Bayesian approaches to mixed effects modeling

Yaqian Huang

yaqian.huang@oeaw.ac.at

September 27, 2023

University of Canterbury | NZILBB

90:00 VS 15:00

wins! wins!(?)

70% (hypothetical) 0 : 1 40% (hypothetical)

P(Hypothesis) P(Hypothesis|Data)

Prior beliefs Data Posterior beliefs

Today’s posterior is tomorrow’s prior. – Lindley (1972:2)

An **A and Z** demo, not A to Z

- Bayesian concepts
- A hands-on example in R
- Resources for A-Z & more advanced topics

Why Bayesian?

- Intuitive, natural way to think about the world
- Philosophical reasons

Frequentist vs. Bayesian

- Null hypothesis significance testing by a cut-off value
- Estimate uncertain
- Truth fixed

"how unusual to get 5 straight heads if one were flipping a 'fair' coin with probability heads being 0.5?"

- Quantifies uncertainty around possible parameter values
- Estimate certain
- Truth unknown

- "how likely is this a 'fair' coin if one got 5 straight heads when flipping it?"

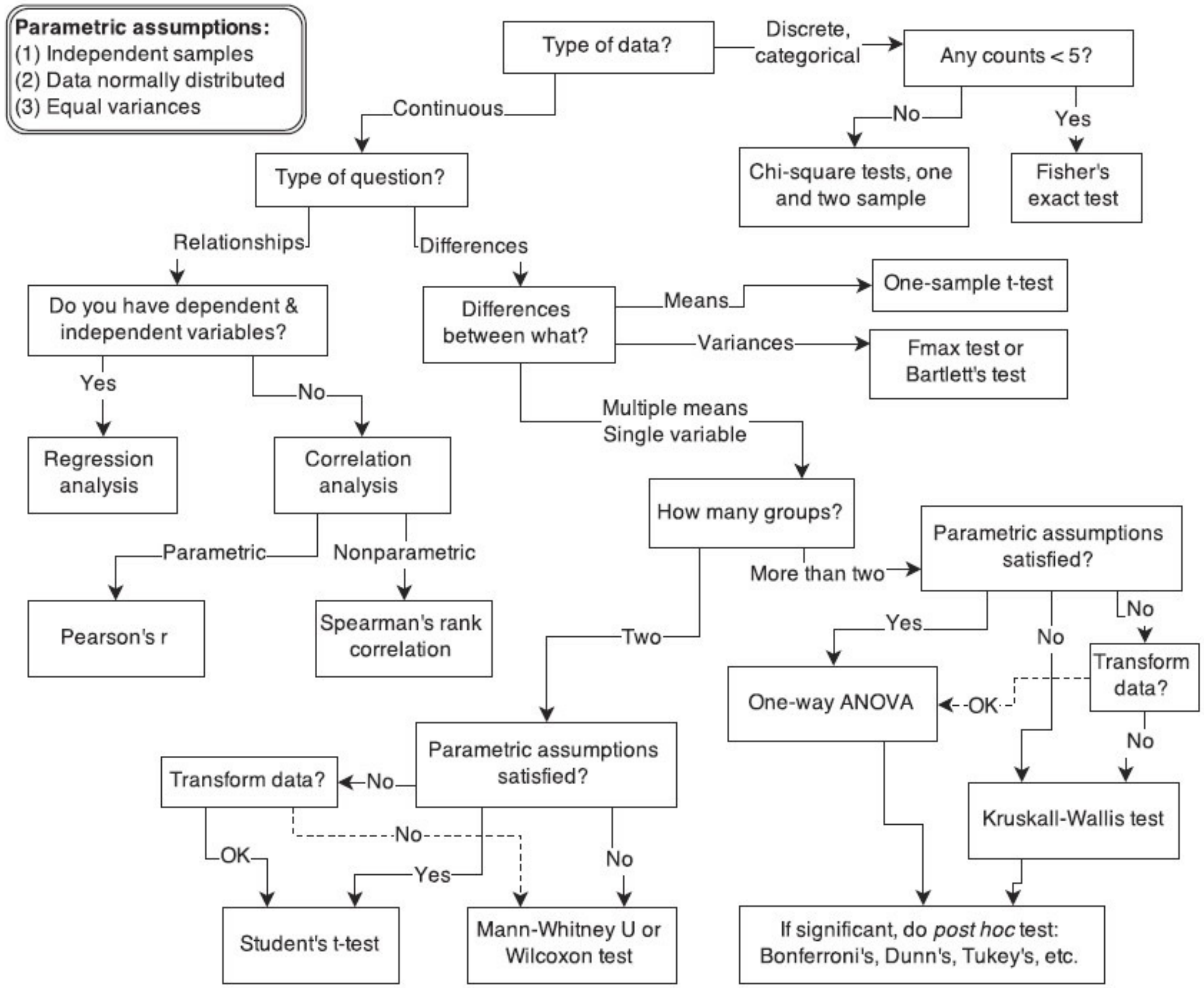
Why Bayesian?

