# Excel – Engineering Statistics

EGN 1006 – Introduction to Engineering

# Analyzing Data

Engineering analysis usually begins with the analysis of data! Engineers gather data to measure VARIABILITY or CONSISTENCY.

Measured Data can tell you a great deal if you know how to interpret the results. Let Excel do the tedious work for you so that you can focus on the interpretation of results.

# Data Characteristics

There are several commonly used parameters that allow us to draw conclusions about the characteristics of a data set. They are:

| | | | |
|---|---|---|---|
| Mean | Median | Mode | Max |
| Min | Variance | Standard Deviation | |

# Mean, Median, and Mode

- Mean – is the arithmetic average of a data set. It represents expected behavior. AVERAGE( ) is used in Excel
- Median – the value where half of the data falls above and half the data falls below. MEDIAN ( ) is used in Excel
- Mode – the value that occurs with the greatest frequency with in data set. Mode ( ) is used in Excel. If a tie results it will always list the FIRST frequent number it encounters

# Min and Max

The min and Max simple represent the extremities of the data set. In Excel ,the MIN ( ) and Max ( ) functions return these values.

NOTE: The MIN and MAX functions return the values that are the smallest and largest *ALGEBRAICALLY*. They do not return values in terms of MAGNITUDE.
Example: ( -5,-2, 1) ; Min = -5 & Max = 1

# Variance

The variance provides an indication of the degree of SPREAD in the data. The greater the variance, the greater the spread. It is determined by the following formula:

Excel uses the VAR( ) function

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$$

$s^2 = \text{variance}$

$n = \#\text{ of data values}$

$x_i = \text{individual data value}$

$\bar{x} = \text{mean}$

# Standard Deviation

The standard deviation is simply the **square root of the variance.**

$$s = \sqrt{s^2} = \text{standard deviation}$$

So why bother with the standard deviation?

The variance is a much more practical value to have but its UNITS are NOT consistent with the mean, median, or mode. Excel use the stdev ( ) command.

# Analyzing a data set

A car manufacturer wishes to determine how accurately the cylinders are being machined in several engine blocks. The design specification call for a cylinder diameter of 3.500 inches, with a tolerance of +/- 0.005 inches. See next slide for data and worksheet

# Manufacturing data for worksheet question #1

| Sample | Diameter (in) | Sample | Diameter (in) |
|---|---|---|---|
| 1 | 3.502 | 11 | 3.497 |
| 2 | 3.497 | 12 | 3.504 |
| 3 | 3.495 | 13 | 3.498 |
| 4 | 3.500 | 14 | 3.499 |
| 5 | 3.496 | 15 | 3.501 |
| 6 | 3.504 | 16 | 3.500 |
| 7 | 3.509 | 17 | 3.503 |
| 8 | 3.497 | 18 | 3.494 |
| 9 | 3.502 | 19 | 3.499 |
| 10 | 3.507 | 20 | 3.508 |

# Histograms

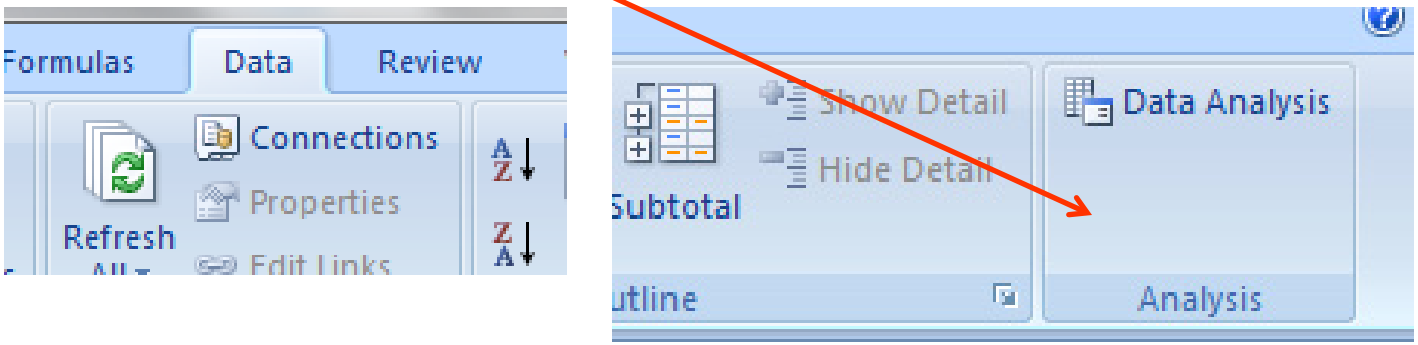
Though the previous statistical characteristics can prove useful in interpreting data, it is often more desirable to the plot the data in a manner that illustrates how the values are distributed within their range. This is called a HISTOGRAM or RELATIVE FREQUENCY plot.

