**Psych 610**

**Homework 6: Due 18 October 2017**

**Reading Questions**

Answer these questions in your Word document.

1. Judd et al. list three questions that we should ask ourselves about the data, in order to detect outliers.
	1. The first question concerns values on the predictors. In this section, Judd et al. discuss how you can think of each point as having its own associated slope. What two points define these individual slopes? Are we more worried about small distance or big distances? Discuss this in terms of how their slopes are weighted.
	2. The second question concerns the data value on Y. In this section, Judd et al. discuss how studentized deleted residuals are related to a very specific model comparison. Describe the compact and augmented models associated with the test of the studentized deleted residuals for a particular observation.
	3. The third question concerns the change in parameter estimates if an observation is deleted. What are Judd et al.’s recommendation for deciding whether an observation requires further attention based on Cook’s D?
2. Judd and colleagues make a very specific recommendation for reporting your model of the data when you’ve decided to exclude an observation that substantially affects your conclusions. What is the reporting recommendation and what is their reason for this suggestion?
3. According to Judd et al., should we be worried about whether the data themselves (Xs and Ys) are distributed normally?
4. Judd et al. suggest examining a normal quantile-quantile plot to detect non-normal distribution of errors. If the errors are normally distributed, what should the plot look like? What would you look for in this plot to decide whether the errors are not normally distributed?
5. If we compare a model with three predictors with a mean-only model, this model comparison (though do-able) may not be very useful. Judd et al. discussed two reasons for this. What are the two reasons?

**Data analysis**

For the data analysis portion of this homework, you will be working with data about implicit biases, which you are familiar with. Below is the study description and codebook for this dataset.

A common theme of the modern study of racial bias is that bias can be both unintentional and subtle. For example, few people say directly that they dislike Muslim people, and most people genuinely want to treat others fairly, regardless of religion. However, when put into an ambiguous situation (such as whether to hire a Muslim applicant whose resume is moderately good, but not outstanding), it is argued that people often fall prey to subtle, unintentional biases.

A group of social psychologists wished to test an intervention they developed that is designed to help people eliminate subtle, unintentional biases by educating them about subtle bias and training them in using a variety of techniques to reduce unintentional bias. First, they measured people’s baseline unintentional anti-Muslim bias using the Implicit Association Test (positive scores indicate an unintentional preference for Non-Muslims). Then, they then randomly assigned people to training or control. Finally, they administered the IAT four weeks and eight weeks after the training manipulation. The experimenters also wished to investigate whether the training was more effective for people who scored high on a scale measuring concern about discrimination in society; high scores on this scale indicate greater concern.

Codebook for the Training dataset

|  |  |  |
| --- | --- | --- |
| Variable name | Description | Values |
| Condition | Whether or not the participant completed the training intervention | 1 = training,0 = control |
| Concern1 | Item 1 of the concern about discrimination scale | 1-10 (reverse-coded) |
| Concern2 | Item 2 of the concern about discrimination scale | 1-10 (reverse-coded) |
| Concern3 | Item 3 of the concern about discrimination scale | 1-10 (reverse-coded) |
| Concern4 | Item 4 of the concern about discrimination scale | 1-10 |
| BaseIAT | Baseline Muslim-Other Implicit Association Test score | -.45 - 1.50 |
| Wk4IAT | Week 4 Muslim-Other Implicit Association Test score | -.83 - 1.45 |
| Wk8IAT | Week 8 Muslim-Other Implicit Association Test score | -1.00- 1.31 |

Preliminaries

1. Read the data into R and inspect them. Based on descriptive statistics and univariate plots, are there any hints that some cases might have been miscoded and should be excluded?
2. Exclude any cases that have clearly been miscoded. While in your own data you might look further into this miscoding to see if you can fix it, for the purposes of this assignment, simply remove the entire participant when you find evidence of clear miscoding. Do a quick check for any additional miscoded values.
3. Create a variable called ConcernM that consists of themean of the 4 concern items, taking reverse coding into account. This is a measure of overall concern about discrimination.
4. Create a variable called PostIAT that consists of themean of the IAT measured at week 4 and week 8.
5. Explore the data set by obtaining the recommended univariate plots on your quantitative variables of interest: ConcernM, BaseIAT, PostIAT
6. Explore the data set by obtaining the bivariate correlations and scatterplot matrix on your all of your variables of interest: Condition, ConcernM, BaseIAT, PostIAT. Note any strong relationships between these variables.
7. Fit a model in which you predict the average of the week 4 and week 8 IAT scores (PostIAT) from baseline IAT scores, condition, and overall concern about discrimination. In a sentence, describe your findings (report *F* and *p* values).

Case Analysis

1. Examine your data for points that have high leverage. In your script, tell us which points have high leverage and the variable(s) they have high hat values on. In your script, write a sentence describing why points of high leverage might be a problem.
2. Examine your data for points that are regression outliers. In your script, tell us which points are regression outliers (if any) and briefly describe whether or not they might be problematic, and why.
3. Examine your data for points with high influence using Cook’s D. In your script, tell us which points have high influence overall (if any) and briefly describe whether or not they might be problematic, and why.
4. In your script, describe what you would do based on the information you gained from your case analysis (i.e., would you remove any cases?).
5. Remove any problematic cases, and store the new dataset in a new data frame.

Data Analysis

1. In Problem 6, you fit a model in which you predict PostIAT from baseline IAT scores, condition, and concern about discrimination, using the original data. Refit the model using your new dataset (d2).
2. Compare the model summary of the original model to the one in which you’ve removed the problematic case. What has changed and why? What has not changed?
3. At the bottom of your script write the number of hours you spent working on this homework assignment.