- Intuitive, natural way to think about the world
- Philosophical reasons

```
Warning messages:
1: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model failed to converge with max|grad| = 0.131568 (tol = 0.001, component 1)
2: In checkConv(attr(opt, "derivs"), opt$par, ctrl = control$checkConv,  :
  Model is nearly unidentifiable: very large eigenvalue
```

Why Bayesian?

- Intuitive, natural way to think about the world
- Philosophical reasons
- Practical and computational reasons
  - Flexible and consistent (no decision trees)
  - Better accuracy in problematic datasets
  - p-value less relevant
  - Incorporate prior knowledge



McElreath (2020) Figure 1.1

What is Bayesian?

Distribution of possible values for the parameters ← Observed data + prior expectation

Prior: initial degree of belief in hypothesis

Likelihood: the probability of the data given the hypothesis

- $\mathrm{P(Hypothesis|Data)} = \dfrac{\mathrm{P(Hypothesis)P(Data|Hypothesis)}}{\mathrm{P(Data)}}$

Posterior: updated degree of belief in hypothesis, after seeing data

1. Start with Prior beliefs.
2. Collect Data.
3. Combine to Posterior beliefs.

How to do Bayesian in R

`library(brms)`   `b~r+(m|s)`

`library(rstan)` <- stan code basis

`library(bayesplot)` <- using `ggplot2`

`library(tidyverse)` <- for tidy dataset

Example: voicing in glottal fricatives

Research question: how does **voiceless** glottal fricative [h] differ in voicing intensity from the **voiced** glottal fricative [ɦ]

335 tokens of [h] in 91 languages and 35 tokens of [ɦ] in 15 languages at the word-initial position

```
> head(df)
# A tibble: 6 × 7
# Groups:   language, Filename, dur, lartype, family, area [6]
  Filename        language   dur   lartype family        area          norm.soe
  <chr>           <chr>      <dbl> <chr>   <chr>         <chr>            <dbl>
1 20-to_go        A'ingae    168.  h       A'ingae       South America   0.0988
2 55-to_go        A'ingae    135.  h       A'ingae       South America   0.172
3 014-hang        Afrikaans   86.4 H       Indo-European Africa          0.329
4 032-hands       Afrikaans  156.  H       Indo-European Africa          0.207
5 148-horizontal  Afrikaans   55.9 H       Indo-European Africa          0.499
6 158-historian   Afrikaans   20.6 H       Indo-European Africa          0.451
```

lme4

```
mod <- lmer(norm.soe~dur+lartype
+(1|family)+(1|area),
data=df)

# running time: 2 s
```

brms

```
mod <- brm(norm.soe~dur+lartype
+(1|family)+(1|area),
data=df,
family = gaussian(),
prior = c(
prior(normal(0.2, 1), class =
Intercept),
prior(normal(0.2, .5), class = b, coef
= lartypeH),
prior(cauchy(0, 0.1), class = sigma)),
iter = 10000, chains = 4,
warmup = 2000)

# running time: 347s (warm-up +
sampling)
```

Model output (1)

```
> summary(mod)
```

Random effects

Rhat < 1.05

```
Group-Level Effects:
~area (Number of levels: 6)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.06      0.04     0.01     0.16 1.00     6161     7157

~family (Number of levels: 24)
              Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)     0.06      0.02     0.03     0.11 1.00     8528    11324
```

ESS: Effective sample size for mean & median, and for 5% and 95% quantiles
Minimum: 10% of all the used iterations (3200)

Model output (2)

```
> summary(mod)
```

Questions: Are these good Rhats? ESSs?

