Q P D O L

Version 1.0

Quantile Plotting Distribution Optimization Library for transforming data

USER GUIDE

Prepared by:
Dr. Ricaro V. Nunes, Universidade Estadual do Oeste do Parana
Dr. Gene M. Pesti, University of Georgia

UNIVERSITY OF GEORGIA
EXTENSION

Presentation Overview

Quantile Plotting Distribution Optimization Library

I started adjusting to 1300 data points and haven't finished

Pronounced: "Kewpie Doll Dot E X E"

Verson 1.1

A workbook to compare methods of transforming data to normalize distributions

Reference: M.S. Bartlett, THE USE OF TRANSFORMATIONS, Biometrics (1947) 3, 39-52

Gene Pesti	Department of Poultry Science The University of Georgia Athens, Georgia, 30602, USA	gpesti@uga.edu
Ricardo Nunes	Department of Agriculture Universidade Estadual do Oeste do Parana Mal. Candido Rondon, Parana, 85960-000, Brazil	nunesrv@pq.cnpq.br
Dmitry Vedenov	Department of Agricultural Economics Texas A&M University College Station, TX 77843 USA	vedenov@tamu.edu
Marek Rychlik	Department of Mathematics The University of Arizona Tucson, AZ 85721 USA	rychlik@math.arizona.edu

This workbook was developed from: http://alamos.math.arizona.edu/~rychlik/math263_old/ExcelAssignments/Assignment2/QQPlot.pdf

Home | Input & Summary | Graphic Summary | Original Data | Log10(x) | Log10(x+1) | Ln(x) | Sq. Root (x) | Square (x) | Arcsin(Root(x)) Degrees | Sq. Root(x+0.5) | Ln(x+1)

Tab Description

1) Home
2) Input & Summary – The place to input data and view the results. All inputs should be entered in this spreadsheet only.
3) Graphic Summary – Shows the frequency distributions of original and transformed data.
4) Original Data – Shows the normal quantile plot and frequency distribution of the original data.
5) $Log_{10}(x)$ – Shows the normal quantile plot and the frequency distribution of the data transformed into Log_{10} space.
6) Worksheets 5 to 20 – Each shows the normal quantile plot and the frequency distribution of the data transformed according to the model for each worksheet.
7) Worksheets 21 and 22 - Are new worksheets to insert additional transformations.
8) Worksheet 23 – Normal distribution generator based on input mean and standard deviation values.
9) Worksheet 24 – Normal distribution generator to view multiple simulations with different numbers of data points.
10) Worksheet 25 – Example data http://alamos.math.arizona.edu/~rychlik/math263_old/ExcelAssignments/Assignment2/QQPlot.pdf used to develop this workbook

[Step 1] Open Input & Summary

❷ Clear the original data from column B

	Original Data		Ranked Data
1		#NUM!	
2		#NUM!	
3		#NUM!	
4		#NUM!	
5		#NUM!	
6		#NUM!	
7		#NUM!	
8		#NUM!	
9		#NUM!	
10		#NUM!	
11		#NUM!	
12		#NUM!	
13		#NUM!	
14		#NUM!	
15		#NUM!	
16		#NUM!	
17		#NUM!	
18		#NUM!	
19		#NUM!	
20		#NUM!	
21		#NUM!	
22		#NUM!	
23		#NUM!	
24		#NUM!	
25		#NUM!	
26		#NUM!	
27		#NUM!	
28		#NUM!	
29		#NUM!	
30		#NUM!	
31		#NUM!	
32		#NUM!	
33		#NUM!	

Input your data into cells A2 to A1301 (up to 1300 points)
QQ Plots can be viewed on the various spreadsheets
Some data will need to be divided by a Factor (Below) to be in the proper range
Some transformations will not work if the data contains any 0's

Min	Max	Requirement	Factor	Transformation	R^2	R^2 Improvement (%)	Qm	CV (%)	Mean	SD
0.00	0.0000		1	Original Data =	#DIV/0!		0.000	#DIV/0!	####	###
0.00	0.0000	x>0	1	$Log_{10}(x)$ =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000		1	$Log_{10}(x+1)$ =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000	x>0	1	Ln(x) =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000		1	Square Root (x) =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000		1	Square (x) =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000	x<100	1	Arcsin(Root(x)) in degrees =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000		1	Square Root (x+0.5)	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000		1	Ln (x+1) =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000	x<1	1	ArcSin(Root(x)) in radians =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000	x<1	1	ln[x/(1-x)] =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000	x<1	1	0.5ln [(1+x)/(1-x)] =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000		1	Cubic Root (x) =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000		1	1 / x^2 =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000		1	1 / x =	#DIV/0!	####	0.000	#DIV/0!	####	###
0.00	0.0000		1	(1/lambda)*asinh(lambda*root(x+0.5))	#DIV/0!	####	0.0000	#DIV/0!	####	###
0.00	0.0000		1	(1/lambda)*asinh(lambda*root(x)) =	#DIV/0!	####	0.0000	#DIV/0!	####	###
0.00	0.0000		1	New 1	#DIV/0!	####	0.0000	#DIV/0!	####	###
0.00	0.0000		1	New 2	#DIV/0!	####	0.0000	#DIV/0!	####	###
				Min or Max =	#DIV/0!	####	0.000	#DIV/0!		

Home | Input & Summary | Graphic Summary | Original Data | Log10(x) | Log10(x+1) | Ln(x) | Sq. Root (x) | Square (x) | Arcsin(Root(x)) Degrees | Sq. Root(x+0.5) | Ln(x+1) | ...

