# How to do statistics and histograms in Excel

## Objective

The objective of this activity is to gain experience using the statistical functions in Excel to process digital signals.

## Intended Learning Outcomes

| | |
|---|---|
| Be able to load and use the Analysis Toolpak in both Excel | |
| Become familiar with using Excel commands: RAND, MAX, MIN, AVERAGE, STDEV, VARIANCE | |
| Use the histogram function to estimate and characterize the probability density function of random variable | |
| Analyze a dataset in terms of its signal to noise ratio | |
| Analyze the histogram of quantization noise for large and small signals relative to the quantization step size | |

## Tips

- For the chart type, use SCATTER.

- For each plot, create a new chart sheet, rather than a new worksheet of data: First make the chart then right click, select MOVE then NEW SHEET and name the sheet P2-…. Etc.

## Procedure

### Chapter 2 section 1

1) Open the excel file on mycourses labeled Data For Statistical analysis of signals using Excel.xlsx

   The spreadsheet contains several pages (tabs at the bottom) containing the exercises. Follow the instructions and use the spreadsheets to find the answers the questions that follow.

2) Sheet Ch2-1. In this sheet, you are given a data signal in column B that contains 730 samples. The sample numbers are contained in column A. From the samples compute the maximum value, minimum value, mean value, the variances, standard deviation, etc. Use the built in excel formulas for these.(MAX, MIN, AVERAGE, STDEV, VARIANCE) Click on the"?" to get help on these excel functions.

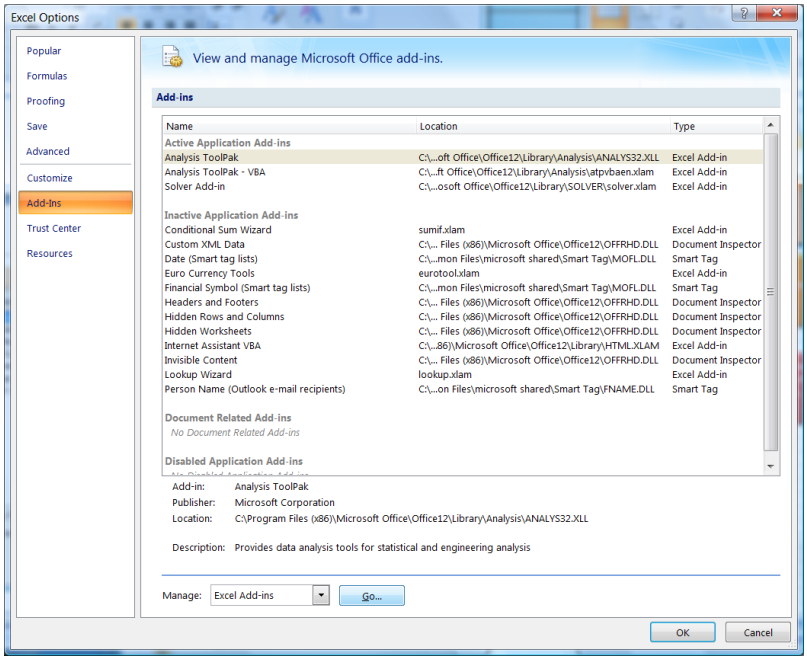
***Q2-1-1.*** Complete table **T2-1-1 on the ANSWER SHEET**

3) Using Sheet Ch2-1.

