Introduction to Bayesian statistics

Part 2 — Application

Jorge N. Tendeiro

Department of Psychometrics and Statistics
Faculty of Behavioral and Social Sciences
University of Groningen

j.n.tendeiro@rug.nl
www.jorgetendeiro.com
https://github.com/jorgetendeiro/GSMS-2020

Worked-out example

I will illustrate Bayesian analyses by means of an example.

General Bayesian workflow:

- Process data, descriptives.
- Build Bayesian models.
- Assess models through *prior* predictive checks.
- Fit the models to the data.
- MCMC diagnostics.
- Assess model fit through *posterior* predictive checks.
- Model comparison, summarize, report inferences.

Running example

Theory of mind in remitted bipolar disorder

Paper:

Espinós, U., Fernandéz-Abascal, E. G., & Ovejero, M. (2019). *Theory of mind in remitted bipolar disorder: Interpersonal accuracy in recognition of dynamic nonverbal signals*. PLoS ONE, 14(9), e0222112. doi: 10.1371/journal.pone.0222112.

Data:

`https://www.kaggle.com/mercheovejero/theory-of-mind-in-remitted-bipolar-disorder`

Study

Goal:
Examine interpersonal accuracy (IPA) in remitted patients with bipolar disorder (BD).

Groups:

- BD I
- BD II
- Unipolar depression (UD)
- Control

Dependent variable:
Number-correct score on the MiniPONS test to assess IPA.

Analysis:
ANCOVA model, with Age as covariate.

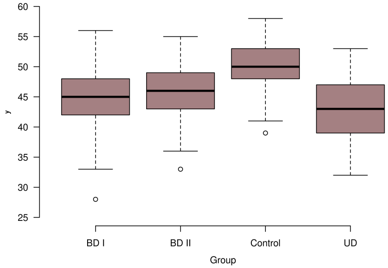
Descriptives

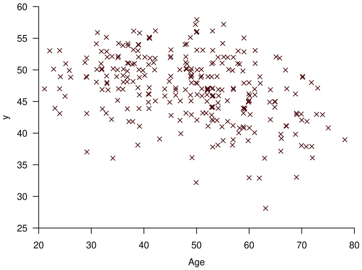
y

Group	n	mean	SD
BD I	70	45.1	4.9
BD II	49	45.7	4.7
Control	119	50.2	3.7
UD	39	42.7	5.0

Age

Group	n	mean	SD
BD I	70	44.5	11.5
BD II	49	49.9	11.5
Control	119	46.1	10.8
UD	39	62.9	9.7





Build Bayesian models

Model	*Formula*	*Obs.*
$\mathcal{M}_1$	$y \sim 1$	baseline
$\mathcal{M}_2$	$y \sim \text{Age}$	simple regression
$\mathcal{M}_3$	$y \sim \text{Group}$	ANOVA
$\mathcal{M}_4$	$y \sim \text{Group} + \text{Age}$	ANCOVA
$\mathcal{M}_5$	$y \sim \text{Group} + \text{Age} + \text{Group} \times \text{Age}$	Heterog. slopes ANCOVA
$\mathcal{M}_6$	$y \sim \text{Group} + \text{Age}$	constrained ANCOVA ($\mu_{\text{Control}} = \mu_{UD}$)

Espinós et al. (2019) focused on the ANCOVA model, $\mathcal{M}_4$.

Here we will also consider the other models and compare them.

Basic Stan code for all models

```
data {
  int<lower=0> N;   // sample size
  int<lower=0> K;   // number of predictors
  matrix[N, K+1] x; // predictor matrix (incl. intercept)
  vector[N] y;      // outcome variable
}

parameters {
  vector[K+1] beta;     // intercept + reg. coeffs.
  real<lower=0> sigma;  // SD residuals
}

model {
  beta  ~ normal(0, 10);            // Prior reg. coeffs.
  sigma ~ cauchy(0, 1);             // Prior sigma
  y     ~ normal(x * beta, sigma); // Likelihood
}
```

Assess models through
prior predictive checks

Prior predictive checks

Ask yourself:
What type of data can my model generate, *before I fit it to my own data*?

