Week 6

- Residual degrees of freedom = Sample size n - model size k -1
- Adjusted R^2 is always greater than R^2.

Calibration Plot

Scatterplot between actual values of Y and expected values of Y

Marginal Slope

slope of explanatory variable in Simple Linear Regression

Partial Slope

slope of explanatory variable in multiple linear regression excluding the effect of other explanatory variables

Path Diagram

schematic drawing of the relationships among the explanatory variables and the response

Interpretation of Regression Output

If p-value is less than alpha value, reject the null hypothesis (regression is not significant)

Variance Inflation Factor

> 💡 quantifies the amount of unique variation in each explanatory variable and measures the effect of collinearity

$$VIF(X_j) = \frac{1}{1 - R_j^2}$$

R_j^2 is coefficient of determination in the regression of X_j on all of the other explanatory variables

Why does VIF matter?

- The standard error in estimation of the partial slope gets inflated due to VIF.
- Typically,

$$se(b_1) = \frac{s_e}{\sqrt{n}} \times \frac{1}{s_{x_1}}$$

- With VIF

$$se(b_1) = \frac{s_e}{\sqrt{n}} \times \frac{1}{s_x} \times \sqrt{VIF(X_1)}$$

- Larger the VIF, larger is collinearity
- Large VIF can substantially increase the standard error in predicting the partial slopes making those predictions unreliable

Collinearity

Signs of Collinearity

- R^2 increases less than we'd expect.
- Slopes of correlated explanatory variables in the model change dramatically.
- The F-statistic is more impressive than individual t-statistics.
- Standard errors for partial slopes are larger than those for marginal slopes.
- Variance inflation factors increase.

Collinearity

- Remedies for Collinearity

  - Remove redundant explanatory variables.

  - Re-express explanatory variables.

  - Do nothing if the explanatory variables are significant with sensible estimates.

Adjusted R-squared

$$1 - \frac{(1 - R^2)(n - 1)}{(n - k - 1)}$$

F-statistic

$$F = \frac{R^2}{1 - R^2} * \frac{n - k - 1}{k}$$