# More on Histograms

To create a histogram, you must first ***subdivide the range of the data into a series of adjacent, equally spaced intervals***. The first interval must begin at or below the smallest value (the min) and the last interval must extend to or beyond the largest data value (the max). These intervals are called CLASS INTERVALS. Then you determine HOW MANY values fall within each interval

# The car manufacturer continued’

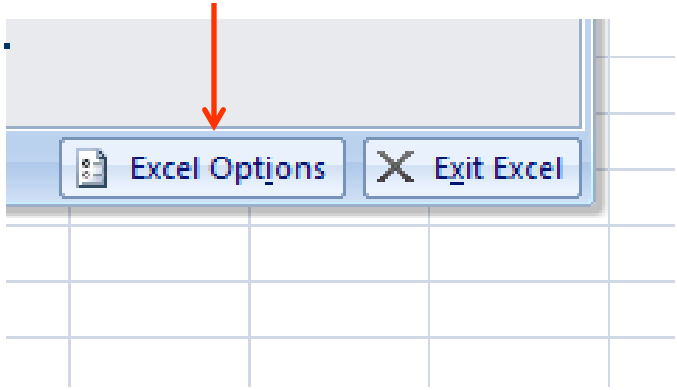
**The histogram feature is found by first choosing the DATA tab. Then choose DATA ANALYSIS from the tool bar.**

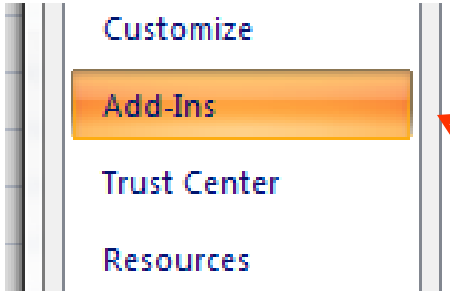



If data analysis is not there, follow the directions on the next slide.

# No data analysis?

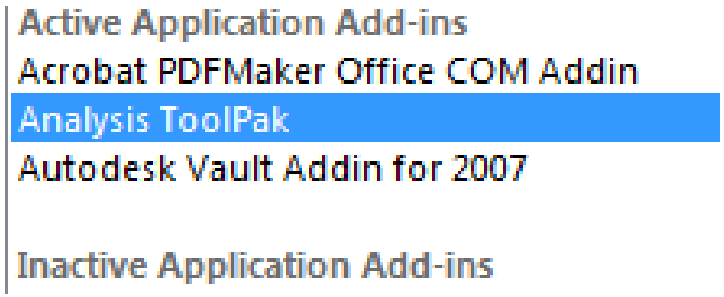

Click the “OFFICE” button in the top left of the page, then click “EXCEL OPTIONS” in the bottom right of the pop-up.



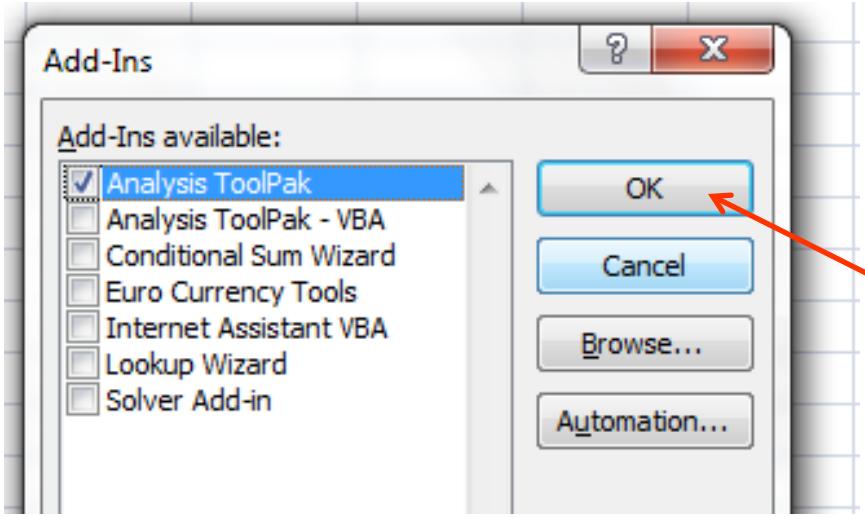




A menu will open and choose “ADD-INS” on the left.

