Introduction to Bayesian Data Analysis

Lecture 2: Analysis Sensitivity and a Bayesian Workflow

Julia Haaf

Summer 2025

"Sensitive Analyses," what does that mean?

What are the causes of sensitive analyses?

1. Data fit different statistical models
   - Different conclusions possible
2. Unclear what is signal and what is noise
   - Random data patterns are interpreted
3. Analyses suggest clear interpretations even though the analysis algorithms did not work reliably.

Sensitive Analyses

What checks for sensitive analyses do you know?

- Check assumptions
  - e.g., repeat analysis without influential outliers
- Multiverse Analysis
  - Identify different reasonable analysis paths and check the robustness of the results

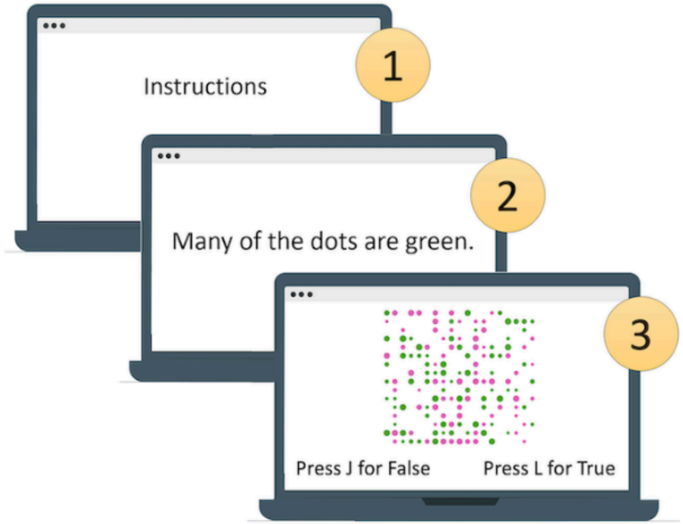
Sensitivity Analysis / Robustness Checks: How robust is my statistical analysis against various decisions (statistical, methodological, theoretical).

Sensitive Analyses

Overview

1. Bayesian Workflow
2. Model Specification
    - Choice of Priors
3. Estimation
    - Convergence
4. Stability and Sensitivity
5. Posterior Predictions

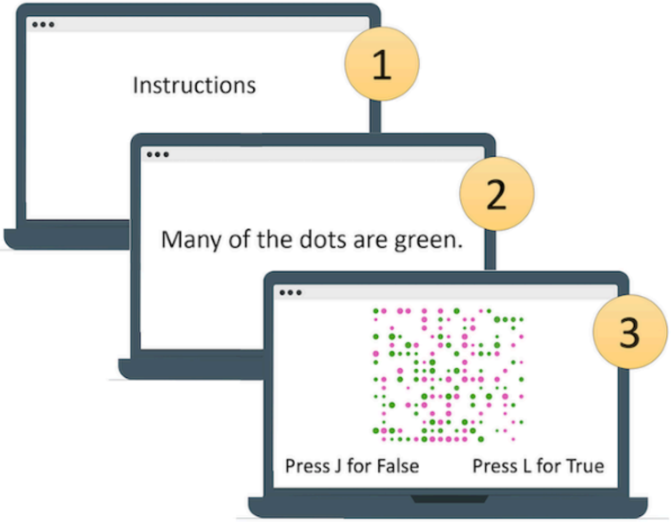
Example: Quantifiers



- Quantifiers: Most, many, more than half, fewer than half, few
- What is the response behavior?
  - Consistent differences between most and more than half
  - Are most and few "symmetric"?
  - Do people differ in their response behavior?
- What is the right model?

Example: Quantifiers

- Model development requires numerous steps that should be well thought out.
- Bayesian Workflow can help



Bayesian Workflow

Bayesian Workflow

Bayesian workflow[*]

Andrew Gelman[†] Aki Vehtari[‡] Daniel Simpson[§] Charles C. Margossian[†]
Bob Carpenter[¶] Yuling Yao[†] Lauren Kennedy[‖] Jonah Gabry[†]
Paul-Christian Bürkner[**] Martin Modrák[††]