Population-Level Effects: Fixed effects

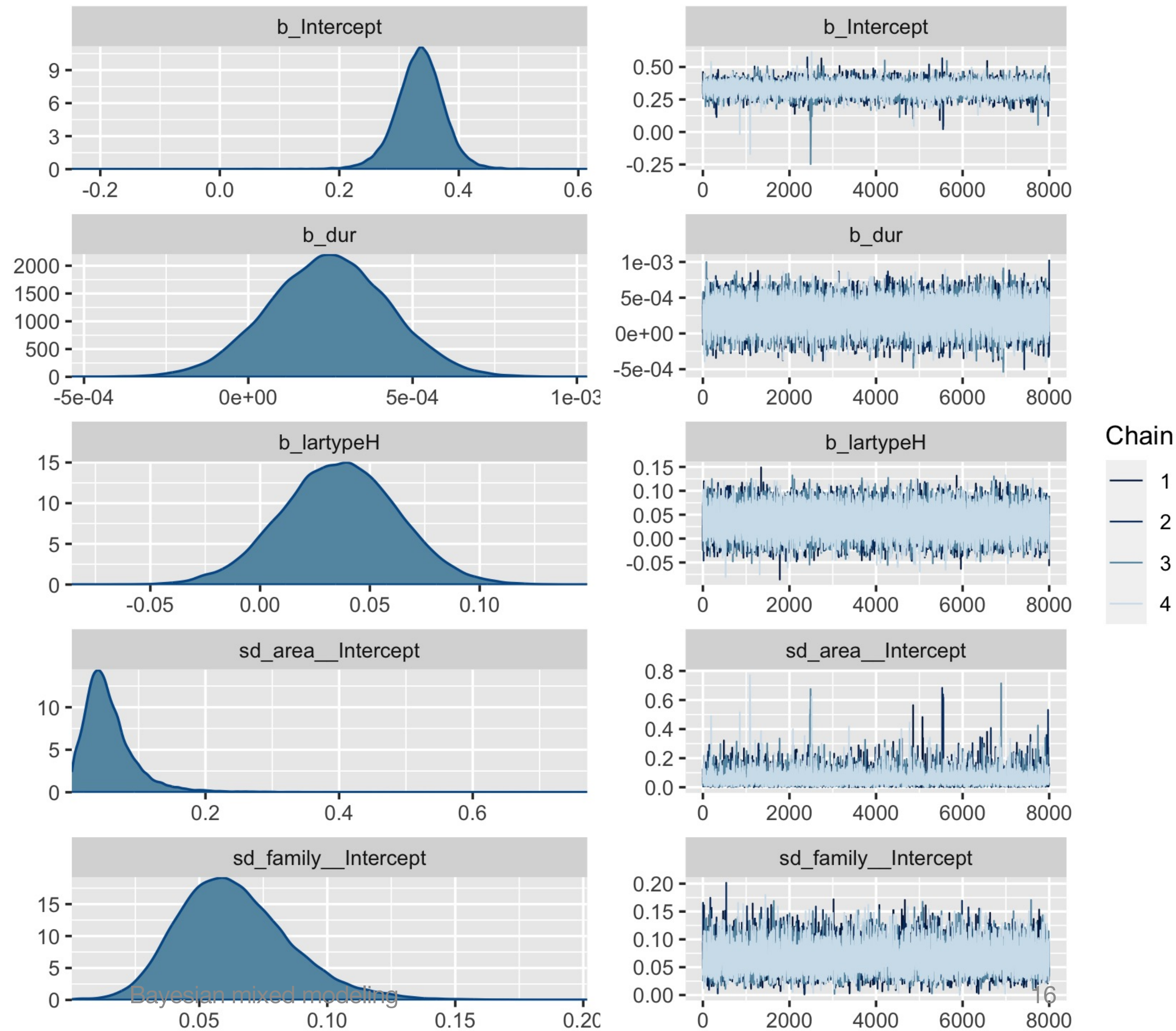
	Estimate (Mean)	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	**0.33**	**0.04**	**0.25**	**0.41**	1.00	11056	12765
dur	0.00	0.00	-0.00	0.00	1.00	32645	20979
lartypeH	**0.04**	**0.03**	**-0.01**	**0.09**	1.00	28550	24317

Family Specific Parameters: Residual variance

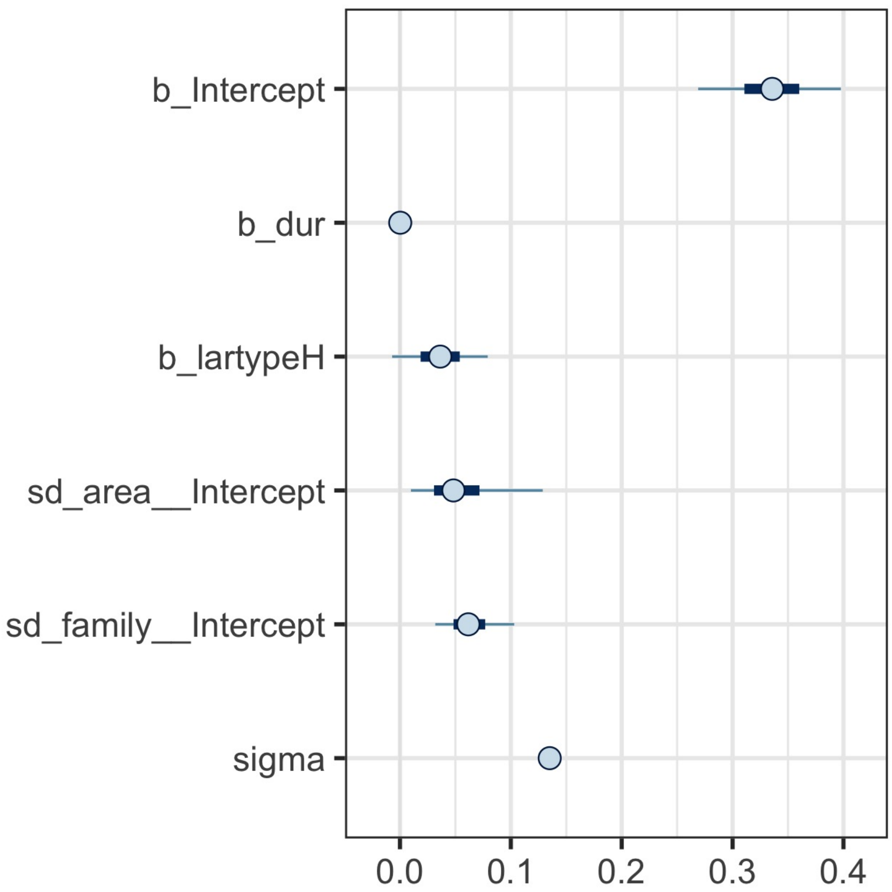
	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
sigma	0.14	0.01	0.13	0.15	1.00	24691	23857

