For Live Session Assignment (FLS)

Unit 1

Philosophy

- On the following slide(s) you will see activities and the estimated / expected time that the student should spend on that activity.
- It is important to note that the goal of the activities is to become familiar with the methods, ideas and implementation involved in that activity so that we can efficiently iron out all the details in live session.
- Analogy: You are building the pieces of the puzzle in the For Live Session Activity and we are putting them together to see the big picture in live session.
- It is ***not*** expected that the student have all the correct answers. The expectation is that each student spend the allotted time (indicated next to the activity) on each activity so that we can discuss the details in live session.
- If you max out the indicated time without finishing the activity and you don't have more time to finish, simply write up what you have learned by that time and record any questions you might have and we will address those in live session!
- We want to develop the questions before live session so that we can use the live session time to effectively answer them!

Question 1 is the
Quick Quiz Questions!
(<= 1 hour)

Quick Quiz Question 1 (QQ1)

In the motivation for the creativity study (page 2 of text), the poems were given to the judges in random order (though all judges graded all poems). Why was that important?

Select one:

- a. Like in ice skating and diving, judges tend to judge early poems more harshly than later ones.
- b. Without randomization, the effect of motivation on creativity would be confounded with order of judging.
- c. It doesn't really matter. The researchers were just being extra careful.
- d. Randomization ensures that each judge will view each poem.

Quick Quiz Question 2 (QQ2)

Choose the correct response regarding random sampling and random assignment to treatment groups:

Select one:

- a. 1) Results from the analysis of a sample taken *randomly* from a larger population can be generalized to this broader population; *randomly* assigning subjects to treatment groups in an experiment allows us to infer that any differences among groups can be attributed to (are caused by and not just associated with) the treatment.
- b. 2) Results from the analysis of a sample taken *randomly* from a larger population allow us to infer that any differences among groups can be attributed to (are caused by and not just associated with) the treatment; *randomly* assigning subjects to treatment groups in an experiment allows us to generalize the reults to a broader population.

Quick Quiz Question 3 (QQ3)

In 1930, an experiment was conducted on 20,000 school children in England. Teachers were responsible for randomly assigning their students to a treatment group – to receive ¾ pints of milk each day – or to a control group – to receive no milk supplement. The study found that children receiving milk gained more weight during the study period. On further investigation, it was also found that the controls were heavier and taller than the treatment group before the milk experiment began. What is the likely explanation and its implication concerning the validity of the experiment?

Select one:

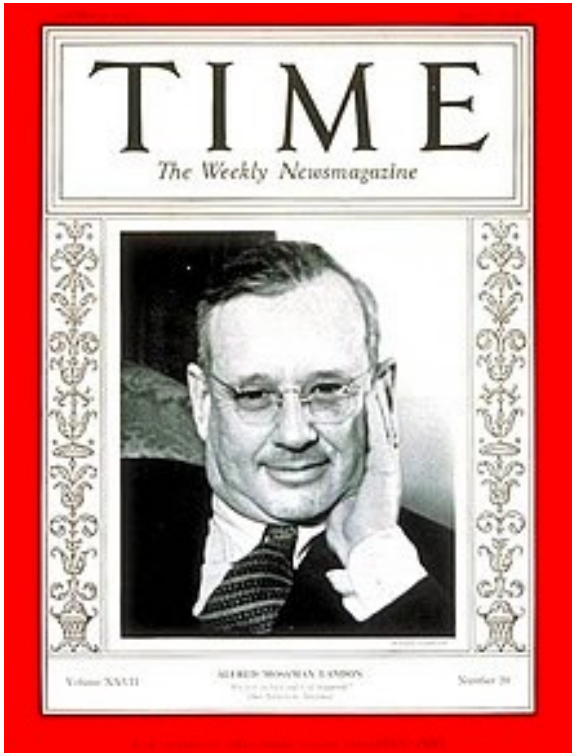
- a. Even when random assignment is done correctly, the resulting groups can be different on important variables. The study is still valid.
- b. Teachers should never be trusted to randomize their students for any type of study.
- c. Even though the random assignment did not work as planned, the study is still valid because there were such a large number of school children involved.
- d. The teachers probably did not really randomize the students. They probably chose students who weighed less to receive the milk supplement; therefore, the results of the study are questionable.

Question 2 ($\leq .5\ hours$)

- What is the difference between a randomized experiment and a random sample? Under what type of study/sample can a causal inference be made?

Questions 3 ($< .5$ *hours*)

- In 1936, the *Literary Digest* polled 1 out of every 4 Americans and concluded that Alfred Landon would win the presidential election in a landon-slide (pun intended:). Of course, history turned out dramatically different (see http://historymatters.gmu.edu/d/5168/ for further details). The magazine combined three sampling sources: subscribers to its magazine, phone number records, and automobile registration records. Comment on the desired population of interest of the survey and what population the magazine actually drew from.



Question 4 (≤1 hour)

- Suppose we have developed a new fertilizer that is supposed to help corn yields. This fertilizer is so potent that a small vial of it sprayed over an entire field is a sufficient dose. We find that the new fertilizer results in an average yield of 60 more bushels over the old fertilizer with a p-value of 0.0001. Write up a scope of inference under the following study designs that generated this data.
  - We offer the new fertilizer at a discount to customers who have purchased the old fertilizer along with a survey for them to fill out. Some farmers send in the survey after the growing season, reporting their crop yield. From our records, we know which of these farmers used the new fertilizer and which used the old one.
  - When a customer makes an order, we randomly send them either the old or new fertilizer. At the end of the season, some of the farmers send us a report of their yield. Again, from our records, we know which of these farmers used the new fertilizer and which used the old.

Question / Activity 5 (2-3 hours)

With respect the Creativity Study you read in the Statistical Sleuth, use the code provided to conduct a permutation test to test for a difference in mean score between those motivated intrinsically and extrinsically.

Recommended approach:

- Reread the material in the book and really think about it.
- Review the code provided and run it a few times … try to make sense of the output and how it is generated.
- Check out the video resource that fully explains the code. VERY IMPORTANT… in the ASYNCH. :)
  - Understanding and adapting code is a HUGE skill in Data Science... This is fun and very useful practice.
- Bring your questions to Live Session and place them at the end of the PowerPoint deck.

Question 6: Takeaways! (~ 1 .5 Hours)

Please provide **at least 4 takeaways** from this section and any questions that you may have. These questions will help design the live session for this unit.

This question will be the last question of every For Live Session Assignment. The idea is that this deck will serve as a document that you can reference in the future to remember what was covered in this section. For instance, this may come in handy for the Capstone and will hopefully become useful in your career. Most immediately, it may become handy as a quick reference for your Midterm and Final! Some students find it very useful and spend a few slides summarizing the asynch material while others learn different ways and only had the minimum 4 takeaways. Either is fine and will earn you full credit for this question.

Question 7: Questions!

Please provide any question or topics of discussion that came up in this Unit! These will help help us optimize our live session for maximum learning and takeaways!

This question will be the last question of every For Live Session Assignment. The idea is that this deck will serve as a document that you can reference in the future to remember what was covered in this section. For instance, this may come in handy for the Capstone and will hopefully become useful in your career. Most immediately, it may become handy as a quick reference for your Midterm and Final! Some students find it very useful and spend a few slides summarizing the asynch material while others learn different ways and only had the minimum 4 takeaways. Either is fine and will earn you full credit for this question.

End of For Live Session: Unit 1

DataScience@SMU