

Lab Challenge 01 – Summarizing and Presenting Data

Due Date: 11:59 pm, three days after class

Each challenge is graded out of 2 points:

- 0 points – no attempt or no progress to a solution
- 1 point – not fully completed or completed with major errors
- 2 points – fully completed with at most a small error

Deliverables

1. A single pdf document containing your solutions to the challenges you completed.
2. An R script (.R extension) containing a *complete* script used to generate your results.

Preliminary Steps

1. *Install* R and RStudio on your laptop.
2. *Save* the Excel file “MATH_1350_Lab_01_Data.xlsx” to your computer. This is a data set based on data from students in MATH 1350 in a previous term.
3. *Cleanse* the data set if there are any obviously wrong entries. For example, if someone wrote “samsing” change it to “Samsung”.
4. *Import* the data into RStudio.

Completing the Lab Challenges

1. The challenges are listed below.
2. Record your R commands in a script as you go.
3. Save the images generated from your R script as jpegs.
4. Write your answers and paste your images into a single document (Word doc).
5. Ensure that your document has your name and ID number.
6. Save your document as a pdf file.
7. Submit both files (.R and .pdf) to the Lab Challenge 01 assignment folder on Learning Hub.

You will likely find it helpful to download the R demo script for Unit 01. (Find it on Learning Hub.)

Challenges

1. In this challenge you will examine the variable *Phone.Brand*.
 - a. Generate a *pie* chart showing the *relative* frequency of different phone brands for students in this class. Ensure that:
 - The chart has an appropriate title and labels.
 - The pie segments are each of a different colour.
 - b. Calculate the sample proportion \hat{p} of students who use an Apple iPhone. Is \hat{p} a good estimate of the proportion p of all BCIT students who use an iPhone? Why or why not?

2. In this challenge you will examine the variable *Age* for students in this sample.
- Generate a *histogram* showing the frequency distribution of *Age* for this class. Ensure that:
 - The class boundaries are 17.5, 19.5, 21.5, 23.5, and so on...
 - There is a tick mark for each year.
 - There are appropriate labels on the axes and a main title.
 - The histogram bars are *pink* (not that pink is better than any other colour...).
 - In your Word doc, complete the following table using the class limits specified:

Class Limits	Upper Boundary	Class Mark	Frequency	Relative Frequency (%)
18-19				
20-21				
22-23				
24-25				
26-27				
28-29				
30-31				
32-33				
34-35				

(For each class, you need to determine the upper boundary and the class mark too.)

- Calculate the probability that a student in this set is at least 22 years old.
3. Calculate the following summary statistics for the variables *Siblings* and *Income.Goal*. Record the values in a table.

sample size (n) – excluding missing values

sample mean (\bar{X})

median (Q_2)

mode

range

sample standard deviation (s)

variance (s^2)

skewness (Sk)

30th Percentile

interquartile range