2 Nov 2020

Abstract

The Bayesian approach to data analysis provides a powerful way to handle uncertainty in all observations, model parameters, and model structure using probability theory. Probabilistic programming languages make it easier to specify and fit Bayesian models, but this still leaves us with many options regarding constructing, evaluating, and using these models, along with many remaining challenges in computation. Using Bayesian inference to solve real-world problems requires not only statistical skills, subject matter knowledge, and programming, but also awareness of the decisions made in the process of data analysis. All of these aspects can be understood as part of a tangled workflow of applied Bayesian statistics. Beyond inference, the workflow also includes iterative model building, model checking, validation and troubleshooting of computational problems, model understanding, and model comparison. We review all these aspects of workflow in the context of several examples, keeping in mind that in practice we will be fitting many models for any given problem, even if only a subset of them will ultimately be relevant for our conclusions.

Bayesian Workflow

Simplified version: Veenman, Stefan, & Haaf (2024)

Bayesian hierarchical modeling: an introduction and reassessment

Myrthe Veenman[1] · Angelika M. Stefan[2] · Julia M. Haaf[3]

Accepted: 18 July 2023 / Published online: 25 September 2023

Abstract
With the recent development of easy-to-use tools for Bayesian analysis, psychologists have started to embrace Bayesian hierarchical modeling. Bayesian hierarchical models provide an intuitive account of inter- and intraindividual variability and are particularly suited for the evaluation of repeated-measures designs. Here, we provide guidance for model specification and interpretation in Bayesian hierarchical modeling and describe common pitfalls that can arise in the process of model fitting and evaluation. Our introduction gives particular emphasis to prior specification and prior sensitivity, as well as to the calculation of Bayes factors for model comparisons. We illustrate the use of state-of-the-art software programs Stan and *brms*. The result is an overview of best practices in Bayesian hierarchical modeling that we hope will aid psychologists in making the best use of Bayesian hierarchical modeling.

Bayesian Workflow

- Model specification
  - Model structure (e.g. Normal vs. Lognormal)
  - Choice of priors (pick priors, prior prediction)
- Estimate model parameters
  - Check convergence via model diagnostics (trace plots, $\hat{R}$, ESS)
  - Visualize posteriors
- Model comparison: The *most Bayesian way* is Bayes factors
- Sensitivity analysis
  - Check estimation stability
  - Check prior sensitivity
- Also sensible: Posterior prediction (posterior predictive checks)

Simple example: Button press

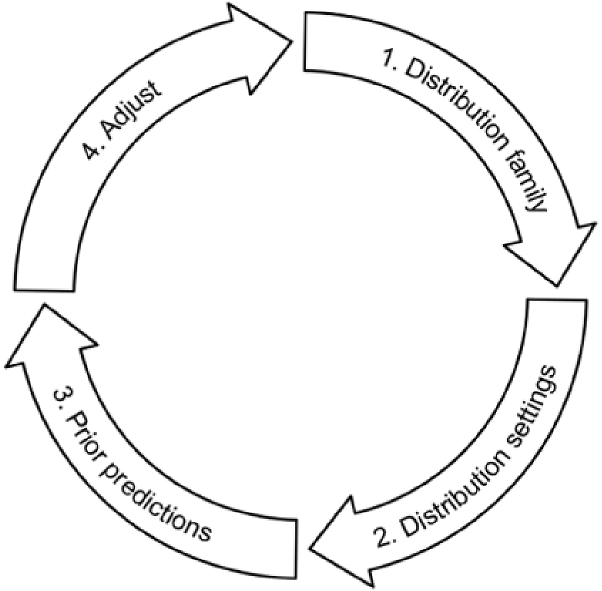
- Finger tapping task (for a review, see Hubel et al. 2013)
- Procedure
    - blank screen (200 ms)
    - cross in the middle of a screen
    - as soon as they see the cross, they tap on the space bar as fast as they can until the experiment is over (361 trials).
- Dependent measure: Finger tapping times in milliseconds.
- Research question is: how long does it take for this particular subject to press a key?

Model specification

What is a good model for these data?

