

Assignment 11

Data Science in BMCE | Summer 2020

1 Introduction

You will continue using RStudio Cloud (<https://rstudio.cloud>) for this assignment.

Please refer to assignment 10 for submission details.

2 Rock-Paper-Scissors

Rock-Paper-Scissors, also called Roshambo, is a popular two-player game often used to quickly determine a winner and loser. In the game, each player puts out a fist (rock), a flat hand (paper), or a hand with two fingers extended (scissors). In the game, rock beats scissors which beats paper which beats rock. The question is: Are the three options selected equally often by players? Knowing the relative frequencies with which the options are selected would give a player a significant advantage. A study observed 119 people playing Rock-Paper-Scissors. Their choices for their first turn are shown in the table.

Option Selected	Frequency
Rock	66
Paper	39
Scissors	14
Total	119

- What is the sample in this case? What is the population? What does the variable measure? Is the variable categorical or numeric?
- You want to perform a hypothesis test to support that someone is more likely to choose rock than the other two options. State the null and alternative hypothesis for testing this claim that individuals are more likely to pick rock than either paper or scissors. For the null hypothesis, think about what happens if they are all equally likely to be picked. Make sure that you include (and define) a parameter in the null and alternative hypothesis.
- What kind of statistical test would you use? Perform this statistical test in R. You will find the statistics section of this [cheat sheet](#) useful. Please be noted that the Chi-squared test is not covered in the cheat sheet. You can find more information on Chi-squared test [here](#).

- (d) Use a significance level of 0.05 (i.e., alpha value). The alpha value is the threshold value that we measure p-values against. It tells us how extreme observed results must be in order to reject the null hypothesis of a significance test. The value of alpha is associated with the confidence level of our test. For results with a 95% level of confidence, the value of alpha is $1 - 0.95 = 0.05$. State your conclusion.

3 Asthma Attack

It has been hypothesised that the probability of having an asthma attack on a particular day may vary across the days of the week. Baibergenova et al. (2005) report a longitudinal study in which the number of emergency department admissions due to asthma were recorded in Ontario between April 2001 and March 2004. Suppose the data were as in the table below:

	Number of visits (Total = 99,054)	(%)
Day of the week		
Monday	15,483	(15.6)
Tuesday	13,748	(13.9)
Wednesday	13,508	(13.6)
Thursday	12,941	(13.1)
Friday	12,511	(12.6)
Saturday	14,082	(14.2)
Sunday	16,781	(16.9)

- (a) Which graphical method (histogram, bar chart, stem-and-leaf plot, boxplot, pie chart) would best display the above data?
- (b) Which of the following best describes the null hypothesis?
- The probability of having an asthma attack in Ontario on a particular day varies across the days of the week.
 - The probability of having an asthma attack in Ontario on a particular day does not vary across the days of the week.
 - The mean number of asthma attacks in Ontario per day varies by day of the week.
 - The number of asthma attacks in Ontario per day varies by day of the week.
 - There is no variation in the daily number of asthma attacks in Ontario.
- (c) Under the null hypothesis, what would be the expected number of emergency admissions due to asthma in Ontario on a Tuesday?

- (d) Select the correct statistical test. Compute the test statistic using R. With a significance level of 0.05, state your conclusion.

Reference:

Cubranic, Davor, Bruce Dunham, and Djun Kim. "On-line homework in probability and statistics: WeBWork incorporating R." Contributed paper at 9th International Conference on Teaching Statistics, Flagstaff, Arizona, USA, 2014.

Baibergenova, Akerke, et al. "Effect of gender, age, and severity of asthma attack on patterns of emergency department visits due to asthma by month and day of the week." *European journal of epidemiology* 20.11 (2005): 947-956.

4 Correlation

Use the built-in iris dataset for studying the correlation between two parameters. You can load the dataset using the command.

```
data("iris")
```

Let's assume we have two random variables X and Y that are Gaussian distributed. The null hypothesis will be that these two random variables have zero correlation:

$$H_0 : \rho(X, Y) = 0$$

- (a) Say you hypothesize that irises with wide petals will also have skinny sepals, and skinny petals will coincide with wide sepals.

```
X = iris$Petal.Width
```

```
Y = iris$Sepal.Width
```

What is the alternative hypothesis here for $\rho(X, Y)$?

- (b) Consider a significance level of $\alpha = 0.05$, sample size $n = 150$, and a degree of freedom ($n - 2 = 148$), what is the critical value for the t statistic? Give the R command (one line) for computing this. What does it return?
- (c) What is the value for the t statistic above?
- (d) Give the R command (one line) for computing the p -value. What value does it return? What is your conclusion on this correlation?