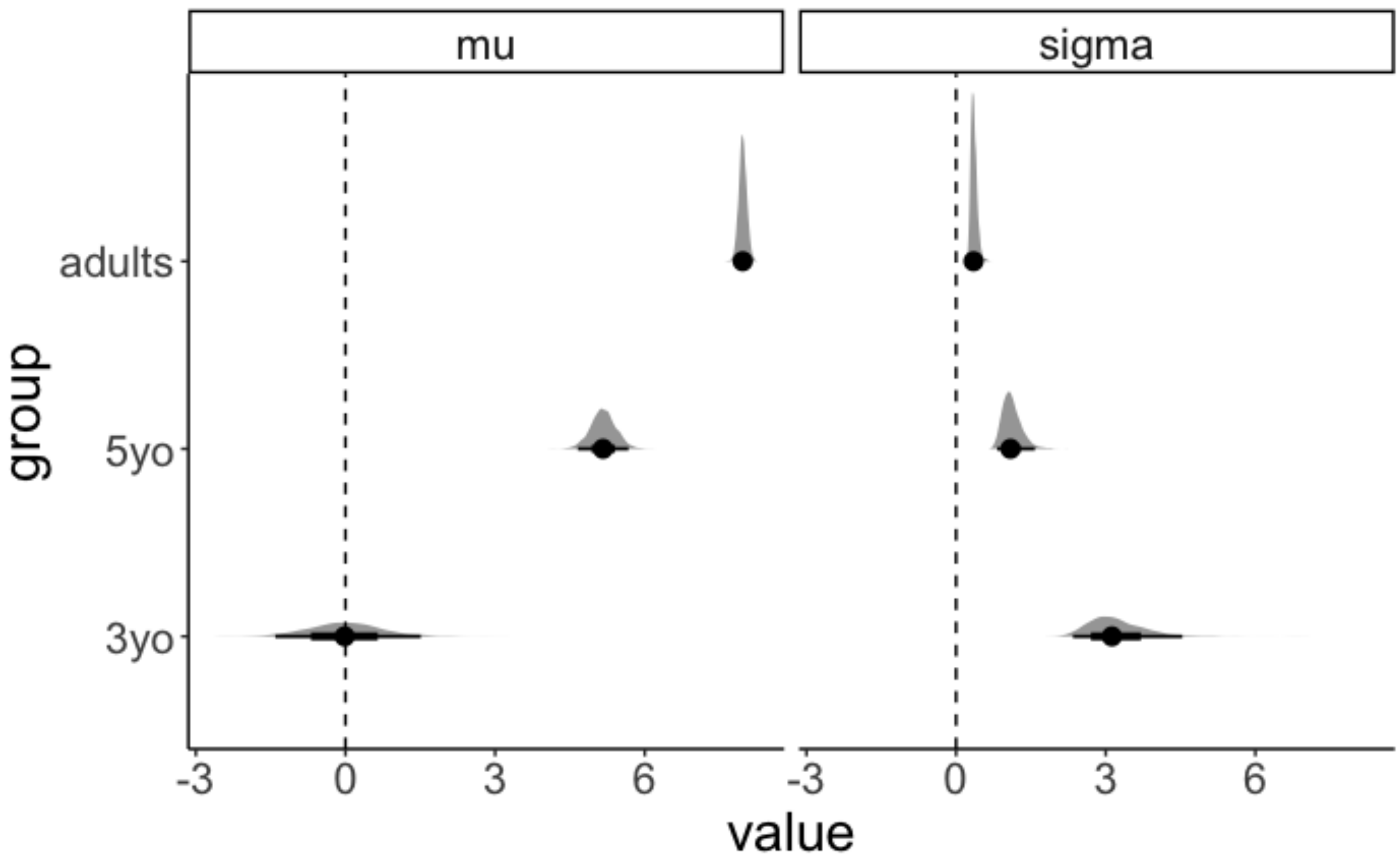
3 year olds $e^{1.15} = 3.16$

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

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

Unequal variance aka heteroscedasticity

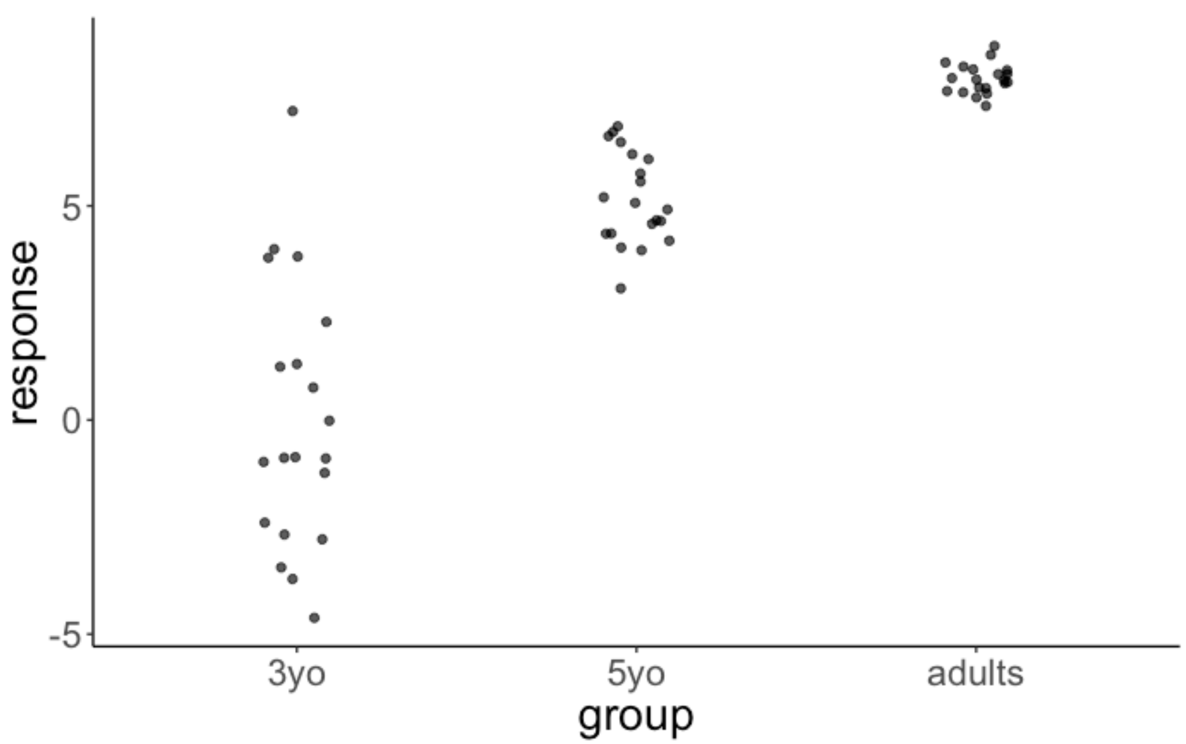
```
df.variance %>%
  expand(group) %>%
  add_fitted_draws(fit.brm1, dpar = TRUE) %>%
  select(group, .row, .draw, posterior = .value, mu, sigma) %>%
  pivot_longer(cols = c(mu, sigma),
               names_to = "index",
               values_to = "value") %>%
  ggplot(aes(x = value, y = group)) +
 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



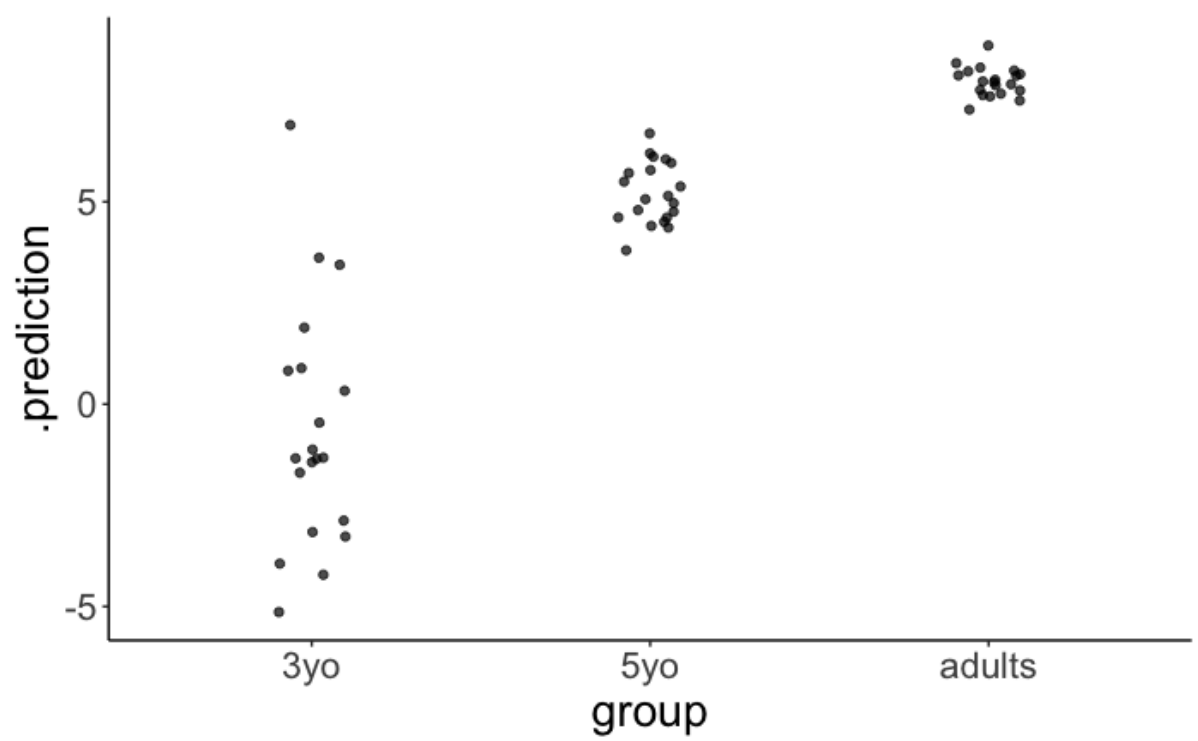
Unequal variance aka heteroscedasticity