❶ Open the spreadsheet "Input and Summary"
The original data will be entered in this worksheet

[Step 2] Open Input & Summary

❶ Copy your "original" data and paste into Column B. This column is standardized for up to 1300 data points

❸ This factor was inserted to help in the transformation of the data. Mathematical functions, such as Log, Ln, Arcsin, among others, require specific numbers to be used. Thus, if necessary, the user can use this factor, such as 1; 10; 100; 0.0001 or other multiple value of 10

	Original Data		Ranked Data
1	5.646	0.915	0.915
2	4.063	0.919	0.919
3	3.401	0.924	0.924
4	4.463	0.925	0.925
5	4.277	0.933	0.933
6	4.301	0.933	0.933
7	5.244	0.936	0.936
8	2.901	0.937	0.937
9	2.234	0.948	0.948
10	4.473	0.953	0.953
11	3.789	0.954	0.954
12	3.875	0.954	0.954
13	2.653	0.964	0.964
14	3.35	0.964	0.964
15	4.005	0.967	0.967
16	3.789	0.968	0.968
17	3.68	0.969	0.969
18	2.364	0.970	0.970
19	2.297	0.971	0.971
20	5.006	0.973	0.973
21	2.284	0.973	0.973
22	3.064	0.977	0.977
23	2.693	0.978	0.978
24	3.101	0.978	0.978
25	2.979	0.980	0.980
26	2.514	0.985	0.985
27	3.399	0.990	0.990
28	2.412	0.991	0.991
29	2.631	1.002	1.002
30	2.15	1.006	1.006
31	1.923	1.010	1.010
32	2.495	1.014	1.014
33	3.338	1.015	1.015

Input your data into cells A2 to A1301 (up to 1300 points)
QQ Plots can be viewed on the various spreadsheets
Some data will need to be divided by a Factor (below) to be in the proper range
Some transformations will not work if the data contains any 0's

Min	Max	Requirement	Factor	Transformation	R^2	R^2 Improvement (%)	Qm	CV (%)	Mean	SD
0.92	6.8610		1	Original Data =	0.9556		126.818	43.0	2.71	1.17
-0.04	0.8364	x>0	1	$Log_{10}(x)$ =	0.9868	3.3	3099.720	48.7	0.39	0.19
0.28	0.8955		1	$Log_{10}(x+1)$ =	0.9863	3.2	-978.661	24.4	0.55	0.13
-0.09	1.9259	x>0	1	Ln(x) =	0.9868	3.3	-978.661	48.7	0.90	0.44
0.96	2.6194		1	Square Root (x) =	0.9835	2.9	-110.018	21.7	1.61	0.35
0.84	47.0733		1	Square (x) =	0.8448	-11.6	-1246.672	85.7	8.72	7.48
#NUM!	#NUM!	x<100	1	Arcsin(Root(x)) in degrees =	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	###
1.19	2.7131		1	Square Root (x+0.5)	0.9799	2.5	-19122.097	18.1	1.76	0.32
0.65	2.0619		1	Ln (x+1) =	0.9863	3.2	3168.094	24.4	1.26	0.31
0.13	0.3534	x<1	1	ArcSin(Root(x)) in radians =	0.9829	2.9	393.875	22.1	0.21	0.05
#NUM!	#NUM!	x<1	1	ln[x/(1-x)] =	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	###
#NUM!	#NUM!	x<1	1	0.5ln [(1+x)/(1-x)] =	#NUM!	#NUM!	#NUM!	#NUM!	#NUM!	###
0.97	1.9002		1	Cubic Root (x) =	0.9875	3.3	1790.929	14.5	1.37	0.2
0.02	1.1944		1	1 / x^2 =	0.7847	-17.9	-60[illegible]973	95.3	0.24	0.23
0.15	1.0929		1	1 / x =	0.9190	-3.8	-478.[illegible]96	45.7	0.45	0.2
0.80	1.1964		1	(1/lambda)*asinh(lambda*root(x+0.5))	0.9872	3.3	3889.24[illegible]	8.9	0.98	0.09
0.70	1.1791		1	(1/lambda)*asinh(lambda*root(x)) =	0.9873	3.3	2380.3896	11.2	0.93	0.1
0.92	6.8610		1	New 1	0.9556	0.0				
0.92	6.8610		1	New 2	0.9556	0.0				
				Min or Max =	#NUM!	#NUM				

❹ When #NUM! appears, the data could not be transformed
Thus, it is necessary to use a factor other than 1 to obtain an acceptable data range

❷ The data will be automatically sorted in ascending order for column D

Home | Input & Summary | Graphic Sur | Ln(x) | Sq. Root (x) | Square (x) | Arcsin(Root(x)) Degre

[Step 2] Open Input & Summary

Some transformations require the data to be in some specific range. Analyzed data may have to be between 0 and 1, for instance. Raw data is divided by the specified factors to get it in an acceptable range. For example 10.