Answer:
Perform prior predictive checks.

What's that?
Look at data generated from your model (i.e., likelihood + priors).
$\longrightarrow$ Akin to test-driving a car before buying it.

What am I looking for?
A model that is flexible enough, but not too wild.

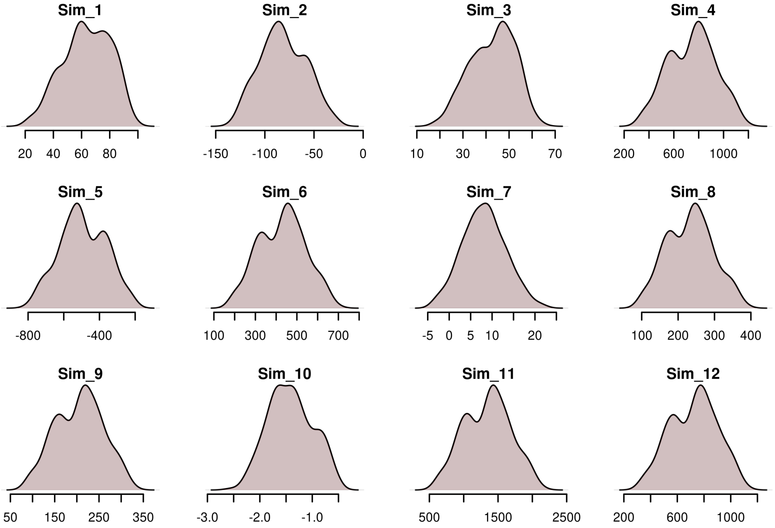
ANCOVA model

```
model {
  beta  ~ normal(0, 10);            // Prior reg. coeffs.
  sigma ~ cauchy(0, 1);             // Prior sigma
  y     ~ normal(x * beta, sigma); // Likelihood
}
```

To sample from the prior predictive distribution, do this a few times:

- Sample beta from its prior $\mathcal{N}(0, 10)$, say beta_i.
- Sample sigma from its prior $\text{Cauchy}(0, 1)$, say sigma_i.
- Sample data from the likelihood $\mathcal{N}(x * \text{beta}_i, \text{sigma}_i)$, say y_i.
- Plot y_i.

ANCOVA model



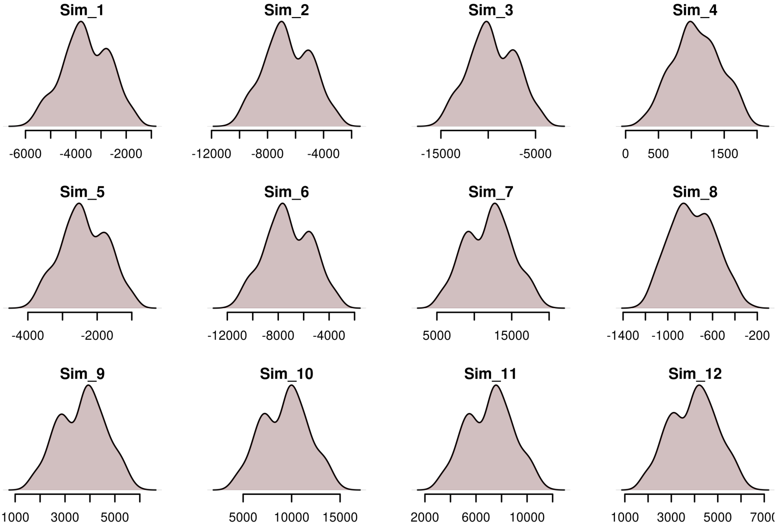
Flexible.

