Probability Plot Tutorial

Probability plot (also known as normal probability plot, NPP) is a tool to determine whether the data follows normal distribution or not. It plots the measured values on the abscissia and the predicted values using normal distribution probablility density function on the ondinate. It is quit simple and effective tool.

First, we need to create some measurement to demonstrate the procedure to make the plot. In this tutorial, we will use Random Number Generator to generate 75 random numbers that are normally distributed. You can also use Randbetween(bottom,top) function from worksheet functions library.

Click **Data/Data Analysis** and select **Random Number Generator** and click **OK**. If you cannot find the Data Analysis ToolPak, please refer to <u>Appendix 1</u> of this tutorial to enable the Analysis ToolPak.

| ОК |
|--------|
| |
| Canada |
| Cancer |
| Help |
| |
| |
| |
| |

Excel will prompt you with Random Number Generator dialog box. Input the following information into its perspective input boxes.

- Number of Variables = 1 "this refers to the number of columns"
- Number of Random Number = 75 "this refers to the number of measurements"
- Distribution = Normal "this refers to normal probability density function i.e. generated data will be normally distributed."
- Mean = 20 "this refers to the sample mean or the central value of the bellcurve"
- Standard deviation = 2 "this refers to the variation in the generated data"
- Output Range =\$A\$1 "generated data will be stored in the active sheet starting at cell A1"

After entering all information above press OK. The generator will fill in A1:A75 with normally distributed random numbers. Now, we are ready to demonstrate NPP.

| lumber of <u>V</u> ariables: | | 1 | ОК |
|---|-----------------|--------|----------------|
| Number of Random Nur | n <u>b</u> ers: | 75 | Cancel |
| Distribution: | Norm | al | • <u>H</u> elp |
| Parameters | | | |
| Mean = | 20 | 24.5 | |
| Chandrad devicetion | 2 | | |
| Standard deviation = | 2 | | |
| Standard deviation = Random Seed: Output options Output Range: | [2 | \$A\$1 | |

Probability plot procedure:

1- Sort data from ascending order "Smallest to Largest". Click Data tab/Sort and set Sort by =Column A, Sort on = Values and Order Smallest to Largest

| | 1 3 9 | ~ (2 ~) = | | - | | | Bool | d - Micro | soft Excel | | | | |
|-----|--------------|---|--|------------------------|-----------|------------------------------|--------------------|-----------------------------|----------------|----------------------------------|---------------------------|------------------------------|---------------------------|
| | Home | Insert | Page Lay | out Form | ulas Dat | ta Revi | iew Vie | ew De | veloper | | | | |
| Get | External Ref | fresh All + Be Connecti | nnections operties it Links ons | AZ↓ AZZ ZA↓ Sort | Filter | Clear Reapply Advanced | Text to Columns | Remove Duplicates Dat | Data Va | lidation → date Analysis → | Group Ungrou Subtot | * 0] up * "] al e 5 | Data Analysis Analysis |
| | A1 | | | | | | | | | | | | |
| 1 | A 🔶 | В | Click h | ere to se | lect colu | umn A | j. | Н | I | J | К | L | M N |
| 1 | 19.39698 | Sort | | | | | | | | 8 | 23 | | |
| 2 | 19.85368 | | | | | | 0.00 | | 4505 | 10.00.00 | | | |
| 3 | 23.54961 | ^Q ₂ 1 <u>A</u> do | Level | (<u>D</u> elete Level | Copy Le | evel 🕆 | 🕘 Op | tions | My | data has <u>h</u> | eaders | | |
| 4 | 20.65463 | Column | | | Sort On | | | Order | G. | | | | |
| 5 | 19.78793 | Sort by | Column A | - | Values | | | ✓ Small | est to Largest | 8 | | | |
| б | 18.59933 | | | | | | | | | | | _ | |
| 7 | 18.94978 | | | | | | | | | | | _ | |
| 8 | 21.11631 | | | | | | | | | | | | |
| 9 | 21.48885 | | | | | | | | | | | | |
| 10 | 16.66301 | | | | | | | | | | | | |
| 11 | 19.07424 | | | | | | | | | | | | |
| 12 | 18.85209 | - | | | | | | ſ | 01/ | | | | |
| 13 | 18.83672 | | | | | | | Į | OK | Can | cei | | |
| 14 | 19.74187 | 6 | | 1 | | 1 | | 1 | | | | | |

2- Fill in a series in column B. This series should represent the count, however if you use 1,2,3...,75, you will get an error later when you are calculating the probability. For this reason, we will start our series at any number less than 1. In cell B1, insert a value of 0.9. While you are selecting cell B1, click Home tab/Fill/Series.

