MATH1P98 - Assignment #1 - Section #1,2

Due: **Monday, September 26 @ 11:59 pm**

Students are expected to complete all questions on the assignment. However, only a subset of questions will be considered for marking. Marks will be deducted for incomplete assignments.

Assignment submissions must be neat, legible, written on one side of the page only, and questions must be submitted in order. A cover page must be attached to the front of the assignment. (See sample cover page on Sakai)

Please submit answers to the following questions to your assignment drop box by 11:59pm of the due date. The text is *Elementary Statistics using Excel* (5^{th} Edition), by M. Triola.

Correction: Question 1(e) should be 'upper class limits'

---

1 a) Use the data for female white blood cell counts found in Appendix B (WHITE column, pg. 785, or FBODY.xls spreadsheet found in the Resources section of Sakai). Starting with a lower class limit of 4.0 (pg. 53) and class width of 1.6, create a frequency table **by hand**. Your table should include columns for frequency, relative frequency and cumulative frequency.

b) Using the frequency table, what percentage of observations is at least 7.2 units?

c) Create a frequency histogram of the table in part (a), using EXCEL and the procedure described below:

- On your worksheet, add a new column anywhere with the upper class limits in order, starting with 5.5. You should have a column of data with 40 rows and column of upper class limits (bins) with 6 rows (*Note: Excel needs the upper class limits, not lower class limits*).
- Load the Data Analysis ToolPak: Click the *File* tab, and then click *Options*. Select *Add-Ins* on the left, click *Analysis ToolPak* in the list on the right, and then in the *Manage* box, select *Excel Add-ins*. Click *Go*. Select *Analysis ToolPak* and click *Ok*.
- Click the *Data* tab in the top menu and click *Data Analysis*
- Select *Histogram* and click *Ok*.
- Click the *Input Range* box and select the column of data with your mouse
- Click the *Bin Range* box and select the upper class limit column data with your mouse
- Select *New Worksheet Ply* and *Chart Output*, and then click *Ok*

Once your histogram is created, edit the main title and x-axis label to be something more informative than the default labels. You can also remove the 'More' column by deleting it from the Excel sheet. Include a printout of the histogram only with your assignment (you do not need to include the other output).

d) Using the histogram, describe the distribution of the data. What is the shape of the distribution? Are there any outliers? If so, approximately how many are there?

e) Repeat part (c), except this time change the number of upper class limits so that there are 15 intervals. To do this, follow the same procedue in part (c), with now 16 rows of upper class limits. Create a histogram. Do this 3 more times, each time selecting the number of intervals to be 10, 5 and 3 respectively. Write a brief summary of what you notice about the shape of the histogram and the number of outliers as the number of intervals changes. What is the disadvantage of having too many or too few intervals? Note: You do not need to include these histograms in your solutions.

2 a) Using the ***Harry Potter*** data from Section 3-3 #9 (pg. 125), calculate the mean and the median. Show your calculations and include units in your response. Based on these two measures, how would you describe the distribution of the data?

b) Calculate the variance and standard deviation by completing a table like the one provided below. Use four decimal places throughout your calculations, and include units in your response. What does $\Sigma(x - \overline{x})$ approximately equal? Explain.

x	$(x - \overline{x})$	$(x - \overline{x})^2$
$\vdots$	$\vdots$	$\vdots$
Total:		
	$s^2 =$	
	$s =$	

c) Calculate a z-score for the observations 58, 29, 4. Show your work. Based on the z-score values, which of these would be considered "unusual" values, and why?

d) If you had to pick two values from the ***Harry Potter*** data set such that the variability was minimized, which two values would you pick?

3 a) A data set for blood platelet counts of women has the summary statistics $\overline{x} = 280$ and $s = 65$. Assuming that the data follows a bell-shaped distribution, what percentage of women has blood platelet counts between 150 and 410?

b) What are the blood platelet counts that are 1.5 standard deviations above and below the mean? What percentage of women has a blood platelet count within this interval?

c) Consider the Harry Potter data set from Exercise 2. Would it be reasonable to use the Empirical Rule on this data set? What about Chebyshevs Theorem? Why or why not?

4 a) Return to the data set used in Exercise 1 (female white blood cell counts data). Using Excel and the procedure described on page 134, find P10 and P84.

b) In which percentile does the data value 10.2 approximately fall? How would you interpret this percentile in the context of female white blood cell counts?

c) Use EXCEL and the procedure described on page 138 to find the 5-Number Summary for this data set. Include the output in your solutions. Use the Procedure on page 139 to construct a **regular** boxplot by hand for this data set (*Note: not a modified boxplot*). On your regular boxplot, label the values from the 5-Number Summary.

5) Obtain a sample of Suncor Energys financial data for a one-year period by the following procedure:

- Go to the website: https://ca.finance.yahoo.com/
- In the top left corner, in the box where it says Enter Symbol, type in **SU.TO**. Click on *Look Up*. You will be directed to the summary page for Suncor Energy Inc.
- In the left side menu, click on the link for *Historical Prices*. You can now select a date range to look up.
  - Start with January 1, and end with December 31. Select any year between 2000 and 2013. For example, if you wanted to look up the price information for the year 2007, you would enter January 1, 2007 and December 31, 2007.
  - On the right hand side, select Weekly. Click Get Prices.
- Export the data to a spreadsheet by scrolling down to the bottom of the page and clicking on Download to Spreadsheet.
- Select one of the columns labeled Open, High, Low, or Close to be your data set (ignore the columns for Date, Volume and Adj. Close). Copy and paste your selected column into a new spreadsheet.
- Complete each of the following. Include your output, and remember to show your work where applicable.
  a) Use EXCEL to find the mean, median, and standard deviation.
  b) By hand, calculate the interquartile range. Show your work.
  c) By hand, use the procedure for Outliers described on page 141 to determine whether your data set contains any outliers. Show your work.
  d) Copy your summary statistics onto the same spreadsheet as your data. Print this page and include it in your solutions. (For the sake of saving paper, once you have finished your calculations you may wish to condense your data into two shorter columns).