ANCOVA model

What if we broaden the prior on beta?

```
model {
  beta  ~ normal(0, 100);           // Prior reg. coeffs.
  sigma ~ cauchy(0, 1);             // Prior sigma
  y     ~ normal(x * beta, sigma); // Likelihood
}
```

ANCOVA model



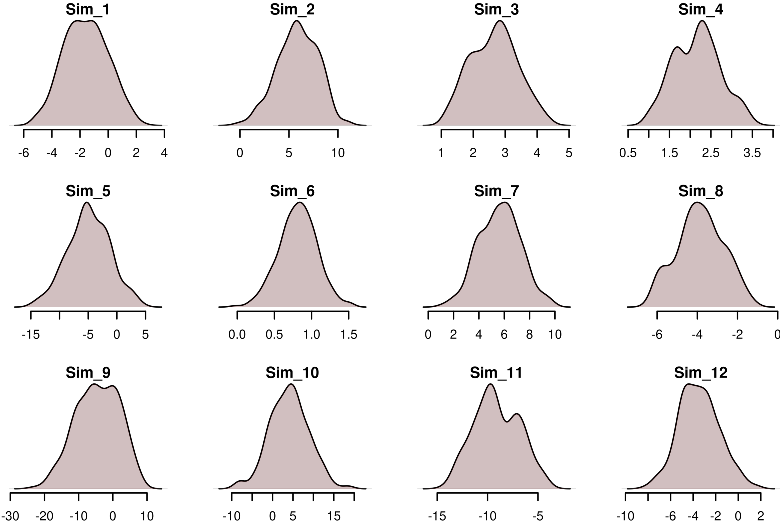
Yikes.

ANCOVA model

What if we shrink the prior on beta?

```
model {
  beta  ~ normal(0, .1);            // Prior reg. coeffs.
  sigma ~ cauchy(0, 1);             // Prior sigma
  y     ~ normal(x * beta, sigma); // Likelihood
}
```

ANCOVA model



Ups.

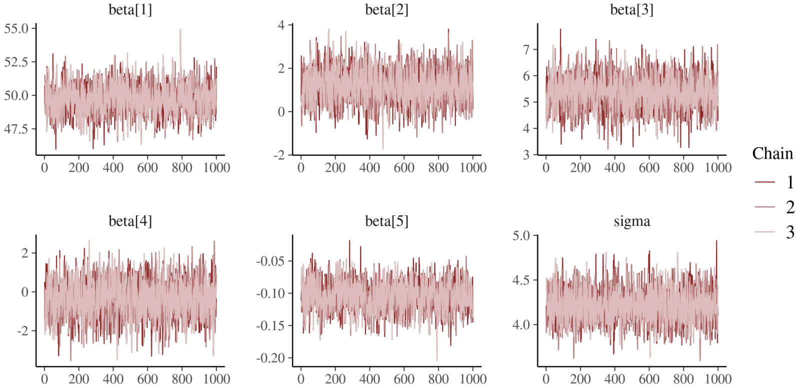
Fit the models to the data

I used R and rstan for the job.

All code is available at:
`https://github.com/jorgetendeiro/GSMS-2020`.

