# Chapter 3: Linear Regression

# Question 1.

Describe the null hypotheses to which the p-values in Table 3.4 correspond. Explain what conclusions you can draw based on these p-values.

## Answer:

Each p-value in Table 3.4 tests the null hypothesis for an individual regression coefficient in the multiple linear regression model:

$$H_0: \beta_j = 0$$

This means we are testing whether the predictor $X_j$ (e.g., TV, radio, newspaper) has **no association** with the response variable (sales), assuming all other predictors are held constant.

- A **small p-value** (typically < 0.05) indicates strong evidence against the null hypothesis, suggesting that the predictor is **statistically significantly associated** with the response.
- A **large p-value** means there is **insufficient evidence** to conclude a relationship between the predictor and the response.

**Conclusions from Table 3.4**:

- The p-values for **TV** and **radio** are small, so we reject the null hypotheses $H_0: \beta_{\text{TV}} = 0$ and $H_0: \beta_{\text{radio}} = 0$. These predictors are significantly associated with sales.
- The p-value for **newspaper** is large, so we **fail to reject** $H_0: \beta_{\text{newspaper}} = 0$. There is **no strong evidence** that newspaper advertising is associated with sales, once TV and radio are accounted for

# Question 2.

Carefully explain the differences between the KNN classifier and KNN regression methods.

**Answer**

**KNN regression**: Predicts a **quantitative** output by averaging the responses of the K nearest neighbors.

$$\hat{f}(x_0) = \frac{1}{K} \sum_{x_i \in \aleph} y_i$$

**KNN classification**: Predicts a **qualitative** output by majority vote among the K nearest neighbors.

# Question 3

$$X_1 = \text{GPA}$$

$$X_2 = \text{IQ}$$

$$X_3 = \text{Level (1 for College, 0 for High School)}$$

$$X_4 = \text{GPA} \times \text{IQ}$$

$$X_5 = \text{GPA} \times \text{Level}$$

$$\hat{Y} = 50 + 20X1 + 0.07X2 + 35X3 + 0.01X4 - 10X5$$

**Answers**:

(a)

It can be shown that:

$$y_{college} = y_{hs} + 35 - 10 \cdot GPA$$

If GPA < 3.5, then college grads earn more.

If GPA > 3.5, then high school grads earn more.

Therefore, the correct answer is actually **iii**:

**High school graduates earn more on average than college graduates provided that the GPA is high enough.**

(b)

$$\hat{Y} = 50 + 20(4.0) + 0.07(110) + 35(1) + 0.01(440) - 10(4.0)$$
$$= 50 + 80 + 7.7 + 35 + 4.4 - 40 = 137.1$$

Predicted Salary is $137.100

(c) False. The size of a coefficient does **not** indicate its statistical significance or the strength of evidence for interaction. Even a small coefficient can be **statistically significant** if its standard error is small.

To determine whether there is an interaction effect, you must look at:

- The **p-value** for the interaction term
- The **context** of the variables: small coefficients can still imply large effects if the variables are on large scales (e.g., GPA × IQ = 4.0 × 110 = 440)

So, without knowing the standard error and p-value, we **cannot** conclude there is little evidence of interaction just because the coefficient is small.

# Question 4

**Answer**

(a) If the true relationship is **linear:**

- **Training error**:
  **Yes**, adding the cubic term $X^3$ will always **reduce** the training error (or at least keep it the same), because the model becomes **more flexible**. A cubic model can fit linear relationships perfectly, and the extra flexibility allows it to capture random fluctuations (overfitting) in the training data.
- **Test error**:
  **No**, the test error is likely to **increase**. Since the true relationship is linear, including a $X^3$ term introduces **unnecessary flexibility**, which can lead the model to fit noise rather than the underlying pattern. This increases the **variance** of the model without reducing its **bias**, leading to worse generalization to new data.

In summary: Adding $X^3$ reduces training error but likely **increases test error** when the true relationship is linear.

**(b)** If the true relationship is not **linear:**

- **Training error**:
  Again, adding the cubic term will **reduce** or **maintain** the same training error due to the added flexibility of the model.

- **Test error**:
  The effect on test error depends on how **non-linear** the true relationship is:

  - If the relationship is only **slightly non-linear**, the cubic term might help capture some curvature and **reduce bias**, leading to better test performance.
  - If the relationship is **not well approximated** by a cubic polynomial, the model may still suffer from **bias** or introduce **variance** due to overfitting.

In this case, adding $X^3$ may improve test error **if the added flexibility captures the structure** of the true relationship. But if it adds unnecessary complexity or doesn't match the form of the true function, it could also hurt generalization.

# Question 5

Consider the fitted values that result from performing linear regression **without an intercept**. In this setting, the *i*th fitted value takes the form:

$$\hat{y}_i = x_i \hat{\beta}$$

where

$$\hat{\beta} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

**What is ai'a_{i'}ai'?**

---

**Answer:**

We are given that:

Substitute $\hat{\beta}$ into the equation for $\hat{y}_i$

$$\hat{y}_i = x_i \cdot \frac{\sum_{j=1}^{n} x_j y_j}{\sum_{j=1}^{n} x_j^2} = \sum_{j=1}^{n} \left( \frac{x_i x_j}{\sum_{k=1}^{n} x_k^2} \right)$$

We can now identify:

$$a_j = \frac{x_i x_j}{\sum_{k=1}^{n} x_k^2}$$

# Question 6

Trivial algebra just substitute what is given

# Question 7

pass