Case: Brain

Story

The dataset `brainweight.txt` contains measurements of the weight of both brain and body for different mammals.

Data

Variable	Description
art	species of mammal
body	weigth of body (kg)
brain	weight of brain (gram)

Exercise

1. Make a scatterplot of body against brain. Do you see any association between the variables?
2. Make a log transform of both body and brain. Make a scatterplot of the transformed variables. Do you see any association now?
3. Fit a regression model describing `log(brain)` by `log(body)`
4. Write up the mathematical model for `log(brain)`. Then take the anti-log on both sides to obtain an expression for `brain` and simplify the equation.

   What is the structure of the relation between brain and body weight? Is it linear?
5. Estimate confidence intervals for the parameters in the model for `brain`.
6. Test the hypothesis that the brain-to-body ratio is constant (does not depend in body weight).
7. Save the code that performs your analysis in a script and add plenty of comments to your code.

Extra

1. Are there any outliers in the data? Take a look at the observations that are the least well fitted by your model. Can you explain these observations?
2. Perform model diagnostics on the model in question 3. and argue that each of the model assumptions are fulfilled.
3. Illustrate the fitted model on the data. Also quantify the model uncertainty.
4. What if a new mammal was discovered with a body weight of 5kg. What would the likely brain weight of this mammal be? To quantify the uncertainty in your estimate, provide an appropriate interval.
5. Investigate the quality of the data source and realize that one mammal appears twice. Which mammal is it and which of the entries is the right one (use google or wikipedia to get the facts)? Are there other problems with the data?