

# Homework 1

Rui Gao

2021-10-03

## 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. Because the treatment “whether an individual own a large yard” is not assigned by the experiment holder. This is an observational study.
  - b) **List two possible confounding factors that may be contributing to the difference.**
    1. allergy, etc. A yard owner may feel uncomfortable with normal pets (cats, dogs). So, maybe there're no relationship between the size of yard and the amount of pets.
    2. Maybe there are some rules about noisy or mental health (scaring). Hence, no matter how large yard a person has, the person is not allowed to raise a pet in the person's community.
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, the treatment was assigned by the scientists. The other factors were the same, and even some factors are uncontrollable, the possibility of they happened on two groups were the same.
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?** The volunteers should be honest and randomly selected by the holders (the holders pick them up and they answered, not they actively tell others. Because they may want to show off if they are rich or earn rewards from the holders if they are poor or like the rewards). The reason they tell their income should not relate to the income.

## 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:** \_\_\_\_\_ “WHY EATING LESS MEAT MATTERS FOR THE PROTECTION OF THE AMAZON. The majority of today’s fires in the Amazon are caused by converting the forests into cattle ranches through the slash and burn technique”

1. **Does the headline of the article imply population inference, causal inference, neither or both?** Be specific about what language implies which inference.

It implies the causal inference because “Less meat” could help us protect the forest. The more forest we destroyed and more pastures we build, the wildfire would happen more on Amazon. But it also says “majority”, so, maybe both? I need to read the whole article before I give the judgement.

2. **What inference is justified by the study?** Justify your answer by including parts of the article that report details of the study crucial to identifying the scope of inference. If the article doesn’t provide enough information, specify what additional information is required.

It’s an observational study. Because the reporters or experiment holders could not decide where be slashed and burned. The language (mater) puts more emphasis on human’s responsibilities, especially the human who eats a lot of meat. So, yes the causal inference is justified. The language, “In Brazil, it is calculated that the majority of the 45 million hectares, or 62% of the total area cleared is to raise cattle”, tells us it’s not a population inference. The author has the data inspecific.

## 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()`?** `x`, `size`, `replace`, `prob`
  - c) **Which arguments to `sample()` are optional and what are their default values?** `replace` and `prob`. In default, `replace = FALSE`, `prob = NULL`
2. **Create a numeric vector called `subjects` that contains the integers 1 to 100.**

```
subjects <- c(1:100)
typeof(subjects)
```

```
## [1] "integer"
```

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

```
sampld_subjects <- sample(subjects, 10)
sampld_subjects
```

```
## [1] 53 1 43 25 72 15 54 46 20 38
```

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

```
sampled_subjects[c(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")
```

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(treatments, 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”.
- Which steps could be used to select units at random? Which object describes the population and which the sample?** Step 3. Subjects means population and `sampled_subjects` means sample.
  - Which steps could be used to allocate units to groups at random? Which object describes which unit gets which treatment?** Step 6. By the order, the experiment holders could decide the assignments. For example, the #78 gets the “B” treatment... until the #62 gets the “A” treatment.