- Y_i is the reaction time in the ith trial
- $Y_i \sim \mathrm{Normal}(\mu, \sigma^2)$
- Priors for μ and σ^2 are needed

Choice of Prior



- What properties should the prior have?
  - e.g., only positive values
- What parameter values do we expect?
  - e.g., by how many units does the dependent variable increase with one unit of the predictor?
- If we choose this prior, does the model make plausible predictions for the data?
- If not, we need to adjust the prior

Choice of Prior

```
library(brms)
bprior <- c(prior(normal(0,2), class = Intercept)
            , prior(student_t(3, 0, 10), class = sigma))
```

- To figure out whether this is a good prior, we can use prior prediction
  - Simulate data from the speficied model
  - Assess whether these data are plausible

Choice of Prior

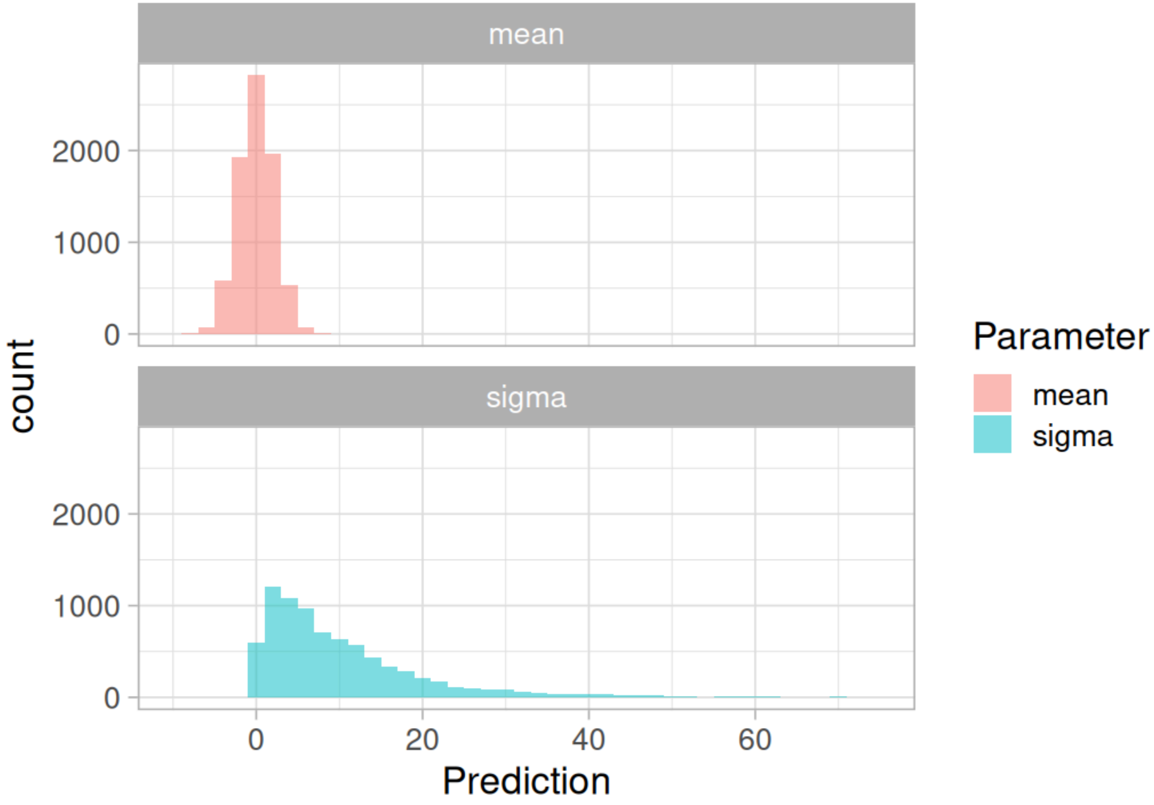
Prior Prediction

```
dat <- data.frame(t = rep(1, 361), trial = 1:361)
model.prior <- brms::brm(t ~ 1
            , data = dat
            , prior = bprior
            , sample_prior = "only"
            , iter = 4000
            , cores = 4)
```

```
prior.predictions <- posterior_predict(model.prior)
```