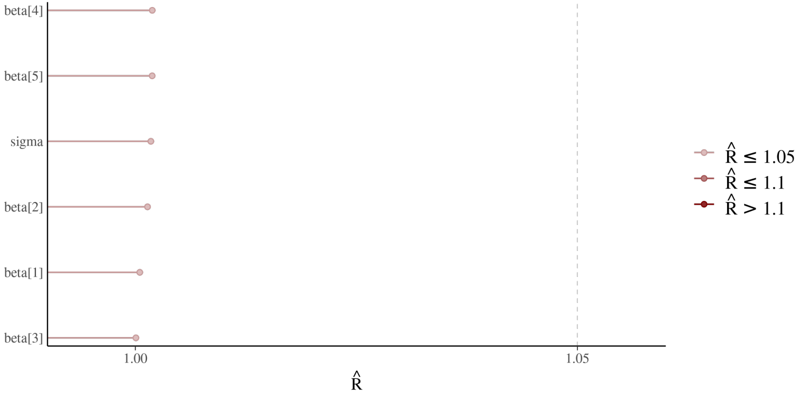
MCMC diagnostics

Trace plot



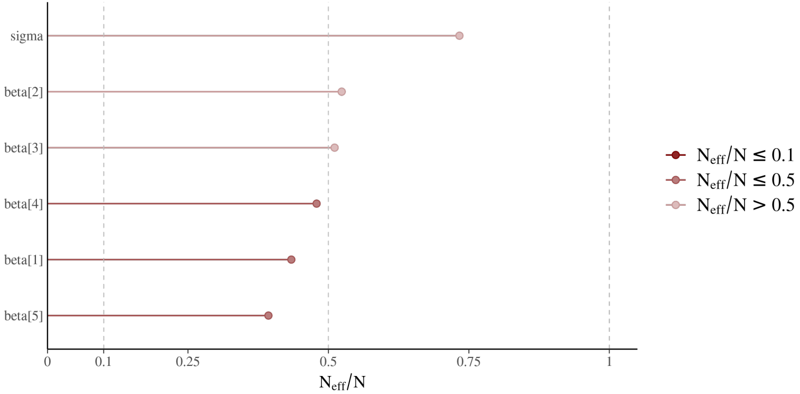
The chains mixed well.

R-hat



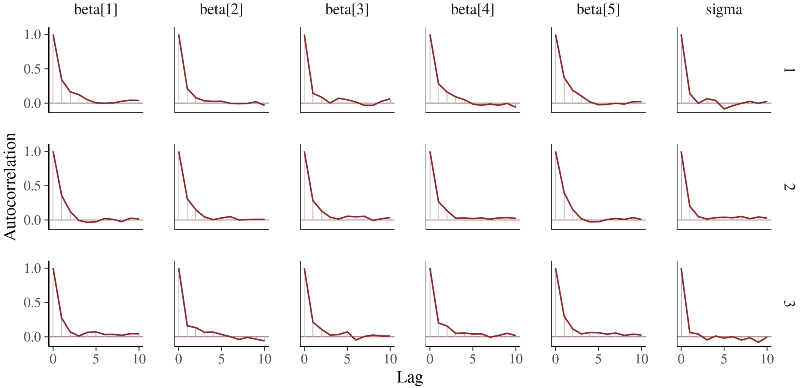
All below, say, 1.05. Good.

Effective sample size



All above, say, 0.1. Good.

Auto-correlation



It approaches 0 rather quickly. Nice.

Assess model fit through
posterior predictive checks

Posterior predictive checks

Ask yourself:
How likely is your fitted model of generating data *like* you collected?

Answer:
Perform posterior predictive checks.

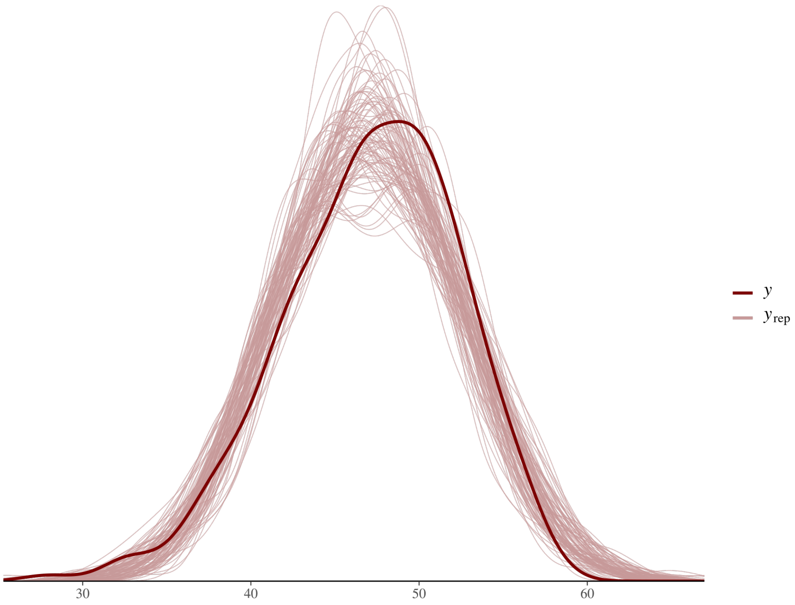
What's that?
Compare observed data to data generated from your *fitted* model.
$\longrightarrow$ Assess model fit.

What am I looking for?
Evidence that your data *could have been* generated from the fitted model.

