LAB 3: DIAGNOSTIC AND MATRIX APPROACH

Problem 1(Question 3.7). Muscle mass. A person's muscle mass is expected to decrease with age. To explore this relationship in women, a nutritionist randomly selected 15 women from each 10-year age group, beginning with age 40 and ending with age 79. The results follow; X is age, and Y is a measure of muscle mass. Assume that first-order regression model (1.1) is appropriate.

i:	1	2	3	...	58	59	60
X_i:	43	41	47	...	76	72	76
Y_i:	106	106	97	...	56	70	74

a. Prepare a histogram for the ages X_i. What information does your plot provide? Is this plot consistent with the random selection of women from each 10-year age group? Explain.

b. Obtain the residuals e_i and prepare a normal probability plot of the residuals. Does the distribution of the residuals appear to be symmetrical?

c. Plot the residuals e_i against $\widehat{Y}_i$ and also against X_i on separate graphs to ascertain whether any departures from regression model (2.1) are evident. Do the two plots provide the same information? State your conclusions.

d. 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 $\alpha = 0.01$. State the alternatives, decision rule, and conclusion. Is your conclusion consistent with your preliminary findings in part (c)?

Problem 2 (Question 3.18). Production time. In a manufacturing study, the production times for 111 recent production runs were obtained. The table below lists for each run the production time in hours (Y) and the production lot size (X).

i:	1	2	3	...	109	110	111
X_i:	15	9	7	...	12	9	15
Y_i:	14.28	8.80	12.49	...	16.37	11.45	15.78

a. Prepare a scatter plot of the data. Does a linear relation appear adequate here? Would a transformation on X or Y be more appropriate here? Why?

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

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

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

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

Homework:

Problem 3. Understand and write the python code for Section 3.11, page 141.

Problem 4. Consumer finance. The data below show, for a consumer finance company operating in six cities, the number of competing loan companies operating in the city (X) and the number per thousand of the company's loans made in that city that are currently delinquent (Y):

i:	1	2	3	4	5	6
X_i:	4	1	2	3	3	4
Y_i:	16	5	10	15	13	22

a. Using matrix methods, obtain the following:

(1) vector of estimated regression coefficients,
(2) vector of residuals,
(3) SSR,
(4) SSE,
(5) estimated variance-covariance matrix of b,
(6) point estimate of $E\{Y_h\}$ when $X_h = 4$,
(7) $s^2\{\text{pred}\}$ when $X_h = 4$.

b. From your estimated variance-covariance matrix. Obtain the following:

(1) $s\{b_0, b_1\}$
(2) $s^2\{b_0\}$
(3) $s^2\{b_1\}$.

c. Find the hat matrix H.

d. Find $s^2\{e\}$.