Choice of Prior

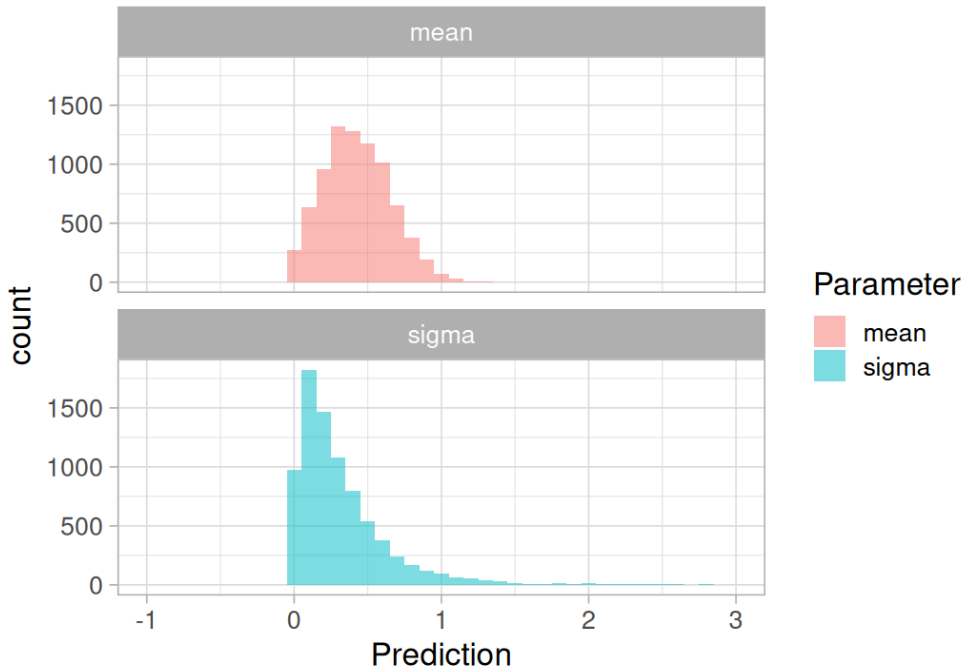
Visualization of predicted data (say the mean)



Choice of Prior

Redo

```
bprior <- c(prior(normal(.4,.25), class = Intercept, lb = 0)
            , prior(student_t(3, 0, .3), class = sigma))
```



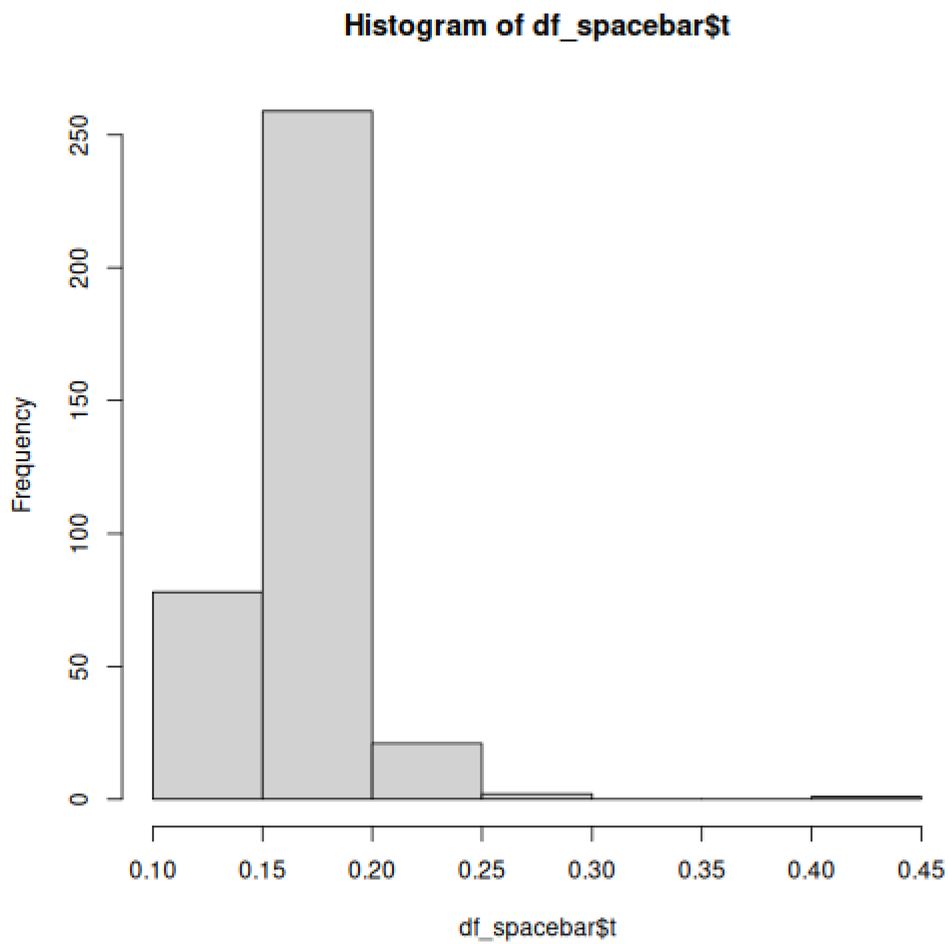
Estimate model parameters

```
load("data/df_spacebar.rda")
head(df_spacebar)
```

```
## # A tibble: 6 × 2
##       t trial
##   <int> <int>
## 1   141     1
## 2   138     2
## 3   128     3
## 4   132     4
## 5   126     5
## 6   134     6
```

```
df_spacebar$t <-  df_spacebar$t / 1000
```