| 6 d 7 | | - | Contract of the local division of the local | 0 | nisi lanno | off Econ | | | | | Cat Cat |
|--|-----------------------------------|-------------------|---|--------------|-------------------------------|----------|----------------------------|-------------------------------|-------------|------------------------------------|--|
| Hore | Incert Pag | e Lasbot Form | ulas (Data) P | calew (| View De | elaper | | | | | ÷ - |
| Patte | Content + 11 B J U + 11 - Font | | = = (⊗~) = = = (≥ = (= Azgravint | 39 (38 - | Seneral \$ - % + Komber | - | Conditions Formatting * | Permat en Table - Ryres | Carlo Carlo | Delete - E | Sort & Find & Ther- Select- Enting |
| 61 A 1 14.43229 | - (* 0.9 | A 0.9 Note tha | t cell | 6 | н | 1 | 9 | к | | Los Lett Banad Wakanesti J., | N O |
| 2 16.11682 3 16.25443 4 16.31456 | | B1 is sele | ected | | | | | | - | Series Fill Series | |

Excel will prompt you with Series dialog box. Input the following information into its perspective input boxes.

- Series in = Columns
- Type = Linear
- Step Value = 1
- Stop Value = 75

| Day Weekday |
|-------------|
| Mookday |
| - iveenuay |
| 🔿 Month |
| O Year |
| value: 75 |
| , |

3- Calculate the percentile of each measurement by dividing the count by the total number of samples. To do so, in cell C1 insert the following formula, then double-click on the right-lower corner of C1 to populate all values in column C.

=B1/count(\$A\$1:\$A\$75)

Note: to fix the reference press F4. f =B1/COUNT(\$A\$1:\$A\$75) C1 - () Note the D Α В С Ε F G formula 14.43229 0.9 0.012 1 2 16.11682 1.9 Double click to 2.9 3 16.25443 fill in values in 3.9 4 16.31456 column C 5 16.66301 4.9

4- Calculate the predicted value. Recall the equation from lecture notes:

 $x_i = z\sigma + x'$

where, σ is the standard deviation which was set to 2 in Random Number generator and x' is the mean and was set to 20.

<u>Note:</u> you can calculate the mean and standard deviation in Excel using =average(*range*) and =stdev(*range*) functions, respectively.

To calculate the z values use =**NORMINV**(*cell,mean,standard_deviation*), which calculate the inverse of the normal distribution. In cell D1 enter the following formula, then double-click on the right-lower corner of D1 to populate all values in column D.

=NORMINV(C1,AVERAGE(\$A\$1:\$A\$75),STDEV(\$A\$1:\$A\$100))

5- Create a Scatter plot (with only Markers) of values in column A (measured values) and column D (predicted values). Note that the plot lie almost on the diagonal because measured values (column A) were generated to be normally distributed.



Appendix 1: 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.

ME335

Dr. Youssef

| annial as | iew and manage interosore cince | addinis | |
|------------|---|--|--|
| orting | Addrins | | |
| nie | Nama | location | Тури |
| hanced | No Active Application Add-tes No Active Application Applies | | |
| istowize | hadive Application 6dd-ins | | |
| tol-thi | Analytic ToolPak | C1_Jo/(CHixe12)Jibrary\Analysis\ANALYSS2.ILL | Escal Add-In Escal Add-In |
| ust Center | Conditional Sum Wittard | samt stern | Escel Add-In |
| n outles | Custom All Data Data Smalt tag list() Euro Currency Tools Financial Synthol (Smart tag lists) Headers and Footers Hidden Rows and Column Hidden Rows and Column Hidden Viorisheets Infarmet Acatatert VBA Invisible Content Lookup Weand Person Name (Outbook e-wail recipients) Solver Add-in Document Related Add-ins Na Cocument Related Add-ins | Cit., Sidecoold Office/Office12/Operational Cit., Sidecoold States (Sear Tag/MOPL.DL) curotool dan Cit., Sidecoold Office/Office12/OPPRID.DLL Cit., Sidecoold Office/Office12/OPPRID.DLL Cit., Sidecoold Office/Office12/OPPRID.DLL Cit., Sidecoold Office/Office12/OPPRID.DLL Cit., States Office/Office12/OPPRID.DLL Sologo.dan Cit., Sidecoold Office/Office12/OPPRID.DLL Sologo.dan Cit., States Office/Office12/OPPRID.DLL Sologo.dan | Document Inspection Smark Tag Encel Add-In Smark Tag Document Inspector Document Inspector Excel Add-In Document Inspector Excel Add-In Smart Tag Excel Add-In |
| | Add Are Analysis Too Pak Publisher: Microsoft Corporation Location: C:Phogram Piles (d50) Micros Description: Provides data analysis tools f Manage: Excel Add-ene | off Officer/Officer.21/Library\Analysic\ANALYS32.811 or statistical and engineering analysis a | |

| OK Cancel Browse |
|---|
| OK Cancel Browse |
| Cancel Browse |
| Browse |
| the second se |
| Autonation |
| |
| |
| |
| |