```
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



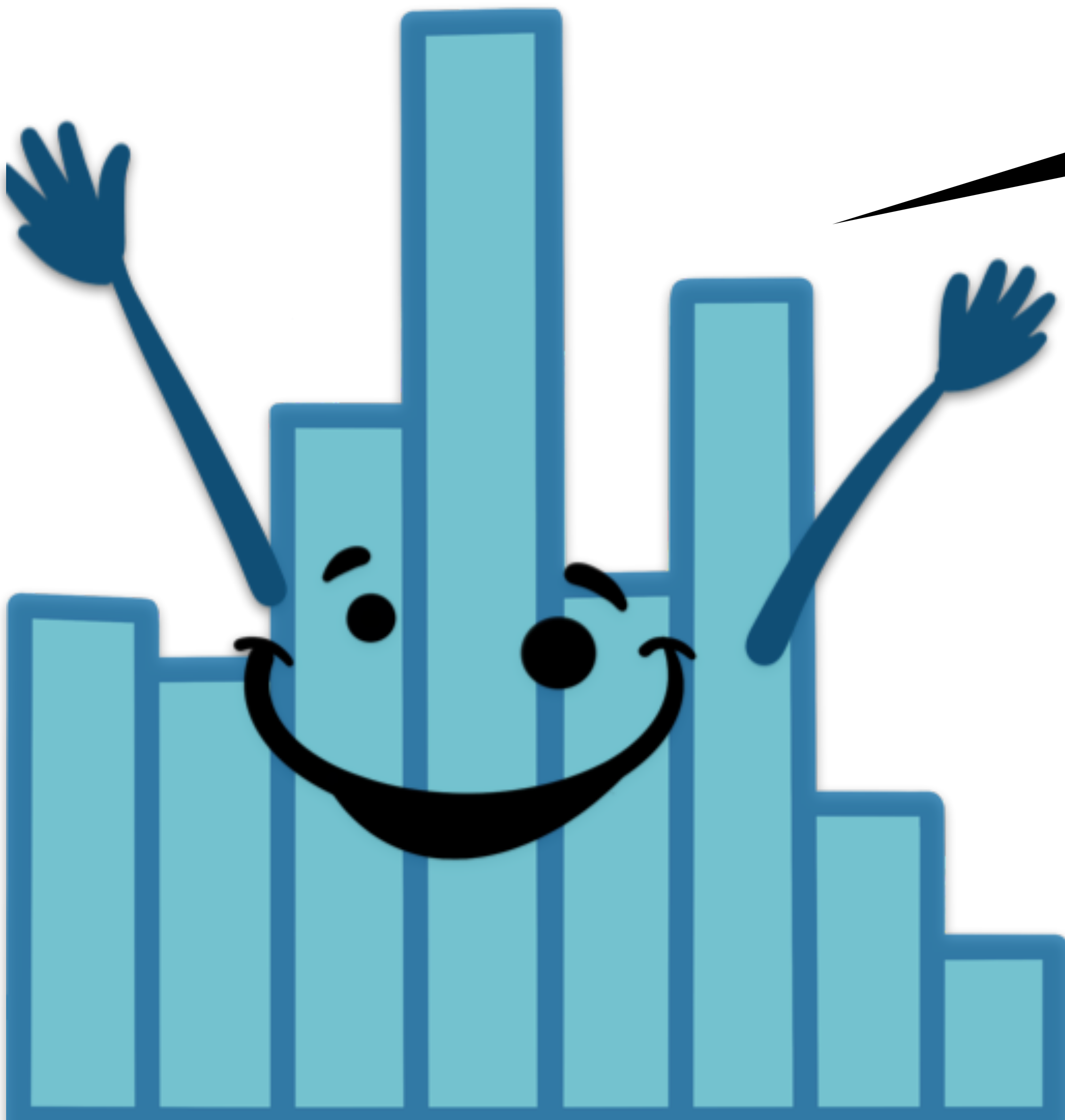
predicted data

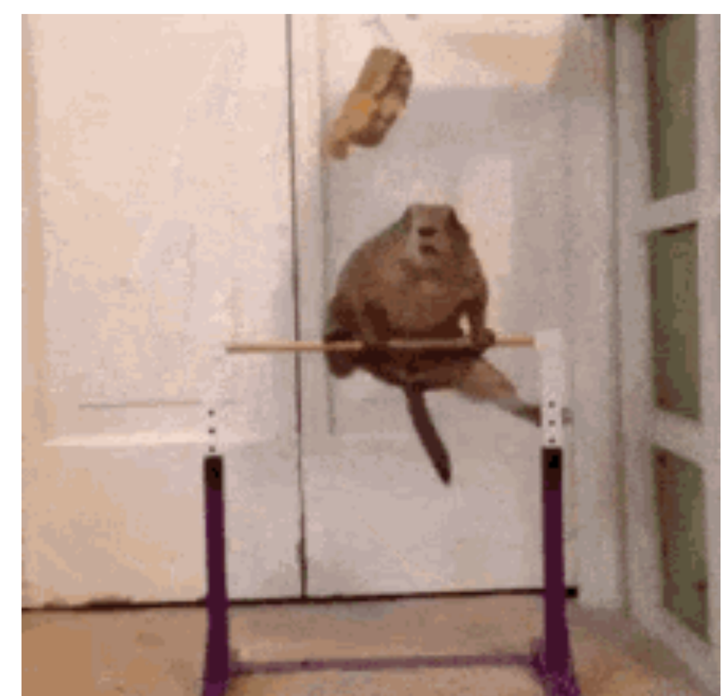


these predictions look good!

01:00

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

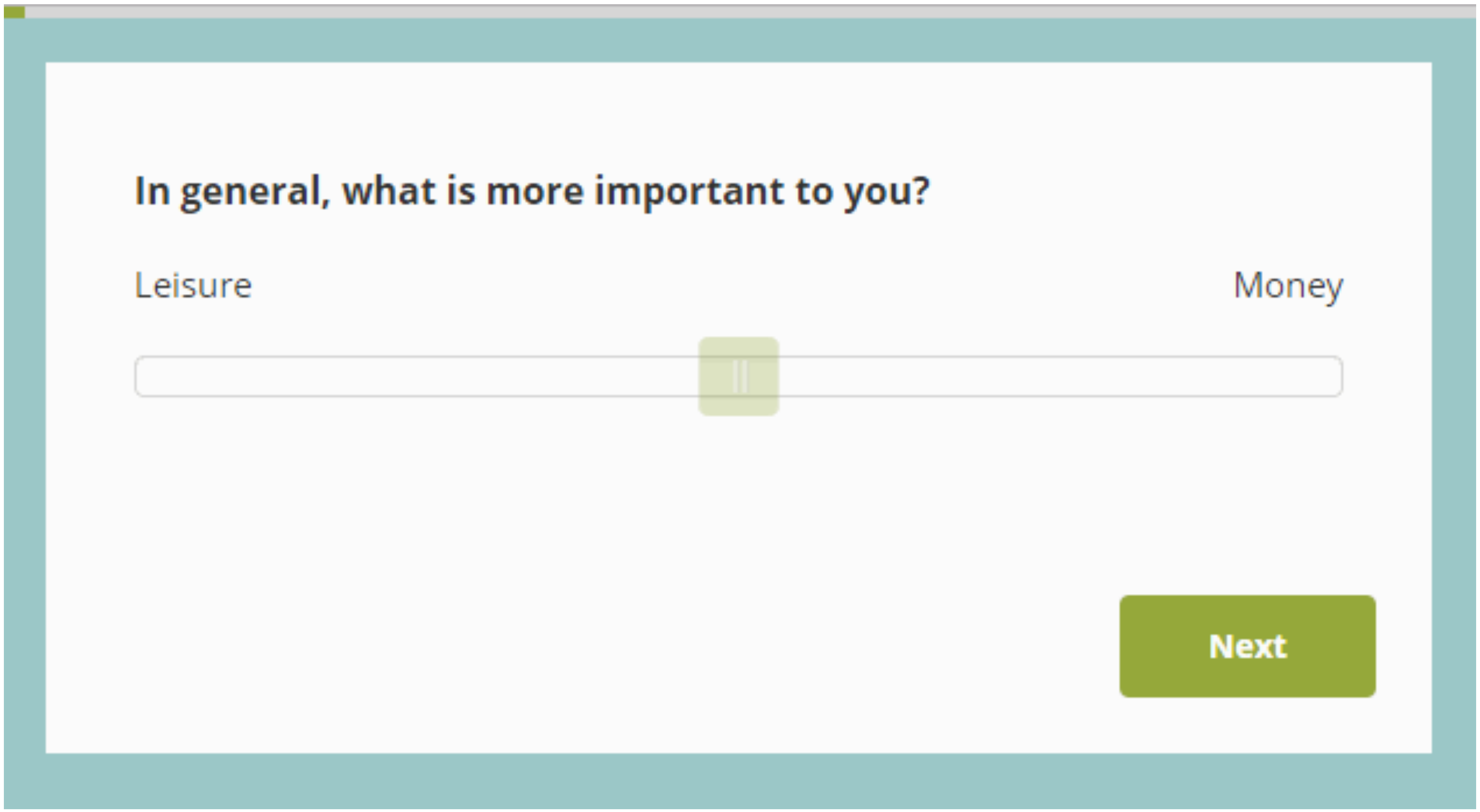
- Introduction
- The zero-one-inflated beta model
- ZOIB regression
- Simulation: Compare ZOIB and t-test performances
- Discussion
- References

Introduction

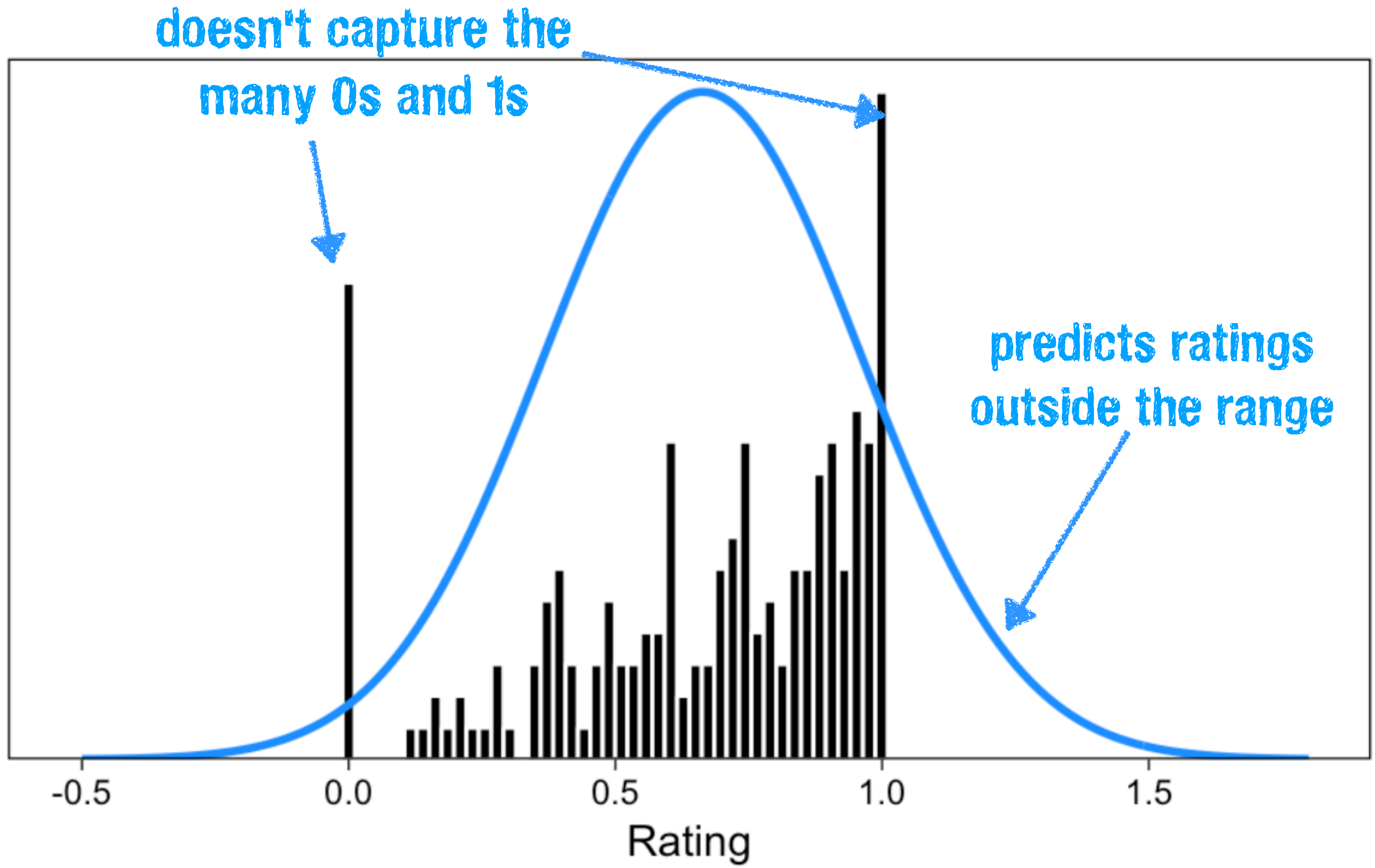
In psychological experiments, subjective responses are often collected using two types of response scales: ordinal and visual analog scales. These scales are unlikely to provide normally distributed data. However, researchers often analyze responses from these scales with models that assume normality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



