Introduction to Bayesian Data Analysis

Assignment 3

Linus Hof

Info

The assignment comprises of 8 tasks. In total, 20 points can be obtained.

By the end of the deadline (18 June, 8pm), upload one `R` script that produces the answers to Moodle:

- Indicate which task the code belongs to by adding a line `#tasknumber` over the code for the respecitve task.
- `# comment` your code by describe what the code is supposed to do.
- At the beginning of the script, specify which packages you used (`library(package)`)

Pre-Processing

1. Load the data set `Aging.csv` and store it in an object. (1 point)
2. Delete all observations (rows) containing missing values, indicated by an `NA`. (We do this only for simplicity - usually, decisions such as removing data must be well justified.) (1 point).

The data stems from 122 people that solved a set of choice problems between a safe and a risky option - e.g., choosing between A) €3 with 100% (safe) or B) €32 with 10% and €0 with 90% (risky). The data set contains the following variables:

- `Age`: Participants' age in years
- `RiskSeeking`: Proportion of choices with the risky option being chosen.
- `Decision Quality`: Proportion of choices with the higher expectation option being chosen.
- `Speed`: Processing speed (measure of fluid cognitive ability)

- `Numeracy`: Ability to operate with numeric information (higher values indicate better ability)
- `NegAffect`: Negative affect = being in a negative emotional state

3. Inspect the distributions of all variables. Add a new variable `AG` to the data frame that groups people into 2 age groups (1 = young and 2 = old). (1 point)

Gaussian Models of Decision Quality and Risk Seeking

4. Build and estimate a Gaussian model for the variables `DecisionQuality` and `RiskSeeking` (separately) using the `quap()` function from the *rethinking* package.

- For the means, use priors that are weakly informative, i.e., not flat (uniform). Defend your prior choice in a few words. (2 points)
- Use the `extract.samples()` and the `HPDI()` function from the `rethinking` package to obtain the 95% highest posterior density intervals for the mean/mu parameter. (1 points)
- Interpret both estimation results with regards to the quality of the decisions and peoples risk attitude.(2 points)

5. Show that young and old people, on average, do not differ in decision quality. To do so:

- Estimate a Gaussian model for each group separately. (1 point)
- Calculate the difference between 10,000 samples from the two posterior distributions of the mean/mu parameter. (1 points)
- Visualize the distribution of difference. (1 point)

Linear Prediction

6. Standardize all variables except age group. (1 point)
7. For both `DecisionQuality` and `RiskSeeking`, build and estimate simple linear prediction models by separately using the variables `Numeracy`, `Speed`, and `NegativeAffect` as the respective predictor (i.e., 2 x 3 models = 6 models in total).

- Briefly describe 2 advantages of using standardized variables for this task. (2 points)
- Which variable has, on average, the strongest total effect on `DecisionQuality`? (1 point)
- Which variable has, on average, the strongest total effect on `RiskSeeking`? (1 point)

8. Build a linear prediction model to estimate the associations among the variables `Speed`, `Numeracy` and `DecisionQuality`.

- Briefly describe your assumptions about how the variables could be causally related, e.g., do you expect some confounding or mediation effects? (2 points)
- Estimate the model and state how the results speak for or against your assumptions. (2 points)