2/?/23 Four: Predictions

Day 5

There are three sorts of uncertainty associated w/ the linear regression prediction:

1) The coefficients $\hat{\beta}_i$ estimate β_i i.e. the least squares plane only estimates the true population plane. Use confidence intervals
- A reducible error

2) Model bias from assuming the true model is linear. Another reducible error.

3) The error from ε. We use prediction intervals to see how much Y will vary from $\hat{Y}$.
- Prediction intervals are wider than confidence intervals.

Recall $Y = f(X) + \varepsilon$.
- Confidence intervals aim to bound $f(X)$
- Prediction intervals aim to bound Y.

2/4/23

3.3: Other Considerations in the Regression Model

3.3.1. Qualitative Predictions

Day 6

Predictors with only two levels
- Regress on an indicator variable.
- Arbitrary to choose 0/1 coding, -1/1 coding, etc.

Qualitative predictors with more than 2 levels

Example: IP ethnicities are Asian, Caucasian, African-American

Let

$$x_{i1} = \begin{cases} 1 & i\text{-th person Asian} \\ 0 & \text{else} \end{cases} \qquad x_{i2} = \begin{cases} 1 & i\text{-th person white} \\ 0 & \text{else} \end{cases}$$

Then the model is

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i = \begin{cases} \beta_0 + \beta_1 + \varepsilon_i & i\text{-th person Asian} \\ \beta_0 + \beta_2 + \varepsilon_i & i\text{-th person white} \\ \beta_0 + \varepsilon_i & i\text{-th person black} \end{cases}$$

Now,

β_0: avg black balance

$\beta_0+\beta_1$: avg Asian balance

$\beta_0+\beta_2$: avg white balance

β_1: avg diff b/w Asian and black

β_2: avg diff b/w white and black

- There is always one fewer dummy variable than levels
- The level with no dummy variable is the <u>baseline</u>.

p-value of coefficients being high implies no difference.

- Use the F-test instead. F-test p-value is 0.96, cannot reject

- This approach works with quantitative predictors mixed in

3.3.2: Extensions of the Linear Model

- Two unrealistic assumptions: <u>additive</u> and <u>linear</u>.

- <u>Additive assumption</u>: The effect of changes in a predictor X_j on the response Y is independent of the other predictor values.
- <u>Linear assumption</u>: The change in Y due to one-unit change in X_j is constant.

<u>Removing the additive assumption</u>

The linear models for the sales data showed that TV and radio seem to be associated with sales, but these models assumed ~~additive~~ effects of ad-mediums were independent

Need to consider interaction effects

$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \varepsilon$

Now $\frac{\partial Y}{\partial X_1}$ depends on X_2 and v.v.

Writing the model as

$$Y = \beta_0 + (\beta_1 + \beta_3 X_2) X_1 + \beta_2 X_2 + \varepsilon$$

tells you how a unit increase of X_1 interacts w/ a given value of X_2

- You could probably use $\partial / \partial X_1$ as well.

- Hierarchical principle: If you include interaction, include main effect terms too.