MATH 106 Statistics Project Instructions (Revised May 2016)

For this assignment, you will implement a project involving statistical procedures. The topic may be something that is related to your work, a hobby, or something you found interesting. If you choose, you may use the example described below.

The project is made of the following tasks. Each task must be addressed in the statistics project report in order to qualify for full credit:

- **Task 1**: identify:
    - Yourself (student's name)
    - name of project
    - purpose of project
- **Task 2**: Conduct Data Collection. Provide:
    - Raw data used (sample size must be at least 10 individual raw scores)
    - source of the data
- **Task 3**: Calculate Measures of Central Tendency and Variability:
    - median, sample mean, range, sample variance, and sample standard deviation (show work)
- **Task 4**: Frequency Distribution. Provide
    - Raw data in frequency table format (1st column of value "intervals", and 2nd column shows "frequency": number of scores falling within each interval)
- **Task 5**: Histogram:
    - Create histogram using frequency table constructed in Task 4
    - NOT a vertical bar chart!
    - *x* – axis must show intervals, *y* – axis must show frequencies
- **Task 6**: Compare Raw Data Distribution to "Standard" Normal Distribution. Using your raw data gathered in Task 2 and the sample mean and sample standard deviation calculated in Task 3, calculate:
    - percentage of your raw data falling within one standard deviation of the mean;
    - percentage of your raw data falling within two standard deviations of the mean;
    - percentage of your raw data falling within three standard deviations of the mean
- **Task 7**: Communicating Evaluation, Analysis, Results, and Conclusions. Provide two to three paragraphs that:
    - interpret your statistics and graphs;
    - answer whether your percentages calculated in Task 6 indicate that your data distribution (shown in the histogram created in Task 5) is the same as the 68/95/99.5% "standard" normal distribution? Be sure to explain why you think your data distribution does or does not match the "standard" normal distribution.
    - relate to the purpose of the project

****

If you choose, you may use the following example for your data.

- Purpose: Compare the amount of sugar in a standard serving size of different brands of cereal. (You may instead choose to compare the amount of fat, protein, salt, or any other category in cereal or some other food.)
- Procedure: Go to the grocery store (or your pantry) and pick at least 10 different brands of cereal. (Instead of choosing a random sample, you might purposely pick from both the "healthy" cereal types and the "sugary" ones.) From the cereal box, record the suggested serving size and the amount of sugar per serving. The raw data is the serving size and amount of sugar per serving for each of the 10 boxes of cereal. Before calculating the statistics on the amount of sugar in each cereal, be sure you are comparing the same serving size. If you use a serving size of 50 grams, you must calculate how much sugar is in 50 grams of each cereal. For example, if the box states that there are 9 grams of sugar in 43 grams of cereal, there would be 50 times 9 divided by 43, or 10.5 grams in 50 grams of cereal. The result of this simple calculation (for each of 10 boxes) is the data you will use in the project statistics and charts.

****

For Task 6: Instructions for Calculating Percentage of Raw Data Falling Within 1, 2, and 3 Standard Deviations of Mean:

1. Determine sample mean $\bar{x}$ and sample standard deviation s for your raw data set (you had to do this to complete Task 3 so they should already be done)
2. Determine the raw score “bounds” for data falling within 1 standard deviation of the mean by subtracting 1 standard deviation from the mean to get the lower bound. Then, add 1 standard deviation to the mean to get the upper bound.
3. Count the number of raw scores in your data set whose values fall between the lower and upper bounds you found in Step 2. Divide that number by n, the total number of scores in your data set, and then multiply the result by 100 to get the percent of raw data falling within one standard deviation of the mean.
4. Now, determine the raw score “bounds” for data falling within 2 standard deviations of the mean by subtracting 2 standard deviations from the mean to get the lower bound. Then, add 2 standard deviations to the mean to get the upper bound.
5. Count the number of raw scores in your data set whose values fall between the lower and upper bounds you found in Step 4. Divide that number by n, the total number of scores in your data set, and then multiply the result by 100 to get the percent of raw data falling within 2 standard deviations of the mean.
6. Now, determine the raw score “bounds” for data falling within 3 standard deviations of the mean by subtracting 3 standard deviations from the mean to get the lower bound. Then, add 3 standard deviations to the mean to get the upper bound.
7. Count the number of raw scores in your data set whose values fall between the lower and upper bounds you found in Step 6. Divide that number by n, the total number of scores in your data set, and then multiply the result by 100 to get the percent of raw data falling within 3 standard deviations of the mean.

Example: The following measurements of grams of fat per ¼ - pound serving of 10 different brands of ground beef were made:

17	11	13	15	18
22	26	23	15	30

Calculate the percentage of raw data falling within 1, 2, and 3 standard deviations of the mean:

SOLUTION:

1. Determine sample mean $\bar{x}$ and sample standard deviation s for your raw data set (you had to do this to complete Task 3 so they should already be done). Sample mean $\bar{x}$ = 19.0 and sample standard deviation s = 6.074
2. Determine raw score "bounds" for data falling within 1 standard deviation of the mean:
    a. subtract 1 standard deviation from mean to get lower bound:
        - 19.0 – 6.074 = 12.926

    b. add 1 standard deviation to mean to get upper bound:
        - 19.0 + 6.074 = 25.074
3. Count number of raw scores in your data set whose values fall between 12.926 (lower bound) and 25.074 (upper bound). 7 out of 10 of the raw scores fall between the bounds ("11", "26", and "30" fall outside the bounds). Divide 7 by total number of scores in data set n = 10 , and then multiply result 0.7 by 100 to get **70 percent of raw data falling within 1 standard deviation of mean**.
4. Now, determine raw score "bounds" for data within 2 standard deviations of the mean:
    a. subtract 2 standard deviations from mean to get lower bound:
        - 19.0 – (2 x 6.074) = 19.0 – 12.148 = 6.852

    b. add 2 standard deviations to mean to get upper bound:
        - 19.0 + (2 x 6.074) = 19.0 + 12.148 = 31.148
5. Count number of raw scores in your data set whose values fall between 6.852 (lower bound) and 31.148 (upper bound). All 10 out of 10 of the raw scores fall between the bounds. Divide 10 by total number of scores in data set n = 10, and then multiply result 1 by 100 to get **100 percent of raw data falling within 2 standard deviations of mean**.
6. Now, determine raw score "bounds" for data within 3 standard deviations of the mean:
    a. subtract 3 standard deviations from mean to get lower bound:
        - 19.0 – (3 x 6.074) = 19.0 – 18.222 = 0.778

    b. add 3 standard deviations to mean to get upper bound:
        - 19.0 + (3 x 6.074) = 19.0 + 18.222 = 37.222
7. Count number of raw scores in your data set whose values fall between 0.778 (lower bound) and 37.222 (upper bound). Again, all 10 out of 10 of the raw scores fall between the bounds. Divide 10 by total number of scores in data set n = 10, and then multiply result 1 by 100 to get **100 percent of raw data falling within 3 standard deviations of mean**.

Now you decide how the raw data distribution of "70/100/100" percent within 1/2/3 std deviations of the mean compares with the "68/95.5/99" percent of the "standard" normal distribution?