Model diagnostics

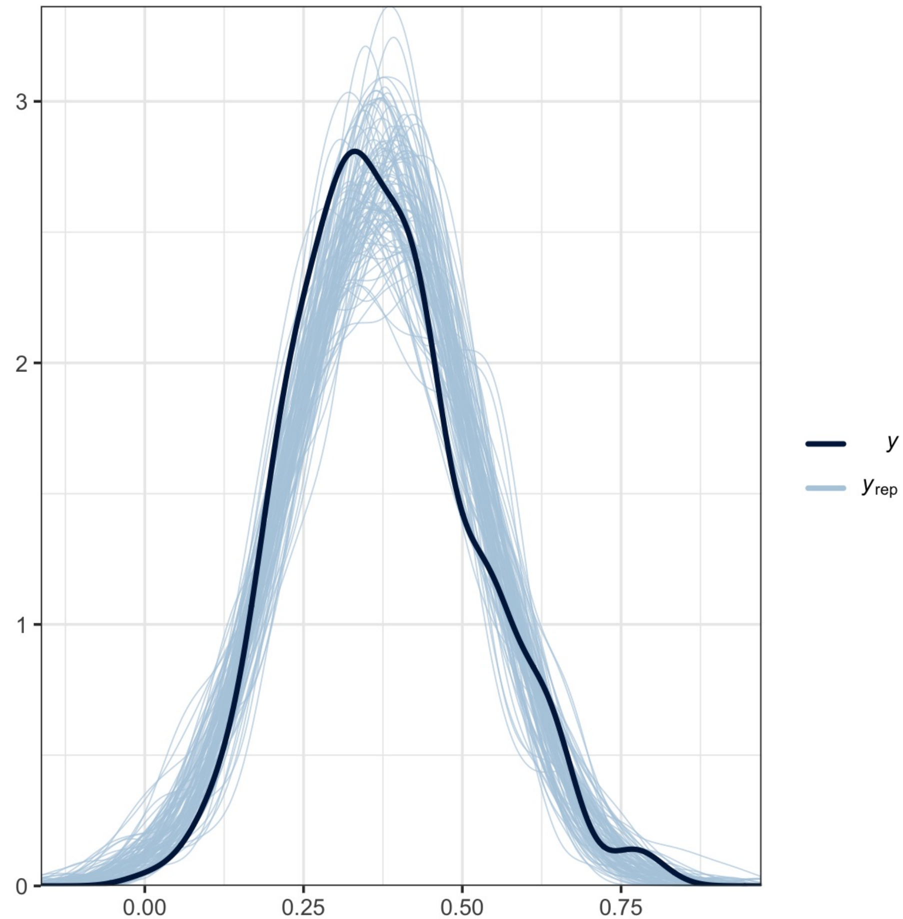
```
> plot(mod)
```



`> mcmc_plot(mod)`



`> pp_check(mod, ndraws=100)`



Indices of effect

- Posterior descriptive stats
  - Mean, median, SD, CI of the posterior distribution
- Effect existence
  - Probability of direction (sign of effect)
- Effect significance
  - Bayes factor
    - Is null or the alternative more likely?
  - Region of practical equivalence (ROPE)
    - the probability of being outside a specific range that can be considered as "practically no effect"

Bayes factor

- The degree by which the mass of the posterior has shifted away from or closer to the null value, relative to the prior distribution

→ the null hypothesis has become less or more likely given the observed data

```
bayes_factor(M1, M2)
```

- summary of evidence for M1 against M2 provided by data

Model comparison: Bayes factor

```
> mod <- brm(
norm.soe~dur+lartype+(1|family
)+(1|area),
data=df, prior = priors
save_pars = save_pars(all =
T),...)

> mod0 <- brm(
norm.soe~dur+(1|family)+(1|are
a),
data=df, prior = priors0
save_pars = save_pars(all =
T),...)
```

```
> bayes_factor(
    x1 = mod,
    x2 = mod0,
    repetitions = 10)
```

```
Estimated Bayes factor (based on
medians of log marginal
likelihood estimates)
in favor of mod over mod0:
0.12752
Range of estimates: 0.12119 to
0.13239
Interquartile range: 0.00331
```

TABLE 15.1: The Bayes factor scale as proposed by Jeffreys (1939). This scale should not be regarded as a hard and fast rule.

BF_{12}	**Interpretation**
> 100	Extreme evidence for $\mathcal{M}_1$.
$30 - 100$	Very strong evidence for $\mathcal{M}_1$.
$10 - 30$	Strong evidence for $\mathcal{M}_1$.
$3 - 10$	Moderate evidence for $\mathcal{M}_1$.
$1 - 3$	Anecdotal evidence for $\mathcal{M}_1$.
1	No evidence.
$\frac{1}{1} - \frac{1}{3}$	Anecdotal evidence for $\mathcal{M}_2$.
$\frac{1}{3} - \frac{1}{10}$	Moderate evidence for $\mathcal{M}_2$.
$\frac{1}{10} - \frac{1}{30}$	Strong evidence for $\mathcal{M}_2$.
$\frac{1}{30} - \frac{1}{100}$	Very strong evidence for $\mathcal{M}_2$.
$< \frac{1}{100}$	Extreme evidence for $\mathcal{M}_2$.

Nicenboim et al. (2023)

Reporting: Bayesian analysis methods

Describe the model like any mixed effects models:

“All Bayesian models were created in Stan accessed with *brms* package in R (Bürkner, 2017). To improve convergence and against overfitting, we specified weakly informative priors.”

Report fitted values shown in `mcmc_plot()`: “the **mean** of posterior distribution and **95% Cr**[edible] **I**[nterval].”

If doing null hypothesis testing, you could also report the **Bayes factor**.

Reporting: results

The posterior distribution of voicing intensity in voiced [ɦ] at the word-initial position has a **mean** of 0.04 with **95% CrI** [-0.01, 0.09].

Optional:

There is moderate evidence for the null model (**BF** = 0.13), thus in favor of the hypothesis that glottal fricatives do not differ in voicing intensity in word-initial positions.