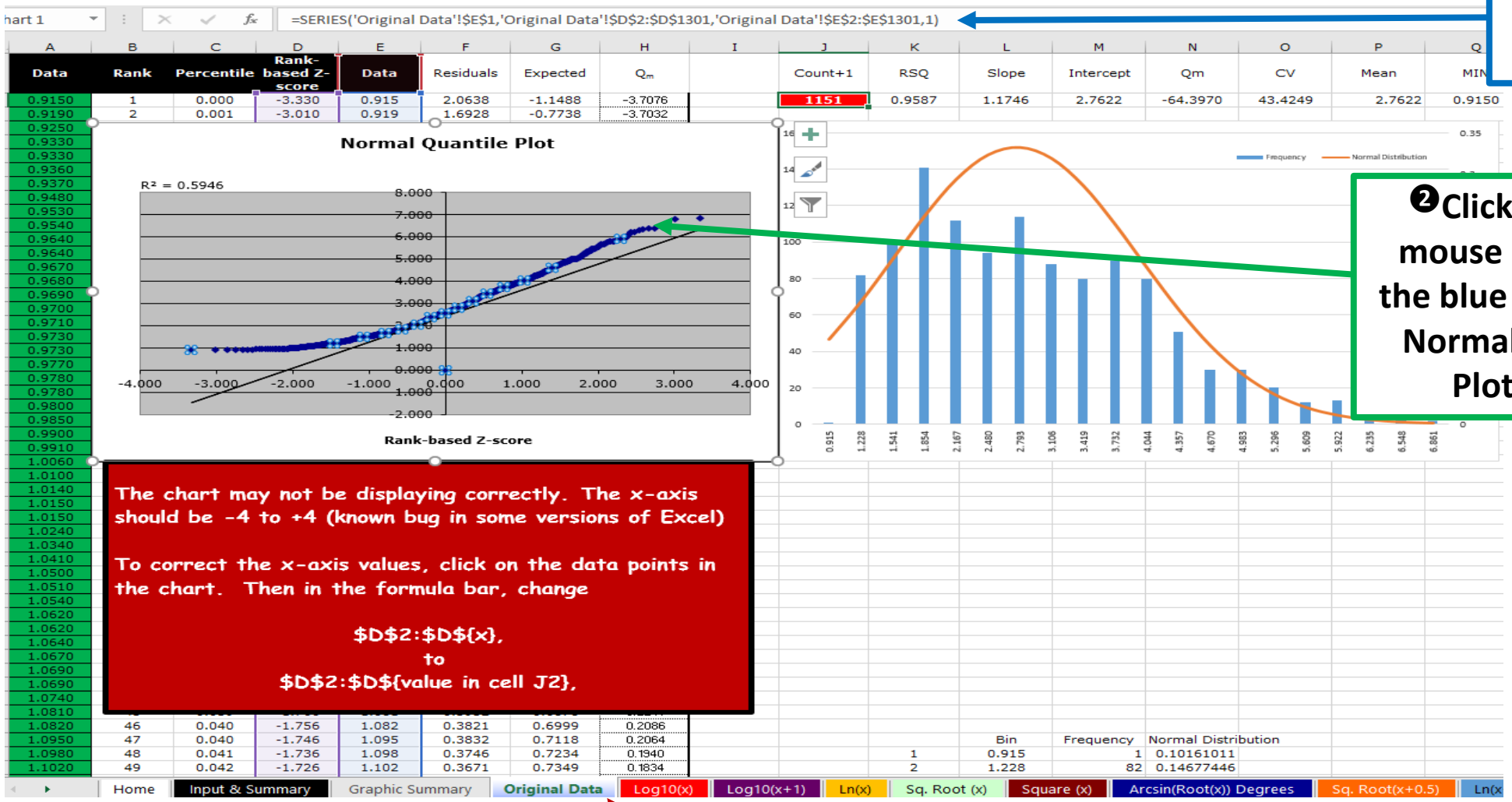
	Original Data		Ranked Data
1	5.646	0.915	0.915
2	4.063	0.919	0.919
3	3.401	0.924	0.924
4	4.463	0.925	0.925
5	4.277	0.933	0.933
6	4.301	0.933	0.933
7	5.244	0.936	0.936
8	2.901	0.937	0.937
9	2.234	0.948	0.948
10	4.473	0.953	0.953
11	3.789	0.954	0.954
12	3.875	0.954	0.954
13	2.653	0.964	0.964
14	3.35	0.964	0.964
15	4.005	0.967	0.967
16	3.789	0.968	0.968
17	3.68	0.969	0.969
18	2.364	0.970	0.970
19	2.297	0.971	0.971
20	5.006	0.973	0.973
21	2.284	0.973	0.973
22	3.064	0.977	0.977
23	2.693	0.978	0.978
24	3.101	0.978	0.978
25	2.979	0.980	0.980
26	2.514	0.985	0.985
27	3.399	0.990	0.990
28	2.412	0.991	0.991
29	2.631	1.002	1.002
30	2.15	1.006	1.006
31	1.923	1.010	1.010
32	2.495	1.014	1.014
33	3.338	1.015	1.015

Input your data into cells A2 to A1301 (up to 1300 points)
QQ Plots can be viewed on the various spreadsheets
Some data will need to be divided by a Factor (Below) to be in the proper range
Some transformations will not work if the data contains any 0's

Min	Max	Requirement	Factor	Transformation	R^2	R^2 Improvement (%)	Qm	CV (%)	Mean	SD
0.92	6.8610		1	Original Data =	0.9556		126.818	43.0	2.71	1.17
-0.04	0.8364	x>0	1	$Log_{10}(x)$ =	0.9868	3.3	3099.720	48.7	0.39	0.19
0.28	0.8955		1	$Log_{10}(x+1)$ =	0.9863	3.2	-978.661	24.4	0.55	0.13
-0.09	1.9259	x>0	1	Ln(x) =	0.9868	3.3	-978.661	48.7	0.90	0.44
0.96	2.6194		1	Square Root (x) =	0.9835	2.9	-110.018	21.7	1.61	0.35
0.84	47.0733		1	Square (x) =	0.8448	-11.6	-1246.672	85.7	8.72	7.48
17.61	55.9257	x<100	10	Arcsin(Root(x)) in degrees =	0.9766	2.2	118.374	24.5	30.91	7.57
1.19	2.7131		1	Square Root (x+0.5)	0.9799	2.5	-19122.097	18.1	1.76	0.32
0.65	2.0619		1	Ln (x+1) =	0.9863	3.2	3168.094	24.4	1.26	0.31
0.13	0.3534	x<1	1	ArcSin(Root(x)) in radians =	0.9829	2.9	393.875	22.1	0.21	0.05
-2.30	0.7819	x<1	10	ln[x/(1-x)] =	0.9902	3.6	2701.983	57.0	-1.07	-0.6
0.09	0.8405	x<1	10	0.5ln [(1+x)/(1-x)] =	0.9373	-1.9	43.008	46.7	0.28	0.13
0.97	1.9002		1	Cubic Root (x) =	0.9875	3.3	1790.929	14.5	1.37	0.2
0.02	1.1944		1	1 / x^2 =	0.7847	-17.9	-603.973	95.3	0.24	0.23
0.15	1.0929		1	1 / x =	0.9190	-3.8	-478.196	45.7	0.45	0.2
0.80	1.1964		1	(1/lambda)*asinh(lambda*root(x+0.5))	0.9872	3.3	3889.2418	8.9	0.98	0.09
0.70	1.1791		1	(1/lambda)*asinh(lambda*root(x)) =	0.9873	3.3	2380.3896	11.2	0.93	0.1
0.92	6.8610		1	New 1	0.9556	0.0	711.9878	102.2	1.14	1.17
0.92	6.8610		1	New 2	0.9556	0.0	711.9878	43.0	2.71	1.17
				Min or Max =	0.9902	3.6	-19122.097	8.9		

Home | Input & Summary | Graphic Summary | Original Data | Log10(x) | Log10(x+1) | Ln(x) | Sq. Root (x) | Square (x) | Arcsin(Root(x)) Degrees | Sq. Root(x+0.5) | Ln(x+1) | ...

