Project 1 - Due Friday, October 18

The focus of this project is to expose you to some of the features in Microsoft Excel that can be used to create some of the objects and to compute some of the values that we have seen in Chapters 1-3 of our text. In particular, in this project you will be asked to create various charts, like a frequency table, a bar chart, an ogive, and a pie chart. Additionally, you will be asked to compute the mean, median, mode, and standard deviation for a set of data.

Although Microsoft Excel is certainly not the only application available (and perhaps is far from the best), it is one of the more common tools used in the "real-world". So, an exposure to some of its computing and displaying capacity can be valuable.

Note that Microsoft Excel is available in computer labs campus-wide. If you do not have a computer with Excel installed, please plan to complete this project in one of the labs or use Microsoft 365 now available for free for CU Boulder students through OIT.

The project is due by class on Friday, October 18. There are two components: A worksheet which you will complete using your results from Excel, and an Excel workbook containing all required components. The worksheet is to be handed in class. The Excel workbook will need to be submitted via the Canvas Project 1 File Upload.

You must work in groups of 2 or 3, handing in one worksheet and submitting one Excel worksheet for the group. On Wednesday, October 2, you will be asked to specify your group members. You will need to upload your Excel worksheet to Canvas by class time on Friday, October 18. You will hand in the worksheet in class on that date, as well.

For those of you unfamiliar with Excel, useful screencasts will be provided in Canvas. There are also many resources online for help and tutorials on the tasks required of you in this project.

To begin. In the folder for Project 1, you should find an Excel file for your section. Download the file to your computer. The data sheet contains the data that you will analyze. The data highlighted in green should NOT be altered in anyway. Changing values or even sorting the data may result in inaccurate results later.

Additionally, the explicit locations for the placement of each of the items you are instructed to create/compute are labeled and highlighted in yellow. You are welcome to use other space in your data sheet if needed (for intermittent computations or analysis), but to earn credit for the required components, they must be in their specified location. Do NOT relocate any of these cells. Even adding a row or column to the data set can result in a loss of points.

Lastly, all objects and computations completed in the Excel file should be done *robustly*, meaning that when any of the original data values are changed, the objects and values that you designed will update accordingly. Part of your grade for this project will be based on the required components updating when some of the data is changed. Use cell references rather than hard-coding any numeric values into your formulas to achieve this robustness.

The data in the spreadsheet are responses from 650 students collected for MATH 2510.

1. Complete the frequency table for the PREFERRED COLOR responses from the sample of 650 respondents in the data sheet. {Hint: COUNTIF}

2. Create a Bar chart to display the distribution of preferred colors. Make sure to label the Bar chart properly.

3. Complete the frequency table for the CLASS STANDING responses from the sample of 650 respondents in the data sheet. {Hint: COUNTIF}

4. Create a Pie chart to display the distribution of class standings. Make sure to label the Pie chart properly.

5. Compute the mean height of the sample of 650 respondents in the data sheet. {Hint: AVERAGE}

6. Determine the five-number summary of the heights of the sample of 650 respondents in the data sheet. {Hint: QUARTILE}

7. Compute the mean height of <u>only the female</u> respondents in the sample of 650 respondents in the data sheet. {Hint: AVERAGEIF}

8. Compute the mean height of <u>only the male</u> respondents in the sample of 650 respondents in the data sheet. {Hint: AVERAGEIF}

9. Complete the FREQUENCY column in the frequency table for the heights of <u>only the female</u> respondents in the sample of 650 respondents in the data sheet. {Hint: COUNTIFS}

10. Unfortunately, there is not (yet) a function called STDEVIF. So, in order to compute the standard deviations for the heights of <u>only the female</u> respondents in the sample of 650 respondents in the data sheet, please use the defining formula and the following steps.

    (a) Complete the column in the frequency table containing the values of $x - \bar{x}$. You have already computed $\bar{x}$, so simply reference the cell containing that value.

    (b) Now complete the column containing the values $(x - \bar{x})^2$.

    (c) Taking in consideration the frequencies of each height, determine the sample variance.

    (d) Lastly, use the sample variance to determine the sample standard deviation.

11. Compute the mean number of coin flips that landed heads as reported by these 650 students. {Hint: AVERAGE}

12. Compute the sample standard deviation for the number of coin flips that landed heads as reported by this sample of 650 students. {Hint: STDEV}

13. Determine the upper and lower bounds of the 84% Chebyshev interval for the number of heads, and then determine what percentage of the outcomes that actually lie within that range. (If it is not 84% or greater, something is wrong.) {Hint: COUNTIFS}

14. Complete the binned frequency table with 10 bins { '1 to 5', '6 to 10', '11 to 15', ..., '46 to 50'} for the number of heads flipped as reported by this sample. Then add to the table values for Relative Frequency, Cumulative Frequency, and Relative Cumulative Frequency. (Be mindful of your class boundaries and how to correctly communicate this to the Excel function.) {Hint: COUNTIFS}

15. In a **single graph**, create an Ogive for the number of heads flipped in each trial into a **single graph**.

Use the information found in the Excel workbook you just created to answer the questions on the following worksheet. Bring the worksheet to class on Friday, October 18 to turn in. Upload your Excel workbook to Canvas by class time on that same day.

Project 1 Worksheet - Due Friday, October 18

NAME 1:_____

NAME 2: _____

NAME 3: _____

1. What was the most popular answer to the "Which of the following colors do you most prefer?" question? How many survey respondents answered with that color?

2. How many students responded that they are Sophomores? Is the sampling method used to collect the data likely a good method to infer the proportion of all students at CU that are Sophomores? Why or why not?

3. What is the mean height of all 650 students? What is the average of the female and male mean heights? You should find that your two values are not exactly equal. Explain why this is the case.

4. What is the standard deviation of the heights of the <u>female</u> respondents in the sample? Explain whether the standard deviation of the heights of all 650 respondents is likely to be greater than or less than the standard deviation for the female respondents alone and why. (You need not compute the standard deviation of the entire 650 students to answer this, but if you do, then an explanation like "It is greater, because the computed value is larger" is not sufficient. Your explanation should relate that you understand what the standard deviation measures.)

5. Although you were not asked to create a histogram for the coin flip data, what does the frequency table indicate about the shape of the distribution? (unimodal, bimodal, symmetric, skewed) How does the ogive also illustrate the shape of the distribution?