Exploration 1.1: Memorizing Letters

You and your classmates will each be asked to study a sequence of letters for 20 seconds and then to write down as many letters as you can remember, in order. Your score will be the number of letters that you remember correctly before your first error of any kind. You will also be asked to report your hours of sleep last night and whether or not you have consumed any caffeine today. After collecting the data, your instructor will tell you more information about the *study protocol*.

Definition: The ***study protocol*** outlines how the study will be conducted, providing enough detail, so that someone else could carry out the same study under identical conditions. It is important to consider the research question when evaluating whether the study protocol will be appropriate.

STEP 1: Ask a research question.

1. What was the *research question* for conducting this study?

STEP 2: Design a study and collect data.

2. Identify the *response variable*. Is this variable quantitative or categorical? (If categorical, note the number of categories. If quantitative variable, note the measurement units.)

3. Was this an *observational study* or an *experiment*? How are you deciding? (*Hint*: Recall from your previous statistics course that the key characteristic of an experiment is that the researcher determines which explanatory group each participant in.)

Definition: The explanatory variable deliberately manipulated in an experiment is often called a ***treatment variable*** or ***factor***. For categorical explanatory variables, the different categories of the treatment variable are often called ***levels***. In experiments, the objects (or "subjects") that we are measuring are often called ***experimental units*** rather than observational units. The conditions we impose on the experimental units (here the levels of the treatment variable) are also called ***treatments.*** Each experimental unit is assigned to one treatment.

4. Identify the *experimental units* in this study. How many are there?

5. Identify the *treatment variable* and its *levels*.

6. If you were planning this study, how would you determine who gets which treatment? What would you try to accomplish?

Key Idea: The goal of *random assignment* is to reduce the chances of there being any confounding variables in the study. By creating groups that are expected to be similar with respect to variables (other than the treatment variable of interest) that may impact the response, random assignment attempts to eliminate confounding. A key consequence of not having variables confounded with the treatment variable in a randomized experiment is the potential to draw cause-and-effect conclusions between the treatment variable and the response variable. See Example 1.1 for more discussion.

7. Identify any other precautions taken in this study to try to make sure the two treatment groups were equally balanced, that is, the treatment conditions were the same for both groups.

Definition: A study is ***double blind*** if
(i) the subjects do not know which treatment condition they are in, and,
(ii) the person evaluating the response variable does not know which treatment condition the subject is in
If only one of the above conditions are true for a study, then the study is said to be ***single blind***.

8. Was this a *double blind* or a *single blind* study? Explain. What other, if any, appropriate precautions were taken in carrying out this study?

9. Were you and your classmates *randomly selected* to participate in this study? Do you think you are representative of a larger population? What would you define that population to be?

10. Are there any individuals that you believe we should exclude from participating in this study? Explain.

Definition: ***Inclusion criteria*** are the set of characteristics that individuals must have in order to participate in a study.

Key Idea: Using inclusion criteria may reduce the variation in the observed response compared to what would otherwise be observed if there were no inclusion criteria. However, by using inclusion criteria we limit the scope of inference for study conclusions. The more inclusion criteria there are, the smaller the population to which the study conclusions will apply.

STEP 3: Explore the data.

Load the ***memory*** data into the **Multiple Variables** applet. Select and Clear the existing data, paste in the four columns of data from your class, press **Use Data**. Drag the *score* variable under the Response header. Check the **Show descriptive** and **Show residuals** boxes.

11. Use the applet to create numerical and graphical summaries of the outcomes of the response variable for your class. Summarize your observations in context.

Recall from the Preliminaries that a statistical model is an equation that predicts the outcome of the response and measures the accuracy of those predictions.

12. Specify a statistical model for predicting future results using the overall mean score for your sample and specifying the standard error of the residuals. (Sometimes referred to as the "single mean" model.)

Drag the *sequence* variable under the **Subset By** header.

13. Now create numerical and graphical summaries comparing the results for the two treatment groups. Based on the group means, did one of the sequence groups tend to score higher than the other? By a lot or just a little? Which sequence group had more variable results? Any other interesting features of the meaningful sequence scores that makes sense in context?