[Step 3] Adjustments Needed



The chart may not be displaying correctly. The x-axis should be -4 to +4 (known bug in some versions of Excel)

To correct the x-axis values, click on the data points in the chart. Then in the formula bar, change

$D$2:D{x},
to
$D$2:D{value in cell J2},

❸ The toolbar will appear

❷ Click the right mouse button on the blue dots of the Normal Quantile Plot graph

❶ On "each" data transformation spreadsheet you are interested in, do the following:

[Step 3] Adjustments Needed

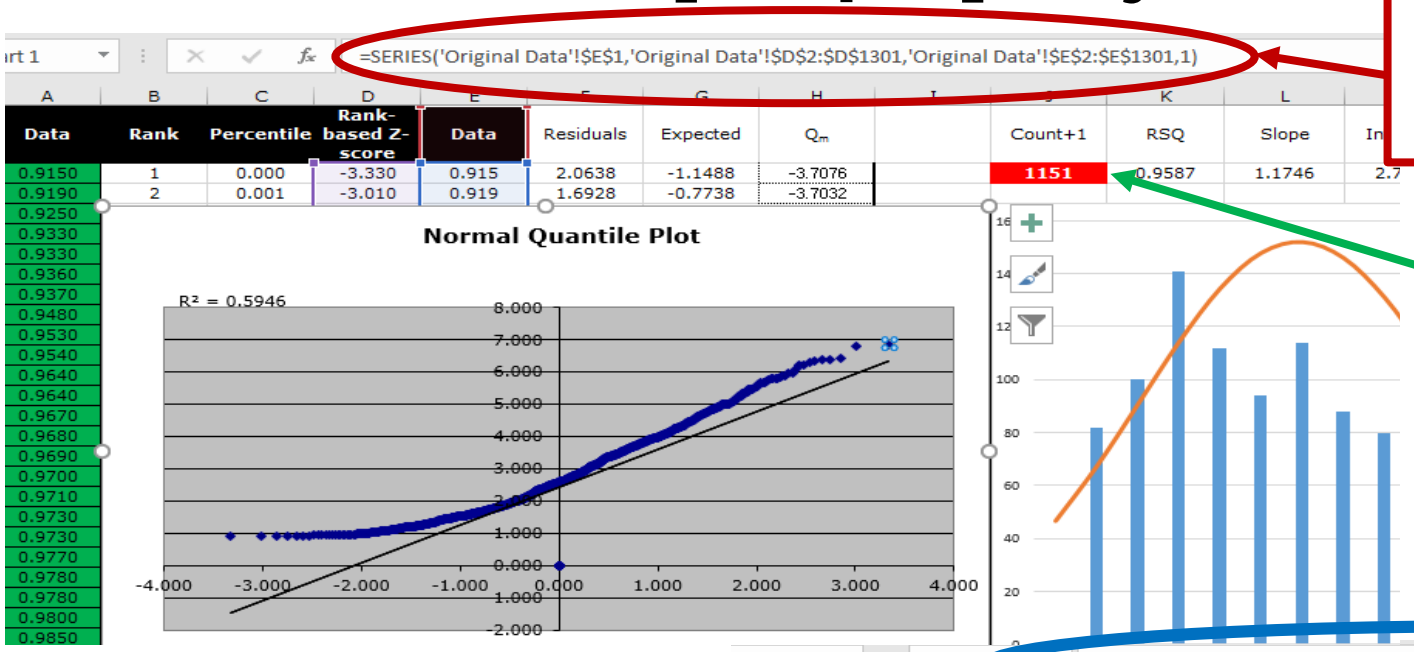
❶ Change the value that appears here (in this case, for example) "=SERIES('Original Data'!$E$1,'Original Data'!$D$2:$D$1301,'Original Data'!$E$2:$E$1301,1)"…

