Project Module 2 – Due November 12, 11:59 AM

STAT-UB.0103 – Statistics for Business Control and Regression Models

Please attach Module 2 at the end of your writeup for Module 1. Otherwise, we won't remember what your project was about

Try to keep the writeup for this module to one or two pages.

In this module, you will do further investigations to see if your data came from a (nearly) normal distribution.

We will focus on the response variable, although in real life you would probably want to do a similar analysis on each of the predictor variables as well. A finding of non-normality implies that caution may be needed in interpreting subsequent findings for this data set. Here, we just want to see if there is any strong evidence of non-normality.

For your response variable, use Minitab to create a Normal Probability Plot. To do this, use the commands **Stat** ⇒ **Basic Statistics** ⇒ **Normality Test**. In the "Variable" box, put your response variable. Make sure the box for "Anderson-Darling" is checked. You can type in a title for your plot.

This plot gives the percentiles of a normal distribution vs. the percentiles of your data set. If the data set came from a normal distribution, the plot should produce a straight line. For guidance, Minitab draws in this ideal straight line. Non-normal distributions will produce curvature in the plot.

1. Describe the pattern you see in your plot. Does this pattern seem to indicate non-normality? If so, in what way? (For example, is the upper tail longer than that of a normal distribution? To diagnose this, see if the points seem to be bending downwards from the line on the right hand side of the plot.)
2. On the upper right hand side of the normal probability plot, you will see a box containing five numbers. The fourth number ("AD") is the Anderson-Darling Statistic. The larger this number is, the stronger the evidence of non-normality. The fifth number ("p-value") gives the probability that we would get such a large value for AD if the distribution were actually normal. The closer this p-value is to zero, the stronger the evidence of non-normality. Does this p-value seem to indicate non-normality in your data?
3. Do your findings based on this plot agree with what you found in Module 1 based on the Descriptive Statistics plot? (If you had deleted the normal curve from the descriptive statistics plot, create another one which does show the normal curve).
4. Repeat questions 1–3 for the logarithm of the response variable. If it's not possible to take logs of the response variable values, skip this question.