

**Economics 472**

**Problem Set #4**

(1) Consider the regression of  $y$  on just a constant term : (no  $x$  included)

$$y_i = \beta_1 + u_i.$$

(1a) Show that the OLS estimate of  $\beta_1$  is just the sample mean of  $y$ :

$$\hat{\beta}_1 = \bar{y}.$$

(Hint: Minimize  $\sum_{i=1}^n u_i^2 = \sum_{i=1}^n (y_i - \beta_1)^2$  by choosing  $\beta_1$ . Simply differentiate this expression with respect to  $\beta_1$ , and set that derivative equal to zero. Then solve for  $\hat{\beta}_1$ . If you are having trouble doing this, just go back to your lecture notes, find our previous expression for the first-order condition (partial derivative) for the *intercept* and set  $\hat{\beta}_2 = 0$ .

(1b) What is the  $R^2$  for this regression? (Use the formulas for  $R^2$  presented in class. Think about the value of your predicted  $y$  for each observation. )

(2) Get the baseball data “bballdat.raw” found on the course web site. Save it to your disk in the “a” drive. This is the same data that was used in class which contains information for 30 teams during the 2001 season. This question will require you to perform some simple analysis of this data using STATA.

To create a log file of your output, I might recommend starting out by typing something like:

**log using a:baseball, replace t**

This will create a log (or history) of all of the results that you will get, and will call that file baseball.log. This file will be saved onto your data disk.

The baseball data contains 4 variables: wins, era, batting avergae, and attend. To load these into STATA, I might type:

**infile wins era batavg attend using a:bballdat.raw**

This loads the data into STATA, and creates the variables wins (Number of wins), era (Earned Run Average), batavg (Batting average) and attend ( attendance in 1,000’s).

Do the same simple regression with this baseball data that is included in your lecture notes. Verify that you obtain the same coefficient estimates.

Present a scatterplot of the data, as well as your estimated regression line, as I have provided in your lecture notes. Follow the same procedure as outlined in our lab discussion to create this plot. Close the log file, and clear your session.

To receive credit for this question, print out your regression output and graph and include it with your problem set.

(3) The data for this question can be obtained from the class website. It is called “unemp.raw” under the ”Unemployment data” link. Save it to your data disk in the a: drive as you did with the baseball data.

The first column of this data set (which you will call “quitrates”) gives the number of workers in the manufacturing sector per 100 employees in the manufacturing sector that leave their jobs *voluntarily* each year. The second column is the unemployment rate in the manufacturing sector. There are 13 observations in total, taken from the years 1960-1972.

Load this data into STATA, using similar methods as those you used for question

I would begin by creating a log file by typing something like:

```
log using a:unemploy, replace t
```

and load the data by typing

```
infile quitrate unempl using a:unemp.raw
```

Run the regression of *quitrates* (the dependent variable) on *unempl*. Present a scatterplot of the data, as well as your estimated regression line, using methods similar to those you used in the previous question.

Close the log file, and clear your session. Include your regression output and graph with your problem set.

Interpret your regression point estimates. Are the signs of the coefficients what you expect? Why?