Bayesian Analysis with BRMS

Yiwen Zhang

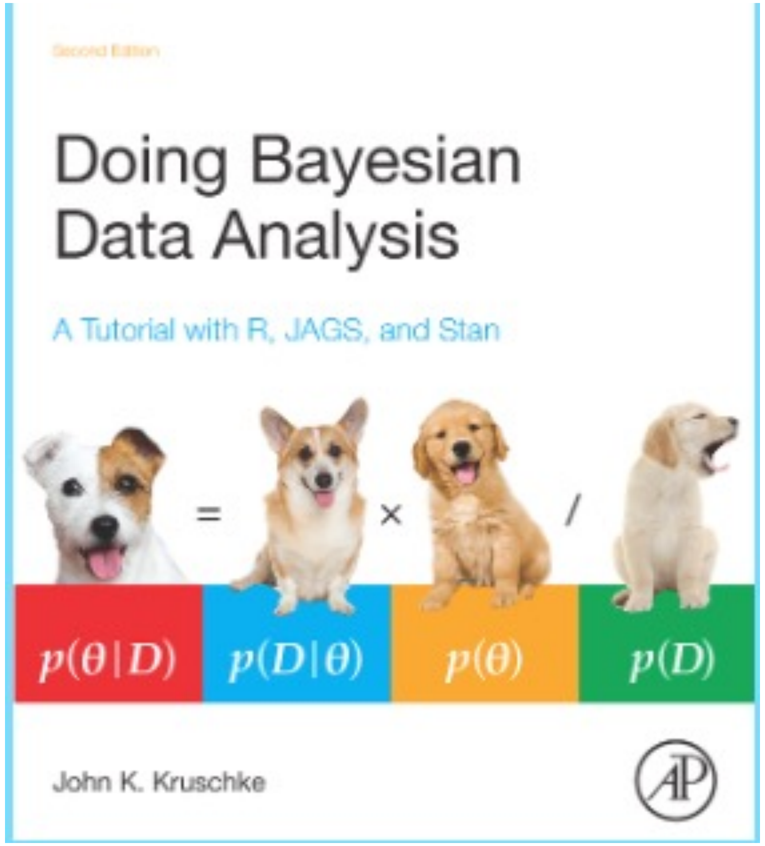
CAMEL Talk 9/21/2023

In this talk…

- Basics Behind Bayesian Analysis
- Practical Benefits I’ve discovered
- BRMS package and R demo

Resources

- BIOST 2063 Bayesian Data Science
- The book: Doing Bayesian Data Analysis, Second Edition: A Tutorial with R, JAGS, and Stan. By John Kruschke
- BayesFactor package documentation
- BRMS package documentation



Bayesian Analysis

- Prior: our prior belief about the parameters of interest P(θ)
- Likelihood: P(observed data | θ)
- Posterior: P (θ| observed data)

