**Lab 1**

The purpose of this lab was to become familiarized with the Minitab 15 statistical package. The tutorial data set Bears.mtw was used for the following exercises. A graphical summary of this data set will not be included in the introduction as some of the following exercises are to produce such summaries.

This data set contains data for bears that were anesthetized and measured to gather data to create table(s) for hunters to use to estimate various characteristics about the bears they hunted. “This would be used because in the forest it is easier to measure the length of a bear, for example, than it is to weigh it,” (Minitab Help).

1. Download the Minitab data set Bears.MTW to your Minitab session.

Completed.

1. Get the description of the data set Bears.MTW. Explain why bear’s body shape probably depends on season.

Bears’ body shape likely depends on the season due to the fact that they hibernate in the winter. Also, all wild bears are born in January, so shape of a bear may depend on its relative development (i.e. youth vs. adult), which would also coincide with season.

1. Construct a histogram and dot plot of bear lengths. (Mark axes in your plots, make informative titles, play with colors, choose the best number of bins for the histogram, etc.) Discuss which graph gives a better representation of the data.



Figure 1: A histogram of bear lengths with a normal distribution curve fitted to the data superimposed on the graph. 15 bins were used to nicely display the data so that the distribution of the data looks more continuous than choppy.



Figure 2: A dot plot representing the frequencies of various observed bear lengths.

The dot plot and histogram perform similar functions: displaying the relative frequencies of various measurement observations. The dot plot displays the data more precisely than does the histogram. However, the histogram displays the data in a more aesthetic manner, and represents the general feel of the distribution of the data (in part due to the fitted curve superimposed on the graph). The dot plot might be superior if the data set had a smaller number of observations and we wanted to see where each individual observation lined up, but as “n” grows, the histogram becomes the preferable choice.

1. Use probability plots to find a distribution that closely approximates the bear lengths. Justify your choice by graphs (do not use formal tests!).



Figure 3: Based on the histogram in Figure 1, the data seems to fit a normal distribution reasonably well. To test this further, the above graph is a probability plot of the bear length data is plotted against what a normal distribution and 95% confidence interval would predict. The data, again, seems to fit reasonable well, with only a few points falling outside the 95% confidence interval, and the general trend of the data points being linear in a manner that follows the linear normal trend.



Figure 4: To try something different and attempt to see if a better distribution to model the data can be found, we use a Lognormal distribution. The data does not fit this distribution as well. The data points seem to make a non-linearly curved trend, and some of the points lie outside the 95% confidence interval. A lognormal distribution does not predict the data as well as a normal one.



Figure 5: Finally, just to try one more distribution, the data is plotted against an exponential distribution. This is clearly a very poor fit. Few of the points fall inside the 95% confidence interval, and the trend of the data points does not match that of the exponential trend.

All in all, it would appear that a normal distribution would be the best for predicting bear lengths, at least for the distributions analyzed above.

1. Use Minitab to compute P(X > 6) for X~N(5, 9).

P( X > 6 ) = 0.455764