MidSemster Data Analysis Synthesis

Neurotech BaM 2020

Commentary on this activity

This activity is an opportunity to demonstrate your skills and understanding in some of the statistics and probability concepts that we have covered so far. This part of the assessment is about actually applying some statistical analysis to data (a pretty darn important part of stats).

All analysis and work shared here should be your own. You may consult course materials, books, and online resources (except for direct communication with people online or other things that don't follow the spirit of independent work). Please do not work together with or discuss this assignment with anyone else within or outside of the class.

This assessment does require writing a bit of code. This is not a programming course, and I want to reduce barriers from coding confusion. Therefore, if you are stuck on a coding or data manipulation problem, you are welcome to email me (Sam) with a description of what you are trying to do or the error you are getting, and I will happily help you with that aspect of the code. You can also come to my office hours with these questions.

This document is view-only. If you have clarifying questions, please feel free to email me. I may make updates to the document to reflect any clarifications. I will generally try to do this as comments so they pop out, unless they are very minor.

Please, please, please read each of the questions carefully and make sure you understand what it is asking.

The "experiment"

On the first day of class, we discussed memory, and I challenged you to memorize 2 lists of 10 words each. Just for fun (no peeking), can you remember any of these words? Feel free to jot them down, and compare to the answers later.

As you may recall, we also conducted a silly experiment where some of you received candy before the memory activity and some of you received no candy (until later in class... I'm not that mean). One might hypothesize that having candy would make you feel happier and therefore make you do better, or perhaps having candy is a distraction and will make you do worse. We will finally seek the answer to this question, which I'm sure has been on your mind all semester.

Now for the twist, there was actually one other potential effect in this study. There is a fairly well studied effect in psychology of inducing false recall in a list of words. This is done by

presenting several semantically related words, but excluding the word that links them together, often called a "lure" or "critical word". For example, if the list of words were "brains, statistics, Sam, data, Olin", the critical word might be "neurotech" (the word that thematically links them together). Our round 1 list of words had a critical word of "book". "Book" was not included in this list, but many of the other words were related to books. To encourage this "false recall" to be even more likely, I included the word "book" in the round 2 list. In this paper (https://pdfs.semanticscholar.org/9425/1836c4c160574b8887362847795da66c2c81.pdf), Roediger and McDermott observed that the critical word was falsely recalled 40% of the time.

The dataset

The data for this activity comes directly from the survey that you submitted with your guesses for each of the 10 words (in each list). The data can be found here:
https://drive.google.com/open?id=1rbBwrv4US82Ff_1JrDpCj8sWvROi56pp
The variables (column headers) are as follows:
- List: 1 means the first list of words and your responses to that prompt; 2 means the 2nd list of words
- W1-W10 are the answers you all submitted for each of the 10 available responses for each list. Some of these are funny… nice work.
- Group: Categorical variable indicating if the person received candy "Candy" or did not receive candy until later "No Candy".
- CodeName: Categorical variable with the code name that you randomly selected for yourself. This allows us to pair up people's responses to the first and second list. These codenames may be people's names (though they might not actually match the person who submitted the response). I chose to preserve the codenames that you submitted (except to make them easier for Matlab to parse). Obviously, please treat each other and the data of others with respect and kindness.

The code provided here (https://drive.google.com/file/d/1g1rn4XEQ1Z-hrYM0ndHmd1R2AqfXaPqB/view?usp=drivesdk) will help you to load this text file into a table, sort the data by code name, and calculate how many correct answers each person submitted (separately for each list). The data are in a table with the following additional variables (not described above):
- CorrectInOrder: Word by word (in the list of words) binary indicating if a particular word from the list was included in a person's response for that list.
- NumCorrect: Number of words in the list that each person correctly included in their response
- List1ContainsBook: Binary (1=yes, 0=no) indicating if a person wrote down "book" as one of the words in list 1. Book was one of the words in the round 2 list, but not the round 1 list. "Book" is considered the critical word or the lure (see context above).
- List1ContainsPencil: indicates if I person wrote down "pencil" as one of the words in list 1. Pencil was in both list 1 and list 2.
- List1ContainsHouse: same as above, but "house", which is from list 2

The questions to answer

1. What is the probability that a person falsely recalled "book" in list 1 (writing it down even though it was only on list 2) if we know that person received candy?
2. What is the probability that a random person selected from our class received candy, given that we know they did NOT falsely recall the word "book" in list 1?
3. If we added 10 more people, all of whom received candy and none of whom falsely recalled the word "book" in list 1, what would the new probability be for question 2?
4. Did people who received candy perform differently than those that did not receive candy on List 1? (Performance here should be indicated by the number of words correctly recalled.) Please select and conduct an appropriate statistical test, calculate effect size, and create one illustrative well-labeled figure to address the question. Please interpret your findings in a sentence or two.
5. The Roediger and McDermott paper observed that people accurately recalled 65% of words. Do our results differ from this significantly?
6. Did people perform differently on their recall of the two lists? Please select and conduct an appropriate statistical test, calculate effect size, and create one illustrative well-labeled figure to address the question. Please interpret your findings in a sentence or two.

What to submit

You can submit a summary document and your raw code. Or you can submit a single pdf generated from a live editor notebook.

Please submit a document that is clean and very easy to read. Be sure to add a semicolon at the end of statements that print out data. Clearly label your work and final answers and which questions they relate to. Since I will be manually grading these, I reserve the right to deduct points for messy and hard to follow submissions (please be kind, rewind… and enjoy this reference from another time period when VHS tapes roamed the earth).