# No data analysis?



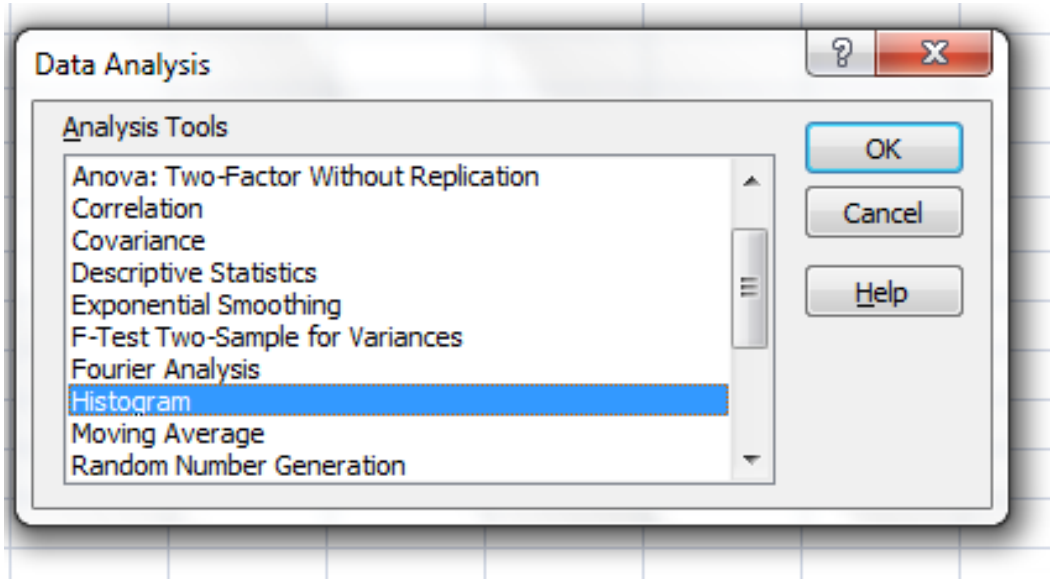

Choose Analysis ToolPak from the list of add-ins (and NOT VBA), then “GO” at the bottom (NOT OK!)

