

Assignment 2

An Intense Course in Data Analysis Using MultiLevel Regression Models Peking University Summer School, 2010

This assignment still uses the sleep data.

1. Choose one predictor and simulate the uncertainty in the line. Make a graph that shows the data, regression line, and uncertainty. How does the uncertainty of that model compare to the one with $\log(\text{body})$ and danger from class?
2. Find a predictor that has only a few possible values and is statistically significant in a model with the predictor that you chose.
 - (a) Change that predictor to a factor and fit an additive model. Do you prefer the additive model or the linear model? Why or why not?
 - (b) Keeping the first predictor and the second as a factor, fit an interaction model. Do you prefer this model? Why or why not?
 - (c) Extending what we did in class, give a formula for an interaction model with only two numeric predictors, neither of which is a factor. How many parameters does it have? Fit that kind of interaction model to your data, treating your factor variable as numeric now. Is the interaction important?
 - (d) Simulate the uncertainty for your best interaction model with 500 trials (more trials is not harder than a few). Compute the estimated mean at a set of 10 values for the X_1 and X_2 in your model for each of the 500 vectors of simulated coefficients. How many of the estimated means for each of the 10 choices of (X_1, X_2) fall into the 90% confidence interval for the mean? (Use the R function `predict` to get the confidence intervals). For each of the ten (X_1, X_2) , compute the .05 and .95 quantiles of the simulated means. Are these close to the endpoints of the confidence intervals from R?