Econ 573 **Juan Alvarez**
Prof. Andrii Babii **February 5th, 2025**

Part 1

Ex. 1, 3, 4 from Chapter 3 of ISL

1. *Describe the null hypotheses to which the p-values given in Table 3.4 correspond. 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.*

The p-values given in Table 3.4 correspond to the individual t-tests for each type of advertising (TV, radio, and newspaper). The null hypothesis for each test is that the corresponding advertising medium does not have a relationship to sales, such that its predicted coefficient is 0. Formally, $H_0 : \widehat{\beta}_1 = 0$ for TV, $H_0 : \widehat{\beta}_2 = 0$ for radio, and $H_0 : \widehat{\beta}_3 = 0$ for newspapers. Since the p-values for TV and radio advertising are both less than 0.0001, we have strong evidence to reject the null hypothesis for these predictors. This implies a statistically significant association between sales and both TV and radio advertising at all conventional significance levels. In contrast, the p-value for newspaper advertising is quite large (0.8599), so we fail to reject the null hypothesis. This suggests there is insufficient evidence to conclude a statistically significant relationship between newspaper advertising and sales.

3. *Suppose we have a data set with five predictors, X_1 = GPA, X_2 = IQ, X_3 = Level (1 for College and 0 for High School), X_4 = Interaction between GPA and IQ, and X_5 = Interaction between GPA and Level. The response is starting salary after graduation (in thousands of dollars). Suppose we use least squares to ft the model, and get $\widehat{\beta}_0$ = 50, $\widehat{\beta}_1$ = 20, $\widehat{\beta}_2$ = 0.07, $\widehat{\beta}_3$ = 35, $\widehat{\beta}_4$ = 0.01, $\widehat{\beta}_5$ = −10.*

(a) *Which answer is correct, and why?*

i. *For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates.*

Incorrect. For college graduates (Level = 1), the coefficients associated with Level and the interaction between GPA and Level are 35 and -10, respectively, resulting in an additional $25,000 on average. For high school graduates (Level = 0), these terms are dropped, making their starting salary \$25,000 less on average compared to college graduates with the same IQ and GPA.

ii. *For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates*

Correct, for the same reasons as before. For college graduates (Level = 1), the combined contribution of the coefficients 35 (for Level) and -10 (for GPA and Level interaction) results in an additional \$25,000 in starting salary compared to high school graduates (Level = 0). Therefore, college graduates earn more, on average.

iii. For a fixed value of IQ and GPA, high school graduates earn more, on average, than college graduates provided that the GPA is high enough.

Correct. In the multivariate model for college graduates we have that GPA interacts with $\widehat{\beta}_1 = 20$, $\widehat{\beta}_4 = 0.01$, and $\widehat{\beta}_5 = -10$. Additionally, we know that college graduates get an extra 35 thousand dollars due to their education level ($\widehat{\beta}_3 = 35$). Provided that IQ is a fixed value, we can define the change in starting salary after graduation for college graduates as a function of GPA such that $f_{college}(GPA) = 10.01 \times GPA + 35$. On the other hand, in the multivariate model for high school graduates GPA only interacts with $\widehat{\beta}_1 = 20$, and $\widehat{\beta}_4 = 0.01$. Again, we can express the change in salary this as a function of GPA such that $f_{hs}(GPA) = 20.01 \times GPA$. We can create an inequality to find the necessary GPA value for high school graduates to earn more, on average, than college graduates: $10.01 \times GPA + 35 < 20.01 \times GPA \rightarrow GPA > 3.5$. In conclusion, high school graduates earn more, on average, than college graduates provided that their GPA is higher than 3.5

iv. For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates provided that the GPA is high enough.

Incorrect. From the previous analysis, we found that when GPA > 3.5, high school graduates earn more on average than college graduates. Therefore, for college graduates to earn more, on average, than high school graduates, their GPA would need to be lower than 3.5. A corrected statement would say: "For a fixed value of IQ and GPA, college graduates earn more, on average, than high school graduates provided that the GPA is low enough."

