

Histogram Tutorial

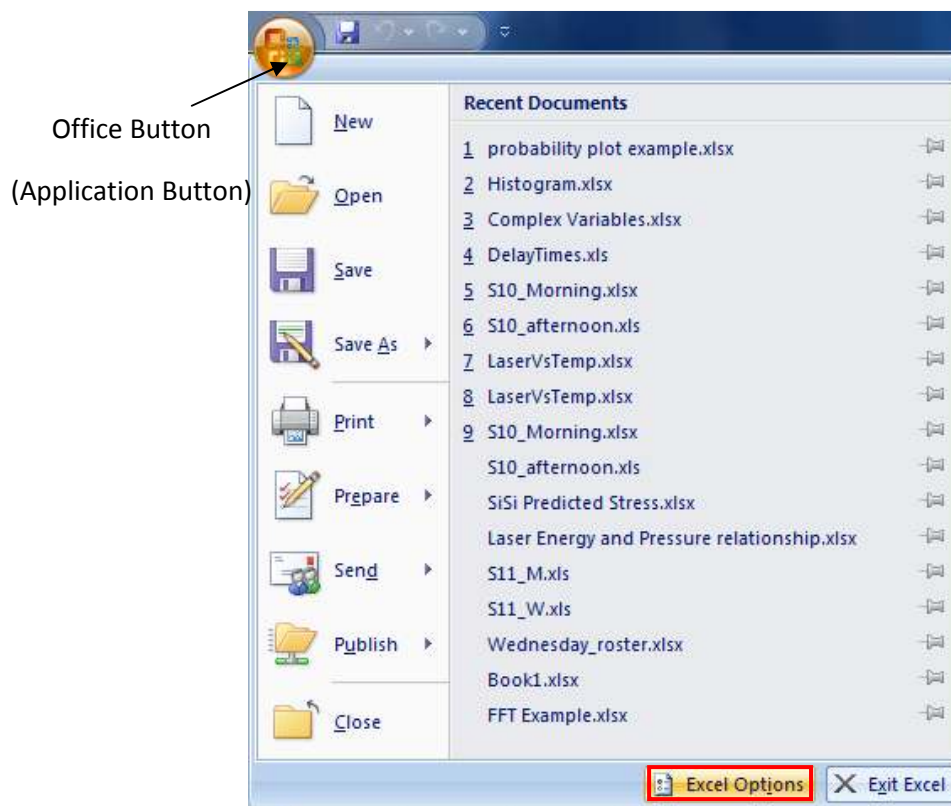
In this tutorial, we will learn how to create a histogram in MS Excel. The tutorial includes how to create Histogram using build-in function and Hand Calculation method discussed in class.

Note: Before starting this tutorial, download *Histogram Example Data.xlsx* file from Moodle.

1- Built-in Function:

Histogram built-in function is included in Data Analysis ToolPak. By default, this add-in is not activated. Therefore the first step is to activate the Analysis ToolPak.