Conduct inference based on Posterior distribution

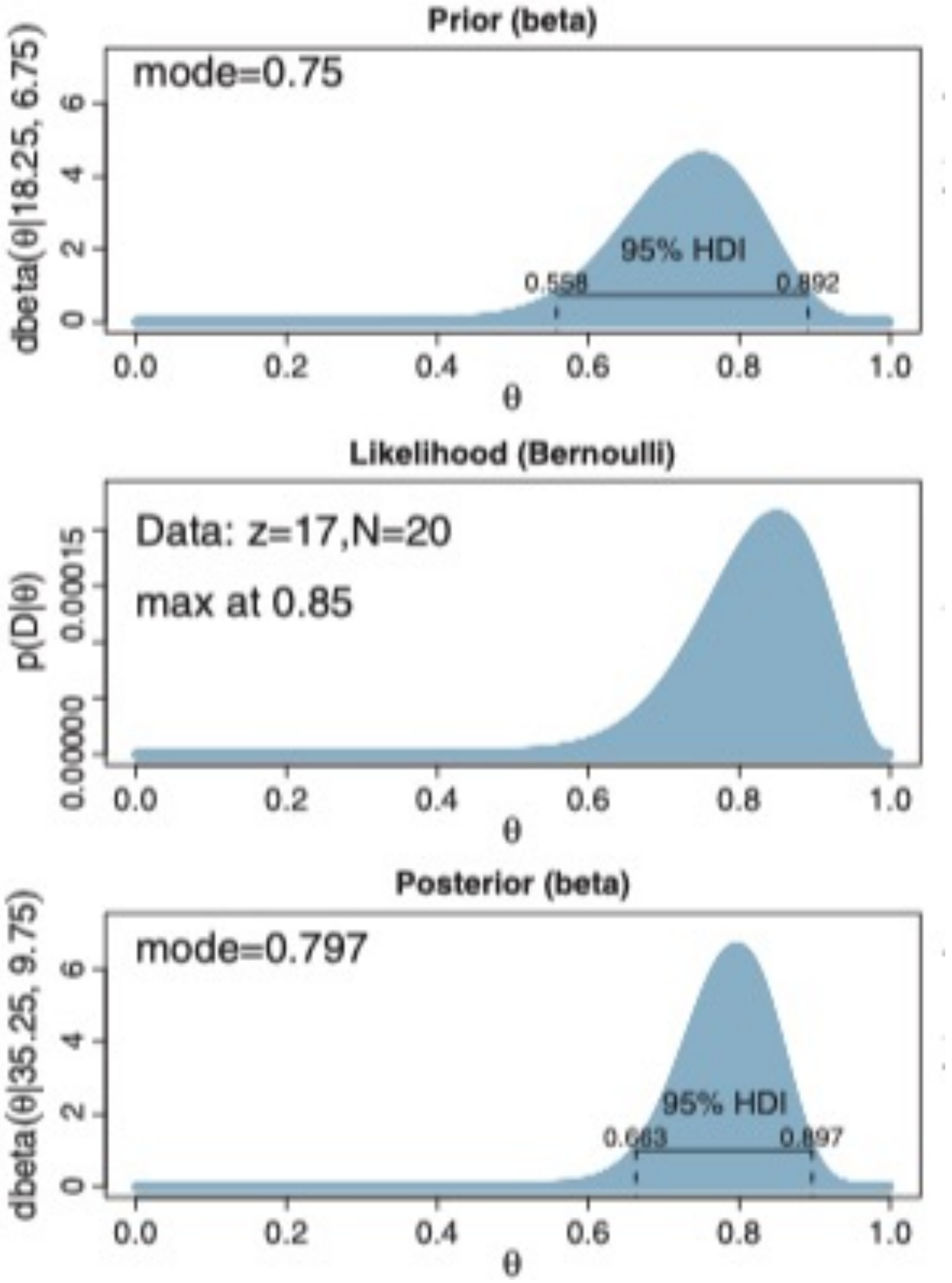
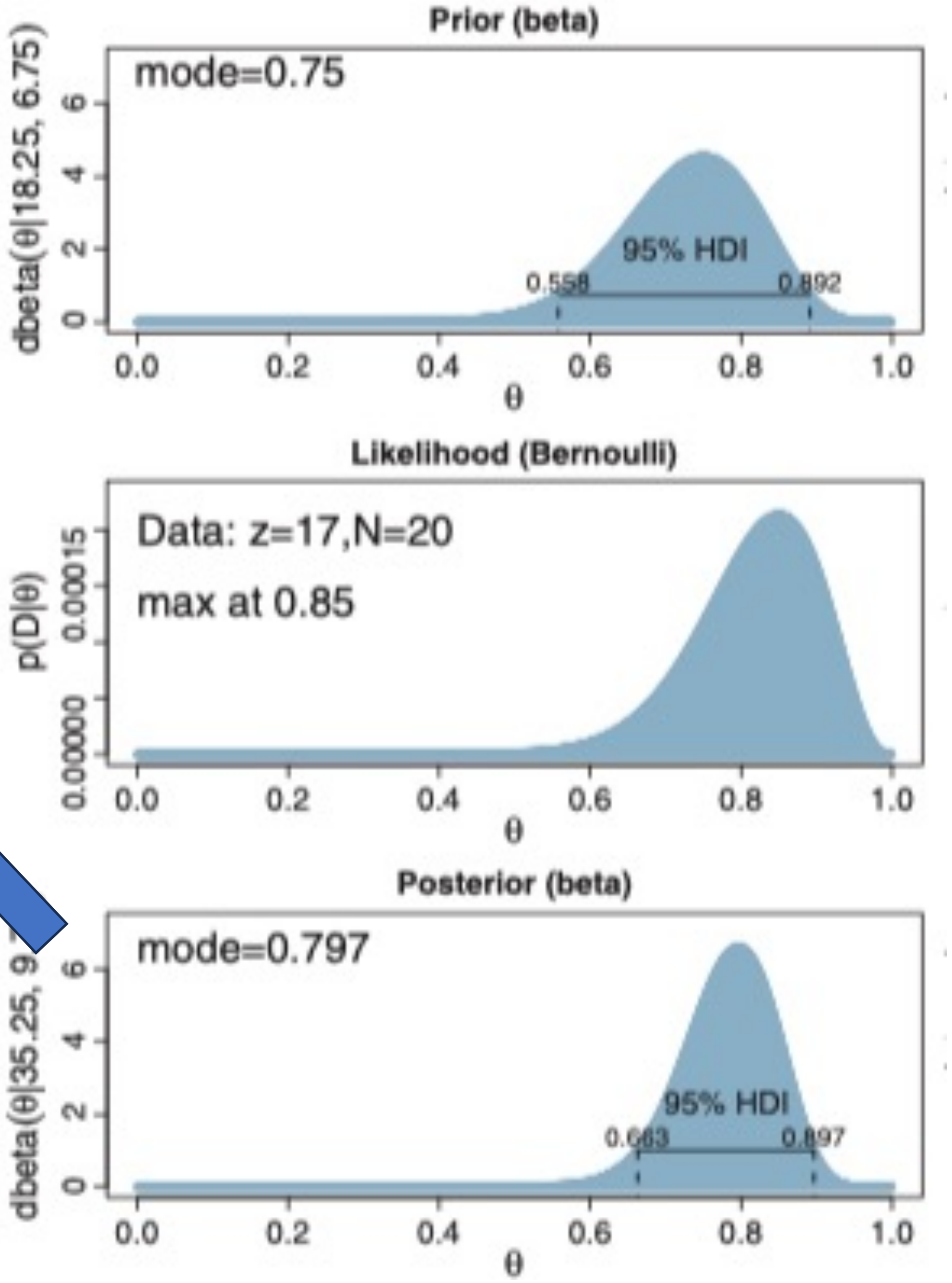