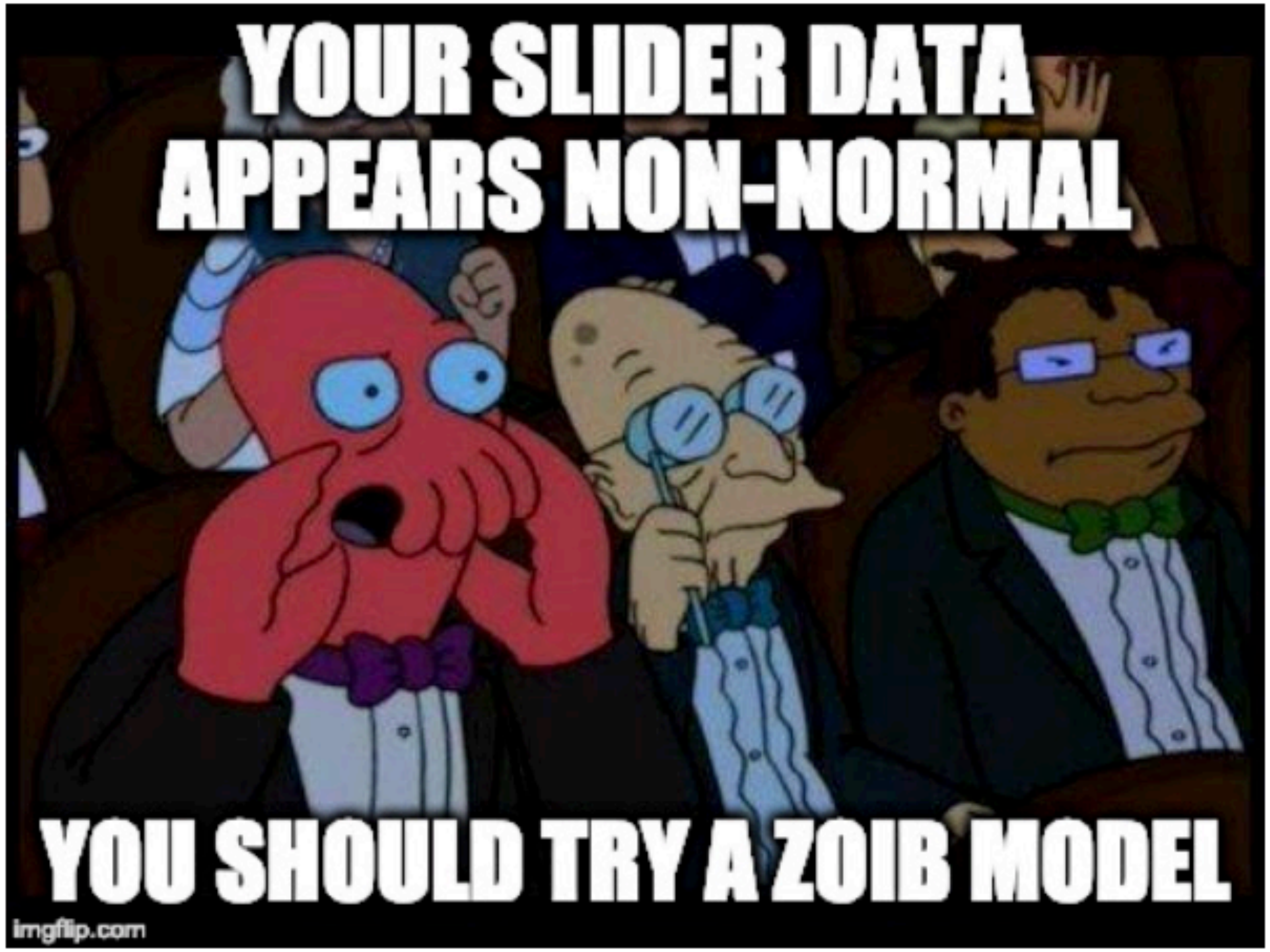
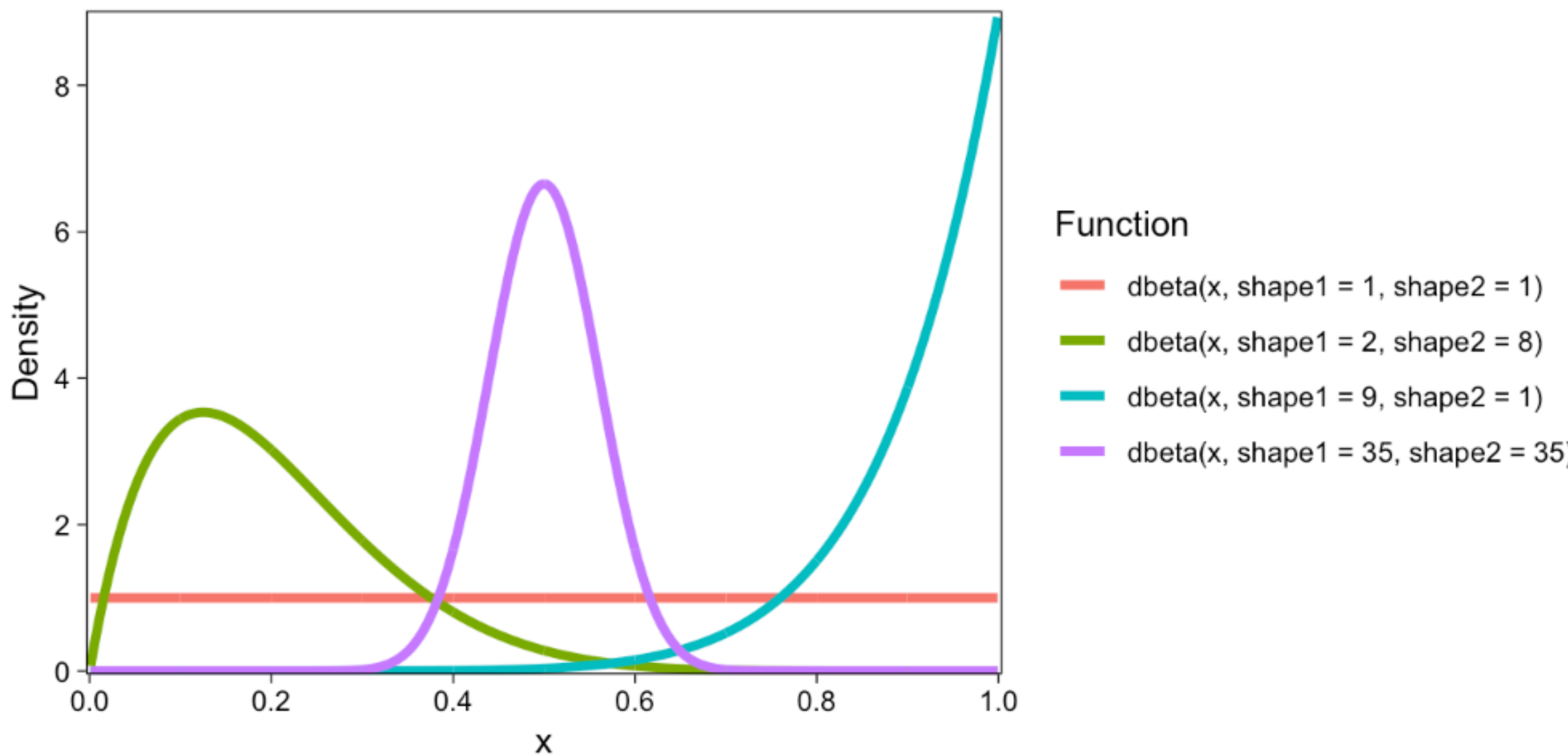


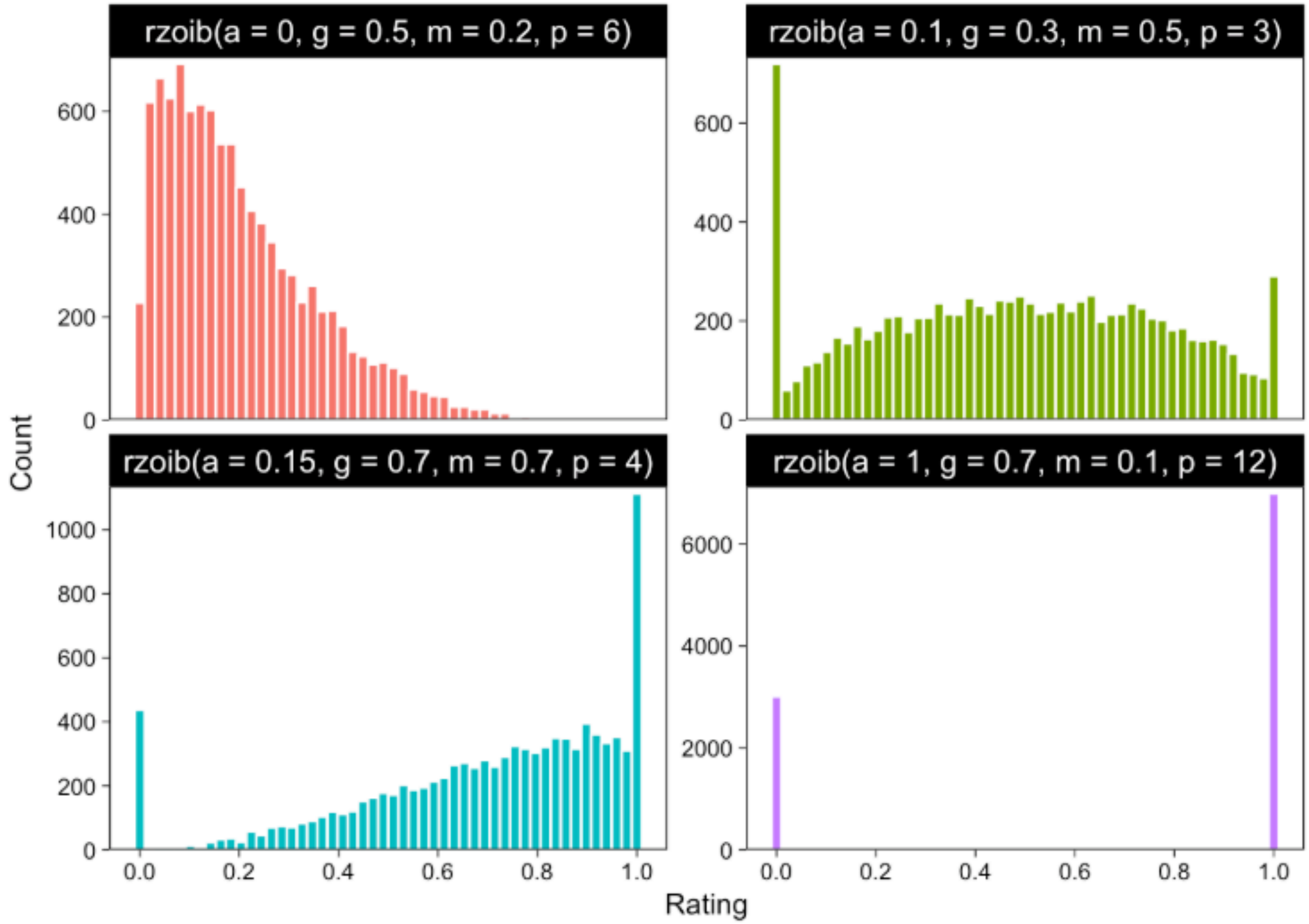
Figure 3: Dr. John A. Zoidberg thinks you should try a ZOIB model on your slider scale data.

