**THE3**

**Assigned: 13 December 2017, 3:00pm**

**Due: 19 December 2017, 11:59pm**

You have come a long way to get to this point: your final take home exam! Great work so far. Remember, you are allowed and encouraged to refer to prior lab scripts, homework assignments, books/articles, and the internet. **You may *not* consult with other students.** Use the same general format as you do for homework assignments (see template document on the website). A few things to note:

* **Submit your files anonymously.** Instead of using the normal Lastname\_HWXX format, use the **last 4 digits of your student ID number** (e.g., “3180\_THE1.R”). *Include no identifying information in your files!*
* The only things that should appear in your Word doc are your write-ups and your figures (no reading questions for this last one). Everything else can just be typed in your R script.
* Follow course norms for naming variables and models (or use names that will be crystal clear to us).
* You don’t need to include the question text next to the question number in your R script; the number is plenty for us.
* You can email us asking questions, but we won’t be as instructive as we would be on a homework assignment. This is an exam, after all.

**Data Analysis**

**Study 1: Toothbrushes (45 points)**  

Researchers were interested in the effects different toothbrushes and tooth brushing habits have on plaque level in the mouth. Users of different toothbrushes with different toothbrush habits were carefully recruited for the study. More specifically, 16 users of Oral B electric toothbrushes, 16 users of Sonicare electric tooth brushes, and 16 users of manual toothbrushes were recruited. Users of each toothbrush type were recruited to be equally representative of people with two brushing habits: half brushed once a day and half brushed twice a day. The researchers then measured plaque level in each person’s mouth using a novel technique that relies on a special camera and software to count the number of plaque “particles” on the teeth.

https://moo.review/electric-toothbrush/

The researchers had the following hypotheses:

1. People who brush twice a day will have significantly less plaque than people who brush once a day.
2. The effect of brushing twice versus once a day will be bigger for people who use electric toothbrushes (both Oral B and Sonicare types) compared to people who use manual brushes.
3. The effect of brushing twice versus once a day will be no different for people who use Oral B compared to people who use Sonicare electric toothbrushes.

Codebook for Study 1

|  |  |  |
| --- | --- | --- |
| **Variable Name** | **Description** | **Values** |
| SubID | Participant (Subject) ID number | 1-48 |
| BrushNumber | Number of times participant brushes a day | 1 = Once2 = Twice |
| BrushType | Type of brush participant uses daily  | Oral B (Electric)Sonicare (Electric)Manual |
| PlaqueLevel | Number of plaque particles found on teeth | 13 – 133 |

**Note: You may pretend that the data for both examples are normal and free from outliers.**

1. Read in the data (THE3\_Data1.dat). For each variable, report if the variable is an integer, a character, or factor. [2]
2. Run a (single) command that shows Plaque Level broken down by each combination of BrushNumber and BrushType. [2]
3. Recode BrushNumber to be “Once” and “Twice” instead of 1 and 2. Then make BrushNumber a factor with these labels. [2]
4. Run a model with the appropriate contrasts to test the researchers’ hypotheses.
	1. Set up a contrast (with label) for BrushNumber. [2]
	2. Set up contrasts (with labels) for BrushType. [2]
	3. Code regressors into your data frame that line up with your BrushType contrasts. [2]
	4. Run the model. [1]
5. List and interpret all stats (including effect size[s]) to test Hypothesis I. [4]
6. List and interpret all stats (including effect size[s]) to test Hypothesis II. [4]
7. List and interpret all stats (including effect size[s]) to test Hypothesis III. [4]
8. Create one publication quality figure that shows these results. Don’t worry about including raw data in the figure. [7]
9. Write a brief results section in your Word doc. Very briefly set up the study and report all statistics that are relevant to the researchers’ three hypotheses (only report the relevant stats!). Make sure it passes Mitch’s “random stranger” test. Refer to your figure where appropriate. Provide a one-sentence summary or conclusion at the end. [10]
10. Now imagine you were interested to see if the benefit of brushing twice versus once a day was significant in each of the three BrushType groups. Use Fisher’s LSD method to determine if testing these pairwise comparisons would be appropriate. [3]