Figure 6.4 from Kruschke (2015)

Summarize Posterior

- Point estimate: mean or median of the posterior distribution
- 95% credible interval: there is 95% probability that θ is between 0.663 and 0.897



Compared to 95% confidence interval: we will be 95% confident that the true value is within … the interval.

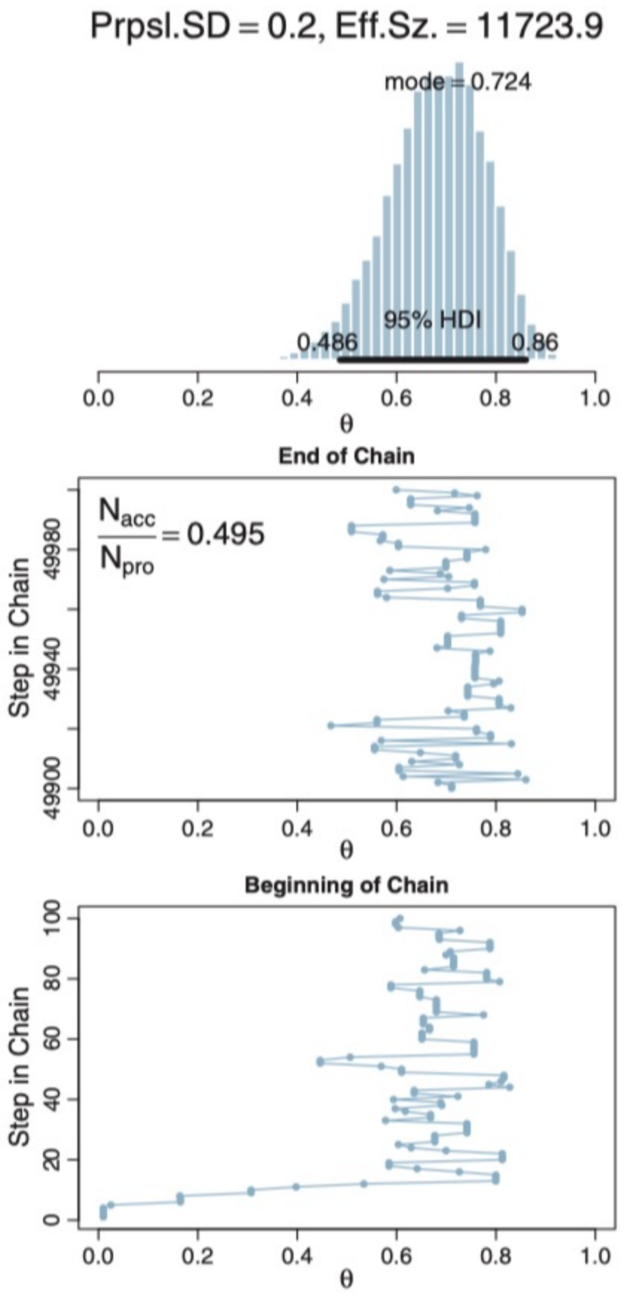
MCMC Sampling to fit model

Markov chain Monte Carlo (MCMC) algorithm

- Random start point of θ
- Each iteration
  - Proposal a move from the current point to a new point
    - The proposed move is generated from a proposal distribution (e.g. normal distribution around the current point)
  - Deciding rule: accept or not accept the proposal
    - This rule use the ratio of relative densities of the proposed to the current value
    - The density is the product of likelihood and prior of the value
- Repeat the above steps until we have a sufficient sample

Random walk

Replaced by more efficient sampling way



Bayes Factor

Likelihood of observed data under H1

$$\frac{P\,(\text{data}\,|\,H1)}{P\,(\text{data}\,|\,H0)}$$

Likelihood of observed data under H0

Bayes Factor Scale (Kass and Raftery, 1995):

BF < 1: Negative (support H0)

1 < BF < 3: Barely worth mentioning

3 < BF < 20: Positive

20 < BF < 150: Strong

BF > 150: Very strong

BF > 1: it favors H1

BF < 1: it favors H0

Equation 10.5 from Kruschke (2015)

Some practical benefits of Bayesian Analysis

- Providing a more direct/intuitive interpretation
  - P values and confidence interval are often mis-interpreted or misunderstood.
  - Bayes factors offer a direct comparison between two hypotheses.

Some practical benefits of Bayesian Analysis

- Providing a more direct/intuitive interpretation

- Better for achieving convergence of linear mixed effect models
  - “One alternative for dealing with small sample sizes and overparameterized/nonconverging models is to switch to Bayesian data analyses.” (Brauer & Curtin, 2017)
    - Setting priors, MCMC sampling…
  - BRMS package provides solutions for convergency problems: Runtime warnings and convergence problems (https://mc-stan.org/misc/warnings.html)

Some practical benefits of Bayesian Analysis

- Providing a more direct/intuitive interpretation

- Achieving convergence of linear mixed effect models

- Don't have to worry about multiple testing corrections.
  - Bayesian inference is based on the posteriors
  - "there is just one posterior distribution" – multiple comparison is just looking at the posterior from different perspective (Kruschke, 2015)