Beta distribution



Zero-one inflated **beta** binomial model

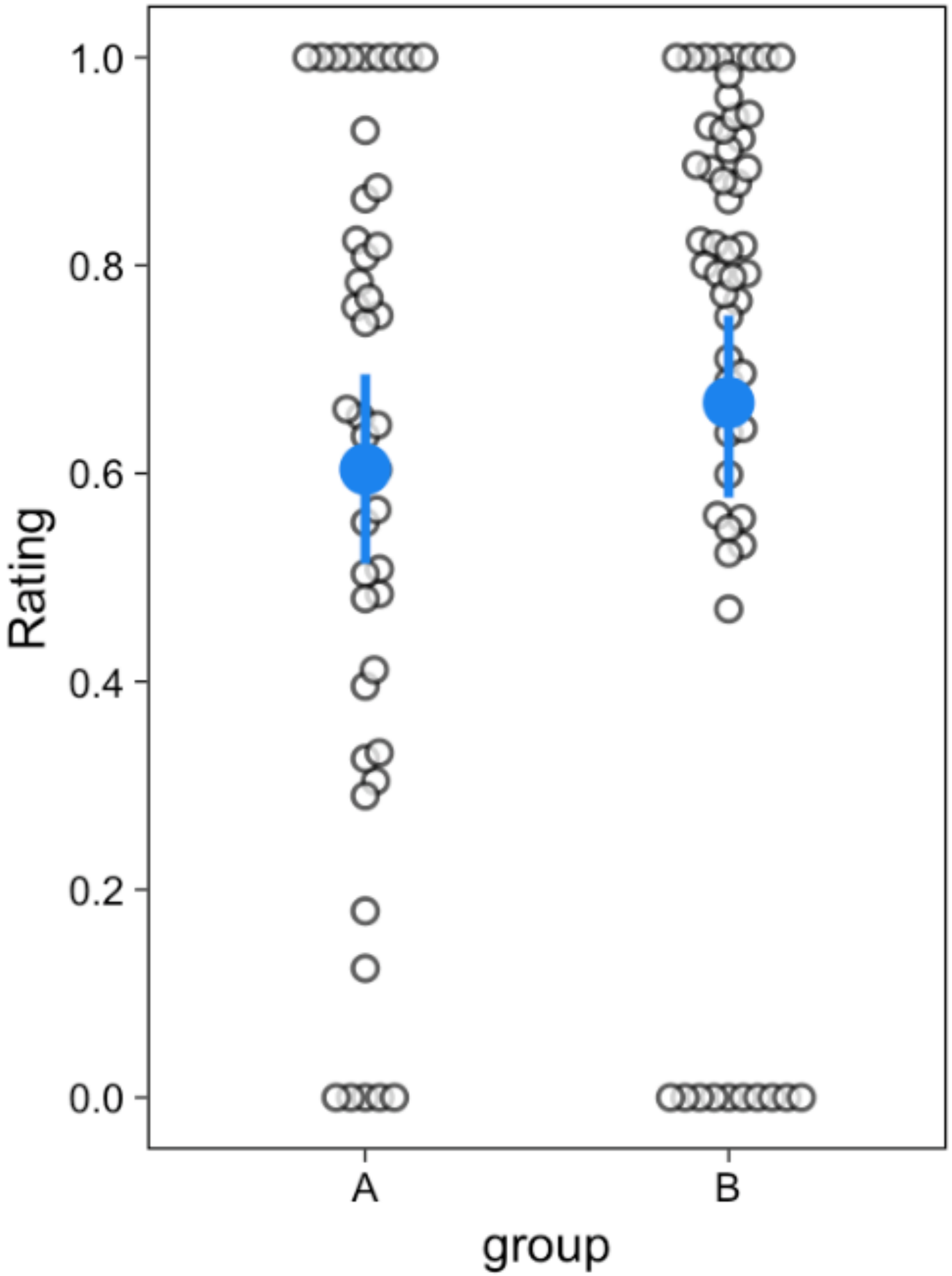
Zero-one inflated beta binomial model



Generative process

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

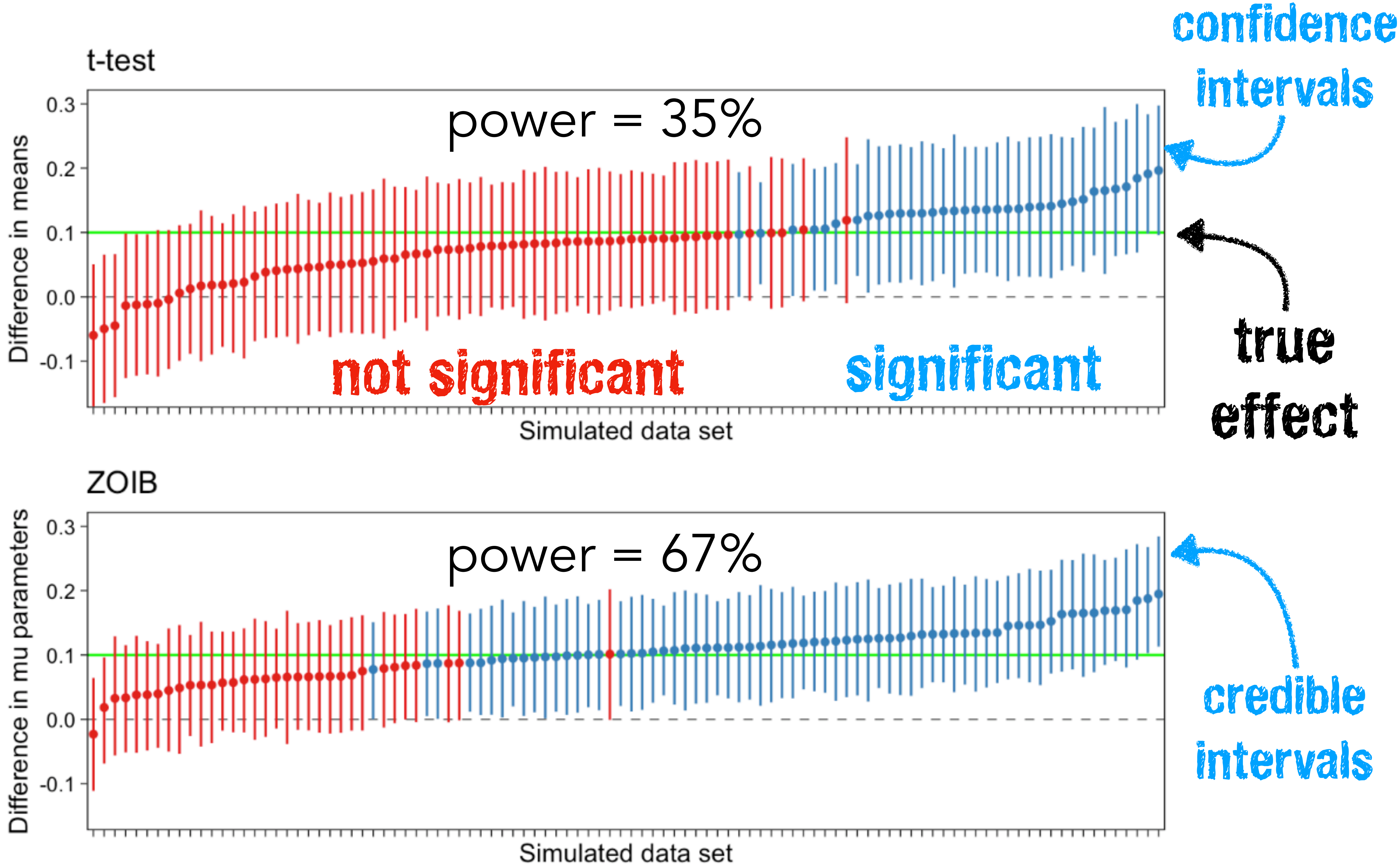
Example data



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



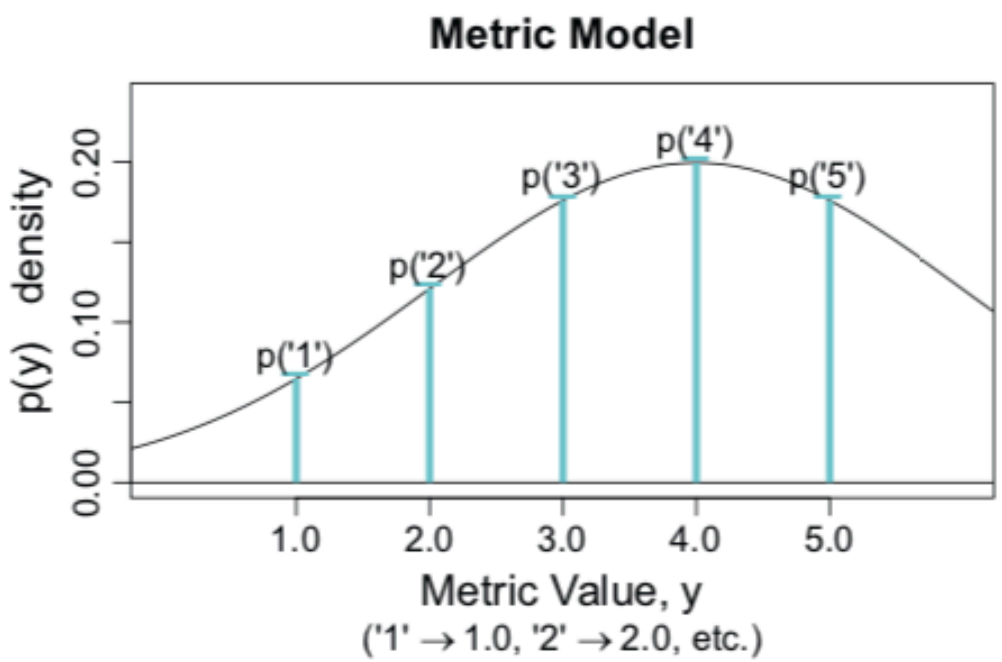
Better modeling Likert scale

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



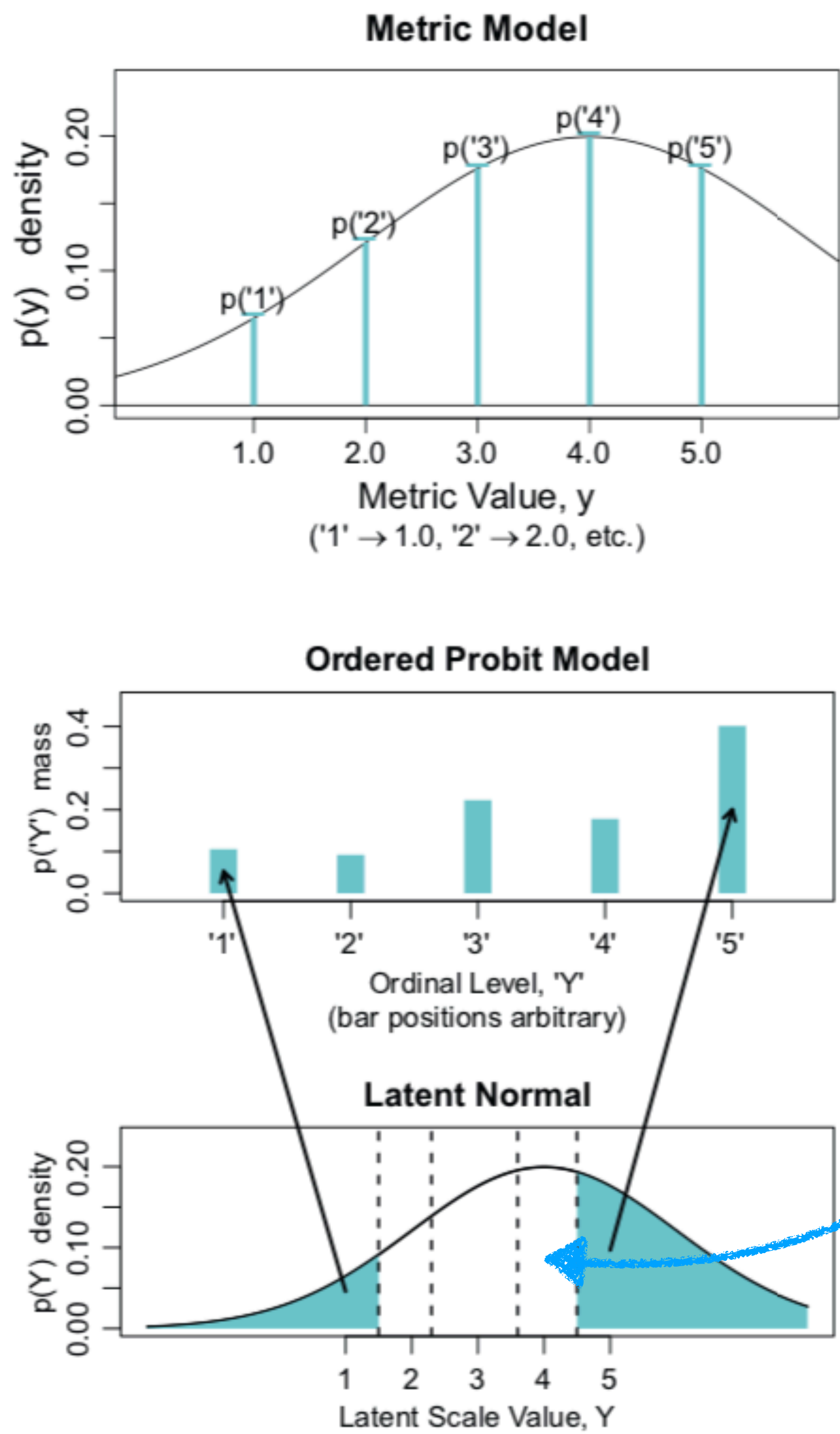
metric model

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

I loooove Bayesian statistics!