Estimate model parameters



Estimate model parameters

```
model.fit <- brms::brm(t ~ 1
            , data = df_spacebar
            , prior = bprior
            , iter = 4000
            , cores = 4)
```

Estimate model parameters

Model Diagnostics for Convergence Checks

Did the algorithm draw sufficient and good samples from the posterior distribution?

- $\hat{R}$: `rhat(model.fit)`
- Effective samples: `neff_ratio(model.fit)`
- Trace plot: `mcmc_plot(model.fit, type = "trace")`

Estimate model parameters

$\hat{R}$

- $\hat{R}$ compares the between- and within-chain estimates.
- $\hat{R}$ is larger than 1 when chains have not mixed well.
- Only rely on the model if the $\hat{R}$s are less than 1.05.

```
brms::rhat(model.fit)
```

```
## b_Intercept       sigma   Intercept      lprior        lp__
##    1.000647    1.001195    1.000647    1.000570    1.000675
```

Estimate model parameters

Effective samples

- Effective samples measure the efficiency of the sampling algorithm.
- Bulk ESS is for the bulk of the posterior distribution (relevant for mean and median estimates)
- Tail ESS indicates the sampling efficiency at the tails of the distribution (relevant for credible intervals)

```
neff_ratio(model.fit)
```

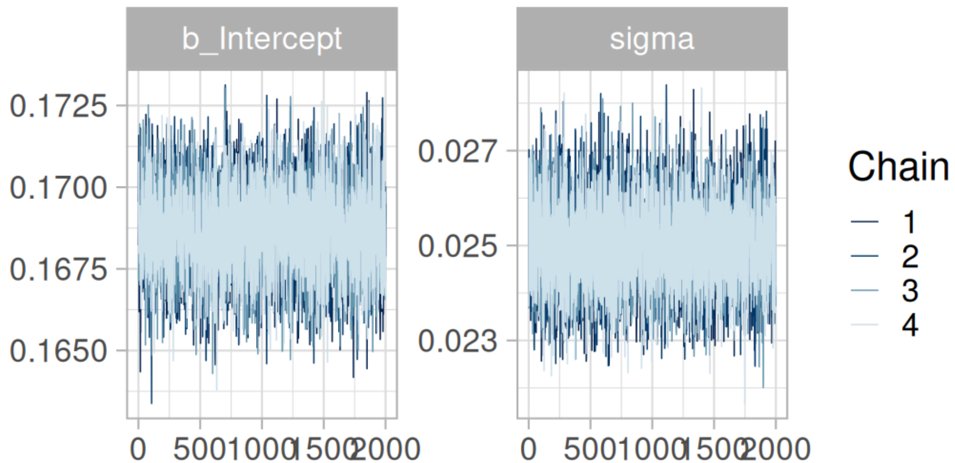
```
## b_Intercept       sigma   Intercept      lprior        lp__
##   0.6001314   0.7009648   0.6001314   0.6059223   0.4880597
```

Estimate model parameters

Trace plot

```
mcmc_plot(model.fit, type = "trace") +
  theme_light(base_size = 20)
```

```
## No divergences to plot.
```



