Problem Set 9 - Due date: 11/13

1) Earnings of baseball players

(a) The dataset X:\PACDATA\ECON70\baseball.dat contains 745 observations on baseball players from 1990. The set contains three columns of data:

1) 1990 earnings for the player
2) year the player first entered the major leagues (ranges from 66 to 90)
3) 1 if the player is a pitcher and 0 if the player is not a pitcher ("batter")

(b) Construct a confidence interval to test whether pitchers and batters earn the same on average. Do you reject or fail to reject this hypothesis? What is its p-value?

(c) Plot earnings vs. experience for all the players. Looking at the graph, does there appear to be a relationship? Now do this for pitchers and batters separately. Does there appear to be a stronger relationship for one group than the other?

(d) Calculate the regression line: earnings = a + b*(years of experience). [Note: if you are doing this in a spreadsheet that does not have a regression function, you can calculate a and b using formulas (11-3) and (11-6) in Chapter 11.] Looking at the regression function, does there appear to be a relationship between earnings and experience? Now do this for pitchers and batters separately. Does there appear to be a stronger relationship for one group than for the other?

(e) Discuss how to go about "improving" on the calculations in (d) between earnings and experience for baseball players. [Think about what other factors you might want to control for to get a "clean" estimate of the effect of experience on earnings for players.]

2) Please do problems 11-2, 12-4 (for problem 11-2 only), 12-6, and 12-8 of the textbook.