Hyunji Kang(20222110017)

Question 6. Describe the null hypotheses to which the p-values given in Table below correspond, with response sales being regressed on predictor TV, radio, and newspaper. Explain what conclusions you can draw based on these p-values. Your explanation should be phrased in terms of sales, TV, radio, and newspaper, rather than in terms of the coefficients of the linear model.

.	Coefficient	Std. error	t-statistic	p-value
Intercept	2.939	0.3119	9.42	< 0.0001
TV	0.046	0.0014	32.81	< 0.0001
radio	0.189	0.0086	21.89	< 0.0001
newspaper	-0.001	0.0059	-0.18	0.8599

Note: For the Advertising data, least squares coefficient estimates of the multiple linear regression of number of units sold on radio, TV, and newspaper advertising budgets.

[Answer]

■ The hypotheses

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 > 0 \end{cases} \quad \begin{cases} H_0 : \beta_2 = 0 \\ H_1 : \beta_2 > 0 \end{cases} \quad \begin{cases} H_0 : \beta_3 = 0 \\ H_1 : \beta_3 > 0 \end{cases}$$

The reason why $H_1 : \beta > 0$ is because we expect an increase in sales due to the advertising effect.

■ Interpretation of Coefficient's sign
The positive sign of TV and Radio - This is interpreted as increasing in sales due to the advertising effect of TV and Radio.
The negative sign of Newspaper - This is interpreted as decreasing in sales due to the advertising effect of Newspaper.

■ Size of S.E
All three variables have small S.E

$$\text{Small S.E} \rightarrow \text{Small } Var(\hat{\beta}) \rightarrow \text{Big } \Sigma\left(x_i - \bar{x}\right)^2 \rightarrow \text{Big } Var(X) \rightarrow \text{Big range} \rightarrow \text{Big swing}$$

$$\begin{aligned} &S.E = \sqrt{\operatorname{Var}(\hat{\beta})} \\ &\operatorname{Var}(\hat{\beta}) = \sigma^2 \left(X'X\right)^{-1} = \frac{\sigma^2}{\sum\left(x_i - \bar{x}\right)^2} \\ &\left(X'X\right)^{-1} = \frac{1}{\sum\left(x_i - \bar{x}\right)^2} \\ &\left(X'X\right) = \Sigma\left(x_i - \bar{x}\right)^2 \leftarrow \text{ Variance of } X \end{aligned}$$

And there seems to be no multicollinearity since the S.E of the three variables are all very small.

■ t-statistic

$$\text{t-stat} = \frac{\hat{\beta} - \beta}{\sqrt{\operatorname{Var}(\hat{\beta})}} = \frac{\hat{\beta}}{\text{S.E}}$$

$$\begin{aligned} &\text{t-stat of Newspaper } = -0.18 < 1.96 \text{ thus, do not reject } H_0 \\ &\text{t-stat of TV, Radio } > 1.96 \quad \text{thus, reject } H_0 \end{aligned}$$

Recall, Small S.E of Newspaper = Big swing of Newspaper
Also, Small S.E = Small denominator in the t-stat formula
It means that advertisers, whether spending less or more on newspaper advertising, did not help sales.

■ p-value
The p-value of Newspaper 0.8599 > 0.05(significant level), thus H_0 is not rejected. The p-value of TV and Radio < 0.05 (significant level), thus H_0 is rejected. This is interpreted as Newspaper is not significant variable, TV and Radio are significant variables.

Question 7. Suppose we have a data set on flights, with three predictors, X_1 = Distance (to destination in kilometers), X_2 = Holiday (1 if Yes, 0 for No) , X_3 = Interaction between Distance and Holiday. The response is the flight ticket price (in US$). Suppose we use least squares to fit the model, and get $\hat{\beta}_0 = 100, \hat{\beta}_1 = 0.2$, $\hat{\beta}_2 = 20, \hat{\beta}_3 = 0.05$.

1. Which answer is correct, and why?
   The model is
   $$Price = \beta_0 + \beta_1 Distance + \beta_2 Holiday + \beta_3[Distance \cdot Holiday] + \epsilon$$
   Holiday's Partial effect = (Holiday = 1) − (Holiday = 0) = $\beta_2 + \beta_3$ Distance
   The following answer is correct because the signs of β_2 and β_3 are both positive.

   - For a fixed value of Distance, on average tickets are more expensive on holidays than on usual days.

2. Predict the average holiday price of a ticket for a flight that travels $1,000$ km to destination.
   $$\begin{aligned}\hat{Y} &= \hat{\beta}_0 + \hat{\beta}_1 Distance + \hat{\beta}_2 Holiday + \hat{\beta}_3[Distance \cdot Holiday] + \hat{\epsilon} \\ &= 100 + 0.2 \cdot 1000 + 20 \cdot 1 + 0.05 \cdot 1000 + 0 \\ &= 100 + 200 + 20 + 50 \\ &= 370(\$)\end{aligned}$$

3. True, false, or uncertain: Since the coefficient for the Distance/Holiday interaction term is pretty small, there is no evidence of an interaction effect. Justify your answer.

   Uncertain: since we don't know Var($\hat{\beta}_3$)