5 Checking Assumptions

Almost all statistical methods make assumptions about the data collection process and the shape of the population distribution. If you reject the null hypothesis in a test, then a reasonable conclusion is that the null hypothesis is false, provided all the distributional assumptions made by the test are satisfied. If the assumptions are not satisfied then that alone might be the cause of rejecting H_0 . Additionally, if you fail to reject H_0 , that could be caused solely by failure to satisfy assumptions also. Hence, you should always check assumptions to the best of your abilities.

Two assumptions that underly the tests and CI procedures that I have discussed are that the data are a random sample, and that the population frequency curve is normal. For the pooled variance two sample test the population variances are also required to be equal.

The random sample assumption can often be assessed from an understanding of the data collection process. Unfortunately, there are few general tests for checking this assumption. I have described exploratory (mostly visual) methods to assess the normality and equal variance assumption. I will now discuss formal methods to assess these assumptions.

Testing Normality

A formal test of normality can be based on a **normal scores plot**, sometimes called a **rankit plot** or a **normal probability plot** or a **normal Q-Q plot**. You plot the data against the normal scores, or **expected normal order statistics** (in a standard normal) for a sample with the given number of observations. The normality assumption is plausible if the plot is fairly linear. I give below several plots often seen with real data, and what they indicate about the underlying distribution.

There are multiple ways to produce normal scores plots in Minitab. The **NSCOR** function available from the Calculator or from the command line produces the desired scores. The shape can depend upon whether you plot the normal scores on the x-axis or the y-axis. SW plot the normal scores on the x-axis (that isn't very conventional) – Minitab wants to plot the normal scores on the y-axis if you use built-in procedures (you can override that, but don't). The information is the same, it's just the orientation of shape that differs.

Graphical displays for **normal** data:

```
Stem-and-Leaf Display: C1

Stem-and-leaf of C1 N = 150

Leaf Unit = 1.0

1 5 6

1 6 69

7 7 0011

13 7 578888

23 8 0002223334

34 8 55667788999

54 9 001111122223334444

73 9 5556677778888999999

(25) 10 000111112222333344444444

52 10 5555556677888899

35 11 00000111123334444

17 11 6688899999

7 12 00134

2 12 68
```



For comparison, consider the plot used by SW. In this case you see very little difference except the little flip on the end is in the opposite direction.



Either way, consider how the outlier shows up in the normal scores plot. You have an isolated point on both ends of the plot, but only on the left side is there an outlier. How could you have identified that the left tail looks longer than the right tail from the normal scores plot?

Examine the first plot (usual orientation). If you lay a straightedge along the bulk of the plot, you see that the most extreme point on the left is a little above the line, and the last few points on the right also are above the line. What does this mean? The point on the left corresponds to a data value *more extreme* than expected from a normal distribution (the straight line is where expected and actual coincide). Extreme points on the left are above the line. What about the right? Extreme points there should be *below* the line – since the deviation from the line is above it on the right, those points are *less extreme* than expected. For the SW orientation you have reverse this – outliers will be below on the left and above on the right. You are much better off sticking with one orientation, and Minitab's default is most common.

There are two considerably better ways to get these plots. We would like the straight line we are aiming for to actually appear on the graph (putting in a regression line is not the right way to do it, even if it is easy). Such a display comes from the menu path Stat > Basic Statistics > Normality Test. In the resulting dialog box you have choices of Variable (the data column), Percentile Lines (use None), and Test for Normality (probably use Ryan-Joiner, don't use Kolmogorov-Smirnov). We'll turn to those tests in a bit. The following graph results from following that path:



It is easier to read the graph with the line drawn for you. In this case the y-axis is labelled percent, but note that it is not a linear scale. This is the same graph as before, but with the normal scores identified with the percentiles to which they correspond. It is useful to do it this way.

Another plot, and probably the most useful of all, adds confidence intervals (point-wise, not family-wise. You will learn the meaning of those terms in the ANOVA section). You don't expect a sample from a normally distributed population to have a normal scores plot that falls exactly on the line, and the amount of deviation depends upon the sample size. Follow the menu path Graph > Probability Plot, click Single, make sure Distribution is Normal (you can use this technique to see if the data appear from lots of possible frequency distributions, not just normal), don't put in Historical Parameters, on Scale don't Transpose Y and X, and on Scale you can choose Y-Scale Type of Percent, Probability, or Score (normal score in this case) — the default is percent, and that works fine.



You only see a couple of data values outside the limits (in the tails, where it usually happens). You expect around 5% outside the limits, so there is no indication of non-normality here. Both

the Ryan-Joiner and Anderson-Darling tests concur with this (we'll discuss those shortly). They should - I *did* sample from a normal population.

Let's turn to examples of sampling from other, non-normal distributions to see how the normal scores plot identifies important features.