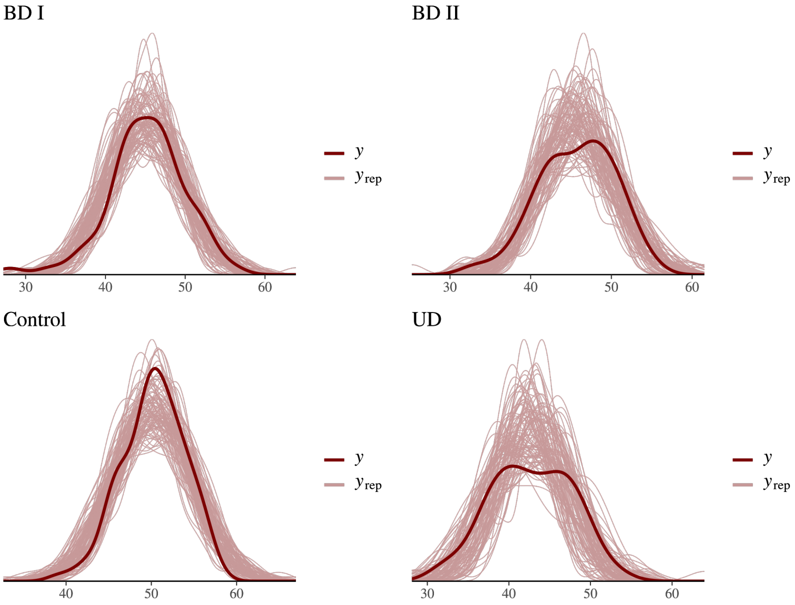
Posterior predictive checks

Let's first focus on the ANCOVA model $\mathcal{M}_4$.

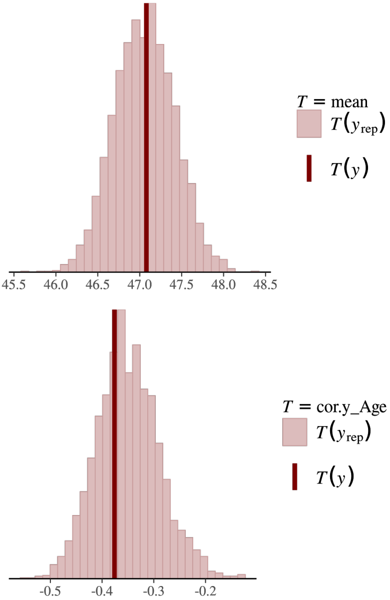
Distribution of y



Distribution of y per group



Various statistics of y

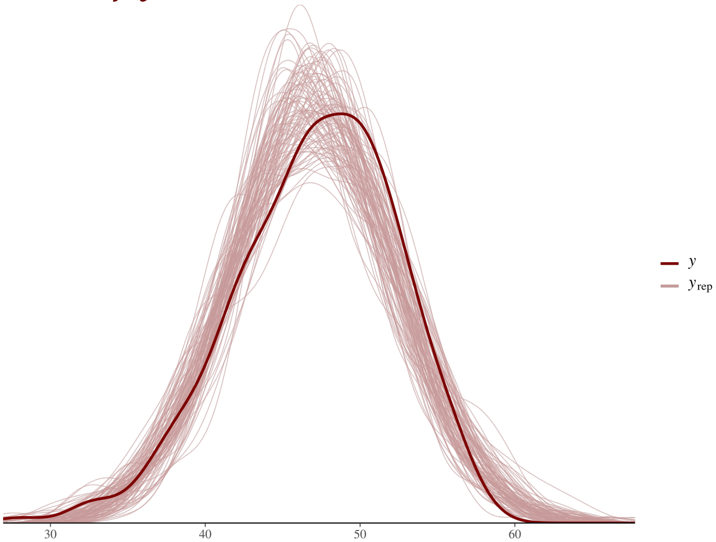


Posterior predictive checks

So the ANCOVA model seems to fit the data well.

