Construct a Normal Probability Plot

Normal Probability Plot is also called a normal quantile-quantile plot): A graph we use to determine whether it is reasonable to believe that a data set was sampled from a normal distribution.

Example: Construct a normal probability plot for the following data.

125	220	290	210	180	300	180	80	125	200	0
135	140	260	150	180	0	150	50	50	210	240
290	0	230	200	140	200					

Column 1: Enter data values into excel sheet, then sort the data Calculate the mean and standard deviation of the sample data. Note that you can use the build in function in excel.

(mean = average(first cell: last cell); standard deviation = STDEV.S(first cell : last cell))

Let n = Number of data values

Column 2: Let i = position value. List all the i-values (1, 2, 3, …, n)

Column 3: Calculate the cumulative areas (cumulative relative frequencies) to the left of the corresponding sample values: $\frac{2i-1}{2n} = \frac{i-0.5}{n}$

Column 4: Find the z-scores corresponding to the area. $z = NormInv(area, 0, 1)$

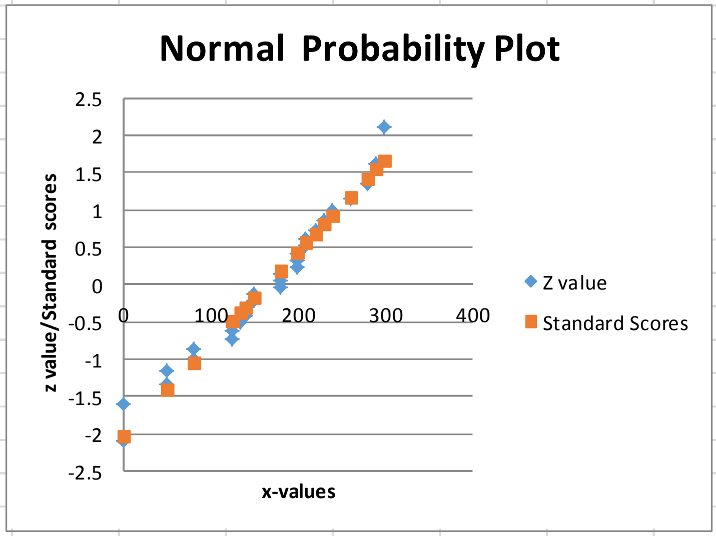
Column 5: Convert data value to standard scores $= \frac{x - mean}{standard\ deviation}$

1. Construct a scatterplot for three columns of data values (x-value, z value, standard value). Highlight these columns (I copied them into a new table). Go to insert, chart, and choose Scatterplot. Create the title for the graph and label the axes.
2. Note that there are two scatterplots.
   Square scatterplot: x-value vs. Standard scores
   Diamond scatterplot: x-value vs. corresponding z-value.

3. The closer the diamonds are to the squares, the more plausible it is that the data were sampled from a normally distributed population.

Values (x)	I (position)	Areas=(i-0.5)/n	z = (area,0,1)	Standardized scores = (x-mean)/standard deviation
0	1	0.017857143	-2.10016549	-2.022617399
0	2	0.053571429	-1.61116916	-2.022617399
50	3	0.089285714	-1.34516663	-1.407706572
50	4	0.125	-1.15034938	-1.407706572
80	5	0.160714286	-0.99152647	-1.038760076
80	6	0.196428571	-0.8544474	-1.038760076
125	7	0.232142857	-0.73180808	-0.485340331
125	8	0.267857143	-0.61930677	-0.485340331
135	9	0.303571429	-0.5141561	-0.362358166
140	10	0.339285714	-0.41441333	-0.300867083
140	11	0.375	-0.31863936	-0.300867083
150	12	0.410714286	-0.22570795	-0.177884918
150	13	0.446428571	-0.13468979	-0.177884918
180	14	0.482142857	-0.04477618	0.191061578
180	15	0.517857143	0.044776177	0.191061578
180	16	0.553571429	0.134689794	0.191061578
200	17	0.589285714	0.225707954	0.437025909
200	18	0.625	0.318639364	0.437025909
200	19	0.660714286	0.41441333	0.437025909
210	20	0.696428571	0.514156101	0.560008075
210	21	0.732142857	0.61930677	0.560008075
220	22	0.767857143	0.731808084	0.68299024
230	23	0.803571429	0.854447399	0.805972405
240	24	0.839285714	0.991526475	0.928954571
260	25	0.875	1.15034938	1.174918902
280	26	0.910714286	1.345166634	1.420883232
290	27	0.946428571	1.611169162	1.543865398
300	28	0.982142857	2.100165493	1.666847563
n	**28**			
Mean:	164.464286			
Standard Devia	81.3126031			