Brms package

- Bayesian Regression Models using ‘Stan’
  - STAN (a probabilistic programming language) to fit models
  - Hamiltonian Monte-Carlo (HMC) sampler and No-U-Turn Sampler (avoid random walk)
- Implement Bayesian models but using lme4-like formula syntax

random effects

Fixed effects

```
m1 <- brms::brm(formula = accuracy ~ 1 + condition_type + num_of_missed_z +
                  (1+ condition_type|participant_id),
                data=test,
                family=bernoulli(link = "logit"),
                cores = 4,
                chains = 4,
                iter = 4000,
                warmup = 1000,
                seed = 123)
```

```r
m1 <- brms::brm(formula = accuracy ~ 1 + condition_type + num_of_missed_z +
                  (1+ condition_type|participant_id),
                data=test,
                family=bernoulli(link = "logit"),
                cores = 4,
                chains = 4,
                iter = 4000,
                warmup = 1000,
                seed = 123)
```

We want to run 4 Markov chains and use 4 cores of my CPU, so the 4 chains can run simultaneously

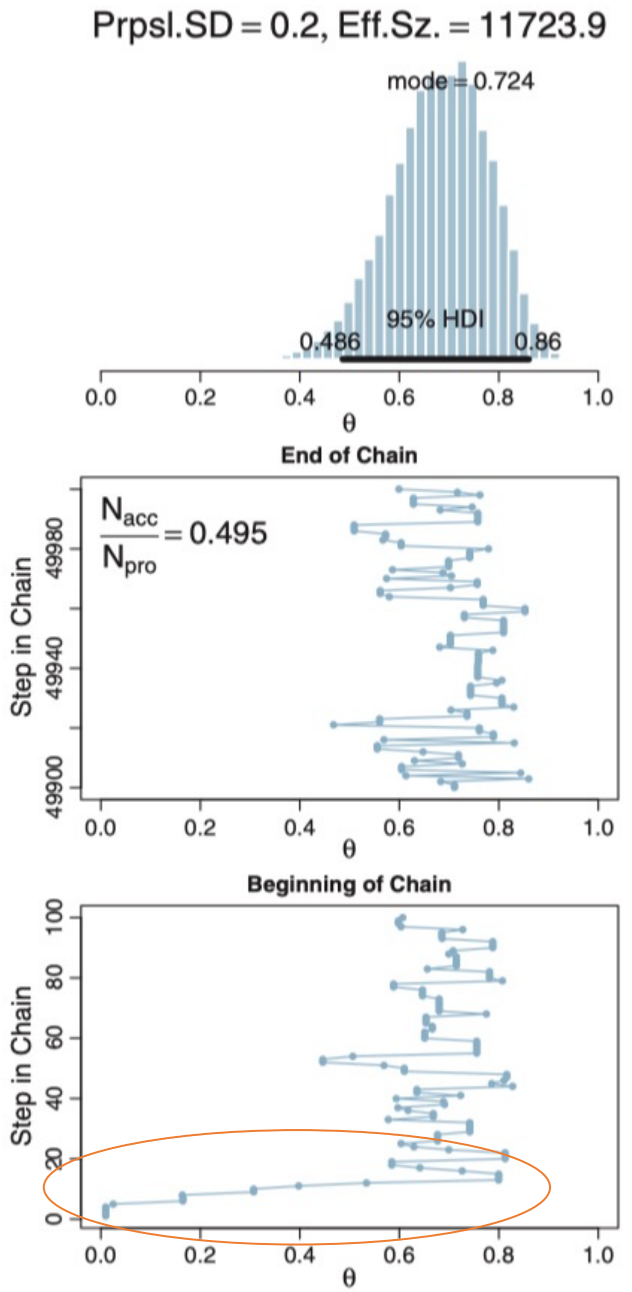
```r
m1 <- brms::brm(formula = accuracy ~ 1 + condition_type + num_of_missed_z +
                 (1+ condition_type|participant_id),
                data=test,
                family=bernoulli(link = "logit"),
                cores = 4,
                chains = 4,
                iter = 4000,
                warmup = 1000,
                seed = 123)
```

