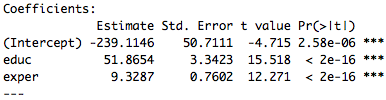
Israel Diego

2/2/16

STATS 500 Homework 2

**Part A**

Problem 1 Regressing weekly wages on years of experience and education gives the following output:



We notice that the beta estimates for ‘educ’ and ‘exper’ are positive and highly statisticly significant exceeding the 1% level. This indicates that a one year increase in education, on average will increase weekly wages by $51.87. A one year increase in work experience will increase weekly wages by $9.33.

Problem 2 The Multiple *R2* statistic is .1348, which means 13.48% of the variation in weekly wages is explained by ‘educ’ and ‘exper.’

Problem 3 Looking at the residual vector from our regression, and sorting it from highest to lowest, we get a value of 7249.1740, which corresponds to observation case number 15387.

Problem 4 The residual mean is -1.381535e-15 and the residual median is -52.1433. Since the mean is higher than the median, the distribution residuals are skewed right, i.e. greater portion of residuals are bunched towards the left.

Problem 5 The difference in weekly wages is the coefficient of ‘exper,’ which is 9.7748.

Problem 6 The correlation between the fitted values and the residuals comes out to 6.35678e-17, which is a very small number close to 0. By construction the residual vector and fitted values vector are orthogonal, which would imply that their correlation would be 0. Indeed what the plot on the next page shows, is that there is an almost no linear trend amongst the data points and the linear trend is only slightly negative.

Macintosh HD:Users:rdiego:Documents:Senior Year:Winter 2016:Stats 500:homeworks:HW2:figures:Residuals Plotted Against fitted values.pdf

**Part B**

Problem 1 For our estimate of we get .

Problem 2 The true variance of can be calculated because we know the true variance , from our random number generator. Thus

Problem 3 Now estimating , our estimate comes out to 1.1741. Intuitively this should be close to 1, and this is indeed true.

Problem 4/5

Macintosh HD:Users:rdiego:Documents:Senior Year:Winter 2016:Stats 500:homeworks:HW2:figures:Histogram of betahat1.pdfMacintosh HD:Users:rdiego:Documents:Senior Year:Winter 2016:Stats 500:homeworks:HW2:figures:Histogram of betaHat2.pdf

Macintosh HD:Users:rdiego:Documents:Senior Year:Winter 2016:Stats 500:homeworks:HW2:figures:Histogram of betaHat3.pdfMacintosh HD:Users:rdiego:Documents:Senior Year:Winter 2016:Stats 500:homeworks:HW2:figures:Histogram of sigmaHat2.pdf

From the graphs above we notice from the histograms that the greatest frequency of values of occur near the true . The is close to the values obtained in problem 2 for each iteration. For the next problem we run another simulation and estimate . I believe we do get a reliable estimate of , since we estimate to be 1 with the highest frequency. This is due to the fact that we use to estimate the residual sum of squares. Since we get reliable estimates of on average, naturally we also get reliable estimates of .

Problem 6

Macintosh HD:Users:rdiego:Documents:Senior Year:Winter 2016:Stats 500:homeworks:HW2:figures:Histogram of betaHat1 (Problem 4 repeat).pdfMacintosh HD:Users:rdiego:Documents:Senior Year:Winter 2016:Stats 500:homeworks:HW2:figures:Histogram of betaHat2 (Problem 4 repeat).pdf

Macintosh HD:Users:rdiego:Documents:Senior Year:Winter 2016:Stats 500:homeworks:HW2:figures:Histogram of betaHat3 (Problem 4 repeat).pdfMacintosh HD:Users:rdiego:Documents:Senior Year:Winter 2016:Stats 500:homeworks:HW2:figures:Histogram of sigmaHat2 (Problem 5 repeat).pdf

In this case the residuals are generated from a demeaned Poisson distribution with parameter . Thus giving us a distribution with mean 0, variance 1. As the plots above show, similar to problem 4, we have the greatest frequency at the true values for , and thus giving us a reliable estimate of . Likewise the bottom right graph, which is a repeat of problem 5, will also gives us a reliable estimate of . The largest frequency of occurs at 1.