ME 303 Lab #0
Exercise #1: MS Excel Gaussian Distribution Exercise

OBJECTIVE

This computer exercise is intended to show you how to use the MS Excel spreadsheet software to create a histogram and along with the corresponding Gaussian distribution.

KEY LABORATORY LEARNINGS

- How to perform professional statistical analysis using MS Excel
- How to create suitable statistical plots as would be put in a professional scientific report or presentation

EQUIPMENT

- Computer with MS Excel

SUMMARY

Experimental data is often too large and cumbersome to be analyzed by hand. Therefore, computer software can assist the experimentalist in analyzing data. Furthermore, computer software can assist in creating professional looking plots suitable for placement in scientific reports or presentations.

In this lab session a dataset will be analyzed statistically and suitable professional plots created. Despite the fact that this data set is "small" and can be analyzed by hand, students are required to use Excel functions to create the appropriate data cells. This will prepare students for future labs in which much larger datasets (where it is unreasonable to analyze by hand) will be obtained.

KEYWORDS

Population, sample, random, probability, probability density, array, class interval, class boundary, class mark, class frequency, frequency distribution, relative frequency distribution, histogram, central tendency, (arithmetic) mean, median, mode, variance, standard deviation, accuracy, precision, confidence interval, accuracy, bias, normal (Gaussian) distribution, variance of the means.

PROCEDURE

Part 1: Raw Data

1. Open MS Excel and enter the data from Table 13.1 in the text into two columns in the spreadsheet. The first column should be "Sample Number" and the second column should be "Yield Strength (ksi)". Note: always include units!

Sample Number	Strength (ksi)
1	65.0
2	68.3
3	72.2
4	73.5
5	74.0
6	75.2
7	76.8
8	77.7
9	78.1
10	78.8
11	79.0
12	79.2
13	79.9
14	80.3
15	81.1
16	82.6
17	84.0
18	85.5
19	87.0
20	89.8

2. Create the following parameters and statistics for the sample set:

MINIMUM FOR INTERVALS (ksi)	**60.00**
MAXIMUM FOR INTERVALS (ksi)	**95.00**
POINTS	**20**
AVERAGE (ksi)	**78.40**
STANDARD DEVIATION (ksi)	**6.04**

MINIMUM FOR INTERVALS is round number chosen that is smaller than minimum measured value, MAXIMUM FOR INTERVALS is a round number chosen that is larger than the maximum measured value, POINTS is the number of data points (use Excel "COUNT" function to calculate this value), AVERAGE is the calculated average of the sample set (use Excel "AVERAGE" function to calculate this value), and STANDARD

DEVIATION is the calculated standard deviation (use Excel "STDEV" function to calculate this value).

Part 2: Frequency Distribution Table

1. In a new worksheet, create the following columns based on the minimum and maximum for intervals and a total of seven intervals:

Group Interval Minimum (ksi)	Group Interval Maximum (ksi)	Group Interval Center Point (ksi)
	60.00	
60.01	65.00	62.50
65.01	70.00	67.50
70.01	75.00	72.50
75.01	80.00	77.50
80.01	85.00	82.50
85.01	90.00	87.50
90.01	95.00	92.50

Note that these columns should all be created using Excel formula functions. In the laboratory your group intervals may be much larger than seven and it is unrealistic to type in the values individually.

2. Use the Excel "FREQUENCY" function to create the "Observations in Group" column:

Group Interval Minimum (ksi)	Group Interval Maximum (ksi)	Group Interval Center Point (ksi)	Observations in Group
	60.00		0
60.01	65.00	62.50	1
65.01	70.00	67.50	1
70.01	75.00	72.50	3
75.01	80.00	77.50	8
80.01	85.00	82.50	4
85.01	90.00	87.50	3
90.01	95.00	92.50	0

Note that the "FREQUENCY" function is an array function and therefore, after entering the formula, you need to press CTRL+SHIFT+ENTER, instead of just ENTER.

You will need to use the Excel "Help" feature (i.e. the "**?**" symbol on the top right).

3. Create the "Relative Frequency" column. Note that this column should be created using Excel formulas.

   You may need to use the "Paste Special" feature.

Group Interval Minimum (ksi)	Group Interval Maximum (ksi)	Group Interval Center Point (ksi)	Observations in Group	Relative Frequency
	60.00		0	0.00
60.01	65.00	62.50	1	0.05
65.01	70.00	67.50	1	0.05
70.01	75.00	72.50	3	0.15
75.01	80.00	77.50	8	0.40
80.01	85.00	82.50	4	0.20
85.01	90.00	87.50	3	0.15
90.01	95.00	92.50	0	0.00

3. In a new worksheet, create a dataset of the ideal Gaussian distribution. This can be done be creating a strength column at even intervals across the range of the data. For example, you may create a column with values 60.0, 60.1, 60.2, …, 94.8, 94.9, 95.0. Using the formula for the Gaussian distribution, as well as the calculated average and standard deviation, create the "Predicted Relative Frequency" column. Note that you may wish to create intermediate columns such as "z", "$-z^2/2$", or "$\exp(-z^2/2)$ to make the formulas easier.

Part 3: Creating the Graphs

1. From the top menu choose "Insert" and then choose the "Insert Column Chart" option under the "Charts" menu. Then choose the "Clustered Column" option under the "2-D Column" heading. A blank white plot should appear on your Excel spreadsheet.

2. The first task is to create the histogram. On this blank white plot, right-click and choose "Select Data." Choose the "Add" option under "Legend Entries (Series)" to select the desired data for the vertical axis. Select the cells corresponding to the Relative Frequency values. Next, choose the "Edit" option under "Horizontal (Category) Axis Labels" to select the data for the horizontal axis, i.e. the "Group Interval Center Point" values. Then choose "OK." You should get a graph that looks like this:

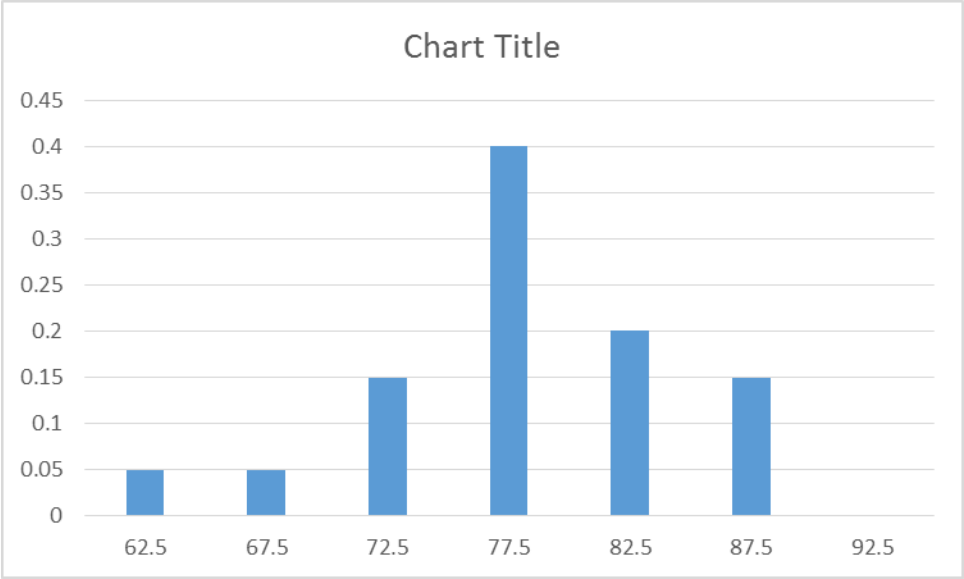
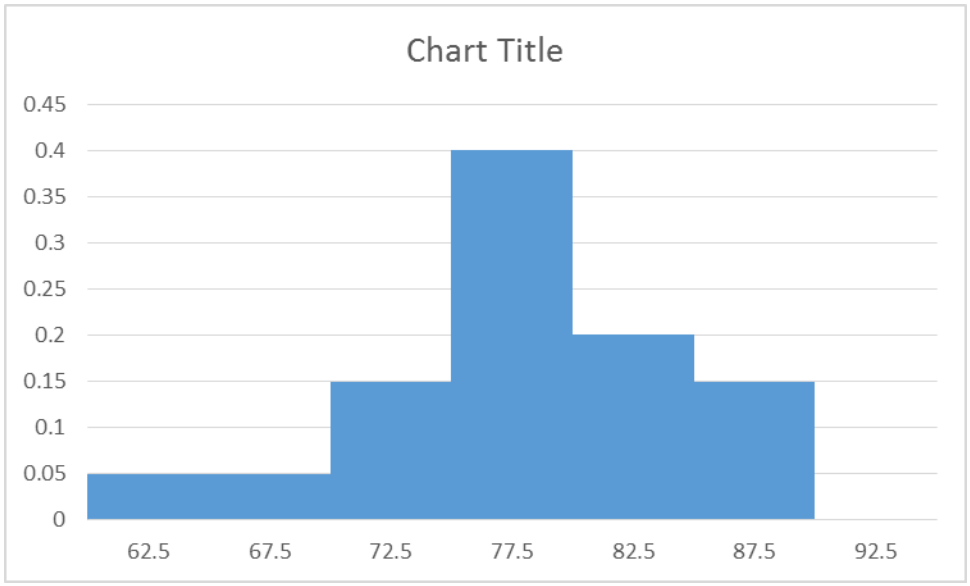


Chart Title
0.45
0.4
0.35
0.3
0.25
0.2
0.15
0.1
0.05
0
62.5
67.5
72.5
77.5
82.5
87.5
92.5

3. The next task is to remove the gaps between the vertical bars on the histogram. Left-click on one of the bars, and then right-click to choose "Format Data Series" and then set "Gap Width" to "0 %." Your graph should look like this:

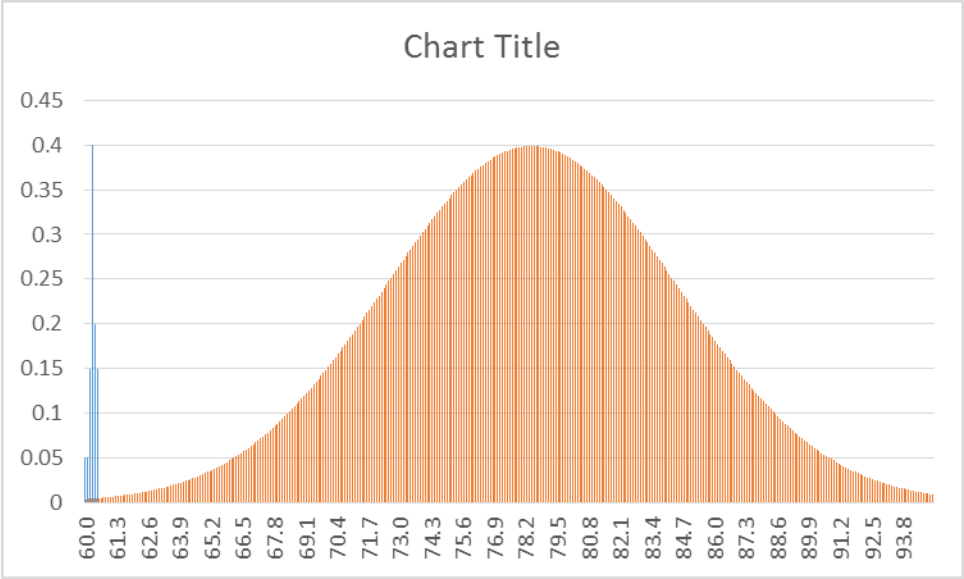


4. The next task is to generate the Gaussian (Normal) Distribution Curve to place it over this histogram.

   Right-click on the graph and choose "Select Data" like before. The same procedure must be followed using the values for this "f(z)" curve that you generated previously.

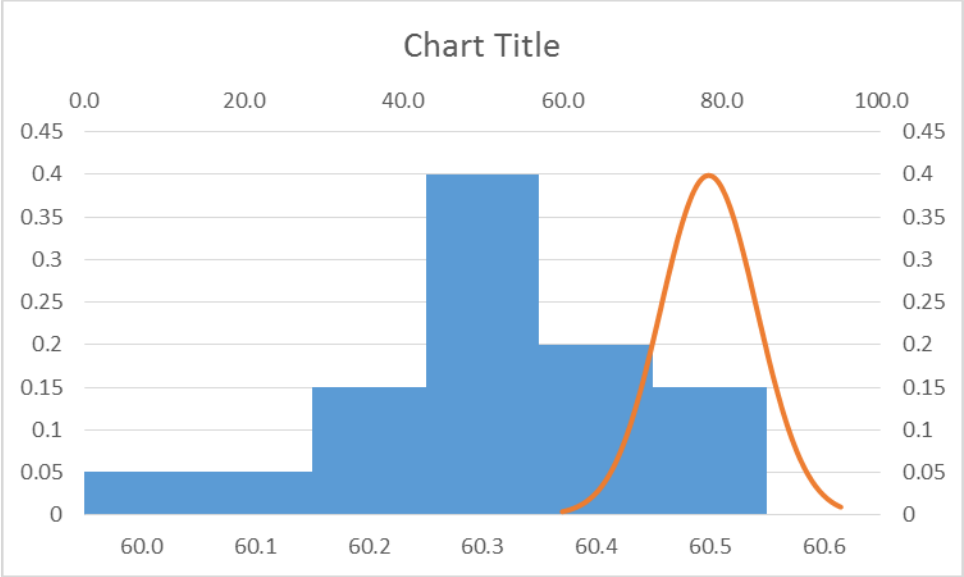
Note: "Ctrl-Shift-↓" can be used to select all numbers in a column after you choose the first data value.

Your graph should like this:



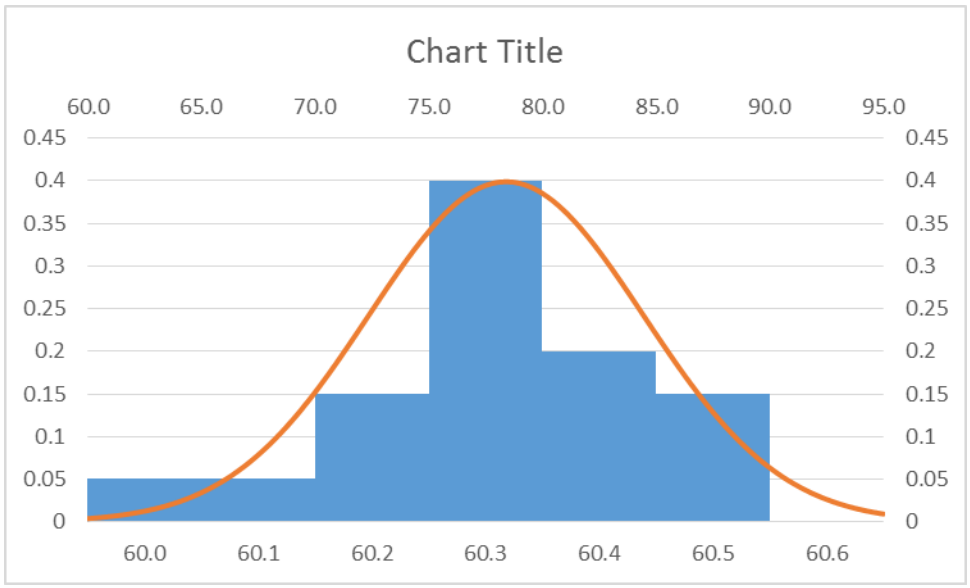
5. Right-click on this new (temporary) histogram and select "Change Series Chart Type" Under "Series2" change the "Chart Type" to "Scatter with Smooth Lines." Select "OK."

Your graph should like this:



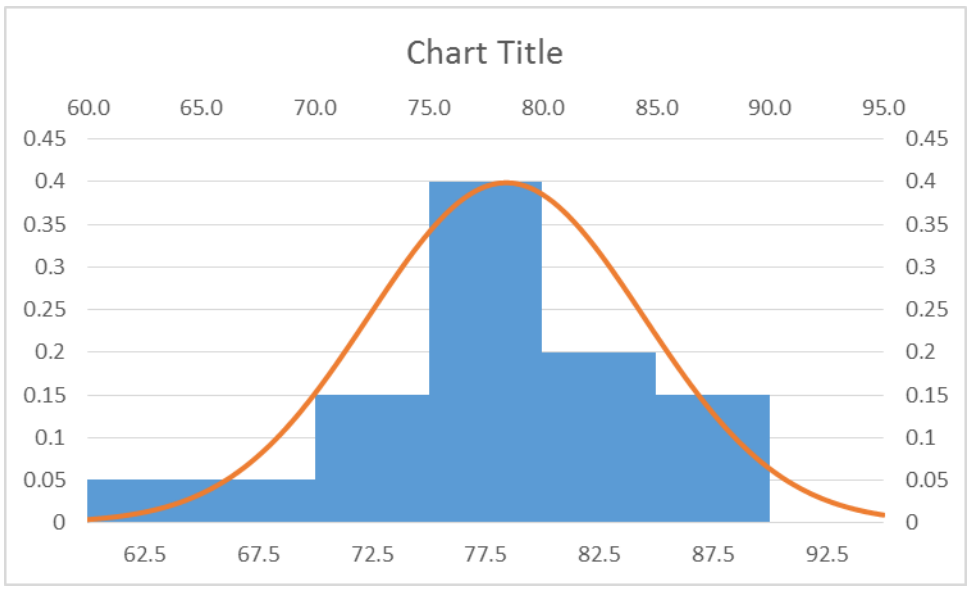
6. The horizontal axis seen on the top is considered the "Secondary Horizontal (Value) Axis." Left-click on this axis and then Right-click to choose "Format Axis." Set the Minimum Bound to "60" and the Maximum Bound to "95."

Your graph should like this:



7. The horizontal axis on the bottom must be re-set to its original values for this histogram. You can right-click on this axis, choose "Select Data" and proceed as before.

Your graph should like this:



8. The last steps are to remove the secondary vertical axis and the secondary horizontal axis.

   a) Left-click on the vertical axis on the right, and choose "+" to activate the "Chart Elements" menu. Left-click on the arrow to the right of "Axes" and remove the "√" mark from the "Secondary Vertical" option.
   At this point, do NOT remove the "√" mark from the "Horizontal Vertical" option.

   b) Left-click on this axis and then Right-click to choose "Format Axis." Left-click on "Labels" and choose "None" to remove this axis from the graph.

   Your final graph should look like this: