GR5205 Section 4 - Homework 3:

Due date: 10/10 by 7:25pm EST.

1. Chapter 3, page 147 3.4.

No need to work on b.c.f., i.e., only work n a.d.e.g.h. below

Use lm in h. directly

*3.4. 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?

c. Prepare a stem-and-leaf plot of the residuals. Are there any noteworthy features in this plot?

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.

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?

g. Assume that (3.10) is applicable and conduct the Breusch-Pagan test to determine whether or not the error variance varies with the level of X. Use $\alpha = .05$. State the alternatives, decision rule, and 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 graphs to ascertain whether the model can be improved by including either or both of these variables. What do you conclude?

i:	1	2	3	...	43	44	45
X_2:	20	19	27	...	28	26	33
X_3:	4	5	4	...	3	3	6

Chapter 3, page 150 3.17.

Only need to work on a, c, d, e below

*3.17. **Sales growth.** A marketing researcher studied annual sales of a product that had been introduced 10 years ago. The data are as follows, where X is the year (coded) and Y is sales in thousands

of units:

i:	1	2	3	4	5	6	7	8	9	10
X_i:	0	1	2	3	4	5	6	7	8	9
Y_i:	98	135	162	178	221	232	283	300	374	395

a. Prepare a scatter plot of the data. Does a linear relation appear adequate here?

b. Use the Box-Cox procedure and standardization (3.36) to find an appropriate power transformation of Y. Evaluate SSE for $\lambda = .3, .4, .5, .6, .7$. What transformation of Y is suggested?

c. Use the transformation $Y' = \sqrt{Y}$ and obtain the estimated linear regression function for the transformed data.

d. Plot the estimated regression line and the transformed data. Does the regression line appear to be a good fit to the transformed data?

e. Obtain the residuals and plot them against the fitted values. Also prepare a normal probability plot. What do your plots show?

f. Express the estimated regression function in the original units.