Case-based Social Statistics I  
CSSS 321  
Professor: Mark S. Handcock  

Solutions to Laboratory 3  
Due Friday, January 21, 2005

Problems to be handed in:

1) Submit on paper exercises 22-31 from Unit A5 of CyberStats entitled “Describing Data Graphically.”

A5 Ex. 22. What is the median height, approximately? How tall is the tallest person? About what percentage of these athletes are taller than six feet? Are these people generally taller than the average male?

Solution: The median height is about 74 inches. Tallest is about 84 inches. We can determine what percentage of these athletes are taller than six feet (72 inches) by counting them in the window created by clicking the “Data” button. 102 out of 132 are taller than 72 inches. This is 77%. Alternatively, the boxplot indicates that the lower quantile is at 72 so about 75% are above 72” (because 72” is Q1). Yes, they are taller than average as the average male is about 5’10” or 70 inches.

A5 Ex. 23. What are the two quartiles for the Weight variable? How heavy is the heaviest person? Are these athletes in general heavier than the average male (in your experience)?

Solution: The lower quartiles is about 215 and the upper quartile about 330. The heaviest person weights about 525 (given by the outlier, not the end of the upper fence!). And yes, in general athletes are heavier than average.

A5 Ex. 24 What is the median BMI number? Approximately what percentage is ”overweight”, i.e., has BMI over 25? What percentage is ”underweight”?

Solution: The median BMI number is about 31. About 75 per cent is overweight (has BMI > 25) and 0 per cent is underweight (has BMI < 20).

A5 Ex. 25 The following is an interactive histogram for the heights. Adjust the number of the bins by moving the slider. Try different values. How many bins revealed the data best? What feature do you see in the data?

Solution: There is a wide range of answers for the number of bins, as long as two clusters show up, and the bins aren’t too narrow (too many unnecessary details), it is a good answer. About 27 bins is a good choice, starting from around bin number 15 you get to see the two clusters. These 2 clusters are the feature that becomes visible.

A5 Ex. 26 The following is an interactive histogram for the weights. How many bins revealed the data best? What feature do you see in the data?

Solution: Just as in the previous question there is a wide range of answers for the number of bins. As long as two clusters show up, and the bins aren’t too narrow (too many unnecessary details), it is a good answer. About 35 bins is a good choice, starting from around bin number 15 you get to see the two clusters. These 2 clusters are the feature that becomes visible.
A5 Ex. 27 About what percentage of these athletes weigh more than 350 pounds?

Solution: About 20 to 25 per cent of the athletes weigh more than 350 pounds. You can find this answer by adding up the relative frequencies (the heights of the bins) on the right of 250.

A5 Ex. 28 The following is an interactive histogram for BMI. Set the number of bins to about 40 so that there are two distinct clusters. What are the medians, approximately, of the two clusters?

Solution: The median of the cluster of the smaller values is around 25, of the larger values around 45.

A5 Ex. 29 Which of the two clusters is more spread out?

Solution: The cluster of the larger values of BMI is more spread out (it has a larger range of values)

A5 Ex. 30 Of the three histograms, which showed most clearly two clusters?

Solution: The last histogram, that of the BMI, showed the two clusters most clearly

A5 Ex. 31 Which showed the clusters better, the histograms or the boxplots?

Solution: Boxplots do not show clustering, the histograms do (if the number of bins is chosen appropriately)