Excel was not intended for scientific data analysis but it and its clones (like OpenOffice Spreadsheet) are generally available, so we use them when necessary. However, on some recent versions of Excel, Microsoft apparently eliminated the useful “Data Analysis Toolpack” that formed the basis for the previously posted set of instructions to create a histogram.

Since histograms are so important for data analysis, we offer some alternative ways of using Excel to create and plot them. Google will find other examples of instructions to create histograms.

All the functions you need to create a histogram are already built in to all versions of Excel, so you can use them in sequence to tabulate data, create histograms and then plot the result in various ways. Here is a step-by-step set of instructions for Windows, Excel 2003. The screen formats will look different in other versions but the functions are the same.

In this example, we will use the function “COUNTIF” to create the actual histogram and then plot the result. How COUNTIF works will be explained later.

**Procedure:**

1. Open Excel

2. Enter the data for the histogram. Here, I’m using the first 100 entries in the Atrium_all_2008.txt file which I opened using the File menu. On the next page, I show the result. The data are in the A column.
3. In the B column, type min, max, avg and sigma into the four top cells.

4. In the C column, find the minimum, maximum, average and standard deviation of the first 100 entries in column A

5. To do this, enter into cells C1-C4 (The “=” says the cell contains a formula)
   =MIN(A1:A100)
   =MAX(A1:A100)
   =AVERAGE(A1:A100)
   =STDEV(A1:A100)
6. Select a suitable X-axis range for the histogram. Note that the minimum and maximum are 1180 and 1570. A convenient range for the histogram might be 1100 to 1600, with steps of 10.

7. Create a column of values in steps of 10 beginning at 1100 and continuing to 1600. To do this, enter 1100 in one cell and 1110 in the next. This sets the pattern for the column. Select both of these cells. Note that there is a small square dot on the lower right corner of the two boxed cells, like this: (think of it as a pull tab)
8. Grab the square dot with the mouse/cursor and pull it down to create the list of BIN VALUES. See

![Image of Excel spreadsheet with bin values]

8. Next, we create the actual histogram. Use the function COUNTIF.

COUNTIF(A,B) counts the number of occurrences of a value in the range A that satisfy the criterion B. For example, COUNTIF(A1:A100, 1) returns zero, because there are no values equal to 1 in cells A1 to A100. We want to know how many values fall in a range of, say 1100 to 1109, which are the bin values in column E. One way to do this is to use the function COUNTIF twice in one cell, first to count the values that are >= 1100 and subtract them from the number that are >=1110.

9. In Column F, cell 2, type the formula
=COUNTIF($A$1:$A$100, ">="&E2) - COUNTIF($A$1:$A$100,">="&E3)

10. Copy cell F2 down the column to F52 and your spreadsheet should look like this:

Columns E and F are your histogram! All you need to do now is to plot it. Here is an example of the many, many ways to do this:

11. Select Insert Chart from the menu. Bar/Columns are fine
12. The next chart option is “**Source Data**”, with options “**Data Range**” or “**Series**”. Either will work, but “Data Range” was selected as shown next.
13. To select the data range, click the mouse in the “Data Range” box and then with the mouse, outline the **Count** column as shown above.

14. Select “Next” and give the data range a name like “Frequency”.

15. “Finish” and the chart will look like the prototype. *However the X axis is mislabeled!*

16. So, **Relabel the X-axis**.

To do this, select the chart with the mouse and select “**Source Data**”. In the “**Category X-axis Labels**” box, use the mouse/cursor to select the **Bin Values** F2:F52.

The chart should look like this:
17. Click **Finish**.

18. Now you can add the “Calculated” column, which is a Gaussian curve or Normal Distribution with the same mean, total of observations and bin range.

\[ FCALC = TOTAL \times BIN\ RANGE \times \text{NORMDIST}(X, \text{MEAN}, \text{SIGMA}, \text{CUMULATIVE}) \]

In our case, this becomes

\[ FCALC=100\times10\times\text{NORMDIST}(\text{bin}, 1371.8, 72.85, \text{FALSE}) \]

Because there are 100 observations, the bin size is 10 and the mean and standard deviation are entered as derived from the data. We use FALSE for the last parameter because we do not want to plot the cumulative distribution.

Enter a new column labeled “calcd” and enter the above formula in G2. The spreadsheet looks like this:
19. Copy that cell down to G52, to get a table of expected counts. It won’t look much like what you observed but the calculated counts should peak at about the same place.

20. Now add this column to the chart. Select the chart, select “Source Data” and add the Series G2:G52
21. The default data representation is bar columns. If you like, you can change the “Chart Type” to X,Y:

Keep in mind that there are other ways to do all this, and that most of these instructions will work for OpenOffice, which is an open-source Excel clone.