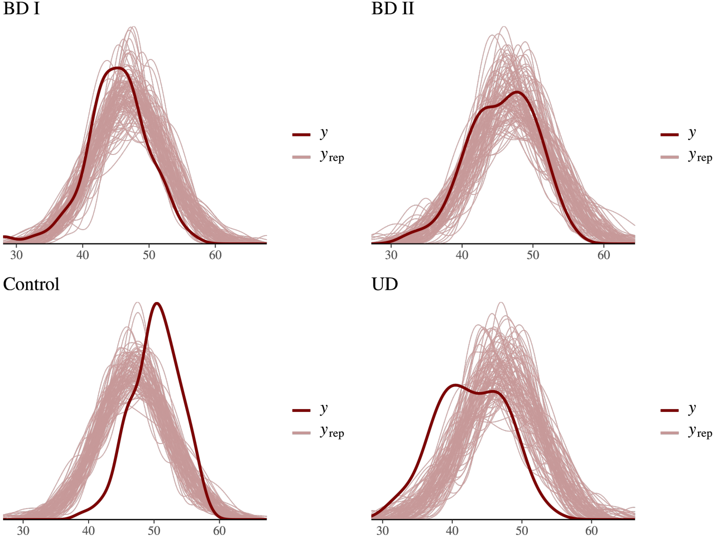
How does the seemingly worse baseline $\mathcal{M}_1$ model do?

Distribution of y



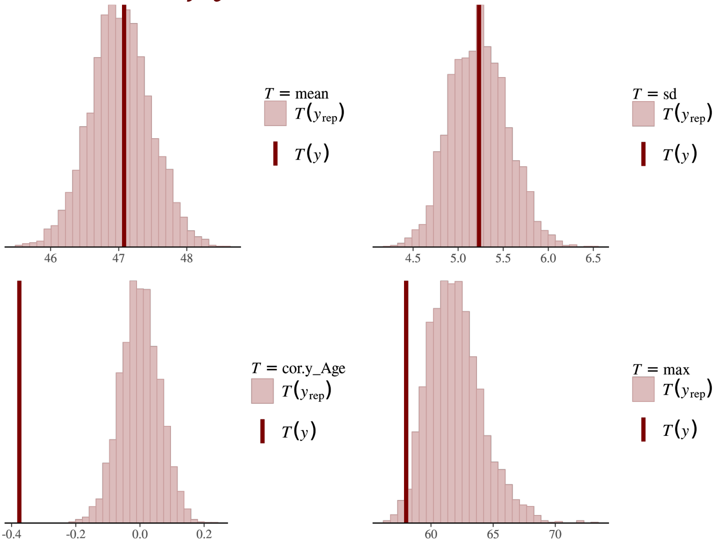
Not that bad!!
(But only because $y \approx \mathcal{N}(\cdot)$, *which need not happen in general*).

Distribution of y per group



Humm, the Control and UD groups are misspecified.
(Of course, 'Group' was not modelled…)

Various statistics of y



cor(y, Age) completely missed.
(Of course, 'Age' was not modelled…)

Model comparison

Leave-one-out cross validation (LOO-CV)

Idea:

- Models are compared based on out-of-sample *expected predictive accuracy*.
- LOO-CV is efficiently approximated by means of PSIS-LOO (Pareto smoothed importance sampling).

Interpretation:

- PSIS-LOO essentially provides a means to rank models.
- It doesn't really quantify differences between models.
- However, as a rule of thumb, consider values of *elpd_diff* at least 4 times as large as its SE as noteworthy.

Leave-one-out cross validation (LOO-CV)

Model	elpd_diff	se_diff	looic
$y \sim$ Group + Age (ANCOVA)	0.0	0.0	1589.8
$y \sim$ Group + Age + Group $\times$ Age	-2.4	1.2	1594.6
$y \sim$ Group	-11.5	4.2	1612.7
$y \sim$ Group + Age, $\mu_{\text{Control}} = \mu_{\text{UD}}$	-18.4	7.1	1626.7
$y \sim$ Age	-38.3	7.9	1666.4
$y \sim 1$	-58.1	8.6	1706.1

- Models are ordered from best to worst.
- Thus, ANCOVA appears to have the best predictive ability.
- Based on the '4SEs' rule of thumb, we discard the last two models.

