

Homework 4

Insert Name

Math 141, Week 4

Due: 11:59pm, Friday February 18

Instructions

Work through the problems below and submit this document as a knitted .pdf to the Math 141 S22 Wells Lecture [gradescope page](#).

For each problem, put your solution between the bars of red stars.

Acknowledgements

If you work with a classmate, please write a note acknowledging this.

Exercise 1

Identify the flaw in reasoning in the following scenarios. Explain what the individuals in the study should have done differently if they wanted to make such strong conclusions.

- a. Students at an elementary school are given a questionnaire that they are required to return after their parents have completed it. One of the questions asked is, “Do you find that your work schedule makes it difficult for you to spend time with your kids after school?” Of the parents who replied, 85% said “no”. Based on these results, the school officials conclude that a great majority of the parents have no difficulty spending time with their kids after school.

- b. A survey is conducted on a simple random sample of 1,000 women who recently gave birth, asking them about whether or not they smoked during pregnancy. A follow-up survey asking if the children have respiratory problems is conducted 3 years later, however, only 567 of these women are reached at the same address. The researcher reports that these 567 women are representative of all mothers.

- c. An orthopedist administers a questionnaire to 30 of his patients who do not have any joint problems and finds that 20 of them regularly go running. He concludes that running decreases the risk of joint problems.

Exercise 2

In a study of the relationship between socio-economic class and unethical behavior, 129 University of California undergraduates at Berkeley were asked to identify themselves as having low or high social-class by comparing themselves to others with the most (least) money, most (least) education, and most (least) respected jobs. They were also presented with a jar of individually wrapped candies and informed that they were for children in a nearby laboratory, but that they could take some if they wanted. Participants completed unrelated tasks and then reported the number of candies they had taken. It was found that those in the upper-class level took more candy than did those in the lower-class level.

- a. Identify the population of interest.

- b. Identify the the variables and specify their types.

- c. Identify the sample in the study.

- d. Identify the research question.

- e. Comment on whether the results of the study can be generalized to the population and if the findings of the study can be used to establish causal relationships.

Exercise 3

Researchers are interested in determining whether coaching for college admissions exams has an effect on exam scores. The researches contact 100 high school junior students who have registered for the SAT exam, offering to provide 20 hours of free SAT coaching by knowledgeable tutors. Students are required to take one SAT practice exam before receiving coaching, and one practice exam after coaching concludes. 60 of the 100 students agree to be coached and complete the 20 hours of coaching. The remaining 40 students did not complete coaching, but still reported scores for a first and second practice test. Data on the scores for these students is stored in the **SAT** dataframe and can be loaded by running the following code chunk. (For this problem, it's **not** necessary to understand what each line of code below is doing)

- a. Does this investigation represent a random experiment or an observational study? Explain.

- b. Explain why the researchers are interested in measuring the difference in first and second exam scores for both coached and non-coached students.

- c. Create a data visualization comparing each student's first and second scores, using an appropriate aesthetic to distinguish coached and non-coached students. Describe any trends observed.

- d. Explain why would would expect to see a relationship between pre- and post-coaching scores, regardless of the effect of coaching.

- e. If coaching had no effect **on average** for SAT scores, what trends would we expect to observe in the scatterplot? Explain.

- f. Based on the information given in the statement of this exercise, do you think it would be reasonable to generalize any conclusions for this investigation to the population of all high school juniors who plan to take the SAT? Why or why not?

Exercise 4

This exercise uses the same data described in the the previous exercise.

- a. Add a new column to the **SAT** dataframe called **difference** consisting of differences in scores for each student.

- b. Compute the mean and standard deviation of the difference in scores for each student among non-coached students and among coached students.

- c. Create a data visualization showing the distribution of the difference in scores, for both the coached and non-coached students.

- d. If coaching had no effect **on average** for SAT scores, what trends would we expect to observe in the histograms? Explain.

- e. Based on your answers to the preceding parts of these exercises, does this data suggest that coaching can improve exam scores for this group of 100 students?