Estimate model parameters

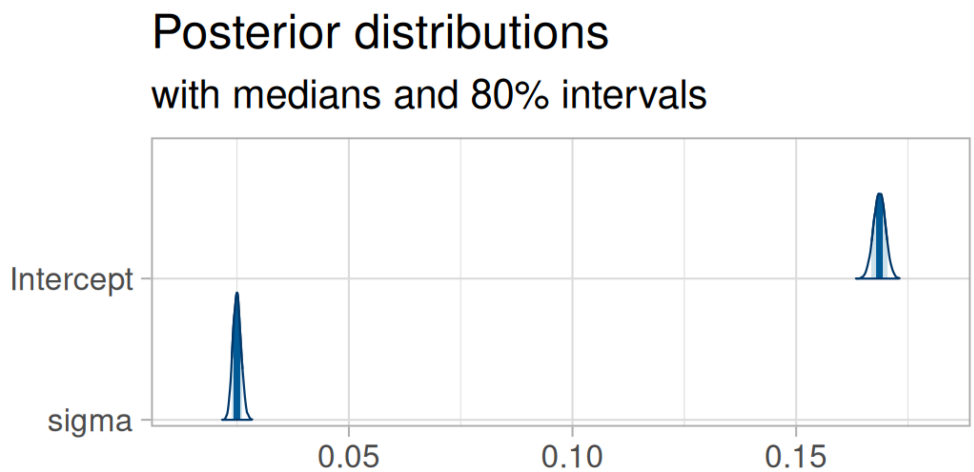
Visualize posteriors

```
model.fit
```

```
##  Family: gaussian
##   Links: mu = identity; sigma = identity
## Formula: t ~ 1
##    Data: df_spacebar (Number of observations: 361)
##   Draws: 4 chains, each with iter = 4000; warmup = 2000; thin = 1;
##          total post-warmup draws = 8000
##
## Regression Coefficients:
##           Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS
Tail_ESS
## Intercept     0.17      0.00     0.17     0.17 1.00     6686
4801
##
## Further Distributional Parameters:
##       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
## sigma     0.02      0.00     0.02     0.03 1.00     6955     5608
##
## Draws were sampled using sampling(NUTS). For each parameter
```

Estimate model parameters

```r
library(bayesplot)
mcmc_areas(model.fit, prob = 0.8
           , pars = c("Intercept", "sigma")) +
  labs(title = "Posterior distributions"
       , subtitle = "with medians and 80% intervals") +
  theme_light(base_size = 20)
```



Model comparison

- To assess how well a model reflects key patterns in the data, one typically compares it to one or more model with alternative specifications.
- That's tomorrow's topic

Sensitivity Analysis

1. Check estimation stability
2. Check prior sensitivity

Estimation stability

- Algorithms for estimating posterior distributions (and Bayes factors) are probabilistic
    - Random fluctuations can affect the results
- Stability Analysis: Repeat all processes with probabilistic estimators.

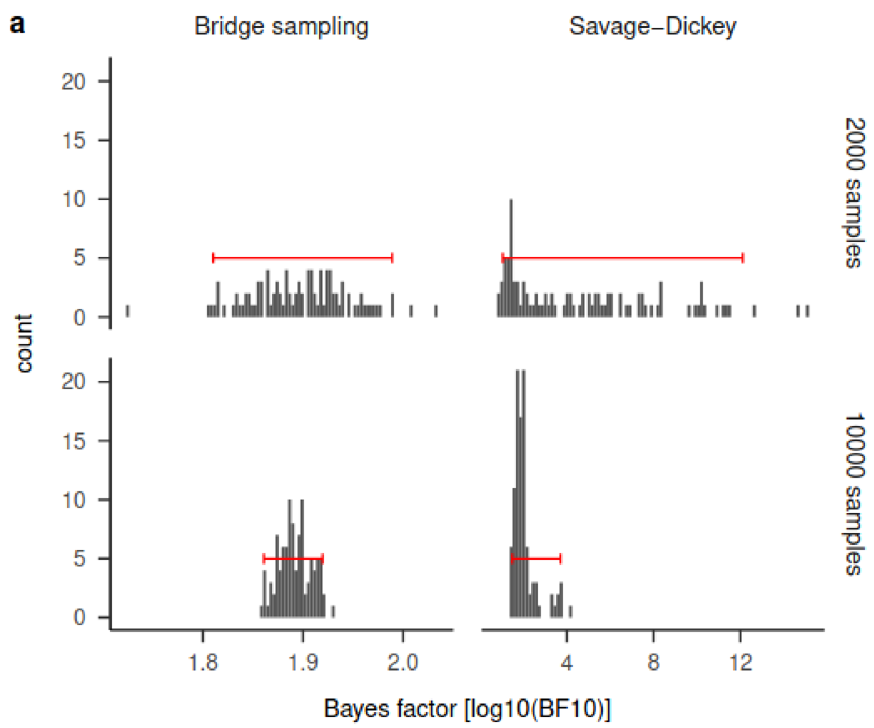


Figure from Schad, et al. (2023) *Psychological Methods*.