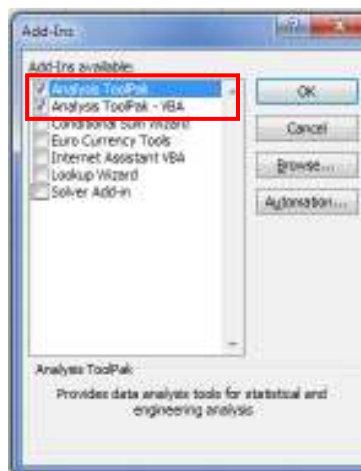
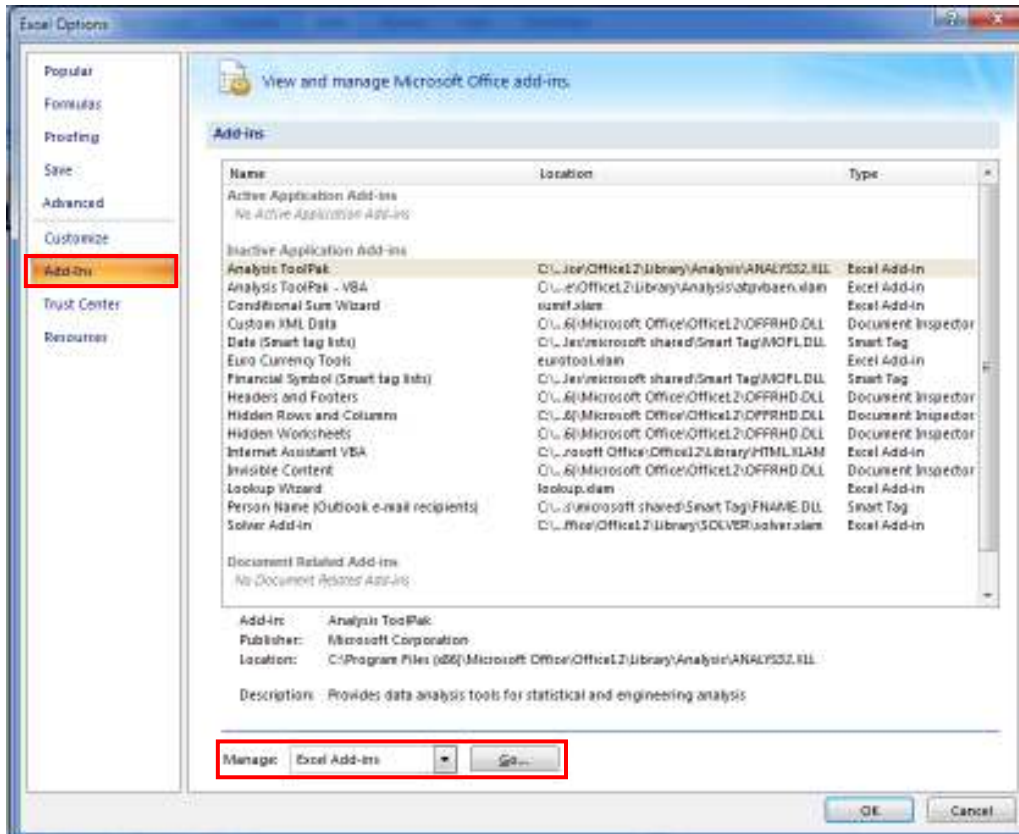
Go to **Office Button** (application Button)/**Excel Options**



Select the **Add-Ins** Tab, then select **Excel Add-Ins** from the **Manage** dropdown menu located at the bottom of the dialog box and click **Go**.

Excel will prompt you with another dialog box “**Add-Ins**” to select the analysis tools you want to activate. Check the box beside **Analysis ToolPak** and **Analysis ToolPak-VBA** and click **OK**.

You don't have to perform these steps every time you use the data analysis tools, you follow these steps if you can't access the Data Analysis tools, which usually in the first time you use PC with freshly installed Office or the first time you use the ToolPak.

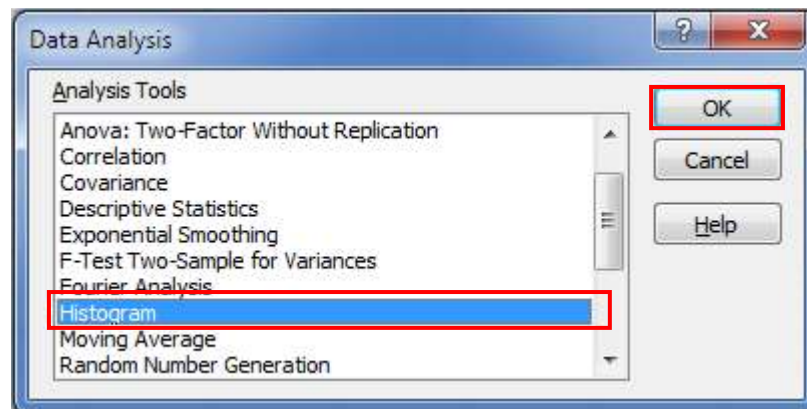


Open Histogram_Example_Data.xlsx file, you should find data points in column A. To create an Excel generated Histogram; follow these simple steps:

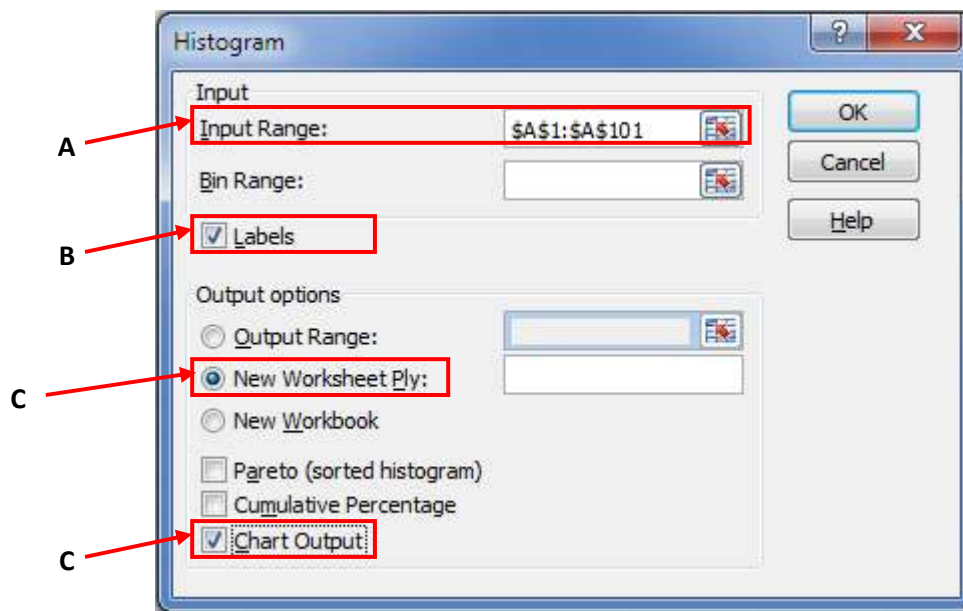
- 1- Select **Data** tab, and click **Data Analysis** from the **Analysis** submenu on the Ribbon. (If you did not perform the steps on page 1-2, you won't be able to access the Analysis submenu).



- 2- Excel will prompt you with Data Analysis dialog box, Select Histogram and click OK.



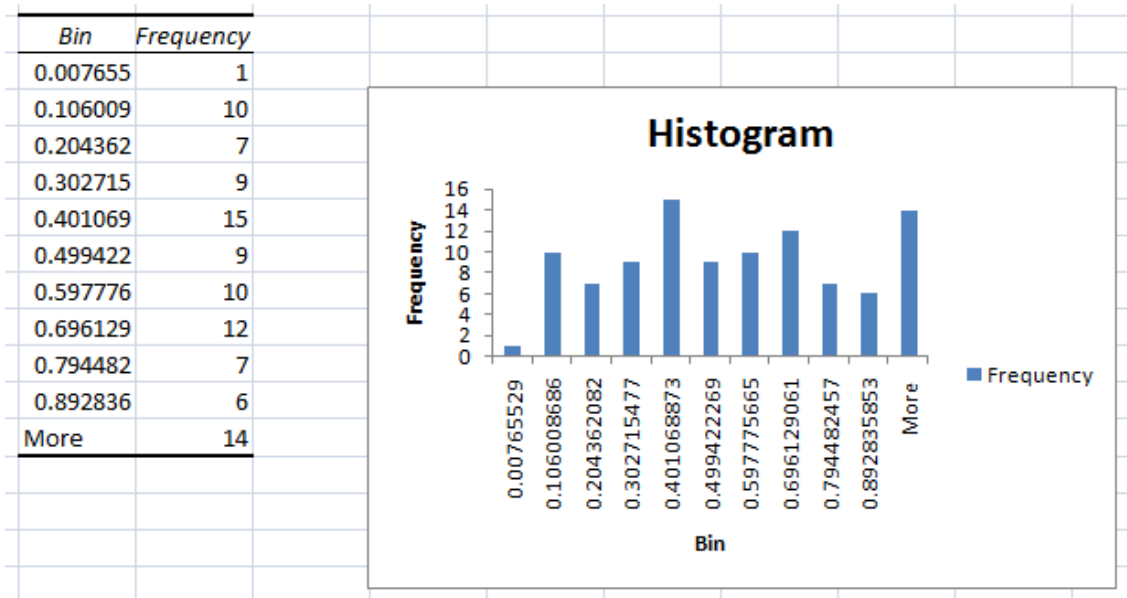
- 3- Excel will prompt you with Histogram dialog box, which looks like:



Input the following information:

- A- Select the input range

- B- If you have column header then check Label box. In our example there is a column header so check that box.
- C- In the Output options, select New Worksheet Ply and check Chart Output.
- D- Excel will create a new worksheet and will display the results. You should get something that looks like:



2- Hand-Calculations:

If you noticed on the above histogram, Excel generates the bins automatically. Sure you can set the bins a prior, and then create the histogram. In this section, we will learn how to build histogram manually using the methodology discussed in class.