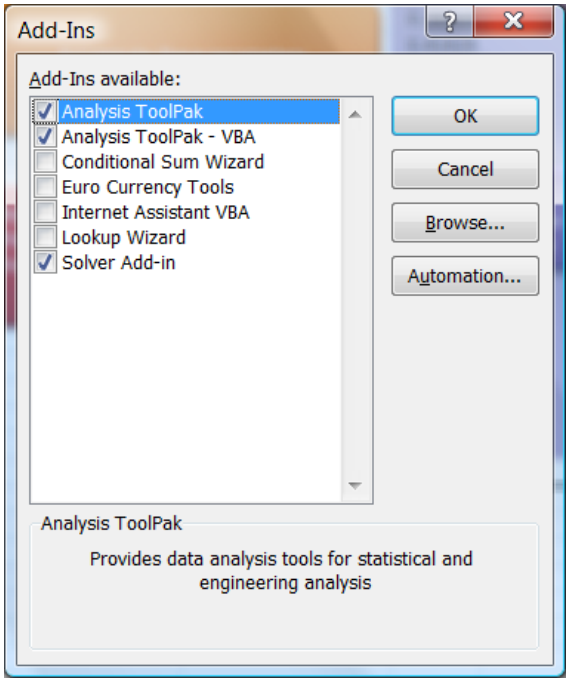
Plot all 730 data points of the signal and label the dependent and independent axes with appropriate labels.

***Q2-1-2.*** How well to the statistical parameters you just calculated agree with the data plot that you just created? What aspects do not agree, if any?

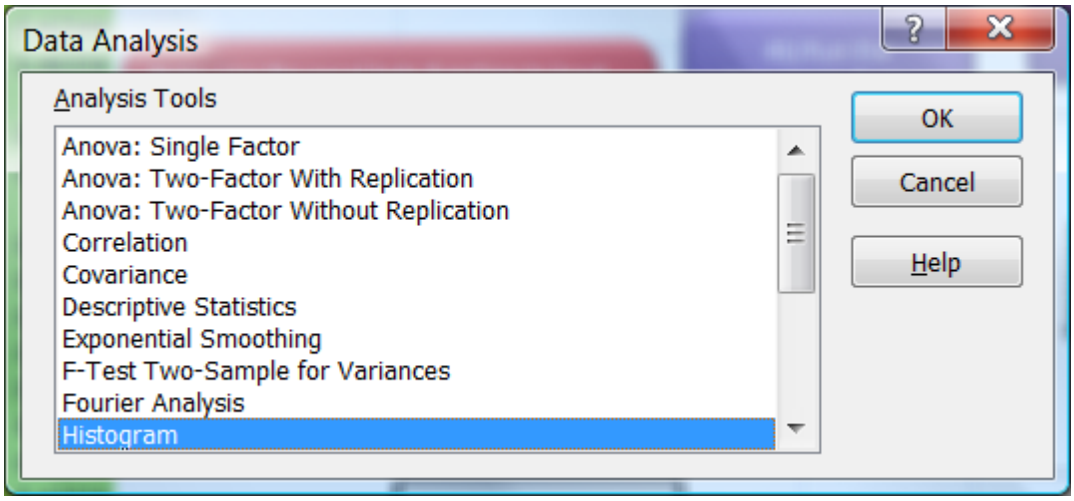
4) Next, create a histogram of all the data points using excel's data analysis tools. First click the office button->select excel options->Add-ins-> Manage (choose excel add ins from the drop down list).



