

Worksheet 1

Author: Ethan Meyers

Discussants:

=====

The purpose of this worksheet is to practice R and to review some of the concepts that we covered in class. Please submit a compiled pdf with your answers to Moodle by midnight on Sunday September 16th. I also highly recommend getting started on this worksheet early since this is the first worksheet and you might run into difficulties on it.

Some functions you will find useful to help you complete the worksheet are: `sqrt()`, `c()`, `sum()`, `table()`, and `prop.table()`, `bar()` and `pie()`. You will also use a function I wrote called `get_sprinkle_sample()`. Please use the Slack channel `#worksheet_01` for any questions that come up, and conversely, if you have completed the worksheet, help others out in the class by answering their questions on Slack.

Part A: Practicing R

Please answer the following questions to get practice with some basic R functions. Make sure you are fluent in answer all these questions since future classes will build on this knowledge.

Exercise 1.1: Let's get started by using R as a calculator. Use R to calculate the square root of 21.32, and then divide this number by 2.71 .

```
# delete the below lines and replace with the correct math (2  
# + 3)~2
```

Exercise 1.2: Create a vector with the numbers 7, 15, 18, 3, 5, 12, and 20 in it and assign this vector to an object called `my_vec`. Multiply this vector by 2 and assign it to the object `my_vec2`. Finally, use the `sum()` function to sum all the values in the vector `my_vec2`.

Exercise 1.3: The code below will get a sample of 233 sprinkles and save them in an object called `sprinkle_sample`. Use the `table()` function to convert this sample into a table that has the count of the number of sprinkles of each color and assign the result to an object called `sprinkle_table`. Below report how many sprinkles there are of the color that had the most sprinkles. Also answer the question, if another sample of sprinkles was taken, do you think the same color would have the most sprinkles? Try it and find out!

```
# the following code creates an object called sprinke_sample  
# that has a sample of 100 sprinkles - do not change this  
# code!  
source("/home/shared/intro_stats/get_sprinkle_sample.R")  
my_sample <- get_sprinkle_sample(72) # check the sample size!
```

```
# use the table() function to create a table that has the
# count of sprinkles of each color and assign the results to
# an object called sprinkle_table.
```

```
# create another sample here and repeat the above steps
```

**** Answer:**** [Which color had the most sprinkles? Also try this a second time and report how many sprinkles were in the category with the most sprinkles].

Exercise 1.4: Now use the function `prop.table()` to convert the counts of sprinkle colors in `sprinkle_table` into the proportion of sprinkles of each color. Save the results to an object called `sprinkle_proportions`. Report what is the proportion of pink sprinkles.

Answer: [Report the proportion of sprinkles that were pink.]

Exercise 1.5: Above we created samples that had $n = 233$ sprinkles. Now repeat the process several times by calculating the proportion of red sprinkles (\hat{p}_{red}) using sample sizes of $n = 1, 10, 100, 1,000,$ and $10,000,$ and $100,000$; i.e., use the functions `get_sprinkle_sample()`, `table()`, `prop.table()` etc. once for each sample size. Do the estimates of the proportion of red sprinkles (\hat{p}_{red}) change a lot depending on the sample size n ? Which sample size (n) seems to give you the most accurate estimate?

Answer:

Part B: Conceptual questions

Please answer the following questions to make sure you have a solid understanding of some of the concepts covered in class. We will build on these concepts over the course of the semester so it is very important you have a solid understanding now. Note: some symbols that might be useful are \hat{p} , \bar{x} , π , μ .

Please define the following terms

In class we discussed the concepts of 1) A *parameter* and 2) A *statistic*. Please describe:

1. Which concept is associated with a sample and which is associated with a population (parameter or statistic?)
2. Which concept is associated with Plato and which with the shadows
3. Which concept are we most interested in to get at the Truth (with a capital T)

4. Which concept is associated with a categorical distribution and which is associated with a bar chart?
5. Which of these vary depending on the particular sample of data collected?
6. When discussing proportions, what is the symbol used to denote the parameter and which one to denote the statistic.

Answers

1. A X is associated with a sample and a Y is associated with a population (replace X and Y with “parameter” or “sample”)
2. etc. for these questions (delete this and fill in)
- 3.
- 4.
- 5.
- 6.

Reflection

Do you feel confident in these concepts and with R. Please briefly reflect below on how you feel and if you have any remaining questions (and if so, bring them up at the start of next class).