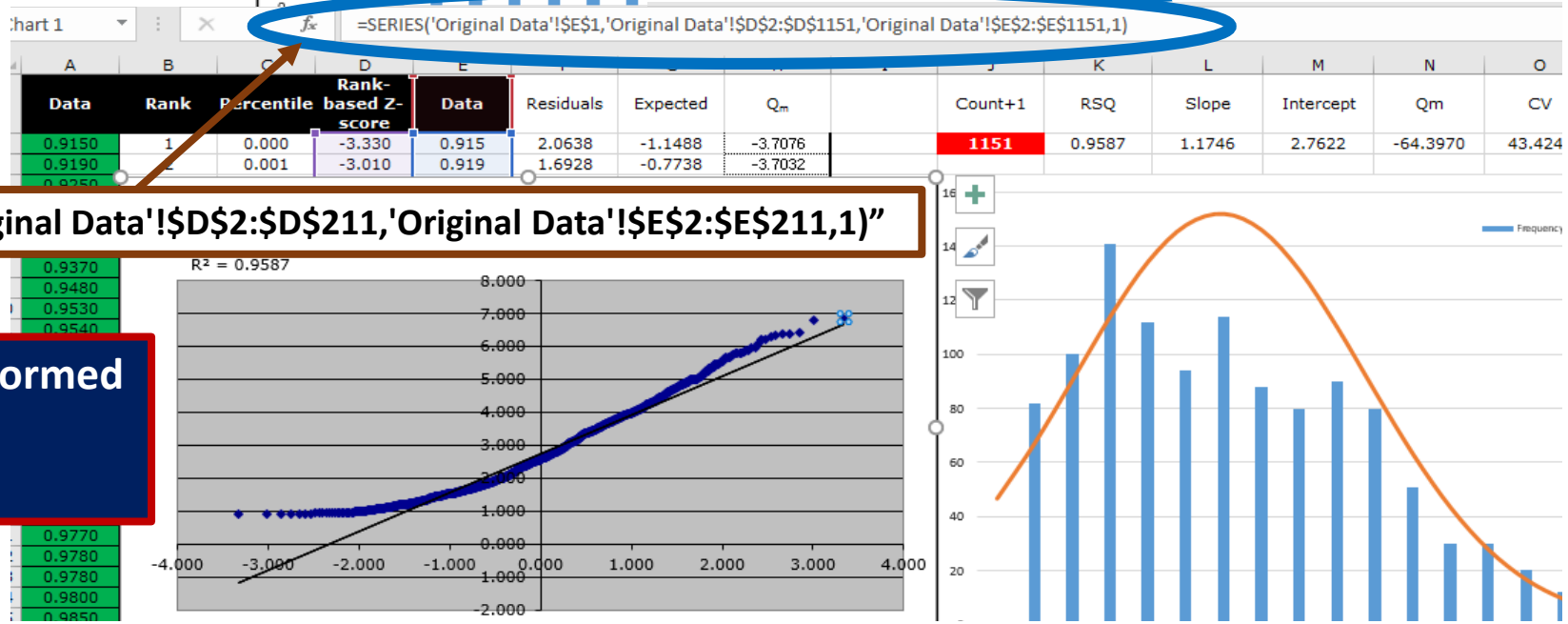
❷…By the value that appears in cell J2…

❸…Resulting in this equation…

❹… "=SERIES('Original Data'!$E$1,'Original Data'!$D$2:$D$211,'Original Data'!$E$2:$E$211,1)"

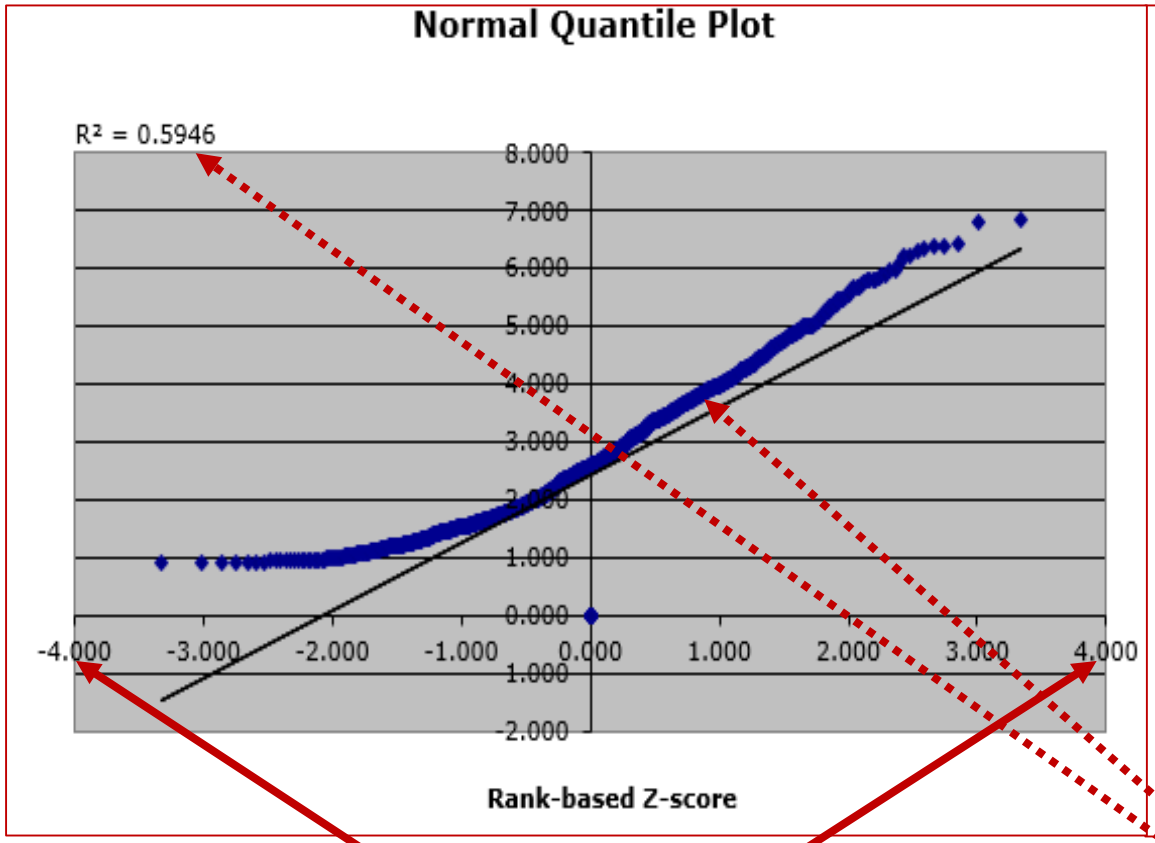
❺This procedure should be performed on all other worksheets "5 and 20 worksheets"



