Homework 1

Nora Quick

2021-09-27

Instructions

Use the R Markdown version of this file to complete and submit your homework. Items in **bold** require an answer. Make sure you change the author in the header to your own name.

Conceptual Questions

1. A study found that individuals who have large yards tend to have pets more often than individuals who do not have large yards.

    a) **Can cause and effect be inferred? Why or why not?** No. This is an obervational study and because of that we cannot make any casual inference on this information.

    b) **List two possible confounding factors that may be contributing to the difference.** One confounding factors could be income. Keeping pets are expensive and depending on the kind of pet (hamster, dog, horse) the cost could be relatively high. Another confounding factor could be house/property size. How many pets can fit within the location?

2. An experiment was performed in which mice were randomly assigned to two groups. One group was fed diet A and the other group was fed diet B. All environmental factors remained the same across both groups. After three months, the scientist measured the weight of the mice. It was found that the mice fed diet A weighed much less on average than the mice fed diet B. **Can cause and effect be inferred? Why or why not?** Yes. We can make a cause and effect inferrence because this is a randomized experiment and therefore we don't need to worry about confounding variables.

3. Random samples of people from New York and Texas are invited to participate in a study comparing income of the two geographic groups. Volunteers participate in the study and their income for the last three years is recorded. **In order to make inference to the population of all New Yorkers and all Texans, what must we assume? Why?** I believe that this would technically be labeled as a randomized experiment, however, this also has volunteers participating. Because of this we can't make an outright population inference. We must assume that there is a wide enough representation of income within the volunteers. We need to assume this because otherwise the data would be skewed.

Inference in the wild

Your task is to find a news article reporting the results of a scientific study and then complete the questions below based on this article.

Article headline: The effect of a halo-type structure on neck muscle activation of an open-cockpit race car driver training under qualifying conditions

1. **Does the headline of the article imply population inference, causal inference, neither or both?** The wording of this headline (and the following article) indicate to me that this is a population inference. I say this because there is clearly a found effect on the drivers and they clearly had to find this effect by studying the population of drivers that use the halo in their driving.

2. **What inference is justified by the study?** The inference is that there is an effect on the driver's neck due to the halo coming down into the vision of the driver. They found this by testing both with and without the halo and acknowledge things such as number of laps drive, length of driving session, temperature, wind speed, etc. It was found that different sections of the neck of a driver were effected in increasing and decreasing ways.

R foundations practice

1. Consider the R function `sample()`.

   a) Open the help page for `sample()`. **What does the "Description" section say `sample()` does?** " 'sample' takes a sample of the specified size from the elements of 'x' using either with or without replacement."

   b) **What are the names of the arguments to `sample()`?** The names of the main arguments are "x", "size", "replace" and "prob".

   c) **Which arguments to `sample()` are optional and what are their default values?** The opetional arguments are "replace" which is set to 'False' and "prob" which is set to 'Null'.

2. **Create a numeric vector called `subjects` that contains the integers 1 to 100.**

```
subjects <- 1:100
```

3. **Use `sample()` to draw a sample of size 10 from `subjects` (without replacement) and save this sample in an object called `sampled_subjects`.**

```
sampled_subjects <- sample(x = subjects, size = 10)
```

4. **Use subsetting to extract the first five elements of `sampled_subjects`.**

```
sampled_subjects[1:5]
```

```
## [1] 53  1 43 25 72
```

5. **Create a character vector called `treatments` that contains ten values: "A" repeated 5 times and "B" repeated 5 times.**

```
treatments <- c("A", "A", "A", "A", "A", "B", "B", "B", "B", "B")
treatments
```

```
##  [1] "A" "A" "A" "A" "A" "B" "B" "B" "B" "B"
```

6. **Use `sample()` to draw a sample of size 10 from `treatments` (without replacement) and save this sample in an object called `treatment_assignment`.**

```
treatment_assignment <- sample(x = treatments, size = 10)
treatment_assignment
```

```
##  [1] "B" "A" "B" "A" "A" "B" "B" "A" "A" "B"
```

7. Now consider the steps above in relation to the ideas of "selection of units at random" and "allocation of units to groups".

   a) **Which steps could be used to select units at random? Which object describes the population and which the sample?** Step 3 is the step where a subset of a group is taken. This could be used to grab units at random. The subjects are the population and the size/sampled_subjects is the sample.

   b) **Which steps could be used to allocate units to groups at random? Which object describes which unit gets which treatment?** Step 6 could be used to randomly allocate units to groups. If we connected this with step 3 then the units would be the sampled_subjects and the object that describes which unit gets which treatment is treatment_assignment.