Number of total iterations per chain

```
m1 <- brms::brm(formula = accuracy ~ 1 + condition_type +
                (1+ condition_type|participant_id),
                data=test,
                family=bernoulli(link = "logit"),
                cores = 4,
                chains = 4,
                iter = 4000,
                warmup = 1000,
                seed = 123)
```

the number of iterations used for stepsize adaptation

We disregard the first several iterations from a chain



```r
m1 <- brms::brm(formula = accuracy ~ 1 + condition_type + num_of_missed_z +
                  (1+ condition_type|participant_id),
                data=test,
                family=bernoulli(link = "logit"),
                cores = 4,
                chains = 4,
                iter = 4000,
                warmup = 1000,
                seed = 123)
```

Set a seed to make results reproducible

Logistic regression

```
 Family: bernoulli
  Links: mu = logit
Formula: accuracy ~ 1 + condition_type + num_of_missed_z + (1 + condition_type | participant_id)
   Data: test (Number of observations: 13944)
  Draws: 4 chains, each with iter = 4000; warmup = 1000; thin = 1;
         total post-warmup draws = 12000

Group-Level Effects:
~participant_id (Number of levels: 296)
                                   Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)                          2.28      0.16     2.00     2.60 1.00     1891     3421
sd(condition_typeD3)                   2.05      0.14     1.80     2.35 1.00     2221     3903
sd(condition_typeD6)                   1.91      0.13     1.67     2.19 1.00     2110     4018
cor(Intercept,condition_typeD3)       -0.60      0.07    -0.71    -0.46 1.00     2575     4627
cor(Intercept,condition_typeD6)       -0.87      0.03    -0.91    -0.81 1.00     2447     4921
cor(condition_typeD3,condition_typeD6) 0.66      0.05     0.54     0.75 1.00     2591     4640

Population-Level Effects:
                  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept             1.35      0.16     1.05     1.67 1.00     1024     2319
condition_typeD3     -0.75      0.16    -1.07    -0.44 1.00     1669     3097
condition_typeD6     -1.25      0.14    -1.54    -0.97 1.00     1319     3237
num_of_missed_z      -0.12      0.05    -0.22    -0.02 1.00     2224     4158
```

Formula and the dataset we used

```
 Family: bernoulli
  Links: mu = logit
Formula: accuracy ~ 1 + condition_type + num_of_missed_z + (1 + condition_type | participant_id)
   Data: test (Number of observations: 13944)
  Draws: 4 chains, each with iter = 4000; warmup = 1000; thin = 1;
         total post-warmup draws = 12000

Group-Level Effects:
~participant_id (Number of levels: 296)
                                     Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)                            2.28      0.16     2.00     2.60 1.00     1891     3421
sd(condition_typeD3)                     2.05      0.14     1.80     2.35 1.00     2221     3903
sd(condition_typeD6)                     1.91      0.13     1.67     2.19 1.00     2110     4018
cor(Intercept,condition_typeD3)         -0.60      0.07    -0.71    -0.46 1.00     2575     4627
cor(Intercept,condition_typeD6)         -0.87      0.03    -0.91    -0.81 1.00     2447     4921
cor(condition_typeD3,condition_typeD6)   0.66      0.05     0.54     0.75 1.00     2591     4640

Population-Level Effects:
                  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept             1.35      0.16     1.05     1.67 1.00     1024     2319
condition_typeD3     -0.75      0.16    -1.07    -0.44 1.00     1669     3097
condition_typeD6     -1.25      0.14    -1.54    -0.97 1.00     1319     3237
num_of_missed_z      -0.12      0.05    -0.22    -0.02 1.00     2224     4158
```

```
 Family: bernoulli
  Links: mu = logit
Formula: accuracy ~ 1 + condition_type + num_of_missed_z + (1 + condition_type | participant_id)
   Data: test (Number of observations: 13944)
  Draws: 4 chains, each with iter = 4000; warmup = 1000; thin = 1;
         total post-warmup draws = 12000
```

