Practical Exercise 1

The goals of this exercise are to (a) to review specific statistical and methodological concepts (i.e., types of variables, reliability and validity), (b) to use R for some basic probability and sampling theory questions, and c) generate some and test some hypotheses using correlation analysis.

### Part A - Knowledge to Discuss Statistical and Methodological concepts

*In your own words*, briefly (a few sentences) describe the difference between the following:

# Task 1. Probability and statistics

# Task 2. Sample statistics and population parameters

# Task 3. Hypothesis and research question

# Task 4. Null hypothesis and alternative hypothesis

# Task 5. The alpha level (α) and the p-value

# Task 6. A one-sided and two-sided test

# Task 7. Type I and type II errors

# Task 8. Correlation and causation

### Part B: Using R for Basic Probability and Sampling Theory Questions

# Task 9. I am going out to a restaurant for sushi, but the sushi I receive is totally up to the chef. He is choosing from salmon, tuna, avocado, eel, krab, or tofu. My favorite is salmon. I go crazy and eat 50 pieces of sushi. What is the probability that 30 of the pieces of sushi I receive are salmon?

# Task 10. Create a variable that contains all the letters of the alphabet to represent your population.

# Task 10)a) Draw 20 random samples of size 10 without replacement and paste the results of your samples.

# Task 10)b) Add a bias to your sampling such that it only draws vowels from the alphabets and then draw 20 samples of size 10 and paste them.

# Task 11. Generate a variable that contains the mean of 10,000 samples of size 200 from a normal distribution with a mean of 30 and a standard deviation of 5. What is the mean value of your sampling distribution of the mean? Now perform the same, but change your sample size to 20. What is the mean now? Why are these different?

### Part C: Using R for Correlation Analyses

For this part of Practical Exercise #1, Tasks indicate things that you need to complete in R/R Studio. You will need to complete the tasks and then provide an answer to the questions that are indicated with a number. Depending on the question, feel free to type your answer or copy the output from R.

# Task 12. Load data from the “The World Almanac and Book of Facts 1993” (OEF8\_dataset.csv)”.

# Task 13. Inspect the data by looking at the first few entries and the last few entries in the dataset. Use the function head() which shows the first N rows of the data frame. Use tail() function that shows the last N rows. You can also open the full dataset and check it out. Be sure to take note of the variables. They include information about the participant’s country, life expectancy, number of people per television, number of people per physician, and then life expectancy for females and males.

# Task 14. State a null hypothesis and an alternative hypothesis about the correlation between life expectancy and the number of televisions per person.

# Task 15. Run a correlation test in R to test your hypothesis

# Task 16. Report the results here. Be sure to include the r-value, the p-value, and the 95% confidence interval. What can you conclude from your results? Do these results seem plausible to you?

# Task 17. State a null hypothesis and an alternative hypothesis about the correlation between life expectancy and the number of physicians per person.

# Task 18. Run a correlation test in R to test your hypothesis.

# Task 19. Report the results here. Be sure to include the r-value, the p-value, and the 95% confidence interval. What can you conclude from your results?