Bayes factors

I also tried to compare models using Bayes factors.
I have a lot to say about BFs, not all of it is good.

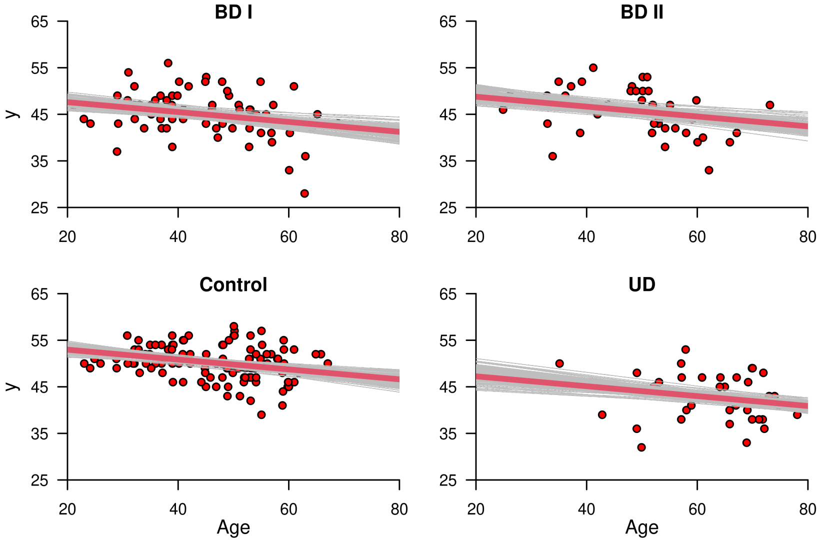
Idea:
Bayes factors compare the models' predictive ability *for the observed data*. Thus:

> *Under which model are the observed data more likely?*

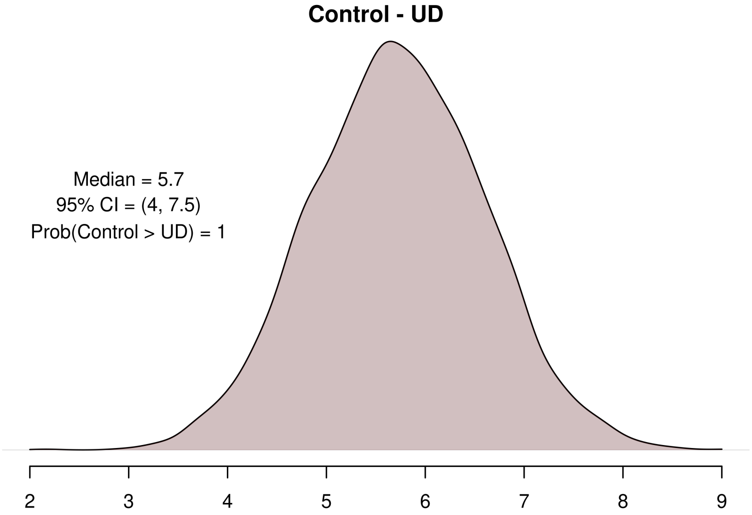
Unfortunately, the results were *tremendously* sensitive to prior specification.
I decided to leave them out.