=(4000-1000)*4

```
Group-Level Effects:
~participant_id (Number of levels: 296)
                                       Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)                              2.28      0.16     2.00     2.60 1.00     1891     3421
sd(condition_typeD3)                       2.05      0.14     1.80     2.35 1.00     2221     3903
sd(condition_typeD6)                       1.91      0.13     1.67     2.19 1.00     2110     4018
cor(Intercept,condition_typeD3)           -0.60      0.07    -0.71    -0.46 1.00     2575     4627
cor(Intercept,condition_typeD6)           -0.87      0.03    -0.91    -0.81 1.00     2447     4921
cor(condition_typeD3,condition_typeD6)     0.66      0.05     0.54     0.75 1.00     2591     4640

Population-Level Effects:
                 Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept            1.35      0.16     1.05     1.67 1.00     1024     2319
condition_typeD3    -0.75      0.16    -1.07    -0.44 1.00     1669     3097
condition_typeD6    -1.25      0.14    -1.54    -0.97 1.00     1319     3237
num_of_missed_z     -0.12      0.05    -0.22    -0.02 1.00     2224     4158
```

```
 Family: bernoulli
  Links: mu = logit
Formula: accuracy ~ 1 + condition_type + num_of_missed_z + (1 + condition_type | participant_id)
   Data: test (Number of observations: 13944)
  Draws: 4 chains, each with iter = 4000; warmup = 1000; thin = 1;
         total post-warmup draws = 12000
```

Random effects

```
Group-Level Effects:
~participant_id (Number of levels: 296)
                                        Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)                               2.28      0.16     2.00     2.60 1.00     1891     3421
sd(condition_typeD3)                        2.05      0.14     1.80     2.35 1.00     2221     3903
sd(condition_typeD6)                        1.91      0.13     1.67     2.19 1.00     2110     4018
cor(Intercept,condition_typeD3)            -0.60      0.07    -0.71    -0.46 1.00     2575     4627
cor(Intercept,condition_typeD6)            -0.87      0.03    -0.91    -0.81 1.00     2447     4921
cor(condition_typeD3,condition_typeD6)      0.66      0.05     0.54     0.75 1.00     2591     4640

Population-Level Effects:
                  Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept             1.35      0.16     1.05     1.67 1.00     1024     2319
condition_typeD3     -0.75      0.16    -1.07    -0.44 1.00     1669     3097
condition_typeD6     -1.25      0.14    -1.54    -0.97 1.00     1319     3237
num_of_missed_z      -0.12      0.05    -0.22    -0.02 1.00     2224     4158
```

```
 Family: bernoulli
  Links: mu = logit
Formula: accuracy ~ 1 + condition_type + num_of_missed_z + (1 + condition_type | participant_id)
   Data: test (Number of observations: 13944)
  Draws: 4 chains, each with iter = 4000; warmup = 1000; thin = 1;
         total post-warmup draws = 12000

Group-Level Effects:
~participant_id (Number of levels: 296)
                                   Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
sd(Intercept)                          2.28      0.16     2.00     2.60 1.00     1891     3421
sd(condition_typeD3)                   2.05      0.14     1.80     2.35 1.00     2221     3903
sd(condition_typeD6)                   1.91      0.13     1.67     2.19 1.00     2110     4018
cor(Intercept,condition_typeD3)       -0.60      0.07    -0.71    -0.46 1.00     2575     4627
cor(Intercept,condition_typeD6)       -0.87      0.03    -0.91    -0.81 1.00     2447     4921
cor(condition_typeD3,condition_typeD6) 0.66      0.05     0.54     0.75 1.00     2591     4640

Population-Level Effects:                                              Fixed effects
                Estimate Est.Error l-95% CI u-95% CI Rhat Bulk_ESS Tail_ESS
Intercept           1.35      0.16     1.05     1.67 1.00     1024     2319
condition_typeD3   -0.75      0.16    -1.07    -0.44 1.00     1669     3097
condition_typeD6   -1.25      0.14    -1.54    -0.97 1.00     1319     3237
num_of_missed_z    -0.12      0.05    -0.22    -0.02 1.00     2224     4158
```