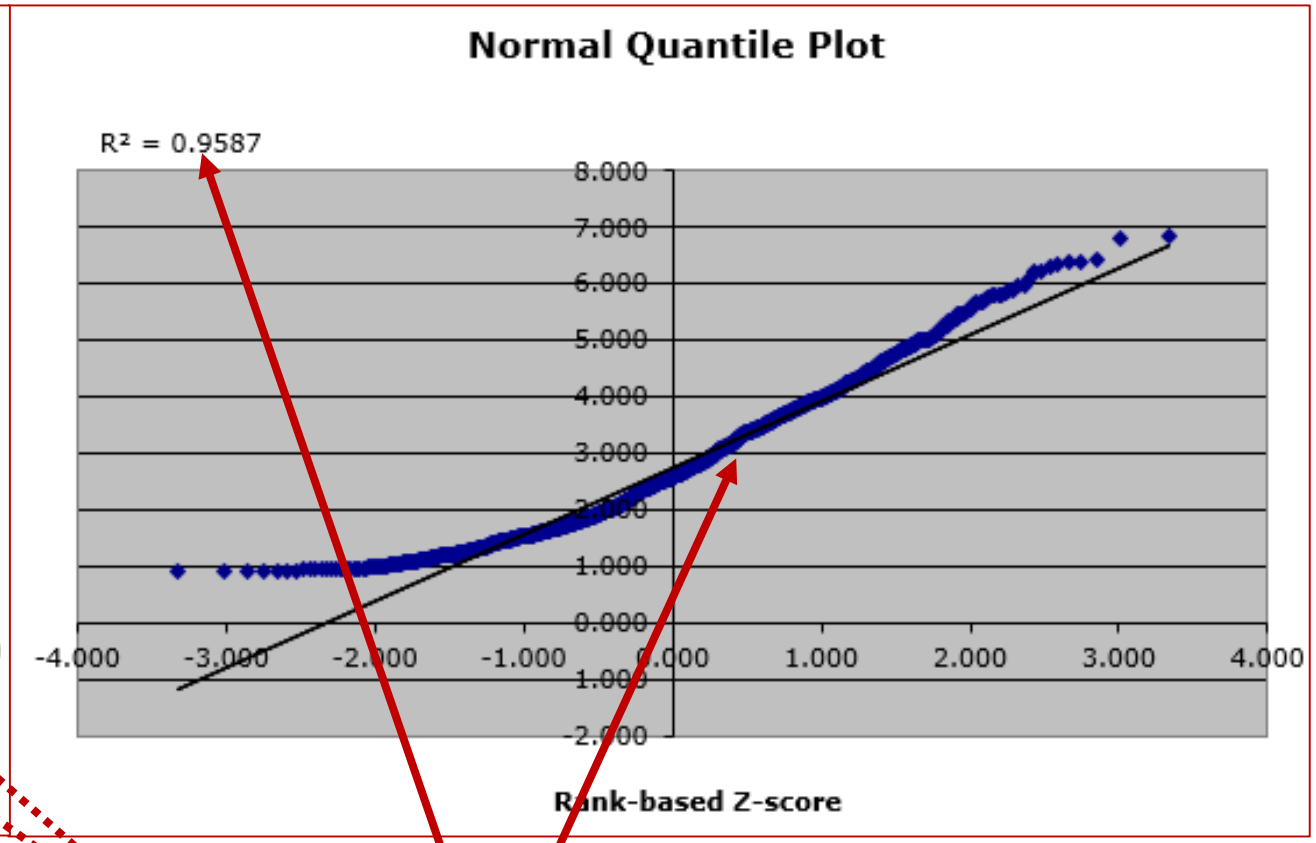


[Step 3] Adjustments Needed

Before the Adjustment



After the Adjustment



Observe that the x-axis should be -4 to +4

CHANGES OCCUR IN GRAPHIC AND IN R^2

[Step 4] Reading Outputs

❶ R^2 and CV values for normal distributions of original and transformed data

❷ Mean and SD of the original and transformed data

	Original Data		Ranked Data
1	5.646	0.915	0.915
2	4.063	0.919	0.919
3	3.401	0.925	0.925
4	4.463	0.933	0.933
5	4.277	0.933	0.933
6	4.301	0.936	0.936
7	5.244	0.937	0.937
8	2.901	0.948	0.948
9	2.234	0.953	0.953
10	4.473	0.954	0.954
11	3.789	0.964	0.964
12	3.875	0.964	0.964
13	2.653	0.967	0.967
14	3.35	0.968	0.968
15	4.005	0.969	0.969
16	3.789	0.970	0.970
17	3.68	0.971	0.971
18	2.364	0.973	0.973
19	2.297	0.973	0.973
20	5.006	0.977	0.977
21	2.284	0.978	0.978
22	3.064	0.978	0.978
23	2.693	0.980	0.980
24	3.101	0.985	0.985
25	2.979	0.990	0.990
26	2.514	0.991	0.991
27	3.399	1.006	1.006
28	2.412	1.010	1.010
29	2.631	1.014	1.014

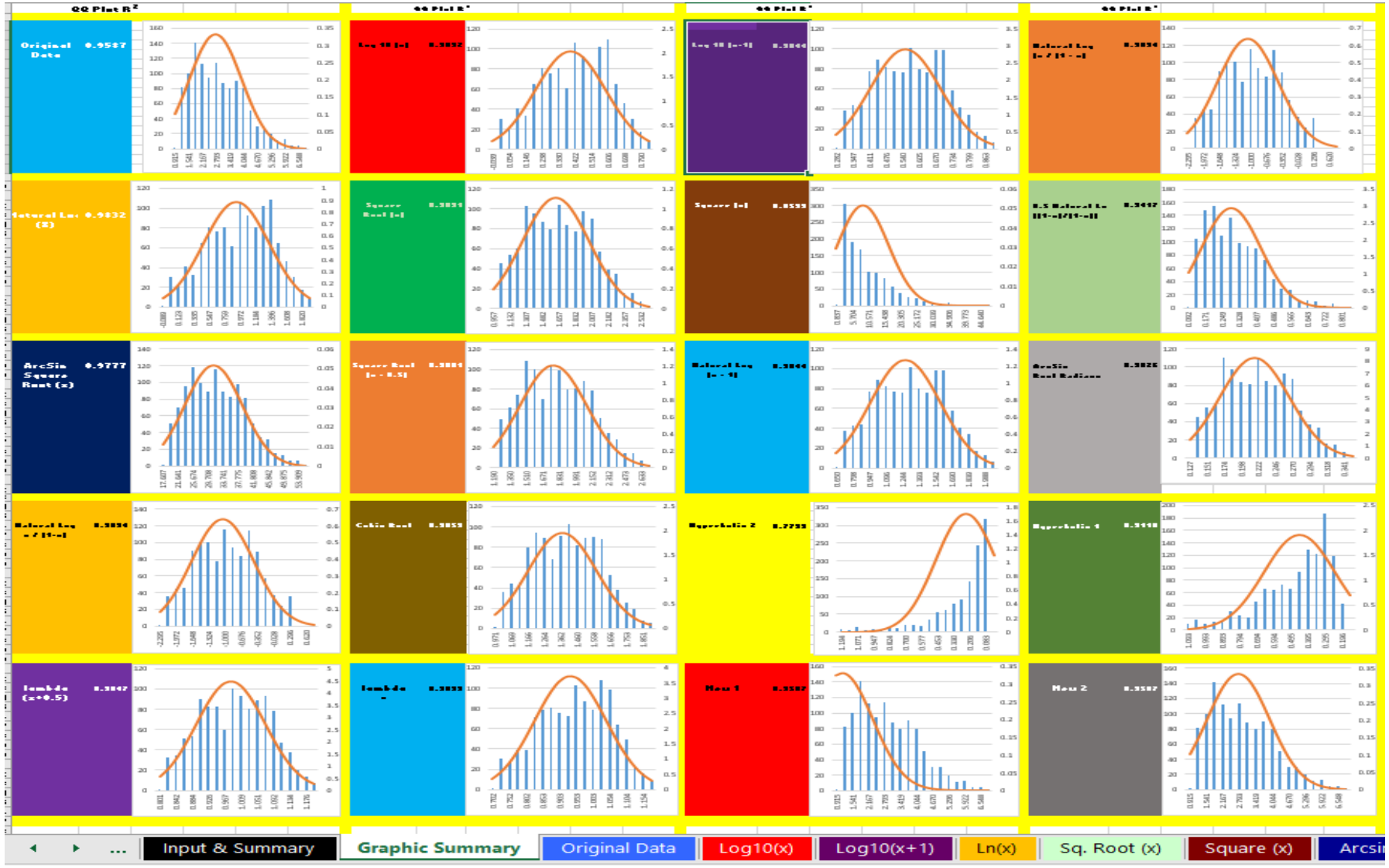
Input your data into cells A2 to A1301 (up to 1300 points)
QQ Plots can be viewed on the various spreadsheets
Some data will need to be divided by a Factor (Below) to be in the proper range
Some transformations will not work if the data contains any 0's

