Carleton University
School of Mathematics and Statistics
STAT 2509 A - Assignment #2

<u>DUE:</u>   May 29th, 2025 (to be submitted on BrightSpace before 10:00am)

1. In order to increase sales, a car dealership has developed a 30-second TV ad to advertise a discount event. The data on the number of ads per day and number of cars sold are provided in the following table:

No. of ads per day	1	1	1	2	2	2	3	3	3	4	4	4
No. of cars sold	21	18	17	24	22	26	34	29	32	45	41	35

$$\sum x_i = 30 \ , \ \sum y_i = 344 \ , \ \sum x_i^2 = 90 \ , \ \sum y_i^2 = 10\ 742 \ , \ \sum x_i y_i = 969$$

a) Identify independent (x) and dependent (y) variables.

b) Construct a scatter plot using **SPSS**. What does the plot suggest about the relationship between the No. of ads per day and the No. of cars sold?

c) State a SLR model making sure you give all assumptions necessary for statistical inference.

d) Find the least squares estimates of β_0 and β_1. Find the least squares fitted regression line.

e) Obtain a plot using **SPSS**, that shows fitted regression line.

For the rest of the question assume that the assumptions hold.

f) Find s, an estimate of σ.

g) Use the t-test to test whether there is a significant linear relationship between the No. of ads per day and the No. of cars sold. Use $\alpha = 0.05$.

h) Find a 95% confidence interval for β_1.

i) Set up the ANOVA table and hence test whether there is a significant linear relationship between the No. of ads per day and the No. of cars sold. Use $\alpha = 0.05$.

j) Find the values of the coefficient of correlation, r, and coefficient of determination, r^2, and interpret their meanings in this problem. What is your conclusion about the goodness of the model?

k) Verify your above results in parts (f) to (j), using **SPSS**.

2. Refers to question 1.

a) Find a 95% Confidence Interval for the average value of the response variable (i.e. of the No. of cars sold) and a 95% Prediction Interval for an individual value of this response variable when the No. of ads per day is 0 (i.e. $x_p = 0$).

What is your conclusion about the widths of these two intervals?

b) Use **SPSS** to compare your results with part (a).

3. Refers to question 1.

Perform a residual analysis using SPSS. That means, obtain following plots: $(\hat{y}_i \ vs \ e_i)$, $(x_i \ vs \ e_i)$, $(Q-Q \ plot \ of \ residuals)$ and $(histogram \ of \ residuals)$. What can you say about the validity of the model assumptions?