Formula: accuracy ~ 1 + condition_type + num_of_missed_z + (1 + condition_type | participant_id)

Categorical variable:
"EOD3", "D3", "D6"

Continuous Variable:
number of missed trials

Variable names

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	1.35	0.16	1.05	1.67	1.00	1024	2319
condition_typeD3	-0.75	0.16	-1.07	-0.44	1.00	1669	3097
condition_typeD6	-1.25	0.14	-1.54	-0.97	1.00	1319	3237
num_of_missed_z	-0.12	0.05	-0.22	-0.02	1.00	2224	4158

EOD3 is the reference group

D3, compared to EOD3
D6, compared to EOD3

```
Formula: accuracy ~ 1 + condition_type + num_of_missed_z + (1 + condition_type | participant_id)
```

Categorical variable:
"EOD3", "D3", "D6"

Continuous Variable:
number of missed trials

Posterior Means

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	1.35	0.16	1.05	1.67	1.00	1024	2319
condition_typeD3	-0.75	0.16	-1.07	-0.44	1.00	1669	3097
condition_typeD6	-1.25	0.14	-1.54	-0.97	1.00	1319	3237
num_of_missed_z	-0.12	0.05	-0.22	-0.02	1.00	2224	4158

```
Formula: accuracy ~ 1 + condition_type + num_of_missed_z + (1 + condition_type | participant_id)
```

Categorical variable:
”EOD3”, “D3”, “D6”

Continuous Variable:
number of missed trials

95% Credible intervals

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	1.35	0.16	1.05	1.67	1.00	1024	2319
condition_typeD3	-0.75	0.16	-1.07	-0.44	1.00	1669	3097
condition_typeD6	-1.25	0.14	-1.54	-0.97	1.00	1319	3237
num_of_missed_z	-0.12	0.05	-0.22	-0.02	1.00	2224	4158

95% Credible intervals that do not contains 0 = statistically credible

Rhat: compares the between- and within-chain estimates for model parameters (< 1.05)

Effective Sample Size (ESS): measure for sampling efficiency; the number of "uncorrelated pieces of information" per chains (>100)

Convergence Diagnostics

Population-Level Effects:

	Estimate	Est.Error	l-95% CI	u-95% CI	Rhat	Bulk_ESS	Tail_ESS
Intercept	1.35	0.16	1.05	1.67	1.00	1024	2319
condition_typeD3	-0.75	0.16	-1.07	-0.44	1.00	1669	3097
condition_typeD6	-1.25	0.14	-1.54	-0.97	1.00	1319	3237
num_of_missed_z	-0.12	0.05	-0.22	-0.02	1.00	2224	4158

Other References

- Assaf, A. G., & Tsionas, M. (2018). Bayes factors vs. P-values. *Tourism Management*, *67*, 17-31.
- Held, L., & Ott, M. (2018). On p-values and Bayes factors. *Annual Review of Statistics and Its Application*, *5*, 393-419.
- Hespanhol, L., Vallio, C. S., Costa, L. M., & Saragiotto, B. T. (2019). Understanding and interpreting confidence and credible intervals around effect estimates. *Brazilian journal of physical therapy*, *23*(4), 290–301. https://doi.org/10.1016/j.bjpt.2018.12.006
- Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the american statistical association*, *90*(430), 773-795.
- Lee M. D. and Wagenmakers, E.-J. (2014) Bayesian cognitive modeling: A practical course, Cambridge University Press.