Prior Sensitivity

What influence does the choice of my prior have on the results?

1. Determine value ranges for realistic priors for relevant parameters
2. Repeat analysis with combinations of realistic priors for different parameters
3. Do the conclusions change (not the values)?

Table 1 Sensitivity of Bayes factors to prior settings

Scale on ν	Scale on ϵ	$\mathcal{M}_0$	$\mathcal{M}_1$	$\mathcal{M}_+$	$\mathcal{M}_{SS}$	$\mathcal{M}_u$
Priming						
1/6	**1/10**	0.12	*	0.01	0.03	0.02
1/12	1/20	0.06	*	0.04	0.06	0.04
1/12	1/5	0.14	*	0.02	0.72	0.04
1/3	1/20	0.06	*	4.79e −8	0	1.03e −5
1/3	1/5	0.13	*	2.25 e −8	7.7e −4	0.98e −5

Figure from Haaf & Rouder (2019) *Psychonomic Bulletin and Review.*

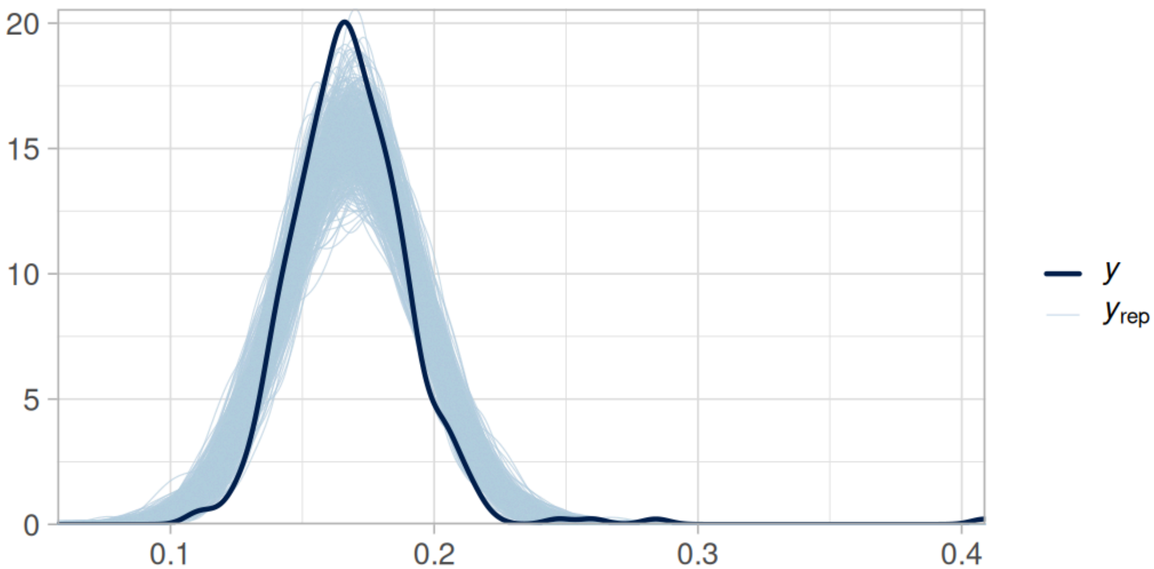
Posterior Prediction

- The posterior distributions can be used to generate predictions for future data from the model.
- Given the data from the current study, the posterior predictive distribution tells us what data we should expect in the future.
- Formally (assuming past and future data areindependent): $p(y_{pred}|y) = \int_\theta p(y_{pred}|\theta)p(\theta|y)d\theta$
- Practically, samples from the posterior predicitve distribution can be drawn using brms.

