Problem Set - 4 (Multiple Linear Regression - Coffee Chain)

ECO 204 (Section 7)
Instructor: Shaikh Tanvir Hossain
TA: Habiba Afroz

Due: 1st Aug (before 10.00 PM), submit in Google Classroom

§

1. **(Coffee Shop Chain Marketing Analysis)** The owner of "Bean There, Done That" coffee shop chain would like to predict monthly sales revenue as a function of marketing expenditures. Historical data for a sample of 30 months across different store locations has been collected and is provided in `CoffeeChain.xlsx`. The variables are:

   - `monthly_sales`: monthly sales revenue ($1000)
   - `digital_ads`: digital advertising expenditure ($1000) - social media, online ads
   - `print_ads`: print advertising expenditure ($1000) - newspapers, magazines
   - `radio_ads`: radio advertising expenditure ($1000) - local radio stations
   - `promotional_events`: in-store promotional events expenditure ($1000)
   - `customer_segment`: primary customer demographic (1 = Young Adults 18-30, 2 = Middle-aged 31-50, 3 = Seniors 51+)

   Note: The customer_segment is a categorical variable representing the predominant customer demographic at each store location. When using Excel, convert this to dummy variables before including in regression models. In R, convert it to a factor variable (factor indicates categorical data). This variable captures the primary age group of customers served by each store location.

   (a) Suppose we want to predict monthly sales revenue as a function of digital advertising expenditures, print advertising expenditures, radio advertising expenditures and promotional events expenditures. What is the population regression model? What is the CEF function here.

   (b) Develop an estimated regression equation to predict monthly sales revenue using all marketing variables (excluding customer segment) as independent variables.

   (c) Write the ANOVA table and calculate SST, SSR, SSE, MSR, and MSE for the multiple regression model.

   (d) Calculate R^2 and Adjusted R^2 for the multiple linear regression model. Provide clear interpretations of these statistics in the context of coffee shop sales prediction.

   (e) With $\alpha = 0.05$, test the individual significance of each marketing variable. Comment on whether any variables should be dropped from the model. Write separate hypotheses for each of the four marketing variables and conduct individual significance testing.

$$H_0 : \beta_j = 0 \text{ for } j = 1, 2, 3, 4$$

   (f) With $\alpha = 0.05$, test the overall model significance. Comment on whether all marketing variables should be dropped simultaneously. Clearly state the hypotheses, test statistic value, and decision criteria.

$$H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$$
$$H_a : \text{at least one } \beta_j \text{ for } j = 1, 2, 3, 4 \text{ is not zero}$$

   (g) With $\alpha = 0.05$, conduct a partial F-test (or restricted / unrestricted F test) to determine whether both radio advertising and promotional events should be dropped from the model. State the appropriate hypotheses and interpret your findings.

   (h) Calculate the predicted monthly sales revenue when the coffee chain spends $3,500 on digital advertising, $2,300 on print advertising, $1,000 on radio advertising, and $500 on promotional events.

(i) Construct a 90% confidence interval for the regression coefficient of digital advertising expenditures. Interpret this interval in practical business terms.

(j) Construct a 96% confidence interval for the regression coefficient of print advertising expenditures. Explain what this interval tells us about the effectiveness of print advertising.

(k) Suppose we include only customer segment as a predictor variable. Write the population regression function using dummy variables to represent the different customer demographics.

(l) Estimate a dummy variable regression using only customer segment variables. Report and interpret the estimated coefficients. Demonstrate using Excel or R that regression analysis is unnecessary for obtaining these estimates by calculating group means directly.

(m) Examine an application of an **interaction model** by analyzing the interaction between customer segment (dummy variables) and print advertising (continuous variable).

The business rationale for this analysis is that *print advertising effectiveness may vary significantly across different customer demographics*. Senior customers typically consume more traditional print media compared to young adults who favor digital channels, potentially resulting in differential responses to print advertising campaigns across age groups. Consequently, we anticipate distinct slope coefficients for print advertising effectiveness within each customer segment.

Conduct a multiple linear regression incorporating print advertising, customer segment dummy variables, and their interaction terms. Report the estimated coefficients and provide detailed interpretations of the results, including practical business implications for marketing strategy.