Summarize and report inferences

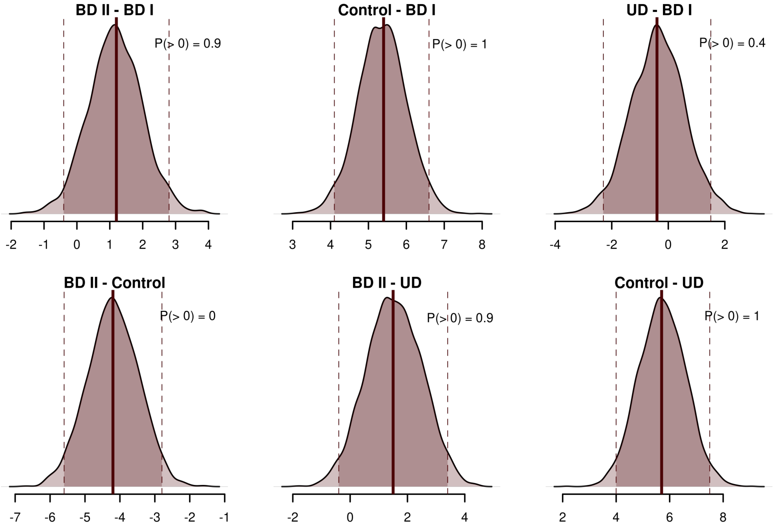
Plots per group



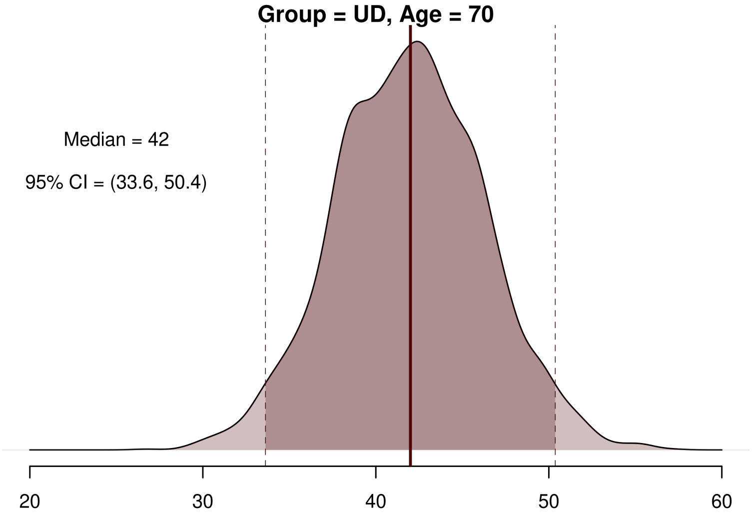
Contrast: $\mu_{Control} - \mu_{UD} = 0$



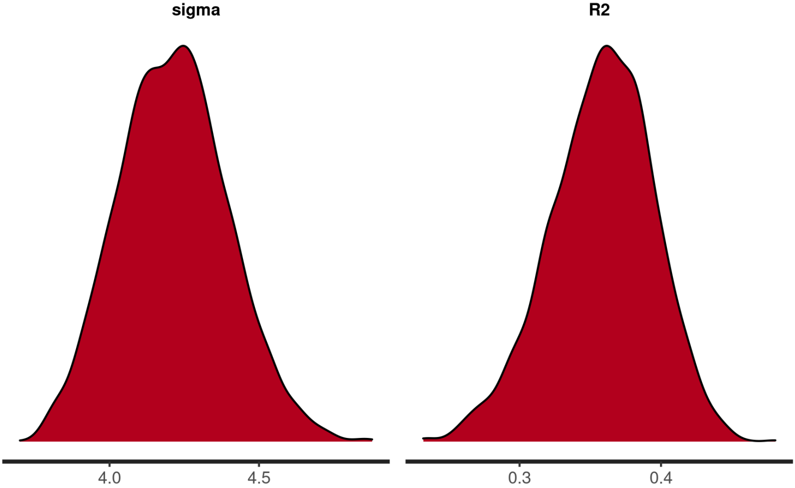
All pairwise contrasts



Prediction for one subject



Posterior dists. σ, R^2



Summary

	Mean	SD	2.5%	97.5%
beta[1]	49.73	1.14	47.46	51.94
beta[2]	1.17	0.81	−0.40	2.79
beta[3]	5.36	0.64	4.08	6.61
beta[4]	−0.36	0.95	−2.27	1.51
beta[5]	−0.11	0.02	−0.15	−0.06
sigma	4.21	0.18	3.88	4.58
R2	0.36	0.04	0.28	0.42

Conclusion

Bayesian modelling is *very* flexible:

- Checking model fit is very *intuitive* and *visual*.
- It is not *that* difficult to adapt the model, if needed be.
- It is possible to perform *any* inference that is a functional form of the data or model parameters.
- It is possible to compare models, for various predictive criteria.
- No statistical significance required.
- All outcomes are *stochastic*:
  You get to report the *uncertainty* in your results.
- The sky is the limit:
  The types of models available are nearly endless.

Now you give it a go!