STAT 554: HW #3
due March 24, 2008

Instructions: For this assignment, I want you to put all of your answers in the appropriate boxes on the answer sheet (distributed in class), attach the three plots requested to the answer sheet, and just turn in that material. Do not turn in any other work with this assignment. I'll truncate scores at 10 for this assignment.

Important: Don't delete any outliers — that is, assume any unusual values are legitimate observations as opposed to being due to mistakes of some sort.

1) Given with this assignment is a data set consisting of 57 $B_{\max}$ values. (*Note*: See the course web site for the data in a form that may be nice if you want to copy and paste it into a Windows software package spreadsheet. In class I will provide a hard copy sheet showing and describing the data sets. If, for some reason, you want me to e-mail you a file containing the data sets, send me an e-mail request.)

(a) (0.5 point) Estimate the mean of the underlying distribution with the sample mean. (Round to the nearest integer (since this is consistent with the 2nd significant digit of the estimated standard error rule).)

(b) (0.5 point) Estimate the median of the underlying distribution with the sample median. (Round to the nearest integer.)

(c) (1.5 points) Estimate the 90th percentile ($\xi_{0.9}$) of the underlying distribution. (Note: For this part, I want you to use the method described in the class notes (p. 115) instead of any of the alternative methods presented in the 7th lecture (that you are to consider using for HW #4).) (Round to the nearest integer.)

(d) (0.5 point) Estimate the skewness of the underlying distribution with the sample skewness. (Round to the nearest tenth.)

(e) (1 point) Give a symmetry plot based on this sample. Be sure to include the proper "comparison line."

2) Given with this assignment is a data set consisting of 41 humerus bone length-to-width ratios.

(a) (0.5 point) Estimate the kurtosis of the underlying distribution with the sample kurtosis. (Round to the nearest tenth.)

(b) (0.5 point) Give a probit plot.

(c) (0.5 point) Give a Q-Q plot based on the $\mathcal{T}_4$ distribution.

3) Now you are to consider six different data sets.

Set 1 — The sample of 57 $B_{\max}$ values (from the first problem).
Set 2 — The sample of 41 ratios (from the second problem).
Set 3 — The sample of 70 pregnancy durations.
Set 4 — The sample of 40 red pine heights.
Set 5 — The sample of 27 solar radiation values.
Set 6 — The sample of 80 sulfur oxide emission values.

In each of parts (a)–(e) below, I describe data sets. Your task to create the best matching of the six data sets and the five descriptions. (Two data sets should be matched to the description for part (e).) (Each part is worth 1 point.)

(a) This data set comes from a distribution which is apparently heavy-tailed and perhaps nearly symmetric.

(b) This data set exhibits no appreciable deviations from normality even though the underlying distribution cannot be exactly normal.

(c) This data set comes from a distribution which is apparently slightly light-tailed and perhaps nearly symmetric.

(d) This data set comes from a distribution which seems to be negatively skewed.

(e) These two data sets come from distributions which seem to be positively skewed.