General Education Statistics
Class Notes
Assessing Normality (Section 7.3)

> How do we know that our random variable is normally distributed?

We use these great statistics to answer all sorts of questions about a random variable. But how can we be sure that we are really dealing with a *normal* distribution. That assumption underlies all we have done recently. We will investigate normal probability plots.

Definition: A **normal probability plot** plots observed data versus **normal scores**.

A **normal score** is the expected z-score of the data value if the distribution of the random variable is normal. The expected z-score of an observed value will depend upon the number of observations in the data set.

Here are the instructions for making this plot. However, we will *not* be doing this by hand. Rather, we will interpret the output given by the calculator or StatCrunch.

Drawing a Normal Probability Plot:
Step 1 Arrange the data in ascending order.

Step 2 Compute $f_i = \dfrac{i - 0.375}{n + 0.25}$ where i is the index of each data value (its position in the list) and n is the number of observations. This f_i is the expected proportion of observations less than or equal to the ith data value.

Step 3 Find the z-score corresponding to f_i from Table V.

Step 4 Plot the observed values on the horizontal axis and the corresponding expected z-scores on the vertical axis.

Once you plot those points, what do you do?

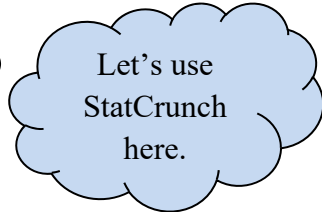**Assessing a Normal Probability Plot:**
If sample data is taken from a population that is normally distributed, a normal probability plot of the actual values versus the expected z-scores will be approximately linear.

We'll perform linear regression on this plot. If the linear correlation coefficient is greater than the critical value found in Table VI in the book, then we can conclude that the data could come from a population that is normally distributed.

Which is great since now we are justified in doing the stats we have been doing.

expl 1: A random sample of weekly work logs at an automobile repair station was obtained, and the average number of customers per day was recorded. Draw a normal probability plot and determine if the population is normally distributed.

Average Number of Customers in a Day				
26	24	22	25	23
24	25	23	25	22

Let's use StatCrunch here.

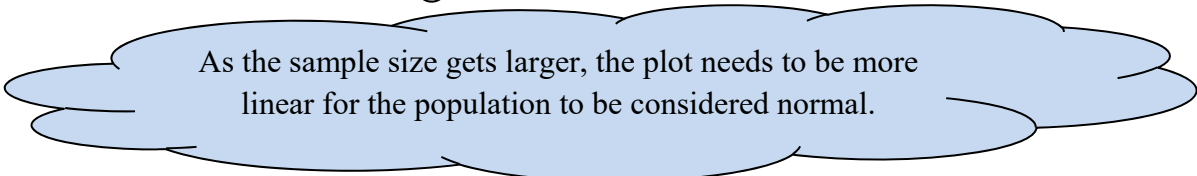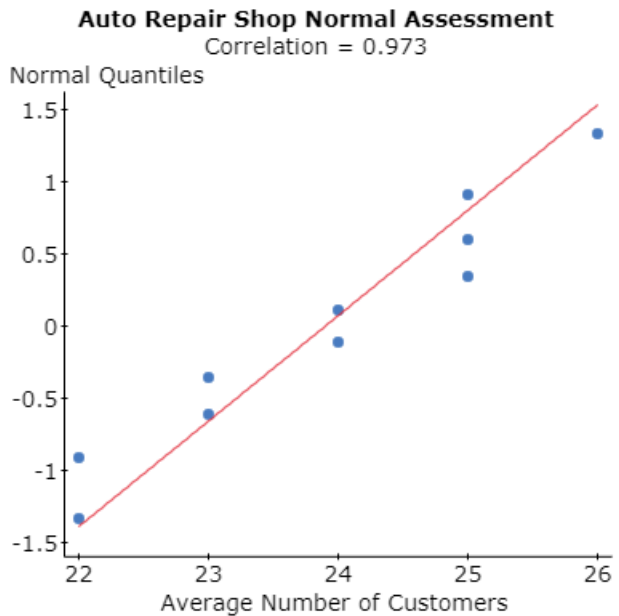**Instructions for Normal Probability Plots on StatCrunch:**

Enter the raw data in column **var1**.

Select **Graph** > **QQ Plot**. Use **var1** as the Selected Column. Check the boxes to add the *Correlation statistic* and *Normal quantiles on y-axis*. You can further add labels for the *x*- and *y*-axes and title your graph. Click **Compute!** when you are ready.

Now that we have this plot and the correlation coefficient, let's do some work ourselves.

Look up the critical value in Table VI for the sample size. This table is reproduced here in part.
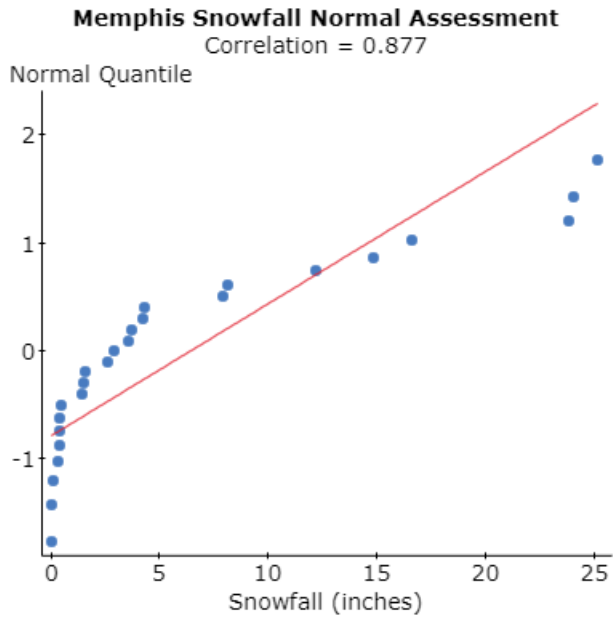
Table VI for Assessing Normality	
Sample size, n	Critical Value
5	0.880
10	0.918
15	0.939
20	0.951
25	0.959
30	0.960

Auto Repair Shop Normal Assessment
Correlation = 0.973

As the sample size gets larger, the plot needs to be more linear for the population to be considered normal.

Compare the correlation coefficient to the critical value. If the correlation coefficient is greater than the critical value, then we say the population is normal. What is your conclusion?

expl 2: A random sample of 25 years between 1890 and 2011 recorded the amount of snowfall, in inches, for Memphis, TN. Below is the data. Look at the normal probability plot provided. What is your conclusion about the normality of this population? Explain.

Inches of Snowfall in a Year				
24.0	7.9	1.5	0.0	0.3
0.4	8.1	4.3	0.0	0.5
3.6	2.9	0.4	2.6	0.1
16.6	1.4	23.8	25.1	1.6
12.2	14.8	0.4	3.7	4.2

Memphis Snowfall Normal Assessment
Correlation = 0.877

In fact, let's look at a histogram (from StatCrunch) of the data. Does the data look normal?

Memphis Snowfall Distribution
Number of Years out of 25 Sampled

Does this have the bell-shaped look of a normal distribution?