- A- Find x_{\max} and x_{\min} using Excel function **=max(range)** and **=min(range)**, respectively. Of course, replace range with selecting the actual cell range where the data are.
- B- Calculate the number of bins:

$$k = 1.87(N-1)^{0.4} + 1$$

Roundup to the next higher bin if necessary; you can use Excel function **=roundup(k,0)**. Replace k with the cell that contains the value of k.

- C- Calculate the bin width:

$$\delta x = \frac{x_{\max} - x_{\min}}{2k}$$

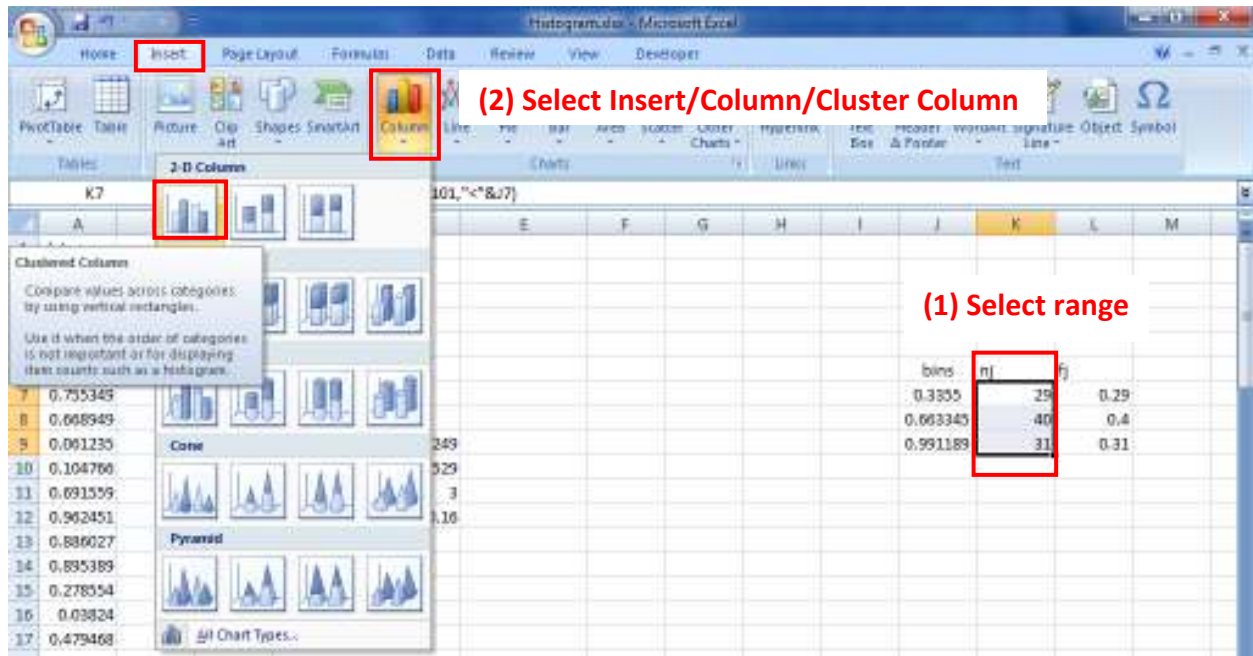
- D- Create the bins: first bin is x_{\min} , second bin is $x_{\min} + 2\delta x$,, last bin is x_{\max} . However, we indicate each bin with one number here, but it should be understood the the bin extents as a range with bin width equals $2\delta x$.
- E- Calculate the number of occurrences (n_j). You can use Excel function **=countif(range,criteria)** to find n_j values. Where, $j=1,2,3 \dots k$ (bin number)
- F- Calculate the frequency distribution