Min	Max	Requirement	Factor	Transformation	R^2	R^2 Improvement (%)	Qm	CV (%)	Mean	SD
0.92	6.8610		1	Original Data =	0.9587		-64.397	43.4	2.76	1.2
-0.04	0.8364	x>0	1	$Log_{10}(x)$ =	0.9832	2.6	718.194	49.0	0.40	0.2
0.28	0.8955		1	$Log_{10}(x+1)$ =	0.9844	2.7	103.725	24.8	0.55	0.14
-0.09	1.9259	x>0	1	Ln(x) =	0.9832	2.6	103.725	49.0	0.92	0.45
0.96	2.6194		1	Square Root (x) =	0.9831	2.5	3538.142	22.1	1.62	0.36
0.84	47.0733		1	Square (x) =	0.8539	-10.9	1426.911	85.3	9.07	7.74
17.61	55.9257	x<100	10	Arcsin(Root(x)) in degrees =	0.9777	2.0	111.396	24.9	31.20	7.77
1.19	2.7131		1	Square Root (x+0.5)	0.9801	2.2	-3711.509	18.4	1.78	0.33
0.65	2.0619		1	Ln (x+1) =	0.9844	2.7	1208.193	24.8	1.28	0.32
0.13	0.3534	x<1	1	ArcSin(Root(x)) in radians =	0.9826	2.5	-428.346	22.4	0.22	0.05
-2.30	0.7819	x<1	10	ln[x/(1-x)] =	0.9894	3.2	2380.507	59.7	-1.05	-0.6
0.09	0.8405	x<1	10	0.5ln [(1+x)/(1-x)] =	0.9417	-1.6	-21.309	47.2	0.29	0.14
0.97	1.9002		1	Cubic Root (x) =	0.9759	2.8	-2929.789	14.8	1.37	0.2
0.02	1.1944		1	1 / x^2 =	0.7793	-18.7	-475.730	97.9	0.24	0.23
0.15	1.0929		1	1 / x =	0.9118	-4.9	-2408.960	47.1	0.44	0.21
0.80	1.1964		1	(1/lambda)*asinh(lambda*root(x+0.5))	0.9847	2.7	-1458.0124	9.0	0.98	0.09
0.70	1.1791		1	(1/lambda)*asinh(lambda*root(x)) =	0.9839	2.6	-262.8376	11.4	0.94	0.11
0.92	6.8610		1	New 1	0.9587	0.0	-5006.6835	103.5	1.16	1.2
0.92	6.8610		1	New 2	0.9587	0.0	-5006.6835	43.4	2.76	1.2
				Min or Max =	0.9894	3.2	-5006.683	9.0		