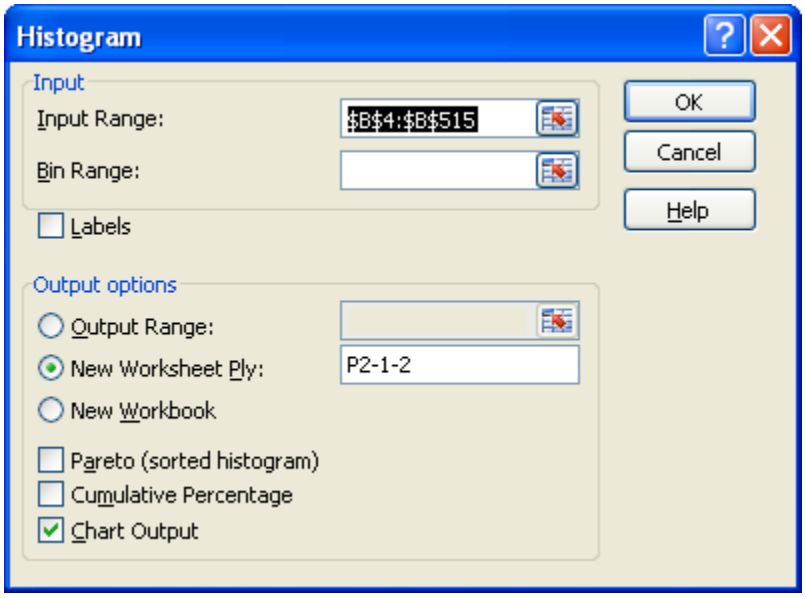

Then check the Analysis ToolPak and Analysis ToolPak - VBA and add these. Click OK




Return to excel and click DATA->DATA ANALYSIS, select HISTOGRAM




The following window will pop up. Select the input range corresponding to your data. Leave the bin range blank. Select the output range to be New Worksheet Ply. Check the chart output box. Click OK.




5) The plot you just created is the histogram of the data. Resize the chart if needed.

***Q2-1-3.*** What distribution best describes the data (uniform, triangular, normal, other)?

**Chapter 2 section 2**

6 ) Using Sheet Ch2-2.

***Q2-2-1*** Fill in the table ***T2-2-1*** by computeing the indicated quantity on each of the three signals. In other words, do not simply add the Max of signal 1 to the Max of signal 2 to get the Max of Signal 1+2.

***Label the X and Y axes and title all graphs with their plot number P-….***

Plot the data of signal 1. Put the plot on its own SHEET TAB labeled ***P2-2-1.***

Plot the data of signal 2. Put the plot on its own SHEET TAB labeled ***P2-2-2.***

Plot the data of signal 3. Put the plot on its own SHEET TAB labeled ***P2-2-3.***

Plot the histogram of the signal 1. Put the histogram data and the plot on its own SHEET TAB labeled ***P2-2-4.***

Plot the histogram of the signal 2. Put the histogram data and the plot on its own SHEET TAB labeled ***P2-2-5.***

Plot the histogram of the signal 3. Put the histogram data and the plot on its own SHEET TAB labeled ***P2-2-6.***

***Q2-2-1***. Answer and address the following: Theoretically, what should the mean of the sum of the signals be equal to? Is this the case for the data given? (exactly or approximately?) (Why, or Why not?)

***Q2-2-2***. Answer and address the following: Theoretically, what should the variance of the sum of the signals be equal to? Is this the case for the data given? (exactly or approximately?) (Why, or Why not?)

***Q2-2-3*** Is the standard deviation of the sum (sig1+sig2) equal to, less than or greater the sum of the standard deviations of the individual components sig1, sig2? (exactly or approximately?) (Why, or Why not?)

***Q2-2-4.*** What distribution is the data in ***P2-2-6*** best described by (uniform, triangular, normal, other)?

It is important to note that in all the analysis you've just done the specific values for standard deviation, mean, and so forth will change as the random numbers change. In the spreadsheet you were given a fixed set of random numbers, but in reality, each time you run such a test the random noise changes and your answers will change slightly. You are only estimating the true mean from a finite set of data points. Thus, you will have some typical error between this value and the true mean.

**Chapter 2 section 3**

7) Using Sheet Ch2-3.

Create a random variable in column B using the RAND() function in excel. In column B the formula will be =RAND() which gives uniform random noise from 0 to 1 with a mean of 0.5

In column C, create a uniform random second random variable again with values from 0 to 1 and mean of 0.5.

In column D, add columns B and C.

Fill out the rows down to have 2000 data points.

The remaining columns contain data from other random noise sources.

Plot the histogram of the signal 1 (include all 2000 data points). Put the histogram data and the plot on its own SHEET TAB labeled ***P2-3-1.***

Plot the histogram of the signal 1+signal 2. Put the histogram data and the plot on its own SHEET TAB labeled ***P2-3-2.***

Plot the histogram of the sum of signals 1-12. Put the histogram data and the plot on its own SHEET TAB labeled ***P2-3-3.***

***Q2-3-1.*** Carefully compare the distributions (from the histogram) of Signal 1, Signal 1+2, and the sum of signal 1-12. How are each of these distributions best described (uniform, triangular, normal, other)? How does this compare to theory? What theory applies to the sum of the signals 12?