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

Ordinal regression



ordered probit model

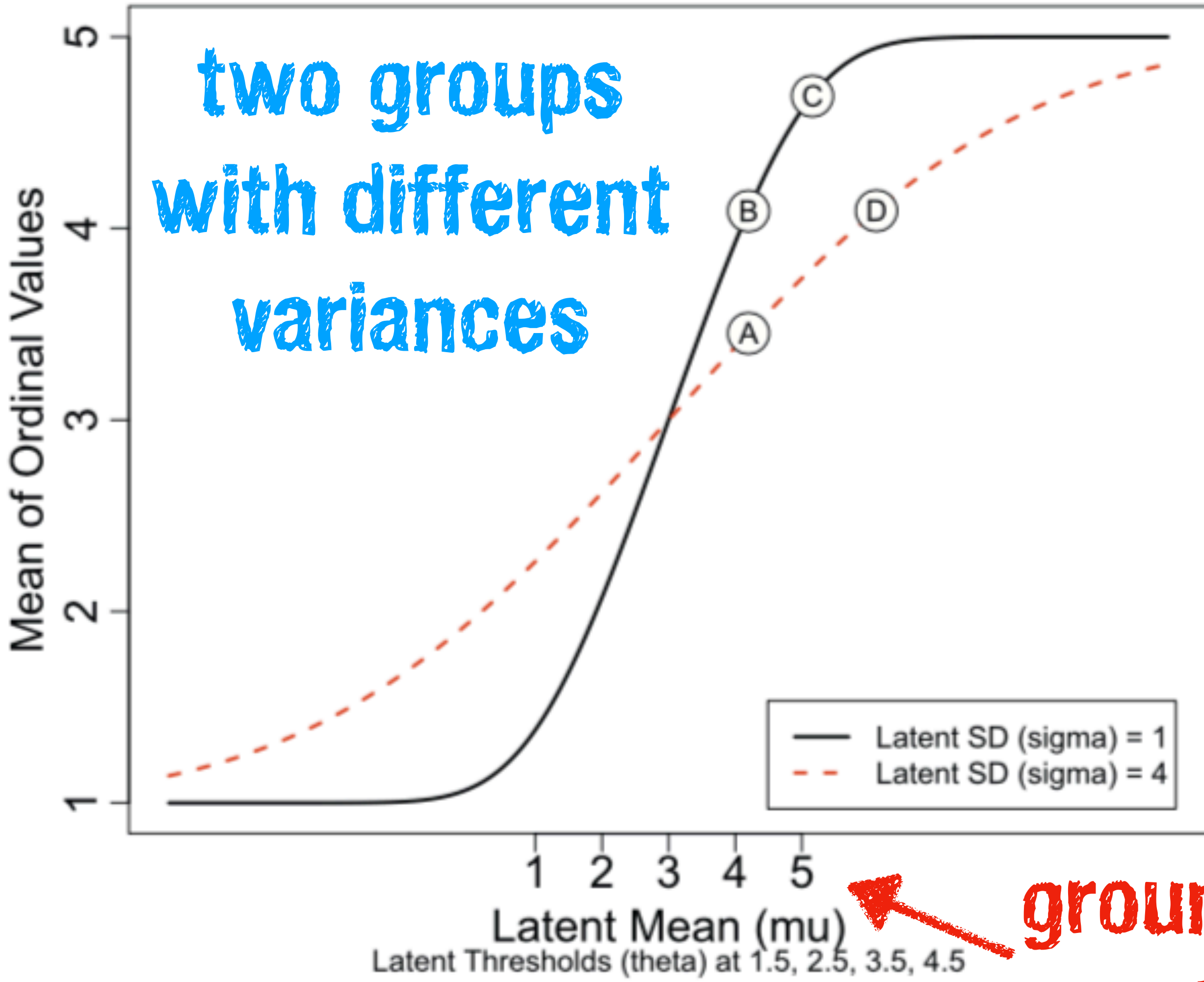
- assumes an ordering of the categories
- assumes a normal distribution in latent space
- finds a good mapping from that normal distribution in latent space to the ordered responses

finds thresholds that separate the categories

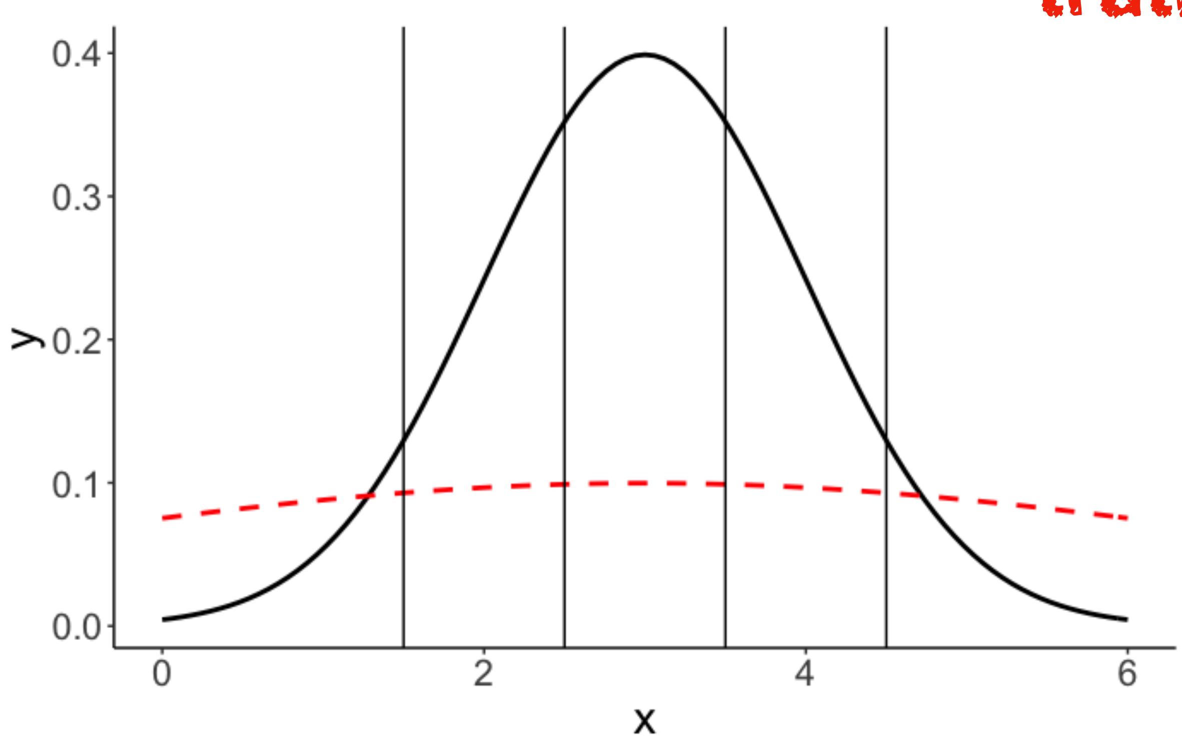
Does it matter?

