65 points

Statistics in Sports (Math 192) - Quiz 1
Spring 2009 - Brad Hartlaub

Name Brad Hartlaub - Solutions

Directions: Please answer all of the questions below. The point values for each problem are indicated in parentheses. Partial credit will be awarded if you show your work. Be careful not to spend too much time on any one part of a question.

1. The file P:\data\math\hartlaub\sportsstats\MadduxGlavin.mtw contains career pitching statistics (W = number of wins, L = number of losses, PCT = winning proportion, and ERA = earned run average) for Greg Maddux and Tom Glavin. The first 6 columns contain the statistics for both pitchers, and then the data were unstacked so that the statistics for Glavin are in columns C7 though C11 and the statistics for Maddux are in columns C12 through C16.

a. What plot would you use to provide a graphical comparison of the season winning percentages for these two pitchers? Explain why you chose this plot. (10)

A side-by-side boxplot is the most useful plot for comparing distributions because you can easily compare centers (line inside the box), variabilities (lengths of the boxes), and extremes (whiskers).

b. Which pitcher tended to win a greater percentage of games? Explain. (10)

The boxplot shows that the medians are almost identical. Both pitchers tend to win about 60-65 percent of their games. Maddux has the higher mean (.624 vs. .609) and median (.656 vs. .645)

c. Which pitcher tended to have a lower season ERA? Explain. (10)

Maddux has on average a lower ERA (3.035 vs 3.541)

d. Find and interpret the z-score for Tom Glavin's lowest winning proportion. Using the 1.5*IQR criterion, should this value be tagged as an outlier? Explain. (10) 5

$z\text{-score} = \frac{.291 - .6089}{.1465} = -2.17085$ ← This value is 2.17 standard deviations below his average.

$1.5 * IQR = 1.5(.14) = .21$ $\quad Q_1 - 1.5\,IQR = .56 - .21 = .35$ Since $.291 < .35$, this observation should be tagged as an outlier.

e. Plot the ERA for Greg Maddux against time. Is there an obvious pattern? Explain. (10)

Yes, the loess smoother clearly shows a decrease from 1985 to about 1995 and then a more gradual increase from 1995 to 2001.

f. Find the regression line for predicting Tom Glavin's winning proportion from year and identify the fitted value and residual for 1995. Does this linear model provide a good fit? Explain. (15)

$\widehat{\text{Winning\_PCT}} = -39.4 + .0201\ \text{Year\_Glavin}$

1995: Fit $= \hat{y}_{1995} = .629005$; Resid $= .695 - .629005 = .065995$

No, this linear model does not provide a good fit. $R^2 = 37.6\%$, so only 37.6% of the variability in winning proportion is explained by year. A plot also shows a clear quadratic pattern