LINEAR REGRESSION MODELS W4315

HOMEWORK 4 QUESTIONS

February 17, 2010

Instructor: Frank Wood

1. (50 points) [1] Refer to **Copier maintenance** Problem 1.20.

a. Prepare a dot plot for the number of copiers serviced X_i. What information is provided by this plot? Are there any outlying cases with respect to this variable?

b. The cases are given in time order. Prepare a time plot for the number of copiers serviced. What does your plot show?

d. Prepare residual plots of e_i versus $\hat{Y}_i$ and e_i versus X_i on separate graphs. Do these plots provide the same information? What departures from regression model (2.1) can be studied from these plots? State your findings. And the model (2.1) is as follows:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

where:

$$\beta_0 \text{ and } \beta_1 \text{ are parameters}$$
$$X_i \text{ are known constants}$$
$$\epsilon_i \text{ are independent } N(0, \sigma^2)$$

e. Prepare a normal probability plot of the residuals. Also obtain the coefficient of correlation between the ordered residuals and their expected values under normality. DOes the normality assumption appear to be tenable here? Use table B.6 and $\alpha = .10$.

f. Prepare a time plot of the residuals to ascertain whether the error terms are correlated over time. What is your conclusion?

h. Information is given below on two variables not included in the regression model, namely, mean operational age of copiers serviced on the call(X_2, in months) and years of experience of the service person making the call(X_3). Plot the residuals against X_2 and X_3 on separate

---

[1]This is problem 3.4 in 'Applied Linear Regression Models(4th edition)' by Kutner etc.

graphs to ascertain whether the model can be improved by including either or both of these variables. What do you conclude?

N.B. 1. If you need any software for this problem, do not use the embedded linear regression commands, say, 'regress' in MATLAB is not allowed. 2. If you are using software, please attach the code at the back of your handed-in homework instead of mixing codes with the results. 3. You don't have to hand in part c and h of this problem.

2. (25 points) [2] A student fitted a linear regression function for a class assignment. The student plotted the residuals e_i against Y_i and found a positive relation. When the residuals were plotted against the fitted values $\hat{Y}_i$, the student found no relation. How could this difference arise? Which is the more meaningful plot?

3. (25 points) [3] If the error terms in a regression model are independent $N(0, \sigma^2)$, what can be said about the error terms after transformation $X' = 1/X$ is used? Is the situation the same after transformation $Y' = 1/Y$ is used?

[2]This is problem 3.19 in 'Applied Linear Regression Models(4th edition)' by Kutner etc.
[3]This is problem 3.20 in 'Applied Linear Regression Models(4th edition)' by Kutner etc.