Assessing a Multiple Regression Model

READING:	3.2
EXERCISES:	CH 3. 19-21
ASSIGNED:	HW 7
PRODUCER:	DR. MARIO

IMG CREDIT: ALEX RIEGERT-WATERS

t-Tests for Coefficients

- Hypotheses:
  - $H_0: \beta_i = 0$
  - $H_a: \beta_i \neq 0$
- Test Statistic:
  - $t^* = {\widehat{\beta}_i}/{SE_{\widehat{\beta}_i}}$
- P-Value: Non-directional and Use t-Distribution with *n-k-1* d.f.

Confidence Intervals for Coefficients

- Formula for 95% Confidence Interval:

$$\hat{\beta}_i \pm \mathrm{t}_{0.025,n-k-1} * SE_{\widehat{\beta_i}}$$

- Critical Value $\mathrm{t}_{0.025,n-k-1}$ is the 97.5 percentile on the t-Distribution with $n - k - 1$ Degrees of Freedom

ANOVA Table

- Updated for Multiple Linear Regression

Source	d.f.	Sum of Squares	Mean Square	F	P-value
Model	k	$SSModel$	$\frac{SSModel}{k}$	$\frac{MSModel}{MSE}$	$F_{k,n-k-1}$
Residual	$n-k-1$	SSE	$\frac{SSE}{n-k-1}$		
Total	$n-1$	$SSTotal$			

- Formulas for *SSModel*, *SSE*, and *SSTotal* are Same

ANOVA F-Test

- Tests the **Overall** Effectiveness of the Linear Model as a Whole
- Hypotheses:
  - $H_0: \beta_1 = \beta_2 = \cdots = \beta_k = 0$
  - $H_a: at\ least\ one\ \beta_i \neq 0$
- Test Statistic: *Make sure you divide by the correct degrees of freedom.*
- P-value: Use F-Distribution with *k* numerator degrees of freedom and *n-k-1* denominator degrees of freedom.

Coefficient of Determination

- Recall Formula:

$$R^2 = \frac{SSModel}{SSTotal} = 1 - \frac{SSE}{SSTotal}$$

- Book Uses $\boldsymbol{R^2}$ instead of $\boldsymbol{r^2}$ Since Finding R-squared is Not as Simple as Just Calculating Correlation and Squaring It
- $\boldsymbol{R^2}$ can be Found by Squaring the Correlation between $\boldsymbol{y}$ and $\boldsymbol{\hat{y}}$
- Problem: *Adding a new predictor variable into your linear regression model* ***will never*** *decrease* R^2 (make it worse)

Adjusted Coefficient of Determination

- Formula for Adjusted R-Squared:

$$R^2_{adj} = 1 - \frac{\left(\frac{SSE}{n-k-1}\right)}{\left(\frac{SSTotal}{n-1}\right)} = 1 - \frac{\hat{\sigma}_\epsilon}{s_y^2}$$

- *SSE* will **Never Increase** as You Add Variables to Your Model
- Sample Variance of *Y* Is **Completely Unaffected** by Model
- Complexity Measured By *k* **Influences** adjusted R-squared

Adjusted Coefficient of Determination

- Adding an Extra Variable to the Model Will Likely Cause *SSE* to Decrease or Stay the Same (Extremely Unusual)

- However, *MSE* may Actually Increase Since the **Error Degrees of Freedom** in the Denominator **May Decrease More**

$$MSE = \left(\frac{SSE}{n-k-1}\right)$$

Confidence and Prediction Intervals

- Confidence and Prediction Intervals are Still More Valuable Than Point Predictions

- More Difficult to Calculate by Hand
  - Reason 1: Uncertainty in Each Slope
  - Reason 2: Model is Not Just a Straight Line on a Cartesian Plane

Supplement for Lecture 14

- Fit Multiple Linear Regression Model to Data
- Interpreting Individual Slopes and t-Tests
- Standard Error of Regression and adjusted R-squared
- ANOVA table and F-Test
- Confidence Intervals and Prediction Intervals

Thank You

Make Reasonable Decisions