Graphical displays for a **light-tailed symmetric** distribution:



Graphical displays for a heavy-tailed (fairly) symmetric distribution:



Graphical displays for a distribution that is **skewed to the right**:



Graphical displays for a distribution that is skewed to the left:



Notice how striking is the lack of linearity in the normal scores plot for all the non-normal distributions, particularly the symmetric light-tailed distribution where the boxplot looks very good. The normal scores plot is a sensitive measure of normality. Let us summarize the patterns we see regarding tails in the plots:

	Tail	
Tail Weight	Left	Right
Light	Left side of plot points down	Right side of plot points up
Heavy	Left side of plot points left	Right side of plot points right

Be careful – plots in the SW orientation will be reverse these patterns.

Formal Tests of Normality

A formal test of normality is based on the **correlation** between the data and the normal scores. The correlation is a measure of the strength of a linear relationship, with the sign of the correlation indicating the direction of the relationship (i.e. + for increasing relationship, and - for decreasing). The correlation varies from -1 to +1. In a normal scores plot, you are looking for a correlation

close to +1. Normality is rejected if the correlation is too small. Critical values for the correlation test of normality, which is commonly called the **Shapiro-Wilk** test, can be found in many tests.

Minitab performs three tests of normality: the **Ryan-Joiner** test, which is closely related to the **Shapiro-Wilk** test, the **Kolmogorov-Smirnov** test, which is commonly used in many scientific disciplines but essentially useless, and the **Anderson-Darling** test (related to the Kolmogorov-Smirnov, but useful).

To implement tests of normality follow the menu path Stat > Basic Statistics > Normality Test. A high quality normal probability plot will be generated, along with the chosen test statistic and p-value. We already did this on p. 57. Further, the Anderson-Darling test is printed automatically with the probability plots we have been producing from the Graph menu.

Tests for normality may have low power in small to moderate sized samples. I always give a visual assessment of normality in addition to a formal test.

Example: Paired Differences on Sleep Remedies

The following boxplot and normal scores plots suggest that the underlying distribution of differences (for the paired sleep data taken from the previous chapter) is reasonably symmetric, but heavy tailed. The p-value for the RJ test of normality is .035, and for the AD test is .029, both of which call into question a normality assumption. A non-parametric test comparing the sleep remedies (one that does not assume normality) is probably more appropriate here. We will return to these data later.

Note: You really only need to present one of the normal scores plots. In order to get both tests you need to produce two plots, but in a paper just present one plot and report the other test's p-value.



Example: Androstenedione Levels

This is an independent two-sample problem, so you must look at normal scores plots for males and females. The data are easier to use UNSTACKED to do the normal scores test on the males and females separately. Boxplots and normal probability plots follow.



The AD test p-value (shown) and the RJ test p-value for testing normality exceeds .10 in each sample. Thus, given the sample sizes (14 for men, 18 for women), we have insufficient evidence (at $\alpha = .05$) to reject normality in either population.

The women's boxplot contains two mild outliers, which is highly unusual when sampling from a normal distribution. The tests are possibly not powerful enough to pick up this type of deviation from normality in such a small sample. In practice, this may not be a big concern. The two mild outliers probably have a small effect on inferences in the sense that non-parametric methods would probably lead to similar conclusions here.

Extreme outliers and skewness have the biggest effects on standard methods based on normality. The Shapiro-Wilk test is better at picking up these problems than the Kolmogorov-Smirnov (K-S) test. The K-S test tends to highlight deviations from normality in the center of the distribution. These types of deviations are rarely important because they do not have a noticeable effect on the operating characteristics of the standard methods. Minitab of course is using the RJ and AD tests, respectively, which are modifications to handle some of these objections.

Most statisticians use graphical methods (boxplot, normal scores plot) to assess normality, and do not carry out formal tests.

Testing Equal Population Variances

In the independent two sample t-test, some researchers test H_0 : $\sigma_1^2 = \sigma_2^2$ as a means to decide between using the pooled variance procedure or Satterthwaite's methods. They suggest the pooled t-test and CI if H_0 is not rejected, and Satterthwaite's methods otherwise. There are a number of well-known tests for equal population variances, of which Bartlett's test and Levene's test are probably the best known. Both are available in Minitab. Bartlett's test assumes normality. Levene's test is popular in many scientific areas because it does not require normality. In practice, unequal variances and non-normality often go hand-in-hand, so you should check normality prior to using Bartlett's test. I will describe Bartlett's test more carefully in our discussion of one-way ANOVA. To implement these tests, follow these steps: Stat > ANOVA > Test for Equal Variances. The data must be STACKED.

Example: Androstenedione Levels

The sample standard deviations and samples sizes are: $s_1 = 42.8$ and $n_1 = 14$ for men and $s_2 = 17.2$ and $n_2 = 18$ for women. The sample standard deviations appear to be very different, so I would not be surprised if the test of equal population variances is highly significant. The Minitab output below confirms this: the p-values for Bartlett's F-test and Levene's Test are both much smaller than .05. An implication is that the standard pooled CI and test on the population means is inappropriate.