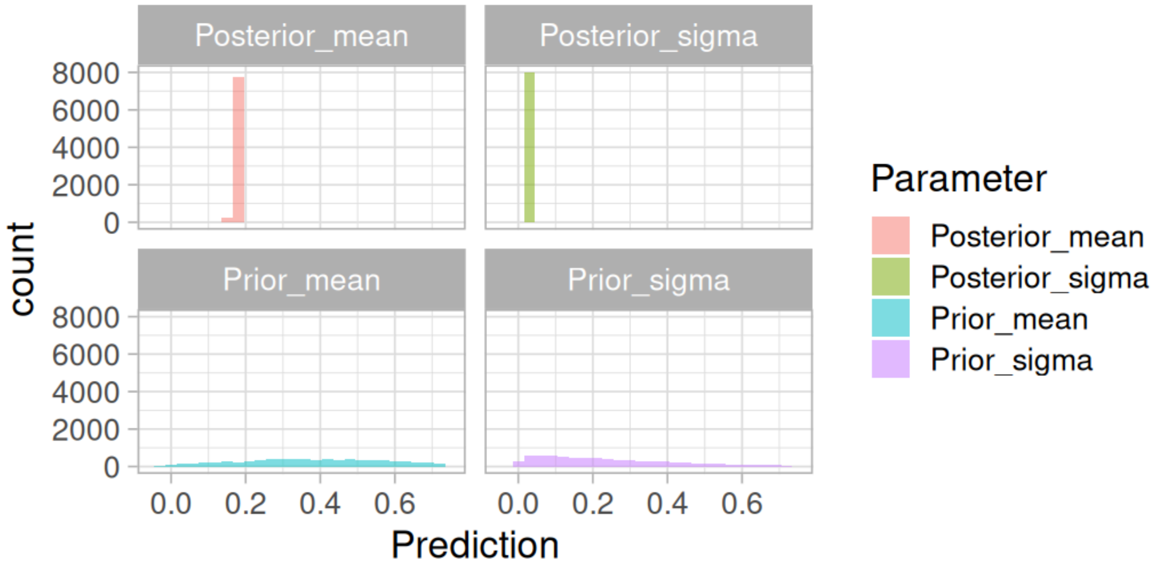
```
posterior.predictions <- posterior_predict(model.fit)
```

Posterior Prediction

```
bayesplot::ppc_dens_overlay(df_spacebar$t,
posterior.predictions[1:500,]) +
  theme_light(base_size = 16)
```



Posterior Prediction



Wrap up

Rethinking needed!

- Uncertainty is not a bad thing in statistics (as long as we can quantify it)
- Making decisions is necessary in statistics (as long as we handle them transparently)
- Workflows can help structure decisions and uncertainties, not prevent them.

Literature

Gelman, A., Vehtari, A., Simpson, D., Margossian, C. C., Carpenter, B., Yao, Y., ... & Modrák, M. (2020). Bayesian workflow. arXiv preprint arXiv:2011.01808.

Schad, D. J., Betancourt, M., & Vasishth, S. (2021). Toward a principled Bayesian workflow in cognitive science. Psychological Methods, 26(1), 103–126. https://doi.org/10.1037/met0000275

Schad, D. J., Nicenboim, B., Bürkner, P. C., Betancourt, M., & Vasishth, S. (2023). Workflow techniques for the robust use of bayes factors. Psychological methods, 28(6), 1404.

Veenman, M., Stefan, A. M., & Haaf, J. M. (2024). Bayesian hierarchical modeling: An introduction and reassessment. Behavior Research Methods, 56(5), 4600-4631.

Ramotowska, S., Haaf, J., Van Maanen, L., & Szymanik, J. (2024). Most quantifiers have many meanings. Psychonomic Bulletin & Review, 1-12.

Sarafoglou, A., Giacobello, A., Godmann, H., Johnson, T., Visser, I., Haaf, J. M., & Szymanik, J. (2024). A Bayesian framework to study individual differences in semantic representations.

Haaf, J. M., & Rouder, J. N. (2019). Some do and some don't? Accounting for variability of individual difference structures. Psychonomic Bulletin & Review, 26(3), 772-789.

Roos, M., Martins, T. G., Held, L., & Rue, H. (2015). Sensitivity analysis for Bayesian hierarchical models. Bayesian Analysis, 10(2), 321–349. https://doi.org/10.1214/14-BA909

:)