Lab Challenge 02 – Normal Distribution and Z-Scores

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 – challenge not fully completed or completed with major errors
- 2 points – challenge fully completed with at most a small error

Deliverables

1. A single pdf document containing your solutions to the challenges you completed.
2. An RStudio file (.R extension) containing a *complete* script used to generate your results.
    - The script must run without errors!

Preliminary Steps

1. Import the Excel file "F2021_MATH_1350_Data.xlsx" into RStudio. (Look in Data Sets on LH.)
2. Complete the Challenges below. Record your R commands in a script as you go.
3. Insert images into a Word doc or Google doc along with your written answers.
4. Ensure that your document has your name and ID number.
5. Save your document as a pdf file and submit it to the Learning Hub assignment folder.

Challenges

1. Plot a histogram for the z-scores based on the variable $X =$ Income.Goal for the population of all students in MATH 1350, where you:
    - use the default "Sturges" method for the breaks.
    - plot probability density on the y-axis.
    - include suitable labels and give it a title.
    - make the rectangles red.

    Then answer the questions below.
    a. What fraction of students in MATH 1350 have Income.Goal values more than one standard deviation above the mean?
    b. If Income.Goal followed a *normal* distribution, what fraction of students would have an income goal more than one standard deviation above the mean?
    c. Find any *unusual* values of Income.Goal.

2. Using the variable $X =$ Siblings for students MATH 1350, plot a histogram where:

    - each whole number has its own class and its own "tick mark"
    - plot frequency on the y-axis
    - use the RGB hex code #CF4371 as the colour

    Then calculate the following:

    a. The quartiles Q_1, Q_2, and Q_3.
    b. The percentiles P_5 and P_{90}.
    c. Determine if there are any outliers for this variable. If so, list them.

3. Using the variable $X =$ Income.Goal, plot side-by-side boxplots (each one extending horizontally) comparing the salary expectations for Males and Females in MATH 1350. Label axes and provide a suitable title. What conclusion can you draw from this graph?