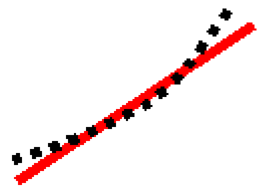
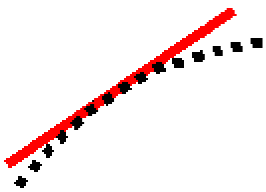
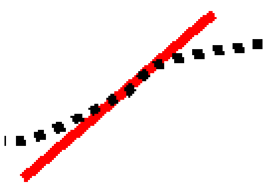
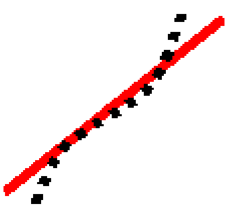
Value x	Z value	Standard Scores
0	-2.10017	-2.022617399
0	-1.61117	-2.022617399
50	-1.34517	-1.407706572
50	-1.15035	-1.407706572
80	-0.99153	-1.038760076
80	-0.85445	-1.038760076
125	-0.73181	-0.485340331
125	-0.61931	-0.485340331
135	-0.51416	-0.362358166
140	-0.41441	-0.300867083
140	-0.31864	-0.300867083
150	-0.22571	-0.177884918
150	-0.13469	-0.177884918
180	-0.04478	0.191061578
180	0.044776	0.191061578
180	0.13469	0.191061578
200	0.225708	0.437025909
200	0.318639	0.437025909
200	0.414413	0.437025909
210	0.514156	0.560008075
210	0.619307	0.560008075
220	0.731808	0.68299024
230	0.854447	0.805972405
240	0.991526	0.928954571
260	1.150349	1.174918902
280	1.345167	1.420883232
290	1.611169	1.543865398
300	2.100165	1.666847563



Note: Since the diamonds are to the squares, we conclude that the data were sampled from a normally distributed population.

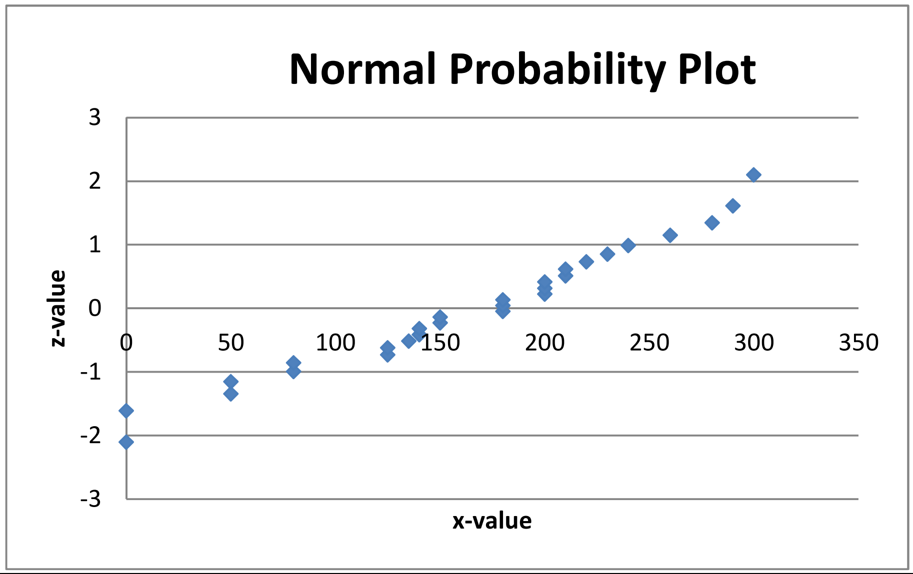
Alternative-- construct the normal probability plot only (original x-value vs. corresponding z-value).

If the pattern of the points is reasonably close to a straight line, then the data appear to come from a population that has a normal distribution. If the points do not lie close to a straight line, or if the points exhibit some systematic pattern that is not a straight-line pattern, then the data appear to come from a population that does not have a normal distribution.

If your chart looks like this:	It indicates that your distribution has:
	Right Skew - If the plotted points appear to bend up and to the left of the normal line that indicates a long tail to the right.
	Left Skew - If the plotted points bend down and to the right of the normal line that indicates a long tail to the left.
	Short Tails - An S shaped-curve indicates shorter than normal tails, i.e. less variance than expected.
	Long Tails - A curve which starts below the normal line, bends to follow it, and ends above it indicates long tails. That is, you are seeing more variance than you would expect in a normal distribution.

Example:

Value x	Z value
0	-2.10017
0	-1.61117
50	-1.34517
50	-1.15035
80	-0.99153
80	-0.85445
125	-0.73181
125	-0.61931
135	-0.51416
140	-0.41441
140	-0.31864
150	-0.22571
150	-0.13469
180	-0.04478
180	0.044776
180	0.13469
200	0.225708
200	0.318639
200	0.414413
210	0.514156
210	0.619307
220	0.731808
230	0.854447
240	0.991526
260	1.150349
280	1.345167
290	1.611169
300	2.100165



Note: Since the pattern of the points is close to a straight line, the data appear to come from a population that has a normal distribution.