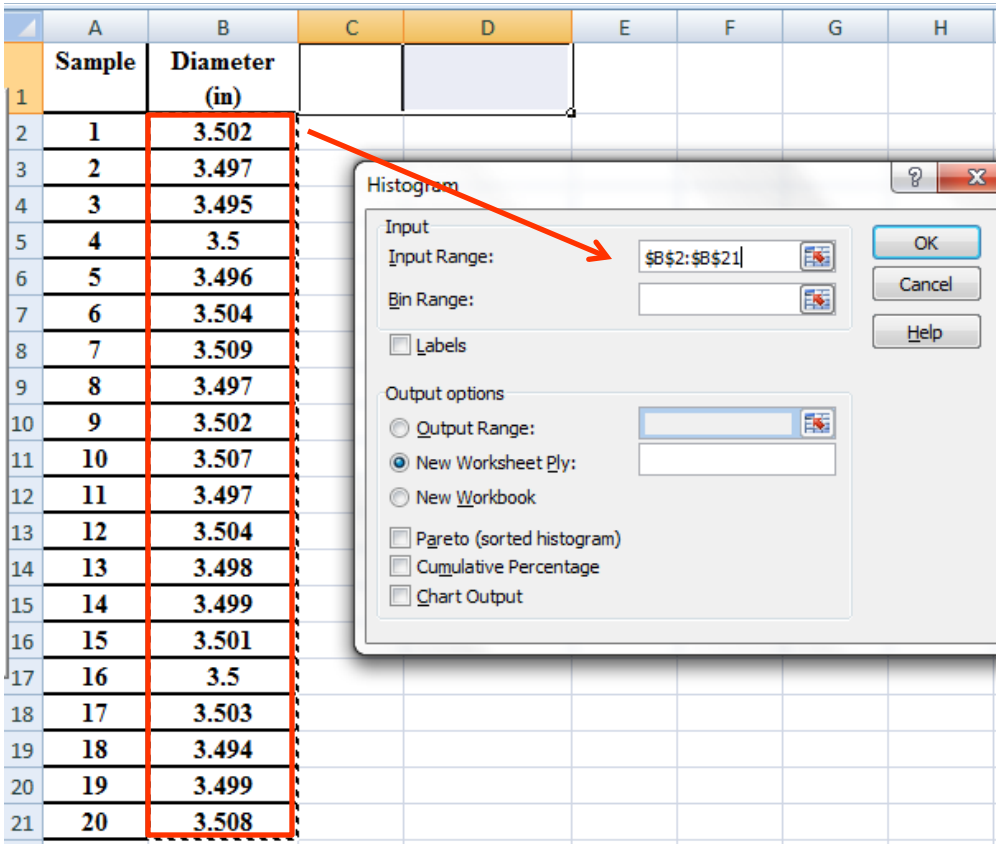

Check “Analysis ToolPak” then OK.

# The car manufacturer continued’

Once data analysis appears in the toolbar, choose Histogram from the list, then OK.



# The input range




An INPUT RANGE – this comes from your data. Click on the Input range box then click on the first cell of the data, hold, and highlight until the last cell is chosen

The data range will automatically fill the input range box if done correctly.

# The Bin Range

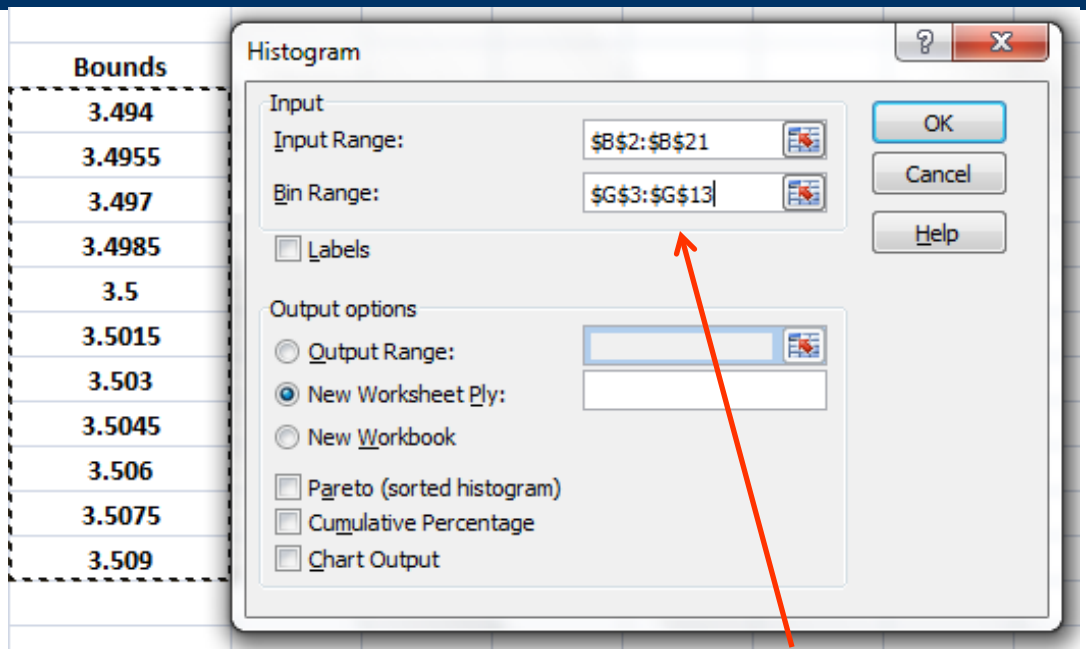
An BIN RANGE – this is the interval bounds or class intervals that you created. Separate your data into 8-10 even spaced intervals and make a separate column called BOUNDS. The best way to do this is to find the min and max. Then find the difference and divide by ten. Add this to the min for the second interval and so on. See next slide for example

# The Bin Range

| | | | | | |
|---|---|---|---|---|---|
| 3.502 | | | | | Bounds |
| 3.497 | | Min | 3.494 | | 3.494 |
| 3.495 | | Max | 3.509 | | 3.4955 |
| 3.5 | | Difference | 0.015 | | 3.497 |
| 3.496 | | Interval width | 0.0015 | | 3.4985 |
| 3.504 | | | | | 3.5 |
| 3.509 | | | | | 3.5015 |
| 3.497 | | | | | 3.503 |
| 3.502 | | | | | 3.5045 |
| 3.507 | | | | | 3.506 |
| 3.497 | | | | | 3.5075 |
| 3.504 | | | | | 3.509 |
| 3.498 | | | | | |
| 3.499 | | | | | |
| 3.501 | | | | | |
| 3.5 | | | | | |
| 3.503 | | | | | |
| 3.494 | | | | | |
| 3.499 | | | | | |
| 3.508 | | | | | |

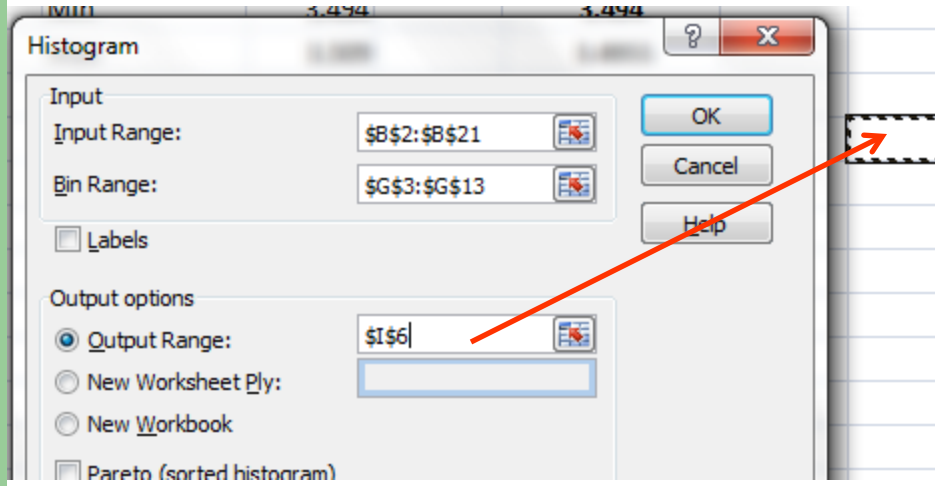
Once the interval width was found it was added to the MIN to get 3.4955. The first 2 bounds were then highlighted and copied down which increase each cell automatically by 0.0015 until we reach the MAX

# The Bin Range (Bounds)




Click inside the Bin Range box then select the BOUNDS by highlighting the data. It will automatically fill the box.

# The Output Range




To see the “frequency” of the data click inside the OUTPUT RANGE box then click on the cell where you want the data to be placed. Then click OK.

# The car manufacturer continued’

Your data should look like this!

| *Bin* | *Frequency* |
|---|---|
| 3.494 | 1 |
| 3.4955 | 1 |
| 3.497 | 4 |
| 3.4985 | 1 |
| 3.5 | 4 |
| 3.5015 | 1 |
| 3.503 | 3 |
| 3.5045 | 2 |
| 3.506 | 0 |
| 3.5075 | 1 |
| 3.509 | 2 |
| More | 0 |

# Making a BAR graph

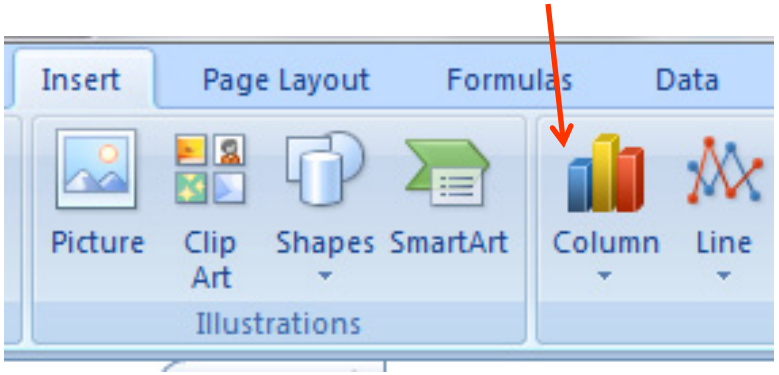
Now we want to GRAPHICALLY display HOW MANY data points fell into each interval. Click on each cell and add the entire interval. So for the first cell we have 3.494. In the formula bar click after the number and add a minus then the next cell # or in this case 3.4955. Don't worry if you think you are subtracting because you are NOT. Do this for each interval

