STA130 - Week 1 Problem Set (Winter 2024)

Nathalie Moon and Josh Speagle

Instructions

How do I hand in my solutions and how do I check my work

You will submit your solutions (.Rmd and .pdf) on MarkUs at the following link: https://markus4.teach.cs.toronto.edu/2024-01/courses/1 Submissions are due at 11:59pm on Thursdays; see (Quercus page)[https://q.utoronto.ca/courses/341136/pages/course-schedule-and-materials] for the specific deadline for each problem set.

Usually when you do an assignment, you don't find out whether your answers are correct until *after* the deadline, when you get your grade back. However, using MarkUs, you can submit your work before the deadline and run tests to check your solutions!

Note: Some parts of some questions may not be covered by tests in MarkUs, but you're still responsible for reviewing the posted solutions and make sure you understand them.

What to do if a test fails on MarkUs

- Take a deep breath! Your work won't really be graded until the deadline, so start early to make sure you have lots of time to resolve issues before the deadline.
- Read the message to get hints about what the problem is. For example "variable X not present" means that you may have a typo in the name of the variable we're looking for - re-read the question carefully and make sure you're following the instructions.
- Search on Piazza to see if other classmates have encountered a similar error (and if not, consider posting a screenshot of the error message)
- Come to TA or instructor office hours with your issue

Question 1

In this question, you'll use R as a calculator to get familiar with the kinds of operations it can do.

(a) What is 3257 times 494. Save the answer in the variable below called Q1a.

```
# Replace NULL with your answer below
Q1a <- NULL
```

(b) What is 112 divided by 4? Save the answer in the variable below called Q1b

```
# replace NULL with your answer below
Q1b <- NULL
```

(c) What is the sum of all positive integers from 1 to 10? Save the answer in the variable below called Q1c.

```
# Replace NULL with your answer
Q1c <- NULL
```

(d) What is 0.05 cubed (that is, the third power of 0.05)? Save the answer in the variable below called Q1d.

```
# Replace NULL with your answer
Q1d <- NULL
```

Question 2

In this question, you'll experiment with the `length()` function. Run the code chunk below (using the green arrow in the top right corner of the gray block) and answer the questions below.

```
# Below, the courses object is a vector with the list of the courses a student is taking this semester
courses <- c("STA130", "MAT135", "ECO101", "CS121", "PHY100")
length(courses)
```

```
## [1] 5
```

(a) Write one sentence describing What the `length()` function does?

Your answer:

(b) Suppose the student decides to take an additional course: BIO100. Create a new vector with all the courses from before, but also this new course. Save the answer in the variable below called Q2 (replace NULL with your answer).

```
Q2 <- NULL
Q2
```

```
## NULL
```

Question 3

For this question we will work with data about the TV show Avatar: The Last Airbender.

a) The name of the data set is `avatar.csv`. Load the data using `read_csv()` and save it under the name "avatar".

```
# Tip: don't forget to put quote marks around the name of the dataset inside the function
avatar <- NULL
```

b) We have learned two functions this week that let us quickly get an idea of our data. Apply both of them to the avatar data.

c) Based on your answer to b) answer the following:

- How many observations does the `avatar` data frame include? Save your answer in the R object below called `num_observations` (replace NULL with your answer).
- How many variables are measured for each observation? Save your answer in the R object below called `num_variables` (replace NULL with your answer).
- What is the name of the third variable in the `avatar` tibble? Save your answer in the R object below called `third_variable_name` (replace NULL with your answer)
- What is the name of the third variable in the `avatar` tibble? Save your answer in the R object below called `third_variable_name` (replace NULL with your answer). Hint: when your answer is a word, make sure to put it in quotation marks.
- What is the value of the `character` variable for the first observation in the `avatar` tibble? Save your answer in the R object below called `first_value_of_character` (replace NULL with your answer). Hint: when your answer is a word, make sure to put it in quotation marks.

```
num_observations <- NULL
num_variables <- NULL
third_variable_name <- NULL
first_value_of_character <- NULL
```

Question 4

In this question, you will consider another example of survivor bias. In 1987, a study published in the Journal of the American Veternary Medicine Association reported that cats that survived falls from higher floors in high-rise buildings suffered fewer injuries than cats who fell from lower floors (e.g. more than 6 stories vs less than 6 stories). While this finding seems counterintuitive, the authors suggested that this might be due to the cats relaxing and repositioning themselves for a relatively safer landing after they reached maximum speed during their fall. The data for this study was collected from cats who suffered falls and were brought to veternary clinics.

(a) Is this sample representative of all cats who suffer falls?

(b) Do you expect that the average number of injuries for cats suffering falls calculated from this sample will be close to the true value?