- 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%

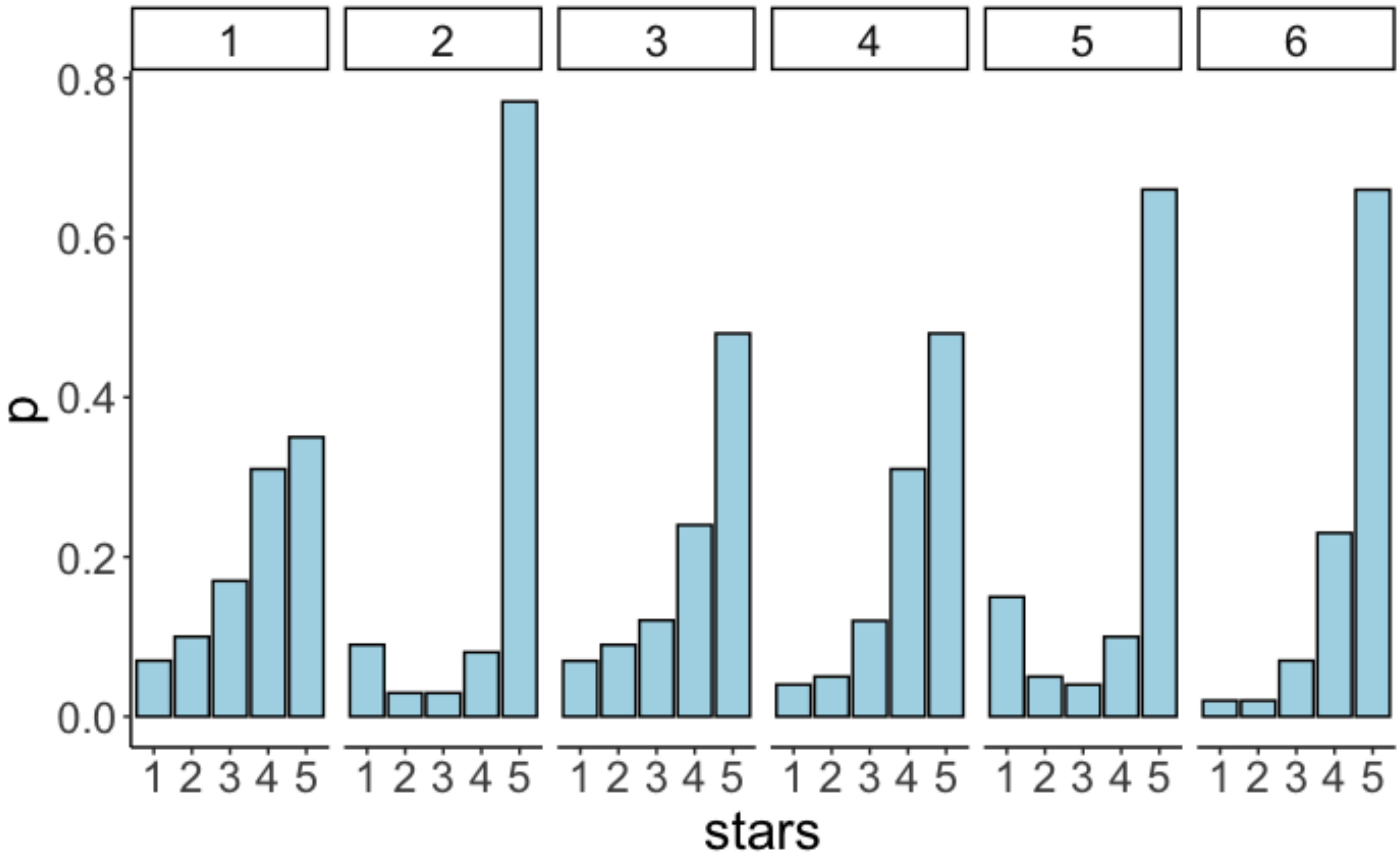
Customer reviews

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

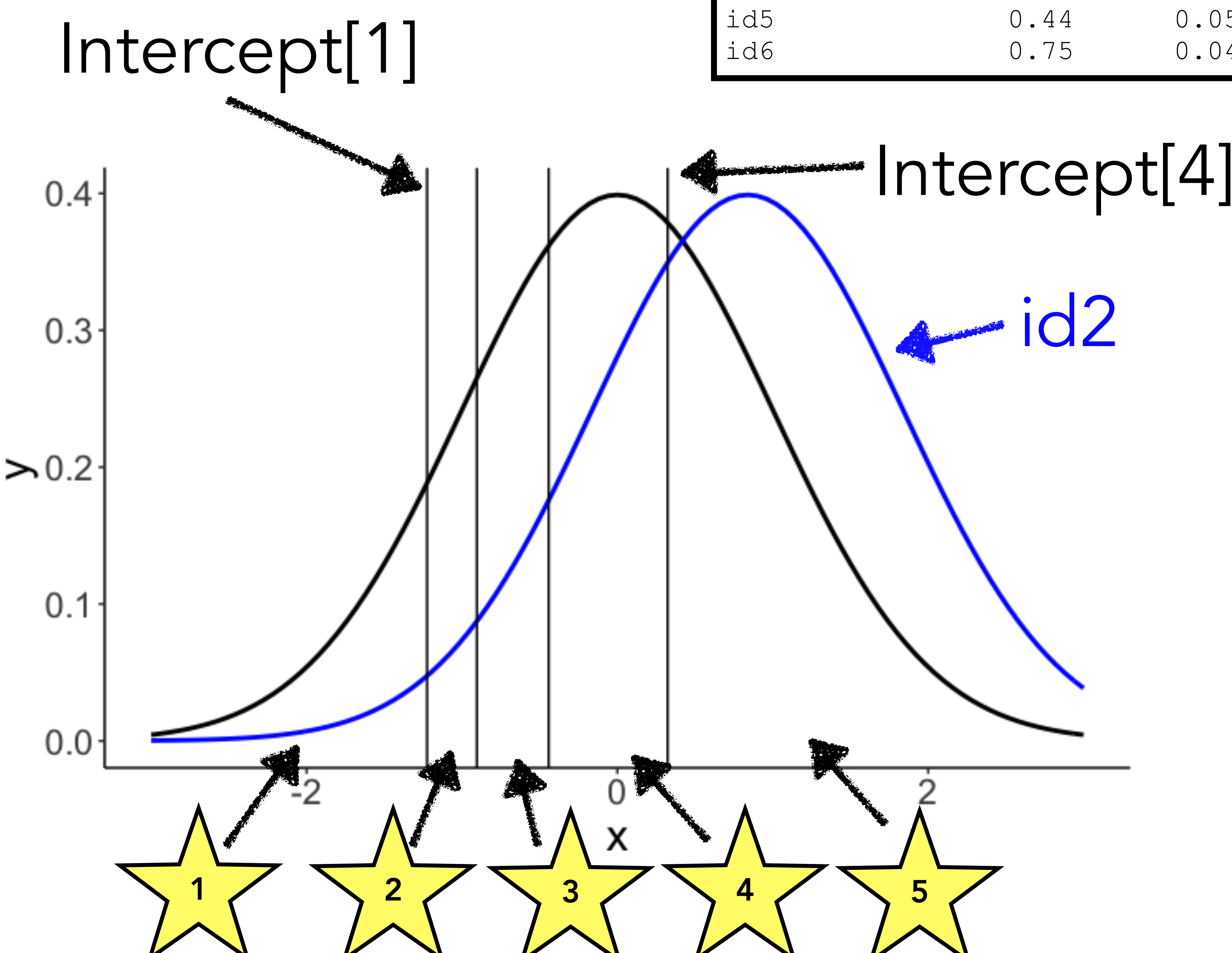
thresholds

difference in mean to reference category

```
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).
```

Interpretation

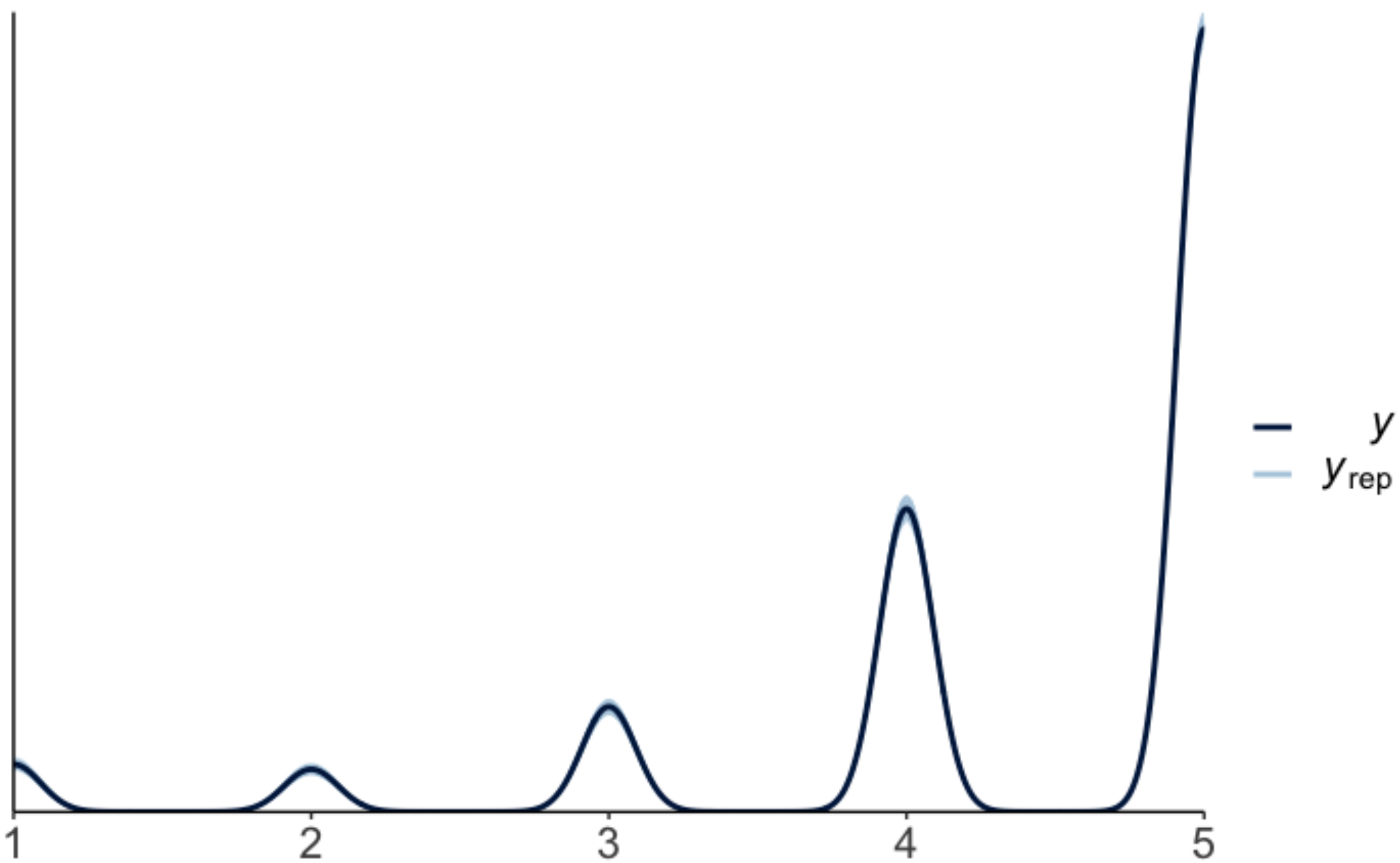
```
Family: cumulative
  Links: mu = probit; disc = identity
Formula: stars ~ 1 + id
   Data: df.movies (Number of observations: 21708)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

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
```

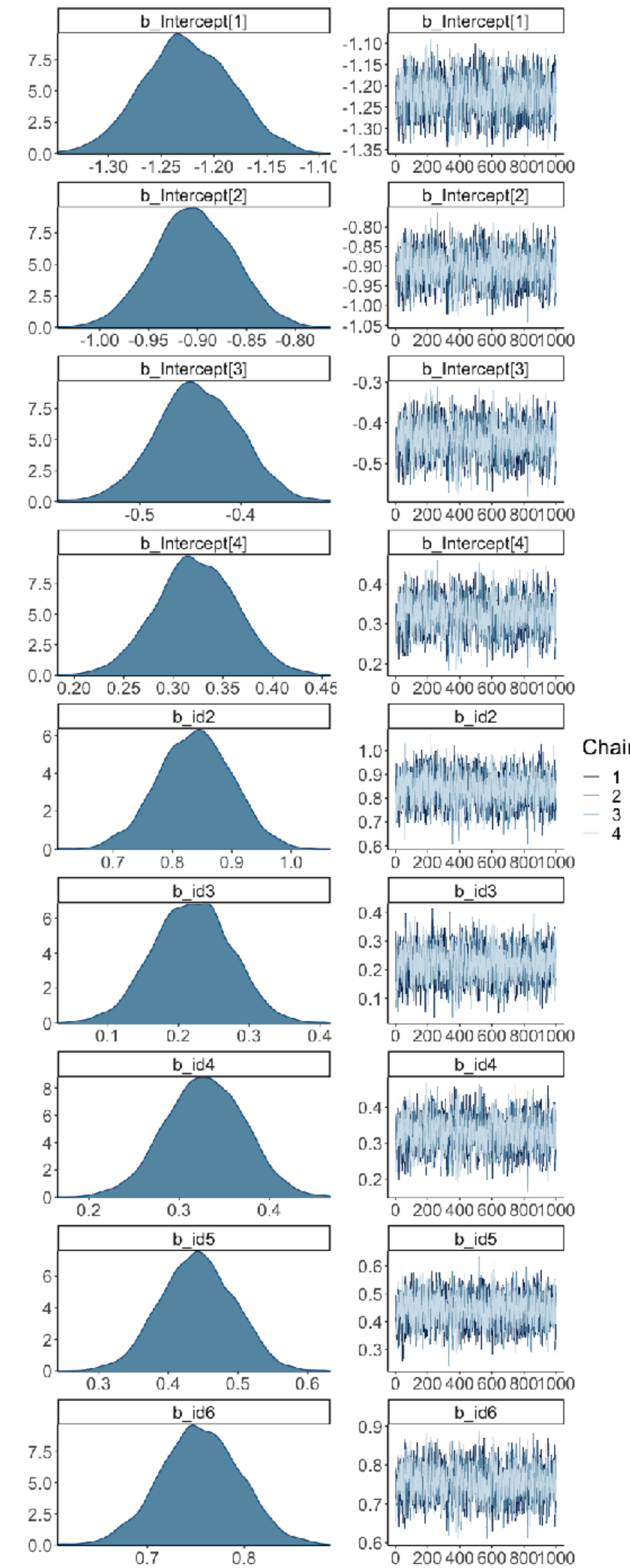


Check the model

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

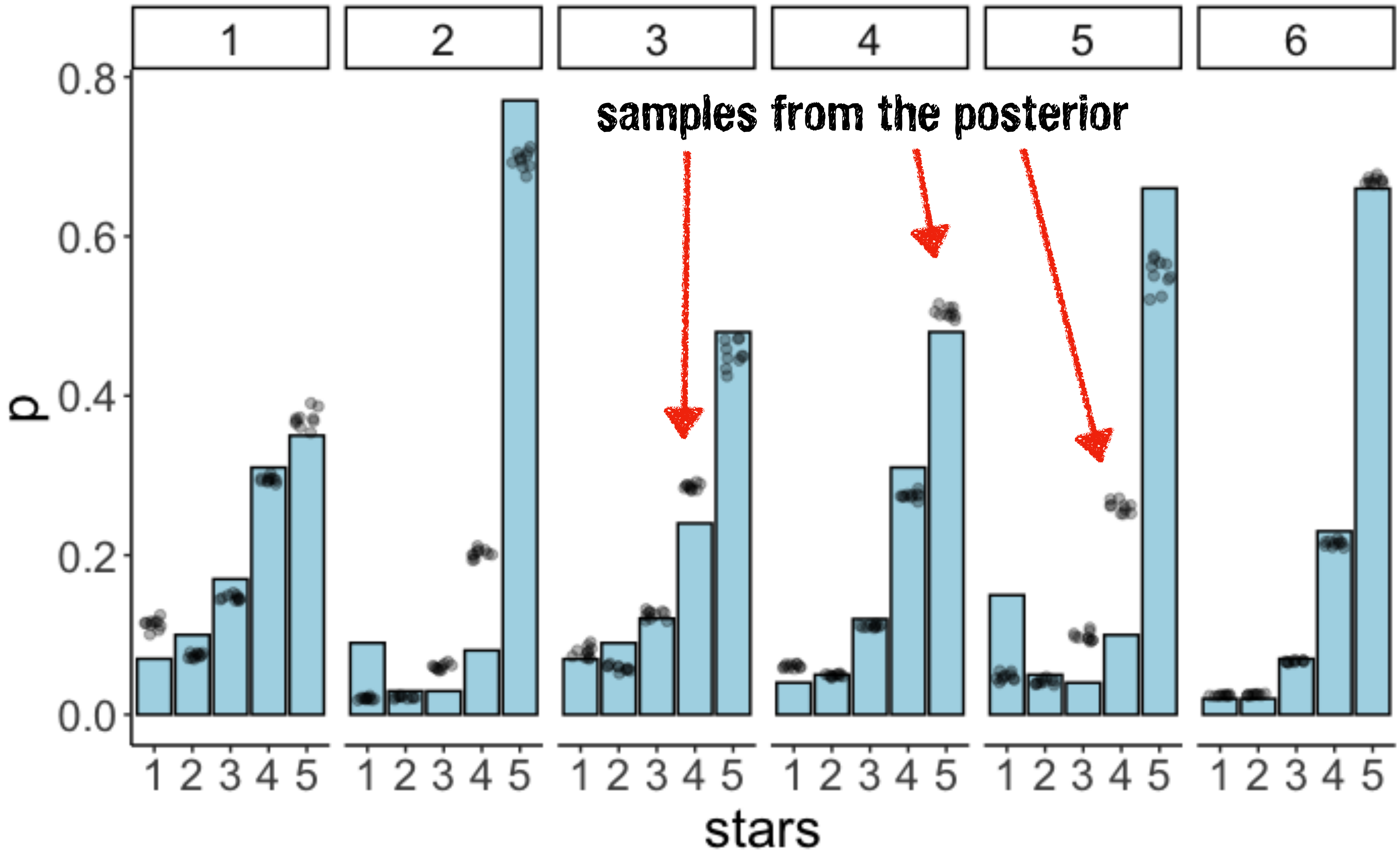
```r
fit.brm5 %>%
  pp_check(nsamples = 100)
```





looking good!