**Chapter 2 section 4**

8) Using Sheet Ch2-4.

This sheet contains the daily average temperature in Rochester NY over a two year period (Column E). Plot this and label as **P2-4-1**. Remember to put it on its own sheet.

The daily average temperature can be estimated by a constant value of 48.2 F plus a sinusoidal variation of amplitude 23 degrees F and phase of -1.9 radians. Putting this all together, the estimated temperature is:

Daily average temperature (F) = 48.2+23*SIN((2*PI()*day/365)-1.9)

The estimator value is in column F. Column G which contains the error in the estimator i.e the difference between the actual data and estimator. Plot the error of the estimator and label as plot **P2-4-2.**

Now take the histogram of the error and plot it. Label as **P2-4-3.**

**Q2-4-1.** How would you describe the distribution of the error of the estimate (normal, uniform, triangular, other)?

**Q2-4-2.** What is the standard deviation of the error?

**Q2-4-3.** If the estimator predicts that the temperature today will be 55 degrees, what is the signal to noise ratio of this estimate, given that the standard deviation of the error was the value you just calculated?

**Chapter 3 section 1**

9) Using Sheet Ch3-1.

This sheet contains a 256 samples of a sinewave (Column B) quantized with a quantization step size of 1 (Column C).

The formula for the sinewave is: =127.5+127.5*SIN(A2/10)

The sinewave is not tightly correlated with the sampling frequency. Plot the histogram of the quantization error (Column D) using the given histogram bins (Column E) and label as **P3-1-1**. Use the excel's histogram plot function and

also check the box to get the cumulative % in addition to the frequency. Remember to put it on its own sheet.

**Q3-1-1.** How would you describe the distribution of the error of the estimate (normal, uniform, triangular, other)?

**Chapter 3 section 2**

10) Using Sheet Ch3-2.

This sheet contains a 256 samples of a sinewave (Column B) quantized with a quantization step size of 1 (Column C).

The formula for the sinewave is: =127.5+127.5*SIN(2*PI()*A2/64)

The sinewave is not tightly correlated with the sampling frequency. Plot the histogram of the quantization error (Column D) using the given histogram bins (Column E) and label as **P3-2-1**. Use the excel's histogram plot function and also check the box to get the cumulative % in addition to the frequency. Remember to put it on its own sheet.

**Q3-2-1.** How would you describe the distribution of the error of the estimate (normal, uniform, triangular, other)?

**Q3-2-2.** Are the distributions of the quantization error significantly different or largely the same between **P3-1-1** and **P3-2-1**?

**Chapter 3 section 3 Dithering**

11) Using Sheet Ch3-3.

The effective resolution of an A/D converter can be increased by oversampling and averaging. However, when this requires sufficient dithering noise to ensure the binary code from the A/D is changing over time.

This sheet contains 256 quantized samples of a constant analog signal to which dithering noise was added before quantization. The dithering noise is zero-mean, uniform random noise with a standard deviation of 1 mV.

From the data, estimate what the true value of the analog signal is, using a minimum number of samples, so that your estimate has a typical error of less than 100 microvolts.

**Q3-3-1.** What is the estimated true value of the constant analog signal?

**Q3-3-2.** How many samples must you use in your estimate?

**Checklist**

Every plot

- ✓ has a label for X axis
- ✓ has a label for Y axis
- ✓ has a title that starts with the plot number P….
- ✓ has a title with a descriptive name e.g. histogram of 12 bit quantization noise with 0 bits dither.
- ✓ is on its own sheet tab

Every sheet tab

- ✓ is named by its plot number P…
- ✓ is sorted in order from first to last plot number.