| |
|---|
| 3.494-3.4955 |
| 3.4955-3.497 |

# Making a BAR graph

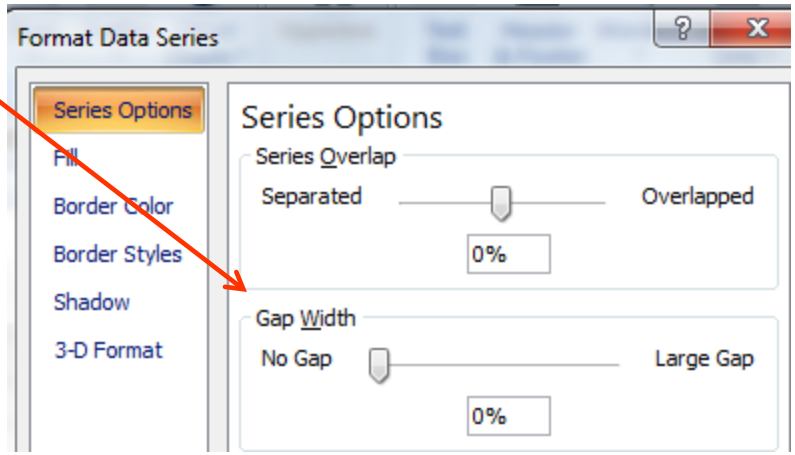
| *Bin* | *Frequency* |
|---|---|
| 3.494-3.4955 | 1 |
| 3.4955-3.497 | 1 |
| 3.497-3.4985 | 4 |
| 3.4985-3.5 | 1 |
| 3.5-3.5015 | 4 |
| 3.50153.503 | 1 |
| 3.503-3.5045 | 3 |
| 3.5045-3.506 | 2 |
| 3.506-3.5075 | 0 |
| 3.5075-3.509 | 1 |
| 3.509-3.5105 | 2 |
| More | 0 |

Your data should look like this. Now highlight all the data, then choose INSERT then COLUMN GRAPH.



# Cleaning up the graph

The first thing we want to do is eliminate the GAPS. Right click on any bar on the graph and choose FORMAT DATA SERIES. Change the GAP WIDTH to 0%.

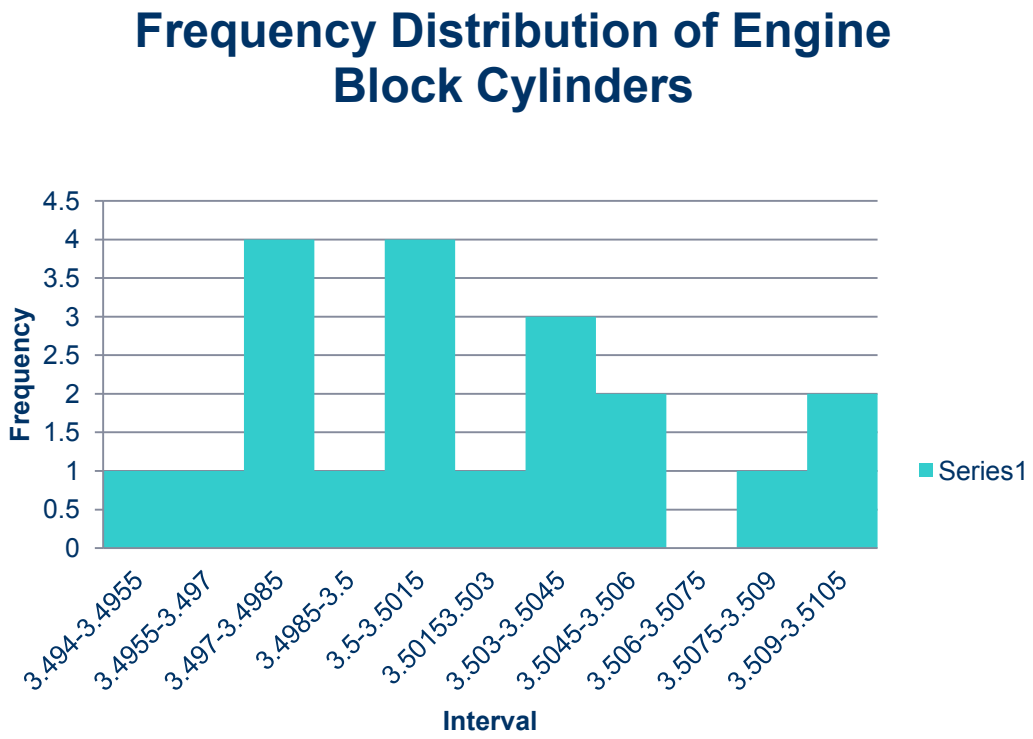


# Graph Layout

Choose CHART LAYOUTS at the top and choose LAYOUT 9. Enter “frequency” on the y-axis and “Interval” on the x –axis. For the title enter Frequency distribution of Engine Block Cylinders.