$$f_j = \frac{n_j}{N}$$

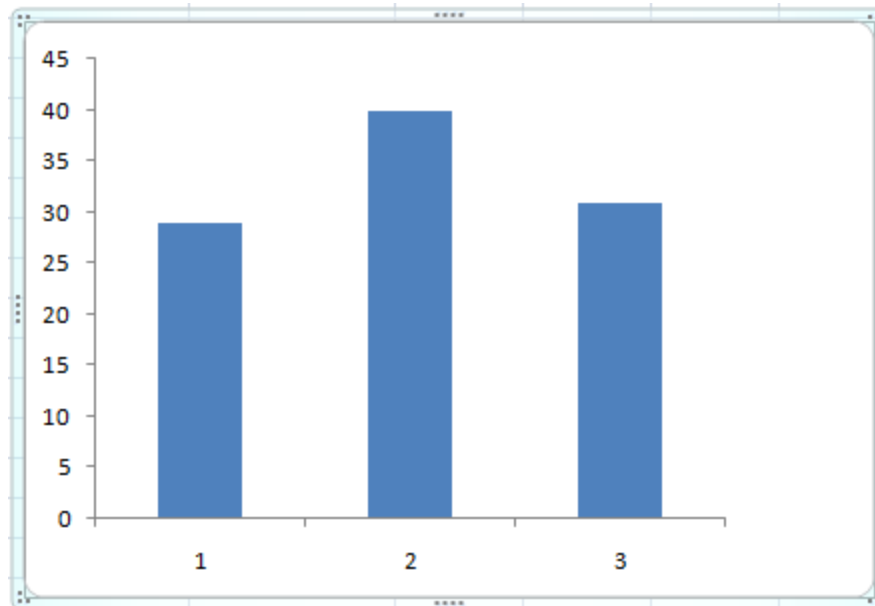
The figure below shows the inserted formulas in Excel.

xmax	=MAX(A2:A101)	
xmin	=MIN(A2:A101)	
k	=ROUNDUP(1.87*(COUNT(A2:A101)-1)^0.4)+1,0)	
δx	=(xmax-xmin)/(2*k)	
bins	nj	fj
=xmin+2*D12	=COUNTIF(A2:A101,"<"&C16)	=D16/COUNT(\$A\$2:\$A\$101)
=C16+2*D12	=COUNTIF(A2:A101,"<"&C17)-(D16)	=D17/COUNT(\$A\$2:\$A\$101)
=xmax	=COUNTIF(A2:A101,"<="&C18)-(D17+D16)	=D18/COUNT(\$A\$2:\$A\$101)

G- Plot: Select the range where nj values are, then click Insert tab/ Column/Clustered Column



Excel will generate the following graph:



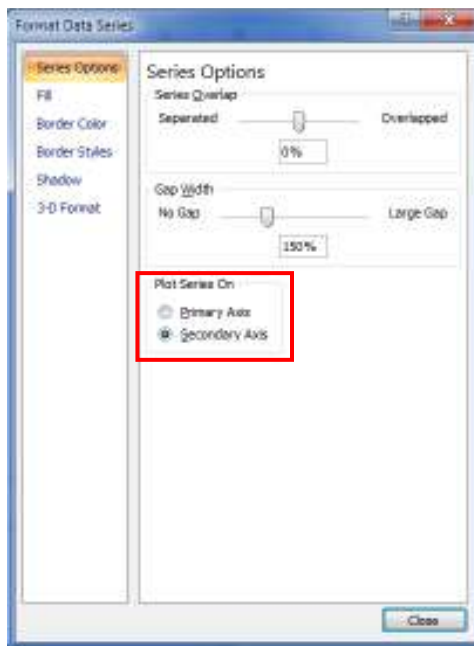
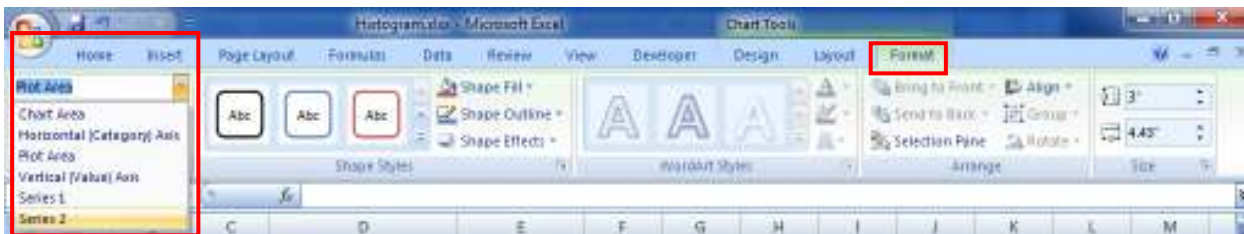
- H- Next step is to add the secondary axis, which represent the frequency distribution. The frequency distribution is included on a Secondary Axis. Select the Chart/Format/Current Selection/ select Chart Area. You will see that n_j range highlight, expand that to include the values for f_j .



bins	n_j	f_j
0.3355	29	0.29
0.663345	40	0.4
0.991189	31	0.31

Click and drag to include f_j

- I- Select the Chart/Format/Current Selection/ Series2 from the dropdown menu and then click Format Select and Select Secondary Axis



- J- Add the correct axes titles to have the histogram look like:

