Lab 8: Linear Regression
Objective: Creating Scatterplots, Calculating Correlation, and Determining the Least-Squares Regression Lines, Check Assumptions, Perform Inference

A. (80 points) House Prices (Data Set: sales.txt - webpage)

Real estate is typically reassessed annually for property tax purposes. This assessed value, however, is not necessarily the same as the fair market value of the property. The data file summarizes an SRS of 30 properties recently sold in a Midwestern city. Both variables Sales Price and Assessed value are measured in thousands of dollars.

TABLE 10.2 Sales Price and Assessed Value (in $ thousands) of 30 Homes in a Midwestern City

Property	Sales price	Assessed value	Property	Sales price	Assessed value	Property	Sales price	Assessed value
1	179.9	188.7	2	240.0	220.4	3	113.5	118.1
4	281.5	232.4	5	186.0	188.1	6	275.0	240.1
7	281.5	232.4	8	210.0	211.8	9	210.0	168.0
10	184.0	180.3	11	186.5	294.7	12	239.0	209.2
13	185.0	162.3	14	251.0	236.8	15	180.0	123.7
16	160.0	191.7	17	255.0	245.6	18	220.0	219.3
19	160.0	181.6	20	200.0	177.4	21	265.0	307.2
22	190.0	229.7	23	150.5	168.9	24	189.0	194.4
25	157.0	143.9	26	171.5	201.4	27	157.0	143.9
28	175.0	181.0	29	159.0	125.1	30	229.0	195.3

Some of the following questions may be done by hand. If done by hand, all work needs to be shown.

1. (5 pts) Make a scatterplot of the data with the assessed value on the *x* axis and the sales price on the *y* axis. Please include the linear regression line on the plot.
2. (5 pts) From the scatterplot in part (1), describe the form, direction, and strength of the relationship. Identify any outliers. Is the relationship approximately linear?
3. (5 pts) Find the correlation between the sales price (Y) and the assessed value (X). Are your conclusions about the strength the same in this part as in part (2)? If they are different, provide a possible explanation for the difference.
4. (5 pts) Look at the scatterplot for these data that you made in part (1). Is the correlation a good numerical summary of the graphical display in the scatterplot? Please explain by discussing the reasons why correlation can or cannot be used.
5. (6 pts) Obtain the equation of the least-squares regression line for predicting the sales price from the assess value. What is r^2 for these data?
6. (5 pts) Predict the sales price for the first case (with Property = 1), and calculate the residual. This part may be done by hand.
7. (5 pts)) Obtain the rest of the residuals and plot them versus the assessed value. There is no need to have a listing of the residuals. Is there anything unusual to report? If so, explain. Are the conclusions from the residual plot the same as from the scatterplot (parts 2 and 4)? If they are different, provide a possible explanation for the difference.
8. (5 pts) Do the residuals appear to be approximately Normal? Explain your answer. Be sure to include the appropriate graph(s) in your answer.
9. (5 pts) Based on your answers to parts, (1), (7), and (8), do the assumptions for the linear regression analysis appear reasonable? Explain your answer.
10. (12 pts) Construct and interpret a 99% confidence interval for the slope and the intercept. What is the significance of the result for the slope? Is the inference on the intercept of interest in this problem? Why or why not?

11. (10 pts) Is there significant evidence that assessed value is associated with the sales price at a 0.01 significance level? Please perform the 4*-step process (state hypotheses, give a test statistic and *P*-value, and state your conclusion).

12. (6 pts.) How are the results from parts 12 and 13 similar? How are they different?

13. (6 pts.) Write a short paragraph in complete English sentences summarizes the results which is understandable to non-statisticians. The summary should contain the following parts: a) is the model appropriate to use, b) What are the effects of switching X and Y in this situation? c) What is the relationship between the strengths of the dominant and non-dominant arms? d) Is this situation good for prediction? e) Is there any causality in this situation? f) Can you generalize this situation to homes in the west?