Normal Probability Plots

If it is not known whether the underlying population is normally distributed, you can use a graphing calculator or software to construct a normal probability plot of the sample data. A **normal probability plot** graphs the data according to the probabilities you would expect if the data are normal, using *z*-scores. If the plot is approximately linear (a straight line), the underlying population can be assumed to be normally distributed.

Using a Graphing Calculator

A toy tricycle comes with this label: “Easy-To-Assemble. An adult can complete this assembly in 20 min or less.” Thirty-six adults were asked to complete the assembly of a tricycle, and record their times. Here are the results:

16	10	20	22	19	14	30	22	12	24	28	11	17	13	18	19	17	21
29	22	16	28	21	15	26	23	24	20	8	17	21	32	18	25	22	20

1. Using a graphing calculator, enter these data in L1. Find the mean and standard deviation of the data.

2. Make a normal probability plot of the data. Using **STAT PLOT**, select 1:Plot1, and the settings shown below.

Based on the plot, are assembly times normally distributed?

3. a) What is the probability that an adult can complete this assembly in 20 min or less?

b) What proportion of adults should complete this assembly within 15 to 30 min?

Using Fathom™

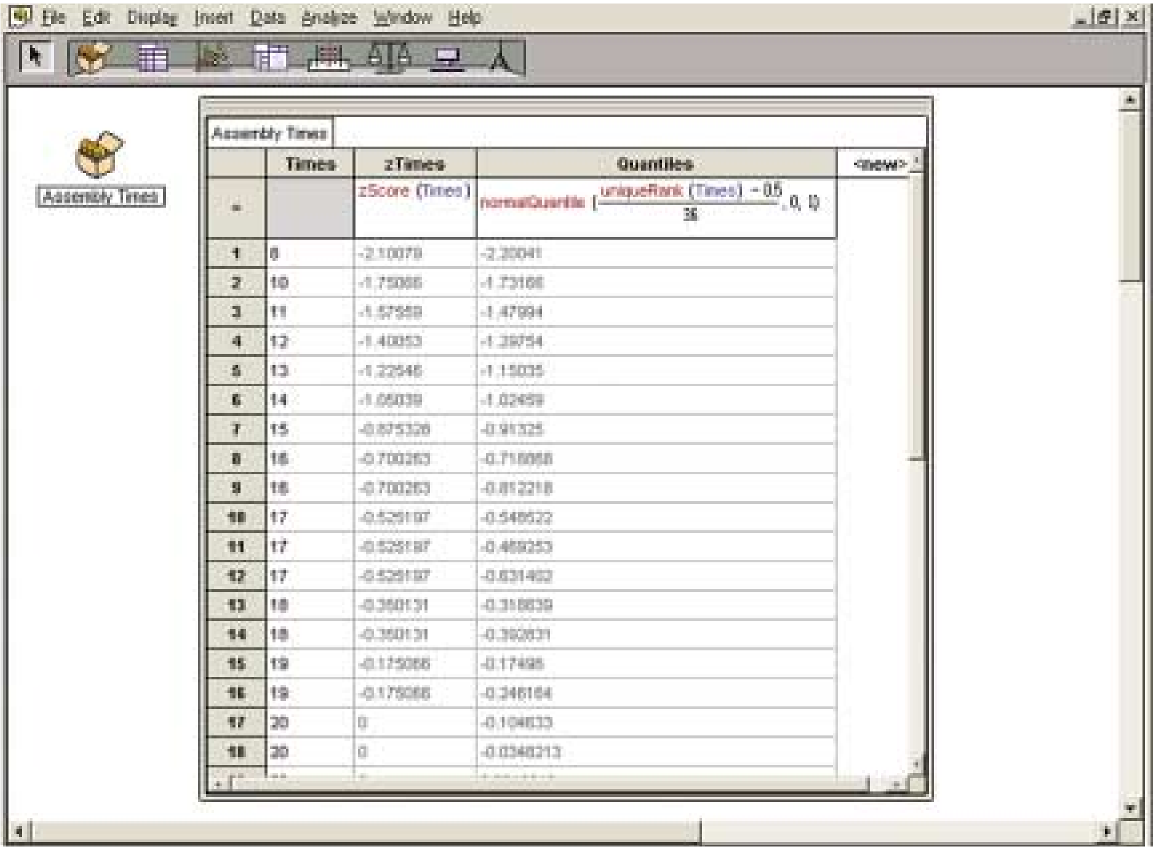
4. a) Open Fathom™, and open a new document if necessary. Drag a new **collection** box to the workspace. Rename the collection Assembly Times, and create 36 new cases.

b) Drag a new **case table** to the workspace. Name the first column Times, the second column zTimes, and the third column Quantiles.

c) **i)** Enter the time data in the first column. **Sort** it in ascending order.

ii) Edit the formula in the second column to **zScore**(Times). This will calculate the z-scores for the data.

iii) Edit the formula in the third column to **normalQuantile**((uniqueRank(Times) – 0.5)/36, 0, 1). This formula will calculate the z-scores of the quantiles corresponding to each entry in the Times column. The **uniqueRank( )** function returns the "row number" of the *sorted* data. Note that most of the quantile z-scores in the screen below are different from the z-scores for the corresponding data.



	Times	zTimes	Quantiles
=		zScore (Times)	normalQuantile ((uniqueRank (Times) – 0.5)/36, 0, 1)
1	8	-2.10079	-2.20041
2	10	-1.75066	-1.73166
3	11	-1.57559	-1.47994
4	12	-1.40053	-1.39754
5	13	-1.22546	-1.15035
6	14	-1.05039	-1.02459
7	15	-0.875326	-0.91325
8	16	-0.700263	-0.716868
9	16	-0.700263	-0.812218
10	17	-0.525197	-0.548522
11	17	-0.525197	-0.469253
12	17	-0.525197	-0.631402
13	18	-0.350131	-0.318639
14	18	-0.350131	-0.392831
15	19	-0.175066	-0.17496
16	19	-0.175066	-0.246164
17	20	0	-0.104633
18	20	0	-0.0348213

d) Drag a new graph to the workspace. Drag the Times title to the horizontal axis, and the Quantiles title to the vertical axis to generate a normal probability plot. Calculate the linear correlation coefficient for Times and Quantiles and comment on how near to linear this graph is. Are the data normally distributed?

e) Double click on the collection to open the **inspector**. Choose the Measures tab. Create four measures: Mean, StdDev, P20orLess, and P15to30. Use the **mean**, **standard deviation**, and **normalCumulative** functions to calculate the mean, the standard deviation, and the answers to question 3.

5. For each of the questions 1 to 6, 8, and 12 of section 8.3, pages 439 to 441, use a normal probability plot to determine how close to a normal distribution each data set is.

6. Let $x_1, x_2, \ldots x_n$ be a set of data, ranked in increasing order so that $x_1 \leq x_2 \leq \ldots \leq x_n$.
For $i = 1, 2, \ldots, n$, define the **quantile** z_i by $P(Z < z_i) = \dfrac{(i - 0.5)}{n}$,

where Z is a standard normal distribution (mean 0, standard deviation 1).

a) For a data set of your choice, plot a graph of z_i against x_i. Remember to sort the x-values into increasing order. Use the **invNorm(function** on your graphing calculator, or the table of Areas Under the Normal Distribution Curve on pages 606 and 607, to calculate the z-values. Notice that these quantile z-values are different from the z-scores in earlier sections.

b) Compare this graph with the normal probability plot for the data set. Explain your findings.

c) Explain why, if the data are normally distributed, a graph of z_i against x_i should be close to a straight line.