Illustrate the predictions



predictions look pretty good but
maybe we can do better?

Before doing better, let's do worse!

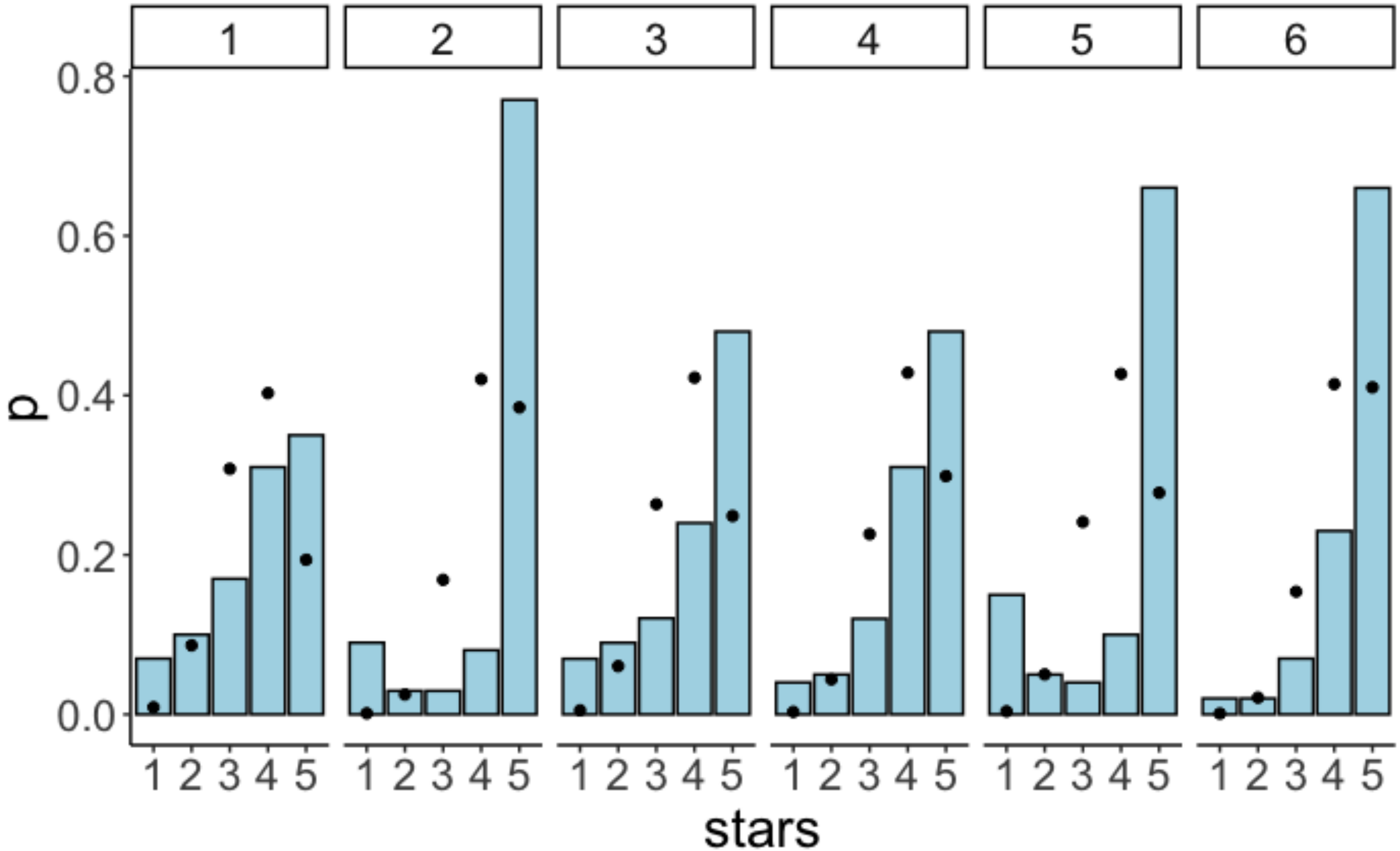
```r
fit.brm6 = brm(formula = stars ~ 1 + id,
               data = df.movies,
               file = "cache/brm6",
               seed = 1)
```

