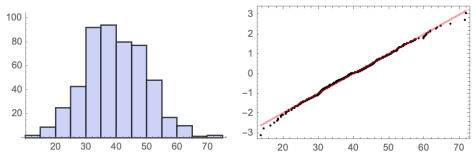
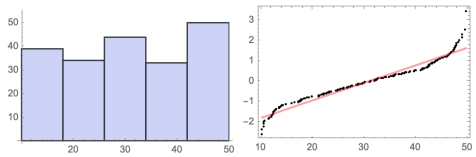
Math 146 6.5 — Assessing Normality

Back in chapter 2 (and Technology Insight 2) we briefly looked at *normal-quantile* plots to determine if a sample data set appeared to be from a normal population. We now have a few more tools to look at this idea a little deeper. If you recall, the closer the points lie to a perfect line, the more we can say the data is normal. If the data points are not close to a line, or show a patter, the data is not normal.

a) 

b) 

Data set (a) appears roughly bell-shaped and it's normal quantile plot is fairly linear. Data set (b) is more uniform and there is an obvious pattern to the data points.

So, what is a normal quantile plot really showing? The horizontal axis are the values of the actual data values, and the vertical axis are *z*-scores.

Steps to Create a Normal-Quantile Plot

(1) Reorder the data in ascending order (from smallest to largest.)

(2) We assign each data value to its respective *percentile*.

(3) For each percentile we calculate the *z-score*. (This is the longest step).

(4) For each data value and the *z*-score we create an order pair (value, z-score). This is why the horizontal axis are the data values, and vertical axis are the expected *z*-scores for each data value.

EXAMPLE 1: Create a quantile plot for the data: 32 33 42 39 20 25 32 37 39 34

Step 1) Order the data: 20 25 29 32 32 33 34 37 39 42

Step 2) Since we have 10 data values, 20 represents the bottom 1/10 of the data or the 10^{th} percentile. The 25 is also 1/10 of the data, so it is at the 20^{th} percentile etc. However, since the 20 is representative of the 10^{th} percentile (or the bottom 10%), that value could be anywhere from the 0% up to the 10%, so we split the difference and assign it **5%**. Same with 25, it represents the 20^{th} percentile, so it could actually be anywhere between about 10% to 20%. Again, split the difference and use **15%**. Continue this for all ten data values.

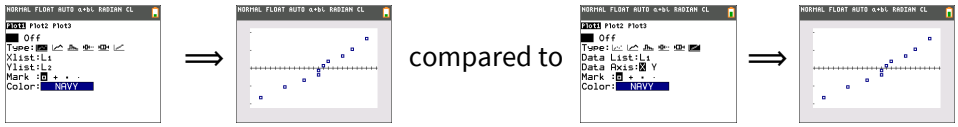
x	% tile	*z – score*
20	0.05	
25	0.15	
29	0.25	
32	0.35	
32	0.45	
33	0.55	
34	0.65	
37	0.75	
39	0.85	
42	0.95	

Step 3) Find the z-score for each of the percentages in the %tile column. Use **invNorm(** for the TI, or **NORM.INV(** for excel.

x	% tile	$z-score$
20	0.05	−1.645
25	0.15	−1.036
29	0.25	−0.674
32	0.35	−0.385
32	0.45	−0.126
33	0.55	0.126
34	0.65	0.385
37	0.75	0.674
39	0.85	1.036
42	0.95	1.645

Notice between the values 32 and 33 would be a z-score of 0, which makes sense because 50% of the data is on the left and 50% of the data is on the right of $z = 0$.

Step 4) Enter the first column of data values into **L1**, and the z-scores in **L2**, and make a scatter plot of the data.



I would say our data looks fairly normal.

Helpful hint: the %tile column is $\frac{1}{2n}, \frac{3}{2n}, \frac{5}{2n}$ etc., where n is the number of data values. For this data, we have $\frac{1}{20} = 0.05$, $\frac{3}{20} = 0.15$, $\frac{5}{20} = 0.25$, $\frac{7}{20} = 0.35$, etc.

EXAMPLE 2: Create a normal-quantile plot and assess the normality of the data:

44 70 68 60 45