HW1 – Math 58B

due: Thursday, January 29, 2026

your name here

```
library(tidyverse)
```

Assignment Summary (Goals)

- (Actually no R code for this assignment!! You should still compile using the Quarto format.)
- experiments vs observational studies & confounding variables

Q1. Eat better, feel better?

In a public health study on the effects of consumption of fruits and vegetables on psychological well-being in young adults, participants were randomly assigned to three groups: (1) diet-as-usual, (2) an ecological momentary intervention involving text message reminders to increase their fruits and vegetable consumption plus a voucher to purchase them, or (3) a fruit and vegetable intervention in which participants were given two additional daily servings of fresh fruits and vegetables to consume on top of their normal diet. Participants were asked to take a nightly survey on their smartphones. Participants were student volunteers at the University of Otago, New Zealand. At the end of the 14-day study, only participants in the third group showed statistically significant improvements to their psychological well-being across the 14-days relative to the other groups.[1]

(a) What type of study is this?
(b) Identify the explanatory and response variables.
(c) Comment on whether the results of the study can be generalized to the population.
(d) Comment on whether the results of the study can be used to establish causal relationships.

---

[1] From OpenIntro Statistics, exercise 1.41. Tamlin S Conner et al. "Let them eat fruit! The effect of fruit and vegetable consumption on psychological well-being in young adults: A randomized controlled trial". In: PloS one 12.2 (2017), e0171206.

(e) A newspaper article reporting on the study states, "The results of this study provide proof that giving young adults fresh fruits and vegetables to eat can have psychological benefits, even over a brief period of time." How would you suggest revising this statement so that it can be supported by the study?

Q2. Have a Nice trip

An area of research in biomechanics and gerontology concerns falls and fall-related injuries, especially for elderly people. Recent studies have focused on how individuals respond to large postural disturbances (e.g., tripping, induced slips). One question is whether subjects can be instructed to improve their recovery from such perturbations. Suppose researchers want to compare two such recovery strategies, lowering (making the next step shorter, but in normal step time) and elevating (using a longer or normal step length with normal step time). Subjects will have first been trained on one of these two recovery strategies, and they will be asked to apply it after they feel themselves tripping. The researchers will then induce the subject to trip while walking (but harnessed for safety), using a concealed mechanical obstacle.[2]

You will need to access the Randomizing Subjects applet (which simulates a study as described above).

Note: while gender is not a binary characteristic in humans, there are times when understanding trends in body dimensions (e.g., center of gravity) is useful for understanding biomechanics. We use gender here as an insufficient proxy and not as a claim that all people will fit this model nor as a claim that gender is necessarily the right proxy measure.

(a) One way to design an experiment for this study would be to assign the eight females to use the elevating strategy and the 16 males to use the lowering strategy. Would this be a reasonable strategy? If not, identify a better method for deciding who uses which strategy.

Use the applet [http://www.rossmanchance.com/applets/2021/sampling/Subjects.html] to randomly assign the participants to the two treatment groups. Repeat the assignment over and over until you get a sense for how the different variables (gender, height, BMI, "gene") are / can be distributed across the two groups.

(b) Does random assignment always equally distribute/balance the men and women between the two groups? Give the range of possible discrepancies for the gender imbalance.

(c) Is there a tendency for there to be a similar proportion of men in the two groups? How are you deciding? What does this tell you about the plausibility of any later difference in the two groups being attributed to females having better balance?

---

[2]From ISCAM, Inv 3.4

(d) Prior research has also shown that the likelihood of falling is related to variables such as walking speed, stride rate, and height, so we would like the random assignment to distribute these variables equally between the groups as well. In the applet, use the pull-down menu to switch from the sex variable to the height variable. The dotplot now displays the differences in average height between Group 1 and Group 2 for these 200 repetitions. In the long-run, does random assignment tend to equally distribute the height variable between the two groups? Explain.

(e) Suppose there is a "balance gene" that is related to people's ability to recover from a trip. We didn't know about this gene ahead of time, but if you select the "Reveal gene?" button and then select "gene" from the pull-down menu, the applet shows you this gene information for each subject and also how the proportions with the gene differ in the two groups. Does this variable tend to equalize between the two groups in the long run? Explain.

(f) Suppose there were other "x-variables" that we could not measure such as BMI, stride rate, or walking speed. Select the "Reveal both?" button and use pull-down menu to display the results for BMI. Does random assignment generally succeed in equalizing this variable between the two groups or is there a tendency for one group to always have higher results for BMI? Explain.

(g) Is the fact that the people volunteered to participate a "confounding variable" in this study? Explain.

(h) Suppose the "tripping" study was able to rule out chance as a mechanism for the differences seen in the two strategies. What conclusion would you draw? For what population? What additional information would you need to know?

Q3. AstraZeneca

In the *Wired* article about the AstraZeneca vaccine they mention collecting information about age (i.e., one of the variables which was measured was the age of the participant).[3]

(a) If the vaccine was randomly allocated to the participants, can age be a confounding variable? (Remember, to be a confounding variable you need to have an associating with both the explanatory variable of interest and the response variable of interest).

(b) Could age be the causal mechanism behind why the **vaccine** is working? Explain.

(c) Why did they collect information about age?

───────────────────────────

───────────────────────────

[3]https://www.wired.com/story/the-astrazeneca-covid-vaccine-data-isnt-up-to-snuff/

HW & Lab assignments will be graded out of 5 points, which are based on a combination of accuracy and effort. Below are rough guidelines for grading.

Score & Description

5 points: All problems completed with detailed solutions provided and 75% or more of the problems are fully correct.

4 points: All problems completed with detailed solutions and 50-75% correct; OR close to all problems completed and 75%-100% correct. An assignment will earn a 4 if there is superfluous information printed out on the assignment.

3 points: Close to all problems completed with less than 75% correct

2 points: More than half but fewer than all problems completed and > 75% correct

1 point: More than half but fewer than all problems completed and < 75% correct; OR less than half of problems completed

0 points: No work submitted, OR half or less than half of the problems submitted and without any detail/work shown to explain the solutions.

General notes on homework assignments (also see syllabus for policies and suggestions):

- please be neat and organized, this will help me, the grader, and you (in the future) to follow your work.

- be sure to include your name on the assignment

- please include at least the number of the problem, or a summary of this question (this will also be helpful to you in the future to prepare for exams).

- for R problems, it is required to use R Markdown. You can write out other problems with pencil and combine pdf as appropriate.

- please do not print errors, messages, warnings, or anything else that makes your homework unwieldy. You will be graded down for superfluous printouts.

- in case of questions, or if you get stuck please don't hesitate to email me or DM on Discord! The sooner (and more often) questions get asked, the better for everyone.