mean for the reference category

difference in mean to reference category

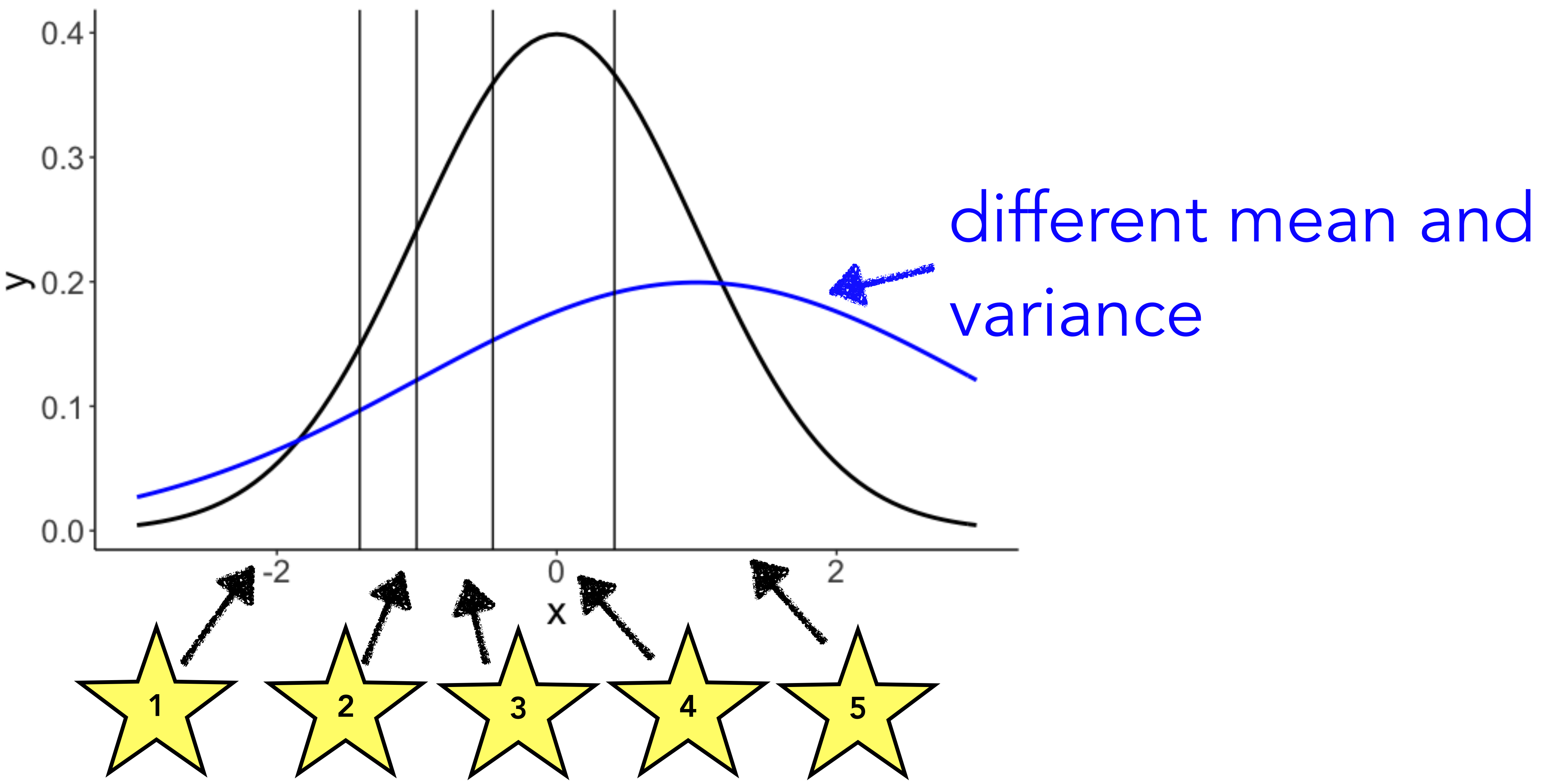
```
 Family: gaussian
  Links: mu = identity; sigma = identity
Formula: stars ~ 1 + id
   Data: df.movies (Number of observations: 21708)
Samples: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;
         total post-warmup samples = 4000

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



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

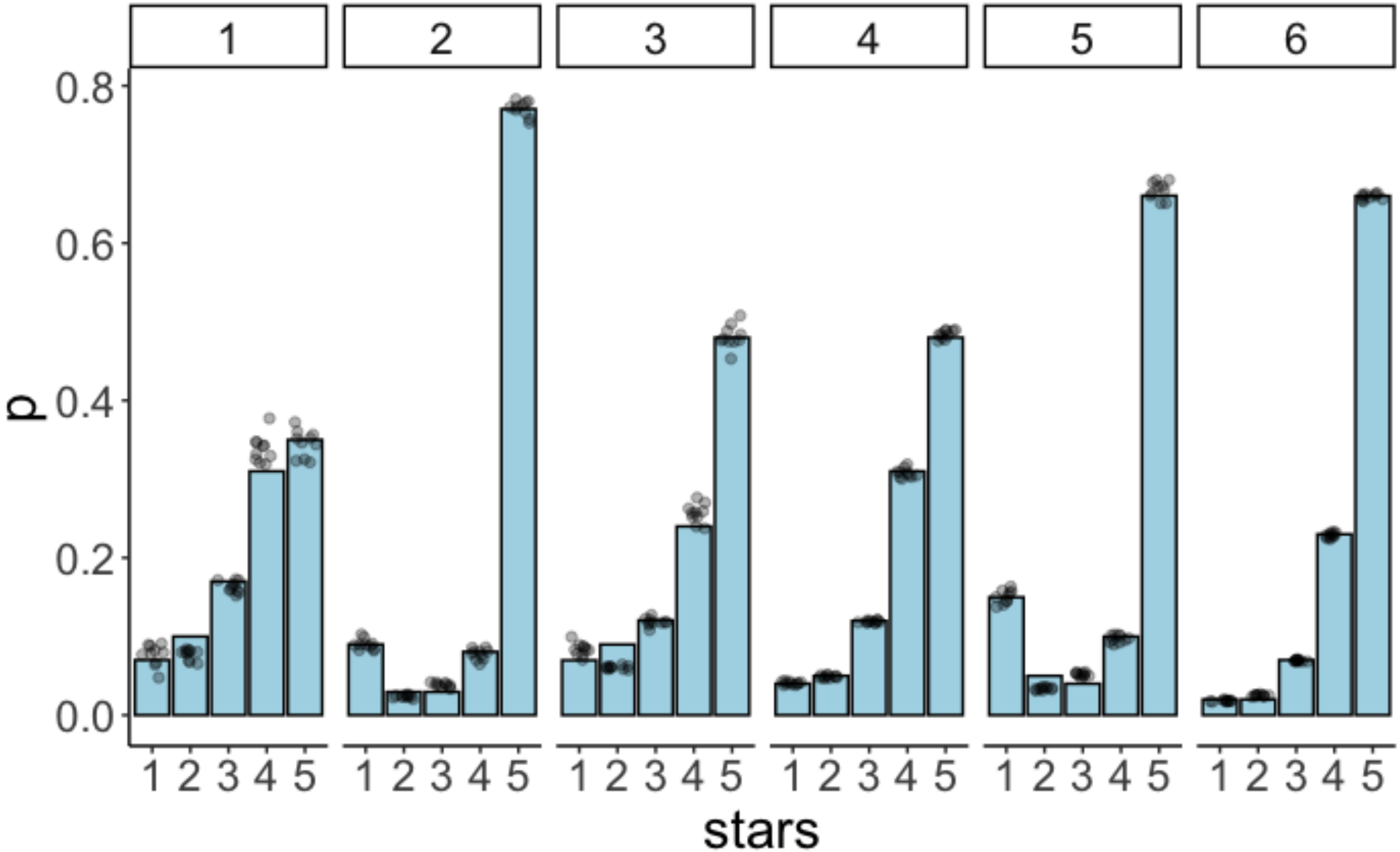
difference in mean

difference in variance

see for details

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

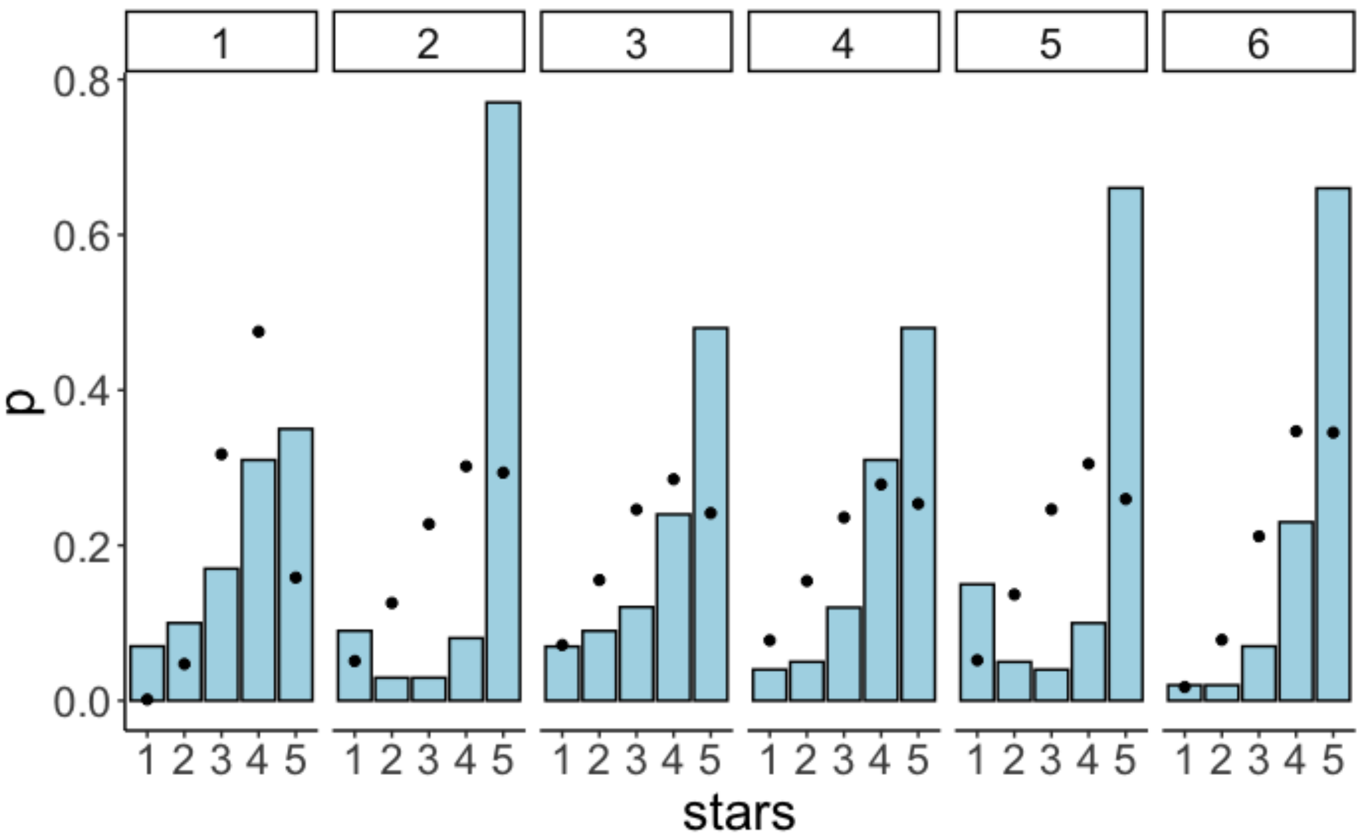
Illustrate the predictions



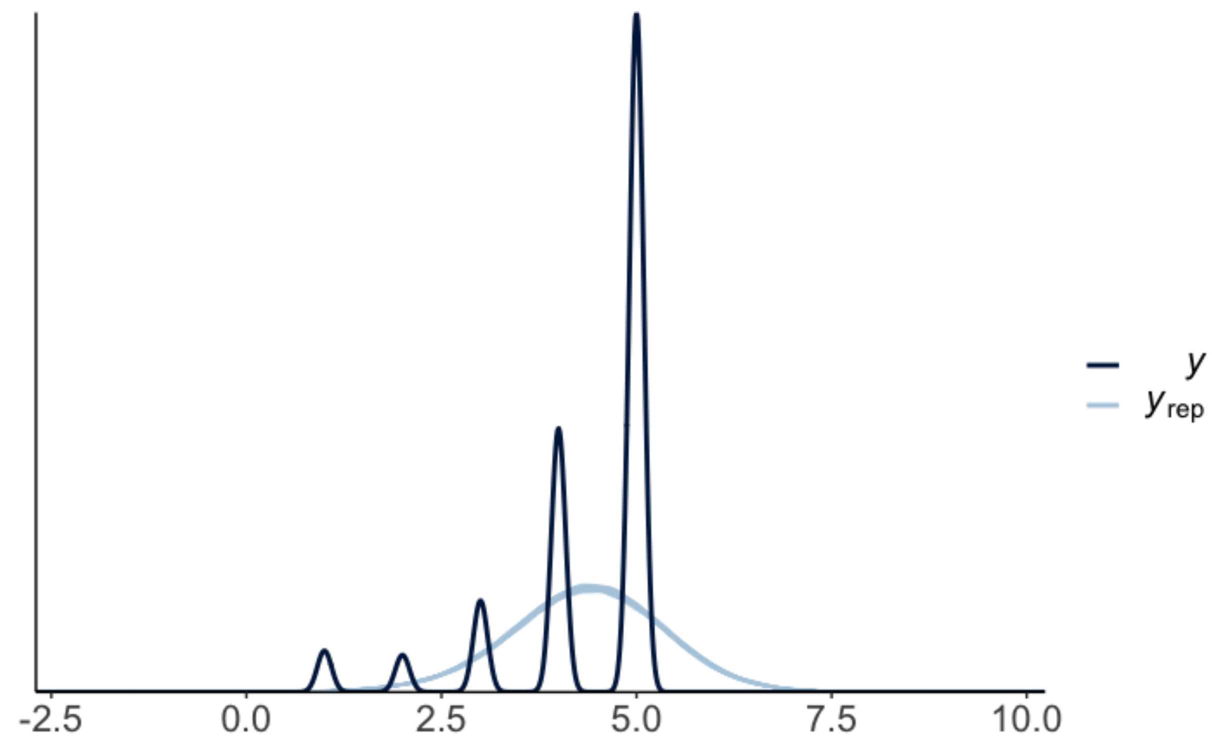
this looks excellent!

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

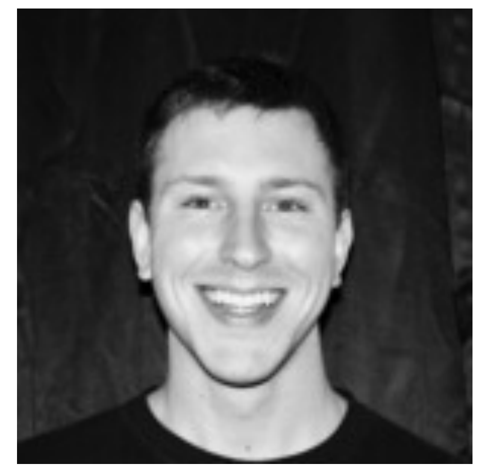
5 star	53%
4 star	13%
3 star	8%
2 star	8%
1 star	17%

Thanks

Psych 252 Team

Andrew Nam

Catherine Thomas



Jon Walters

Dan Yamins

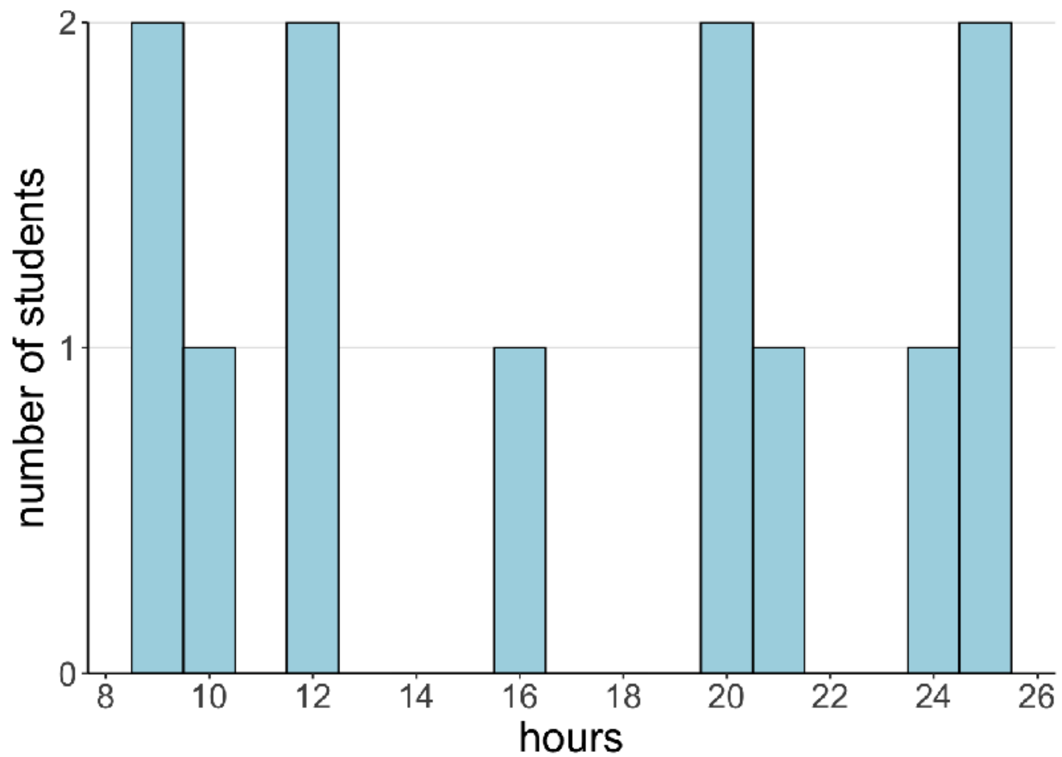
All of you!

Veronica Boyce
Eric Reynolds Brubaker
Madison Bunderson
Anjie Cao
Kevin Chi
Catie Connolly
Lawrence Domingo
Kris Maurice Evans
Dan Fan
Geraldine Fauville
Satchel Grant
Eugy Han
Hsiaolin Hsieh
Emily Hu
Mahnoor Hyat
Kevin Ji
Candice Kim
Christopher John Knight
Tzu-Sheng Kuo
Khuyen Le
Okkeun Lee
Bryce Linford
Marijn Nura Mado
Jessica Jean Mankewitz
Hannah Eve Marshall
Tyler J Matteson
Douglas Steven Miller
Jamie Mitchell
Roza Nalbandyan
Eric Neumann
Joseph Outa
David Rose
Ana Maria Saavedra Pineda
Bianca Silva Santos
Jaehwan Song
Tara Srirangarajan
Aparajitha Suresh
Kayla Chelsea Thomas
Omar Patricio Vasquez Duque
Huan Wang
Jenny Yi-Fang Yang
Yan Echo Zhou

~~Fear~~ of statistics

no more

How long did you work on the midterm?



Thanks for adopting a growth mindset!

fixed mindset:
students believe their basic abilities, their intelligence, their talents, are just fixed traits.

growth mindset:
students *understand* that their talents and abilities can be developed through effort, good teaching and persistence

Vision for this class

In “A Vision for Stanford”, university president Marc Tessier-Lavigne states that Stanford wants to be

“an inspired, inclusive and collaborative community of diverse scholars, students and staff, where all are supported and empowered to thrive.”

Thanks for making it happen!

We're looking forward to your presentations!

Thanks to you!