14. Write out a statistical model for predicting outcomes depending on which treatment condition someone is assigned to, using the treatment-specific mean scores. This could be called the "letter grouping model" or the "separate means model." **Note**: The "separate means model" allows us to assign a predicted response, the group mean, to each treatment in contrast with the "single mean model" which uses the same overall mean response to make our predictions regardless of treatment.

15. Is the standard error of the residuals for the letter grouping model much smaller than the standard deviation of the residuals in the single mean model?

16. Does knowing which treatment group each person was assigned to explain all of the variation in the responses? How are you deciding?

STEP 4: Draw inferences beyond the data.

Recall from the Preliminaries that this step entails investigating whether the difference in average scores detected between the two treatment groups reflects a genuine tendency, and if so, estimating the size of that tendency. Section 1.3 will look at Step 4 in detail. For now, let's move on to Step 5. In this step, we need to review the study protocol to determine the "scope of conclusions" we can draw from the study.

STEP 5: Formulate conclusions.

17. Summarize your "letter grouping model" with a Sources of Variation diagram, including brainstorming some possible sources of the unexplained variation.

Observed Variation in:	**Sources of explained variation**	**Sources of unexplained variation**
Inclusion criteria: *Design:*		

One potential source of variation in how many letters someone can memorize is the amount of sleep he or she got the night before. Ideally, the study protocol has balanced this variable between the two treatment groups (i.e., there is a similar mix of those with lots and little sleep in both groups), but let's check. Remember, to be a confounding variable, the variable must explain variation in the observed memory scores AND it must be associated with the explanatory variable, type of letter grouping.

Remove the *score* variable from the Response box and instead move the *sleep* variable there.

18. Use the applet to examine the distributions of sleep hours for the two treatment groups. Does *amount of sleep* appear to be a confounding variable in this study? How are you deciding?

Now remove the *sleep* variable and move the *caffeine* variable to the Response box – this graph is a little strange, what do you think it and the group means tell you?

19. Is *caffeine* a confounding variable in this study? How are you deciding? (Be clear what steps you took/graphs you examined.)

20. Could there be another explanation, apart from the grouping of the letters, that could explain the difference in the group means that we found? Explain.

Key Idea: *Random assignment* is a critical component of a well-designed experiment that allows us to potentially draw *cause-and-effect* conclusions. How the experimental units are selected for the study (e.g., inclusion criteria, *random sampling*) is a key component of how broadly we can *generalize* the results.

Hopefully you found above that the conditional distributions of the amount of sleep for the two treatment groups and the conditional distributional distributions of caffeine for the two treatment groups were pretty similar, preventing amount of sleep and caffeine category from being confounding variables. With random assignment, we will often trust that this is true for all other potential confounding variables as well.

21. In order to make predictions about how many letters students will be able to memorize, do you recommend the “single mean” model or the “letter grouping” model? Explain. How accurate is the letter grouping (or separate means) model? Does the difference in group means seem meaningful to you in this context? Explain. What population are you willing to generalize these observations to? Are you willing to draw a cause-and-effect conclusion between the type of sequences and the ability to memorize more letters on average in this population? Explain.

STEP 6: Look back and ahead.

22. Suggest at least one way you would improve this study if you were to carry it out yourself.

Section 1.1 Summary

In this section, you reviewed some key principles of experiments. In particular, using *random assignment* to create the two treatment groups substantially reduces the possibility of confounding variables, although it may not be possible to remove this possibility entirely. But if you then decide the observed difference between the groups is meaningful (e.g., *statistically significant* as will be discussed in Section 1.3), you can draw a *cause-and-effect* conclusion between the treatment variable and thee response. In other words, we can use “action verbs” like, using a pleasing scent *leads to* more favorable store ratings, on average. The ***inclusion criteria*** used in the study give us information about how we might *generalize* any conclusions we make to a larger population. In particular, if the experimental units are selected from a larger population using *random sampling*, then you believe that any patterns you do or don’t see in these data will apply to the larger population as well. Neither of these studies uses random sampling but instead were “convenience samples” of college students. For Example 1.1, any conclusions about the effectiveness of scent exposure will apply only to undergraduate business majors similar (e.g., similar background, age) to those who volunteered to participate in this study.