Clip 1: Introduction (Sales spreadsheet on screen)

The Excel Data Analysis Package is able to easily produce for you a wide variety of statistics. You will just scratch the surface of these capabilities in your next assignment

This video uses data in the COGS (Cost of Goods Sold) Column and what you will see in this video is very similar to what you will do in the assignment except you will be using the Revenue Column instead of the COGS column.

To make it easier to write formulas using the COGS Column, we will define a range called Cost. The data is in rows 2-1127. There are 1126 rows of data.

Clip 2

Using this range name, it is easy to use the Data Analysis Package to run descriptive statistics on the Cost information. Select Data, then Data Analysis, and then Select “Descriptive Statistics”. Tell Excel that we want to use the range name Cost. Since the Cost range starts in row 2, we do not need to indicate and there are labels in the first row.

We need to tell Excel what statistics we want and here we select Summary Statistics and we are asking to have the results placed on a separate sheet.

Here are the results. The count is 1126. You can also see here the other statistics that Excel produces automatically without you have to code functions yourself such as Median, Mode, Minimum, Maximum and so on.

Clip 3

We are interested in doing a further analysis of cost data by doing frequency counts and then preparing a histogram. This is done by using the Data Analysis package and selecting Histogram. Again we use the range Cost. When we do this, we see that we must tell Excel how to define the bins that will be used to group the counts. Let’s cancel the analysis for now so we can define the bin information. We wish to start by using five bins.

We have some simple calculations to do to define the bins.

We know the minimum cost and the maximum Cost from the descriptive statistics we just created. Let’s copy that information from the descriptive statistics onto the main sheet at cell N3. It does not matter what cell you use as long as it is outside of the Sales table. N3 is just a convenient location.

Clip 4

We can calculate the difference between the maximum and the minimum and call it difference. We want our first histogram to have 5 bins. The width of each bin is the difference divided by the number of bins. Then we can calculate the starting point of each bin by starting with zero and adding the bin width. To make the histogram more readable, I'll round the bin width to the nearest hundred.

Using the simple calculations that we have done; we can construct a column containing the lowest number in each bin. Start with zero and then and then increment by the Rounded (Bin) Width. Keep going until the maximum number has been exceeded. This is the column of bin information that we will feed to Excel Data Analysis.

Clip 4.5
Now we have all the information we need to create a histogram of Cost information. We use Data Analysis and Histogram. The input range is Cost. The Bin Range we created is in M3 through M9. We specify that we want the output on a new sheet and here is the output. It gives us a count of the number of rows in each bin. Now we can turn that into a histogram by Inserting a Recommend Chart of type Frequency. We can see that most of the sales transactions show a cost of $7200 or less.

Clip 5

Let's do the same analysis only this time we will use 10 bins.

We start by copying the work that we did before and then change the number of bins from five to 10. This time we will create the bin information starting in M13. Again, start at zero and increase by the Rounded Bin Width until the maximum is exceeded. Our bin information is in M13 through M24 and we are ready to do another histogram. This time, the input range is again Cost and the bin information is in M13 to M24. We ask for the frequency information on a new sheet and then ask Excel to Draw the Histogram.

When dealing with a new set of data for the first time, you may need to experiment with the number of bins to see what would be best.

Clip 6

The Percentile.inc and Quartile.inc functions are useful, for example, in finding the minimum GPA required to be in the top 10% or the top 25% of the class.

Let's use these two functions to find the 20$^{th}$ percentile, the first quartile (25%) and the 80$^{th}$ percentile of Cost.

First let's compute the 20$^{th}$ percentile. This means that 20% of the transactions have a cost less than this number and 80% of the transactions have a cost greater than this number

Next let's computer the first quartile. This means that 25% of the transactions have a cost less than this number and 75% of the transactions have a cost greater than this number.

Finally, let's compute the 80% percentile.

Clip 7:

So far we have been dealing with numbers of transactions or counts. What can we say about the actual cost numbers themselves?

Let's compute the sum of the Cost column. We know that the value of the 25$^{th}$ percentile is 2480. Let's do a sum if to add up all the Cost numbers that are less than 2480. The criteria less than 2480 is now in cell K14. So, to do the addition, we use sumif (Cost,K14) Cost is the range and cell K14 contains the criteria. By dividing the number we just obtained by the sum of all cost numbers, we can see that only 7% of the cost is contained in the first quartile. By changing the criteria to greater than 2480, we can see that 92% of the cost is in the top three quartiles.

Clip 8:
Let's see what happens when we change just one number to a very low number. The first row of data has been modified to have 0 in the Cost column. After that minor change, let's run descriptive statistics again.

By copying the results onto the same page of statistics we generated earlier we can see what changed and what has remained the same. In this case only maximum, count, median and mode remain the same.

Clip 9:
In this video you have seen how to use a small part of the Excel Data Analysis package to produce descriptive statistics and histograms. You have also learned how to calculate values for percentiles and quartiles for an array (Excel column) of data. Finally, you saw how even a small change in the data can impact descriptive statistics.

This is the end of the video.