Quiz 2. Linear Regression

Name: Lisa Liubovich Attempt (circle one): 1 (2)

A real estate appraiser is interested in predicting residential home prices in a small city as a function of various features.

- SalesPrice: Sales price of residence in $1000s
- FinishedArea: Finished area of residence (square feet)
- BedroomsTotal: number of bedrooms in residence
- Pool: YES if present, = NO if absent
- LotSize: Lot size (square feet)

The following R code was used with the following output of regression analysis.

```
> reg = lm(SALESPRICE ~ FINISHEDAREA + BEDROOMS + LOTSIZE*POOL)
> summary(reg)

Coefficients:
                Estimate  Std.  Error  t value  Pr(>|t|)
(Intercept)     -9.4e+01      1.4e+01     -6.5  1.80e-10  ***
FINISHEDAREA     1.6e-01      5.8e-03     28.0   < 2e-16  ***
BEDROOMS        -8.6e+00      4.1e+00     -2.1    0.0346  *
LOTSIZE          1.3e-03      3.0e-04      4.5  1.02e-05  ***
POOLYES          8.9e+01      3.5e+01      2.6    0.0105  *
LOTSIZE:POOLYES -3.4e-03      1.4e-03     -2.4    0.0172  *
```

(a) Predict sales price for a 1600 sq. ft. home with 3 bedrooms, a 20,000 sq. ft. lot, and a pool.

Y = -25.55 + 0.59X1 - 8.6X2 + 0.07X3 + 24.19X4 – 0.17X5

Y = -25.55 + 0.59(1600) - 8.6(3) + 0.07(20000) + 24.19(1) – 0.17(20000)

(b) Interpret the estimated slope of the variable *Bedrooms* in the context of the problem. Note that it is a negative value. What does it suggest about the home buyers and housing price?

For every additional bedroom in a house, the price of a house is $8,600 less on average, controlling for finished area, lot size, and whether or not there is a pool. It suggests that perhaps home buyers who want less rooms are willing to pay more.

(Continue on the back.)

(c) Suppose that the relation between the house price, area, lot size, and number of bedrooms actually does ***NOT*** depend on the Pool. If we remove variable *Pool* from our regression model, how would you ***expect*** it to affect:

- the ***training*** residual sum of squares (RSS, aka. SSE, Sum of Squared Errors)?

The model becomes simpler, with fewer parameters to estimate. This can slightly increase the training RSS because the model now has one less variable to account for variations in house prices. However, since Pool does not actually influence house prices, its inclusion in the model does not contribute to reducing the error. Therefore, the increase in training RSS should be minimal, if noticeable at all.

- the ***testing*** residual sum of squares?

Removing an irrelevant variable like Pool can help in reducing overfitting. Overfitting occurs when the model captures noise or random fluctuations in the training data rather than the true underlying pattern. By excluding Pool, the model avoids fitting to noise associated with this variable, leading to potentially better generalization. Since the Pool variable does not actually affect house prices, its removal should not negatively impact the model's predictive power. On the contrary, it can improve the model's performance on the test data by focusing only on the relevant variables (area, lot size, number of bedrooms).

(d) (Stat 627) Based on the given output, can we state that the sales price of a house is expected to increase when its owners build a pool on their lot? Does this depend on the lot size? Give a recommendation to homeowners who plan to increase the sales price by building a pool.

The positive coefficient suggests that having a pool is associated with an increase in house price. The p-value is 0.0105, which indicates that there is strong evidence to suggest an association between the presence of a pool and house price. A negative coefficient for the interaction term suggests that as lot size increases, the positive impact of having a pool on the sales price decreases. The p-value is 0.0172, which indicates that the interaction term is statistically significant and we have strong evidence to believe that the effect of a pool on house price does indeed depend on the lot size.

Recommendation: Building a pool is likely to increase the sales price of the house. Evaluate the cost of building the pool against the estimated increase in house price. Given the significant interaction term, ensure the increase justifies the investment. The benefit of building a pool is reduced as the lot size increases. For very large lots, the negative interaction term could mean that adding a pool does not significantly increase the house price or could even decrease it. Homeowners should carefully consider whether the potential price increase justifies the investment in a pool. Consulting with local real estate experts can provide additional market-specific insights.