

Final Exam, Take-Home Portion

Math 246

May 2, 2018

The Data

The data you'll be investigating in this final exam comes from a recent New York Times article (April 19, 2018) about a study published April 12, 2018 on the possible health effects of sitting.

The article and study are linked on the course website, in the articles section. You do not need to read the articles to perform this analysis, but you may find them interesting.

The dataset is in the data section of the website.

The variables

From the study: "Atrophy of the medial temporal lobe (MTL) occurs with aging, resulting in impaired episodic memory."

We will use the total MTL thickness in mm, given the variable name `TOTAL`.

From the study: "Sitting data from the IPAQ-E was reported as average number of hours spent sitting per weekday over the past week." This is reported as `Sitting`.

From the study: "Total physical activity is quantified by weighting each type of activity by its energy requirements defined in metabolic equivalent units (METs) to yield a score in MET-minutes per week." `METminwk` reports the number of MET minutes per week.

From the study: "For the current analyses, physical activity was examined both as a continuous variable and as a categorical variable: participants were dichotomized into either "lower" or "higher" activity groups based on a standard metabolic equivalent (MET) cutoff of 1500 MET-minutes per week. This cutoff was chosen as a mid-point between no activity (0 METs) and high activity (3000 METs), as defined on the IPAQ website (www.ipaq.ki.se)." `IPAQgrp` is the low/high version of this variable.

The questions

You should submit answers to each of the following questions. Your answers do not need to be in an rmd file, an R script is fine. You should submit your code as well as answers to the question - the answers to the questions can be comments in a script, can be handwritten, typed in another document, or along with code in an rmd file. Any of these options are fine.

1. We are interested in predicting `TOTAL` as this is related to cognitive decline.
 - a. Construct a model using `Sitting` to predict `TOTAL`. What does the sign of the coefficient of sitting tell you?
 - b. Construct a model using `METminwk` to predict `TOTAL`. What is the value of the adjusted R^2 ?
 - c. Construct a model using both `Sitting`, `Sex`, and an interaction term. Which of the terms are statistically significant?
2. `IPAQgrp` is a variable describing high and low activity during the week. Construct a model which predicts `IPAQgrp` using `Sitting` as a predictor. What is the p-value for the `Sitting` variable?
3. Use a non-parametric model to determine whether there is a difference in `TOTAL` when comparing the two groups in `IPAQgrp`. What is the p-value for your model?

4. Suppose you receive feedback that there should have been three groups in IPAQgrp. Instead of greater than 1500 or less than 1500, it should have been less than 1000, between 1000 and 2000, and above 2000 (called low, medium, and high). Perform a non-parametric test determining whether there is a difference in TOTAL based on these three groups. What is the p-value for your model?