Thursday Exam 3 - Solutions - STAT 252

## Question 8:

1. Write out the full regression equation for Model 4.
2. Calculate and interpret the meaning of the regression coefficients regarding the line for El Dorado Corridor (Group 3).

Make sure to label each part to earn full credit.

### Solution

### **a) Write out the full regression equation for Model 4**

Let:

* : predicted price
* : square footage (sqft)
* : dummy for RegionGroup 2 – Roseville Area
* : dummy for RegionGroup 3 – El Dorado Corridor
* : dummy for RegionGroup 4 – Auburn & Foothills
* : dummy for RegionGroup 5 – South & Rural

[You did not have to specify the above for full credit]

*(Group 1 – Sacramento Core is the reference group, so all dummies = 0)*

The full regression equation is:

[this was needed]

### **b) Calculate and interpret the meaning of the regression coefficients for Group 3 – El Dorado Corridor**

For a home in **Group 3**, we have:

Plug into the equation:

#### **Interpretation:**

* **115,380**: This is the **expected (or average) price** of a home in **RegionGroup 3 – El Dorado Corridor** that has **0 square feet**.
* **126.8**: For every additional **1 square foot** of living space of a home in **RegionGroup 3 – El Dorado Corridor**, the **expected (average) price increases by $126.80**, regardless of region (since Model 4 does not include interaction terms between sqft and region).

PTS: 16

## Question 9:

Assume you want to make a prediction using Model 5 for the South & Rural Group.

1. What would the predicted price be if the square footage is 2,000? Interpret the result in the context of the model.
2. What did you assume above this new house to make this prediction?

Make sure to label each part to earn full credit.

### Solution

#### **a) Predicted Price When Square Footage = 2,000**

We are using the following model:

Let:

* (square footage)
* , (Group 5 dummy)

Substitute into the model:

**Predicted Price: $ 295 204.8**

#####nterpretation\*\*

* A home in the **South & Rural region** with **2,000 sqft** is predicted to cost approximately **$294,203**.
* The model accounts for both:
  + A **lower starting price** for homes in this region (–$70,973)
  + A **higher price per square foot** than the baseline region (119.564 + 48.46 = 168.02 per sqft)

#### **b) Assumptions Made**

The **square footage is exactly 2,000 sqft**, which is assumed to be **within the observed range** of square footage values in the dataset for homes in this group.

PTS: 16

## Question 10:

1. Which of the five models would you choose to predict price? Explain your reasoning.
2. What additional steps or criteria would you use to compare the models?

Make sure to label each part to earn full credit.

### Solution

#### **a) Which of the five models would you choose to predict price? Explain your reasoning.**

I would choose **Model 5: price ~ sqft \* RegionGroup** for the following reasons:

1. **Highest Adjusted R²**: Model 5 has the highest adjusted , meaning it explains the most variation in housing prices while accounting for model complexity.
2. **Improved Residual Standard Error**: Model 5 has the **lowest residual standard error** (79,070), indicating better overall predictive accuracy.

#### **b) What additional steps or criteria would you use to compare the models?**

1. Assumptions regarding residuals
2. Residuals Sum of Squares

[Something along these lines]

PTS: 12