

# Problem Set 1

## Introduction to R

Casey Bree

### Problem Set 1

Please complete the following questions in R in a quarto document.

#### Exercise 1 – Assignment, arithmetic, logical expressions

1. Assign `x` and `y` to take values 3 and 4.
2. Create a new variable `z` as the product of variables `x` and `y`.
3. Write code to calculate the square of 3. Assign this to a variable `three_squared`.
4. Write a logical expression on whether `x` is greater than 10. When might you need to filter data based on a condition?
5. Write a logical expression testing whether `x` is *not* greater than 10.

#### Exercise 2 – Sequencing

1. Generate vectors containing the numbers 100, 101, 102, 103, 104, and 105 using 3 different methods (e.g., `c()`, `seq()`, `:`). In what scenarios might each method be most convenient?
2. Generate a sequences of all **even** numbers between 0 and 100.
3. Create a descending sequence of numbers from 100 to 1, and assign it to a variable.

#### Exercise 3 – Data generation and basic statistical analysis

1. Generate a sample of 100 observations drawn from a normal distribution with a mean of 10 and a standard deviation of 2. How is this type of random sampling useful in statistical analysis?
2. Calculate the mean of this generated sample. How does this sample mean relate to the population mean of the distribution?

3. Calculate the difference between the sample mean and the population mean. Why the discrepancy?
4. Repeat steps 1--3 with a sample of 10,000. Did the difference between the sample mean and the population mean decrease? Will this always be the case?

#### **Exercise 4 – Work with real-world data**

For this exercise, we will work with real-world data. Please download the CenSoc-Numident Demo file (as .CSV) and accompanying codebook (as PDF) from:

[<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QVDPM9>](<https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/QVDPM9>).

1. Read in the dataset using `read_csv()` from the `tidyverse` package
2. How many columns does that dataset have?
3. How many rows the dataset have?
4. What are the column names? What type of research question could we use this dataset for?