Wrap-up: what we learned

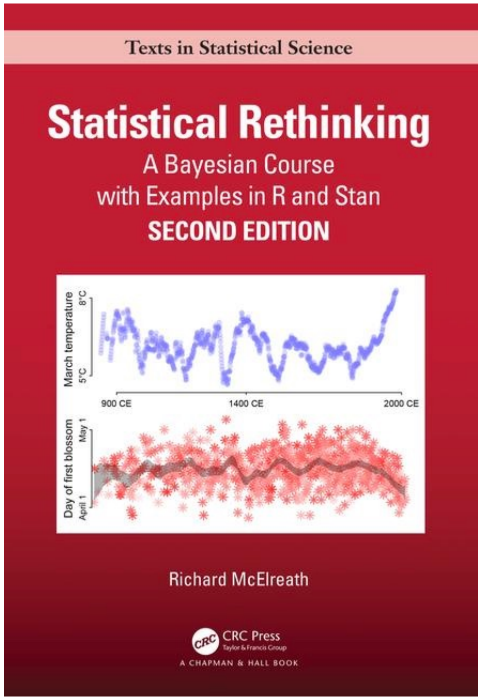
- Bayesian concepts
- Implementation using **`brms`** and other packages
- Inference and reporting

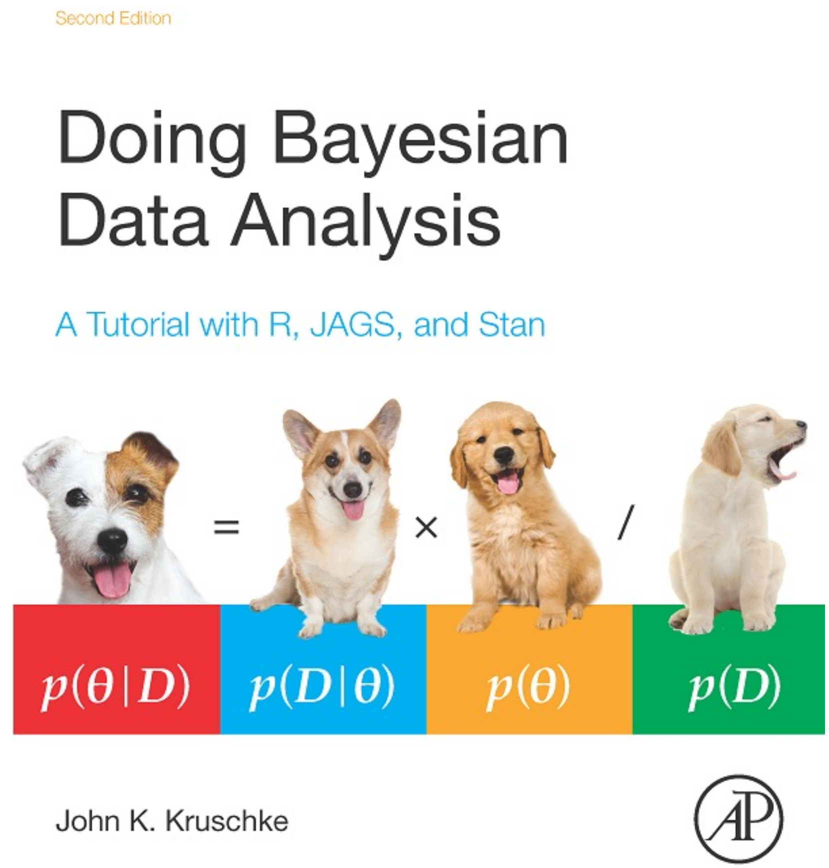
Wrap-up: what to learn next

- What is a prior and why and how should I use one
- Does my model make sense? Posterior distributions, predictions
- More complex models

Resources

- Bürkner, P. C. (2017). Advanced Bayesian multilevel modeling with the R package brms. *arXiv preprint arXiv:1705.11123*.
- Makowski, D., Ben-Shachar, M. S., Chen, S. A., & Lüdecke, D. (2019). Indices of effect existence and significance in the Bayesian framework. *Frontiers in psychology*, 10, 2767.
- Nicenboim, B., Schad, D., & Vasishth, S. (2023). *An introduction to Bayesian data analysis for cognitive science*. Under contract with Chapman and Hall/CRC statistics in the social and behavioral sciences series.
- Vasishth, S., Nicenboim, B., Beckman, M. E., Li, F., & Kong, E. J. (2018). Bayesian data analysis in the phonetic sciences: A tutorial introduction. *Journal of phonetics*, 71, 147-161.





With brms exps: https://bookdown.org/content/4857/

Discussion

- What kinds of data do you encounter in your research?
- What are your ‘go-to’ models?
- How often do you think you have an intuition about the possibility of your hypothesis for your research question (even if they turn out to be wrong)?

Questions?

All the materials can be accessed:
https://github.com/yaqianhuang/Stats-workshops