**Study 2: Facial Hair (48 points)** 

After class one day, Daniel, Mitch and John are approached by a team of educators with a research question that they need tested. The team wants to know how to make instructors more likeable and thus more engaging for nontraditional students (i.e., older students with delayed post-secondary education). The educators think an instructor’s facial hair may be an important factor as non-traditional students may think facial hair is unprofessional and thus unlikable, while traditional students may be more accepting of facial hair. To test this, Daniel, Mitch, and John recruited 58 traditional and 61 non-traditional students. Participants were shown pictures of someone with a full beard (Daniel), someone with an “interesting” mustache (Mitch), and someone with barely detectable stubble (John). Each picture was shown multiple times in a counterbalanced fashion. All participants were hooked up to facial electromyography (EMG) electrodes that could detect subtle muscle activity associated with smiling. The mean EMG activity across trials of each picture type was recorded as an index of how likable the participants found each instructor.

The researchers have the following hypothesis:

1. The difference in smiling between traditional and non-traditional students in response to the pictures may be bigger when viewing pictures of Daniel than when viewing pictures of John because of the clear amount of difference in facial hair between the two pictures.
2. Similarly, the difference in smiling between traditional and non-traditional students in response to the pictures may be bigger when viewing pictures of Mitch than when viewing pictures of John given the facial hair differences.
3. There may or may not be a difference in smiling between traditional and non-traditional students when viewing pictures of Daniel compared to when viewing pictures of Mitch since both Daniel and Mitch have some facial hair. This will show whether absolute presence of facial hair or the amount of facial hair is the more salient determinant of instructor liking.

Codebook for Study 2

|  |  |  |
| --- | --- | --- |
| **Variable name** | **Description** | **Values** |
| SubID | Participant (Subject) ID number | 1 – 128 |
| StudentGroup  | Whether the student is traditional or non-traditional | TraditionalNon-Traditional |
| Picture | Whether the picture is of Daniel, Mitch, or John | DanielMitchJohn |
| EMG | Electromyograph (EMG) smile reading, proxy of liking | 0.08-122.87 |

1. Read in the data (THE3\_Data2.dat) and run a command to see how EMG readings vary by the subject of the photo. [2]
2. Convert the data from long to wide format. [2]
3. Set up the appropriate variables and contrasts to test the researchers’ hypotheses.
	1. Set up a between-subjects contrast (with labels) for StudentGroup. [2]
	2. Create a new variable that allows testing of hypothesis I. [2]
	3. Create a new variable that allows testing of hypothesis II. [2]
	4. Create a new variable that allows testing of hypothesis III. [2]
4. Test hypothesis I and list and interpret all stats (including effect size[s]) for this test. [5]
5. Test hypothesis II and list and interpret all stats (including effect size[s]) for this test. [5]
6. Test hypothesis III and list and interpret all stats (including effect size[s]) for this test. [5]
7. Correct all *p* values for multiple comparisons using the Holm-Bonferroni approach. [3]
8. Make a publication quality figure that shows the interaction between student group and picture type. Don’t worry about including raw data in the figure. [8]
9. Write a brief results section in your Word doc. Very briefly set up the study and report all statistics that are relevant to the researchers’ three hypotheses (only report the relevant stats!). Use only the corrected *p* values. Refer to your figure where appropriate. Make sure it passes Mitch’s “random stranger” test. Provide a one-sentence summary or conclusion at the end. [10]

**Study 3: Online Class (hypothetical) (7 points)**

Some researchers want to see if an online class is effective at improving statistics knowledge. A large number of participants are recruited, and their initial statistics knowledge tested. Then, the participants are randomly assigned to either take the new online stats course or an online control course focused on baking. Your variables are ‘InitialStatsKnowledge’ ‘PostStatsKnowledge’, and ‘ClassType’.

1. Without worrying about setting contrasts, type the lm command you would use to run a model designed to test if the online stats class was effective at increasing stats knowledge (relative to the baking class) given the research design and available variables. [2]
2. Which coefficient (b0, b1, b2, etc.) would you want to interpret from the output of this model in order to see if the knowledge was increased in the online stats class relative to the baking class? [1]
3. The first two datasets you analyzed had small Ns and were likely underpowered. Make sure that this new online class study does not have that problem. How many participants would you need to recruit to obtain 80% power if you believe the class will have a moderate effect on stats knowledge (i.e., $η\_{p}^{2}$ = .08) using a two-tailed alpha of .05? [3]
4. In your word doc, write how long it took you to complete this exam. [1]