(b) *Predict the salary of a college graduate with IQ of 110 and a GPA of 4.0.*

We can consider the provided OLS estimations for the model on starting salary after graduating:

$$\widehat{salary} = 50 + 20 \times GPA + 0.07 \times IQ + 35 \times Level + 0.01 \times (GPA \times IQ) - 10 \times (GPA \times Level)$$

And predict:

$$\widehat{salary} = 50 + 20 \times (4.0) + 0.07 \times (110) + 35 \times (1) + 0.01 \times (4.0 \times 110) - 10 \times (4.0 \times 1) \rightarrow 50 + 80 + 7.7 + 35 + 4.4 - 40 = 137.1$$

Such that a college graduate with an IQ of 110 and a GPA of 4.0 will have a salary of 137.1 thousand dollars.

(c) *True or false: Since the coefficient for the GPA/IQ interaction term is very small, there is very little evidence of an interaction effect. Justify your answer.*

The magnitude of the coefficient for the GPA/IQ interaction term indicates the size of the effect of the interaction, not the evidence for its existence. A small coefficient does not necessarily mean there is little evidence for an interaction effect; it only suggests that the effect size is small. To determine the statistical significance of this interaction, we would need to conduct a t-test or examine the p-value associated with the coefficient from the OLS estimation.

4. *I collect a set of data (n = 100 observations) containing a single predictor and a quantitative response. I then fit a linear regression model to the data, as well as a separate cubic regression, i.e.* $Y = \beta_0 X + \beta_1 X^2 + \beta_3 X^3 + \epsilon$.

(a) *Suppose that the true relationship between X and Y is linear, i.e. Y = β0 + β1X + ϵ. Consider the training residual sum of squares (RSS) for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer.*

We would expect the training RSS for the cubic regression to be lower than the training RSS for the linear regression. Even though the true relationship is linear, the cubic regression can capture random variations (including noise) better than the simple linear model, resulting in a lower training RSS due to overfitting.This happens because the cubic model has more parameters and greater capacity to fit the training data exactly, even if those additional terms are unnecessary for representing the true underlying relationship.

(b) *Answer (a) using test rather than training RSS*

In this case, we would expect the test RSS for the linear regression to be lower than the cubic regression due to the fact that the true relationship between X and Y is linear. The cubic regression model, although more flexible, is likely to overfit the training data by capturing noise rather than the true relationship. This overfitting typically results in higher test RSS compared to the correctly specified linear model.

(c) *Suppose that the true relationship between X and Y is not linear, but we don't know how far it is from linear. Consider the training RSS for the linear regression, and also the training RSS for the cubic regression. Would we expect one to be lower than the other, would we expect them to be the same, or is there not enough information to tell? Justify your answer*.

Similarly to (b), we would expect the cubic regression to have a lower training RSS for the cubic regression. This is because, as the flexibility of the model increases, the model can fit the training data more closely, which typically results in a lower training RSS. A cubic regression, being more flexible, can capture more complex patterns in the data, thereby reducing the training error.

(d) *Answer (c) using test rather than training RSS*.

If the true relationship between X and Y is not too far from linear, then we would expect the linear regression to have a lower test RSS in comparison to the cubic regression, as it is more likely to accurately capture the true relationship. On the other hand, if the true relationship is highly non-linear, the cubic regression might have a lower test RSS due to its ability to capture these complex patterns. In this case, the cubic model's flexibility allows it to fit the underlying structure of the data more accurately. However, there is still a risk of overfitting, especially if the sample size is too small to properly estimate all the coefficients without fitting noise. Additionally, this becomes a subject of trade-off between variance and bias. The linear model produces interpretable and more stable predictions that might exhibit low variance and high bias. While the opposite is true for cubic regression. Lastly, our sample size of $n = 100$ can be considered large enough for the linear regression to detect large effects in the population. However, it might be too small for the cubic regression to avoid overfitting.