❸ The best results will appear highlighted

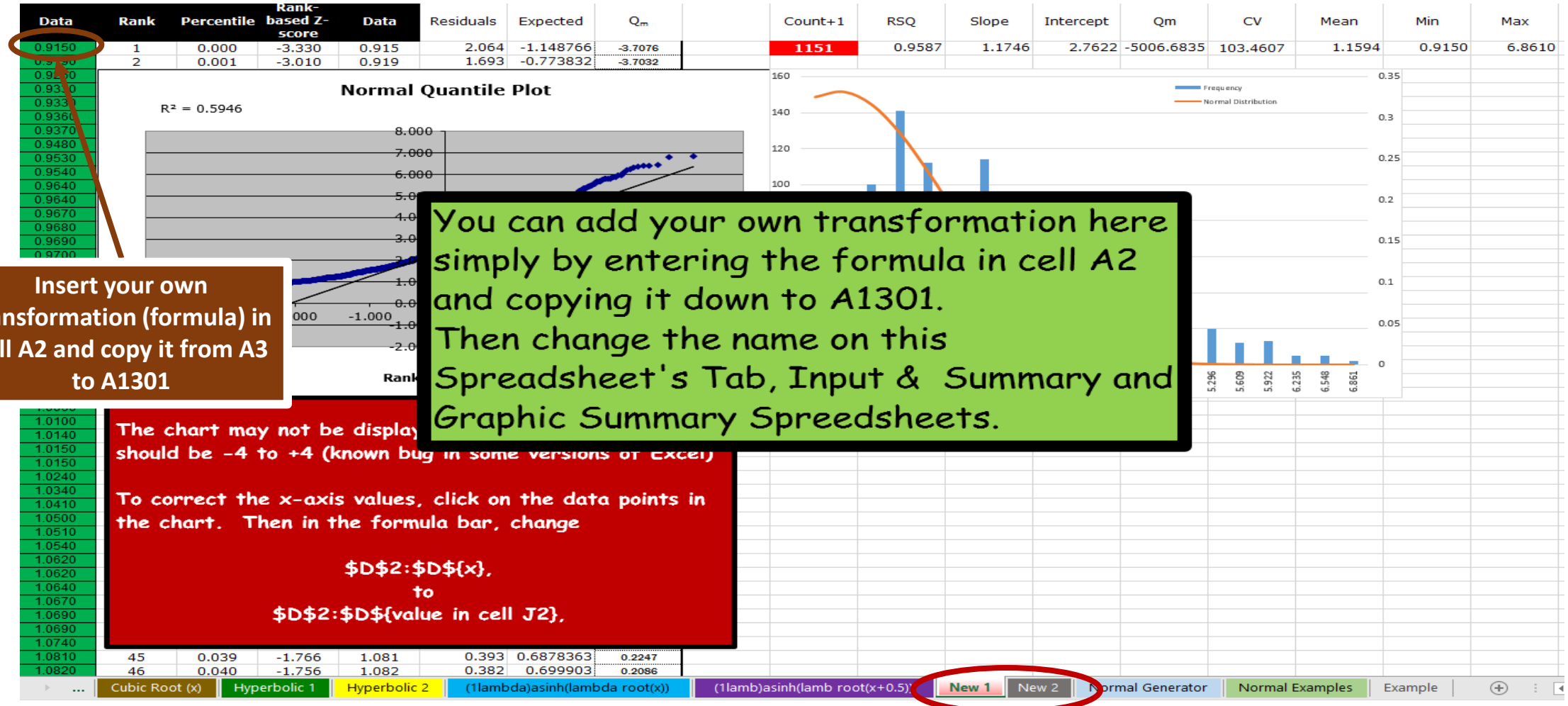
❹ Summary of the best results

[Step 4] Reading Outputs



Graphics of normal distributions for each transformation

[Step 5] New transformations



Insert your own transformation (formula) in cell A2 and copy it from A3 to A1301

You can add your own transformation here simply by entering the formula in cell A2 and copying it down to A1301.
Then change the name on this Spreadsheet's Tab, Input & Summary and Graphic Summary Spreedsheets.

The chart may not be display
should be -4 to +4 (known bug in some versions of Excel)

To correct the x-axis values, click on the data points in the chart. Then in the formula bar, change

$D$2:D{x},
to
$D$2:D{value in cell J2},

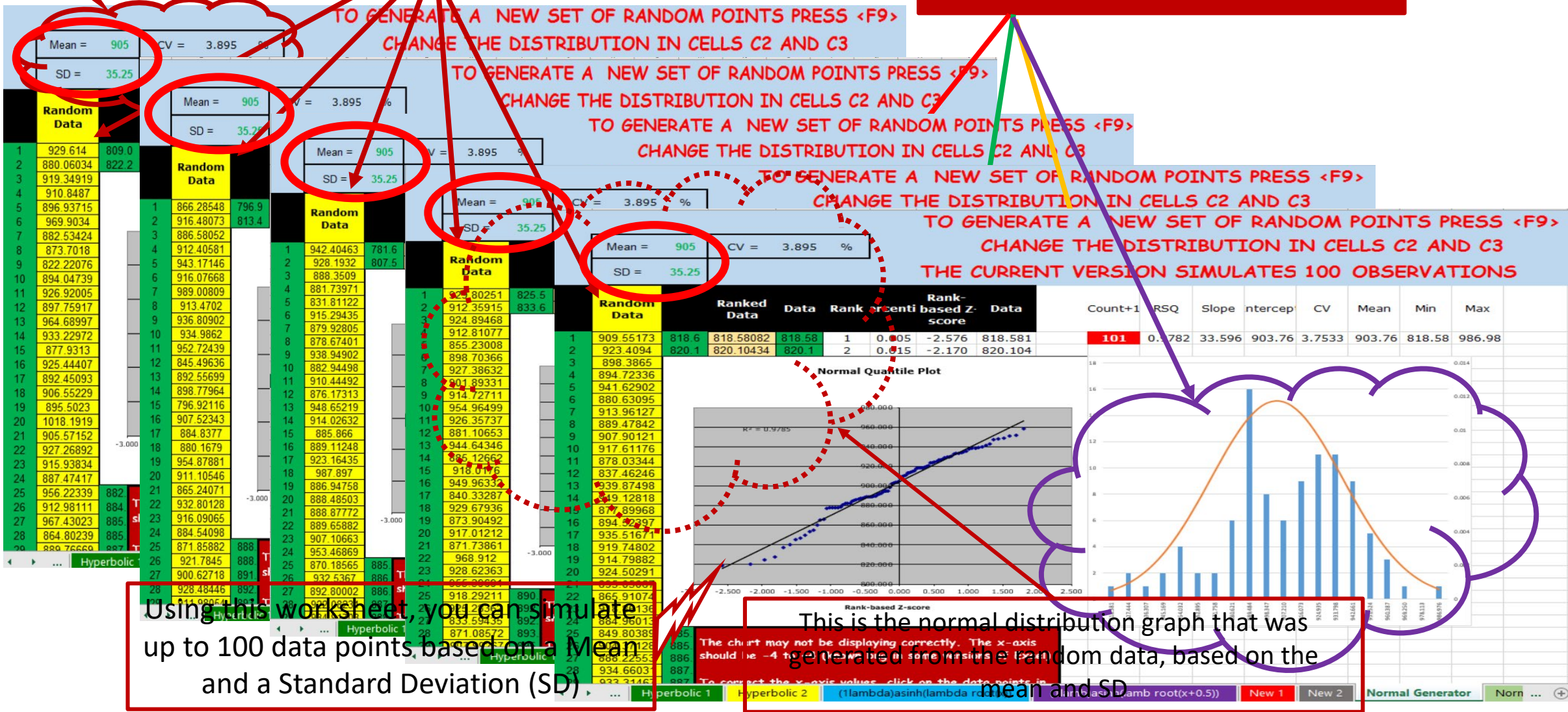
There are two worksheets for inserting transformations which aren`t presented here

The Current Version Simulates

Enter the values of the Mean and SD

By Pressing <F9>
New data will be generated with the same Mean and SD

Note that the data may differ in their normal distribution



Using this worksheet, you can simulate up to 100 data points based on a Mean and a Standard Deviation (SD)

This is the normal distribution graph that was generated from the random data, based on the mean and SD

Important points

❶ This workbook is a tool for interpreting the impact of transforming data for analysis and planning purposes.

❷ There is no objective criteria for deciding that one transformation is better than another.

❸ If you decide it is more appropriate to use transformed data when planning an experiment, the workbook ITSEPG may be helpful for graphing the inverse transformed data power curves.