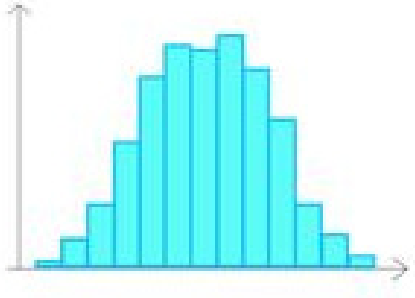
# The graph

Your graph should look like this.




Answer the rest of the questions on the worksheet.

# Drawing inferences



Knowing how to translate the values of the data set gathered from surveys, observations or discovery can help the user monitor a process, determine the profitability of an activity or to analyze the intensity of an element. It also involves determining the acceptable levels and limits, to ascertain which aspects of the processes have to go through further analysis for purposes of improvement.

# Case Study of Call Center's Failure for Rapid Response – Question 2

A call center agency wants to improve its system as an inbound call center and service agency. Although there’s no specific industry standard to which handling of calls should adhere, the agency's goal is to sustain its ability to respond rapidly to customers' calls. Currently, the hold-time it takes for their agents to respond has been falling short of the previous track record of less than a minute.

In line with this, a study was conducted in order to determine if the agency needs more staff to attend to the customers. Fifty (50) inbound calls for technical support took more than a minute before a help-desk technician could attend to the customers' requests for assistance. In fact, it was observed that half of the fifty calls took more than five minutes, which can be gleaned from the following data gathered during the course of the observations.

# Number of Minutes the Customers Were On Hold

- 0-1 minute = This is the rapid-response time by which phone calls should be answered.
- 1-2 minutes = Two (2) customers had hung up their phones.
- 2-3 minutes = Five (5) more customers had hung up.
- 3-4 minutes = Eight (8) customers also gave up on holding.
- 4-5 minutes = Ten (10) customers stayed on the line and were attended to.
- 5-6 minutes = Another set of (10) customers who waited while on hold, had been given assistance.
- 6-7 minutes = Seven (7) more customers had been provided with technical assistance.
- 7-8 minutes = Another set of four (4) customers waited for their turn to be served.
- 8-9 minutes = Three (3) of the customers simply gave up and decided to hang-up.
- 9-10 minutes = One (1) customer's patience paid-off and he was finally served.

# Create a COLUMN Graph

In Excel, set up 2 columns. One for the Bin Range and one for the frequency. This is a bit different from out first example as we can skip over the actual “Histogram” step.

Make the graph and set the GAP WIDTH to 0.

Answer the questions on the worksheet

# A teacher's gradebook: question #3

| Student # | Exam Score | Student # | Exam Score |
|---|---|---|---|
| 1 | 87 | 16 | 71 |
| 2 | 64 | 17 | 41 |
| 3 | 74 | 18 | 77 |
| 4 | 56 | 19 | 74 |
| 5 | 95 | 20 | 56 |
| 6 | 74 | 21 | 79 |
| 7 | 76 | 22 | 90 |
| 8 | 67 | 23 | 47 |
| 9 | 82 | 24 | 44 |
| 10 | 67 | 25 | 79 |
| 11 | 91 | 26 | 96 |
| 12 | 64 | 27 | 69 |
| 13 | 71 | 28 | 66 |
| 14 | 41 | 29 | 50 |
| 15 | 78 | 30 | 77 |

# EPA Data – Question #4

| Sample | Mileage (mpg) | Sample | Mileage (mpg) |
|---|---|---|---|
| **1** | 22.9 | **13** | 25.5 |
| **2** | 23.9 | **14** | 22.2 |
| **3** | 21.4 | **15** | 21.7 |
| **4** | 25.4 | **16** | 23.5 |
| **5** | 23.9 | **17** | 27.1 |
| **6** | 24.4 | **18** | 23.0 |
| **7** | 23.1 | **19** | 23.9 |
| **8** | 22.0 | **20** | 23.6 |
| **9** | 25.4 | **21** | 19.2 |
| **10** | 20.7 | **22** | 22.7 |
| **11** | 21.4 | **23** | 26.0 |
| **12** | 22.8 | **24** | 21.3 |