

READING:

EXERCISES:

ASSIGNED:

PRODUCER:

DR. MARIO



IMG CREDIT: ALEX RIEGERT-WATERS

Effectiveness of a Model

- Idea 1: A simple linear regression model is **not effective** if we **cannot conclude** with confidence that the slope $\beta_1 \neq 0$
- Idea 2: A simple linear regression model is **not effective** if we determine that the **predictions** from the fitted model only explain a **small** amount of the **total variability** of the response variable Y
- Analysis of Variance (ANOVA) = Method for Implementing Idea 2

Partitioning Variability

- Textbook Explanation of Statistical Modeling

$$Data = Model + Error$$

- Partitioning Variability

TOTAL Variation in
Response Variable Y

=

Variation Explained
by MODEL

+

Unexplained Variation
in RESIDUALS

- ANOVA Makes Inference Based Off This Partition

Partitioning Variability

- Mathematical Proof

$$y - \bar{y} = (\hat{y} - \bar{y}) + (y - \hat{y})$$

$$(y - \bar{y})^2 = (\hat{y} - \bar{y})^2 + 2(\hat{y} - \bar{y})(y - \hat{y}) + (y - \hat{y})^2$$

$$\Sigma(y - \bar{y})^2 = \Sigma(\hat{y} - \bar{y})^2 + \Sigma 2(\hat{y} - \bar{y})(y - \hat{y}) + \Sigma(y - \hat{y})^2$$

$$= \Sigma(\hat{y} - \bar{y})^2 + 2\Sigma(\hat{y}y - \hat{y}^2 - \bar{y}y + \bar{y}\hat{y}) + \Sigma(y - \hat{y})^2$$

$$= \Sigma(\hat{y} - \bar{y})^2 + 2\Sigma\hat{y}(y - \hat{y}) + 2\Sigma\bar{y}(\hat{y} - y) + \Sigma(y - \hat{y})^2$$

$$= \Sigma(\hat{y} - \bar{y})^2 + 2\Sigma\hat{y}(y - \hat{y}) - 2\bar{y}\Sigma(y - \hat{y}) + \Sigma(y - \hat{y})^2$$

$$= \Sigma(\hat{y} - \bar{y})^2 + \Sigma(y - \hat{y})^2$$

Partitioning Variability

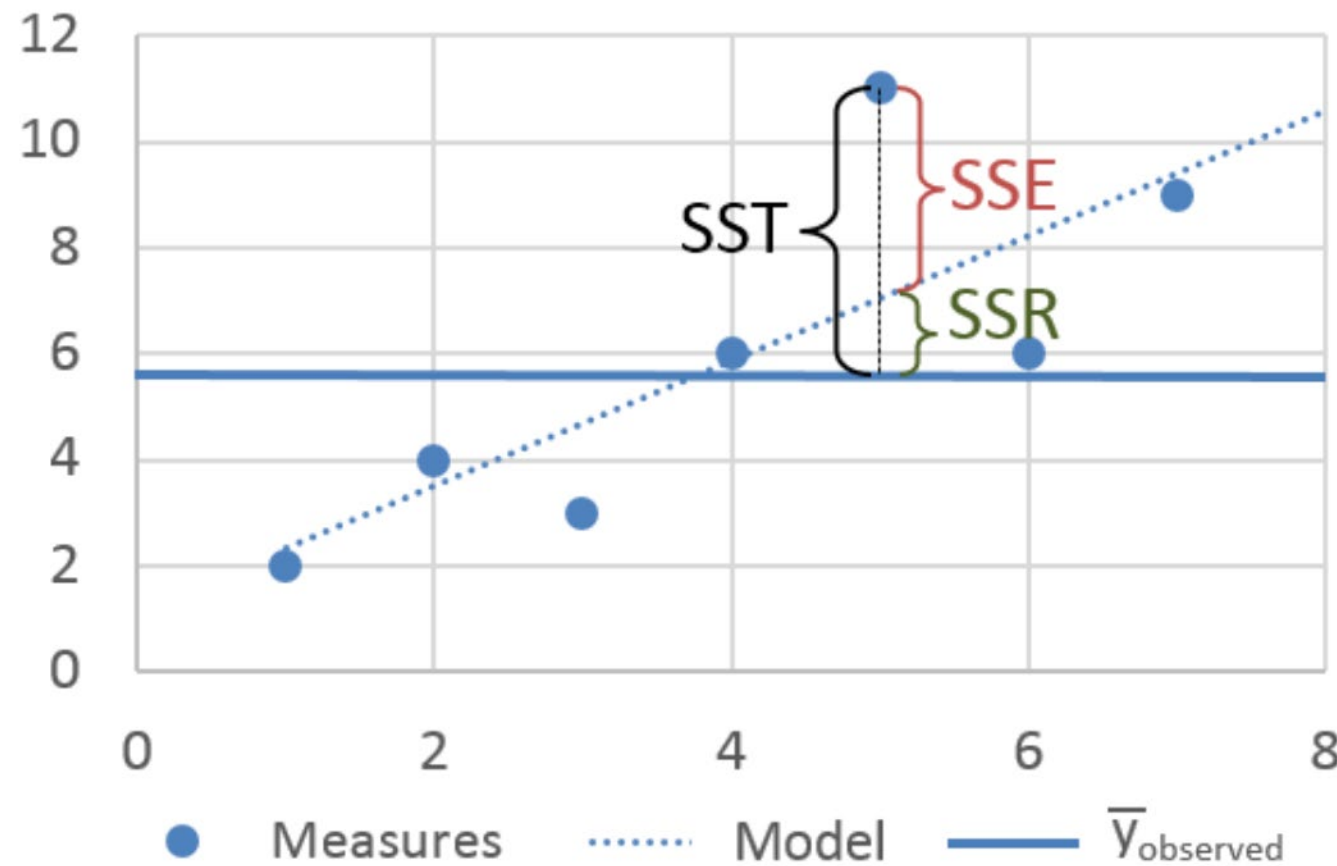
- Conclusion

$$\Sigma(y - \bar{y})^2 = \Sigma(\hat{y} - \bar{y})^2 + \Sigma(y - \hat{y})^2$$

$$SSTotal = SSModel + SSE$$

- In General, We Want *SSModel* to be **Larger Than** *SSE*
- In Many Textbooks, *SSModel* is sometimes called *SSR* for Sum of Squares Regression

Partitioning Variability



IMG CREDIT: [CHRISTIAN GOLD](#)

ANOVA Table

- Classic Table

Source	d.f.	Sum of Squares	Mean Square	F	P-value
Model	1	SS_{Model}	$SS_{Model}/1$	$\frac{MS_{Model}}{MSE}$	$F_{1,n-2}$
Residual	$n - 2$	SSE	$SSE/(n - 2)$		
Total	$n - 1$	$SSTotal$			

- Degrees of Freedom and Sum of Squares for Total are Both a Summation of Values for Model and Residual

ANOVA Test for Simple Linear Regression

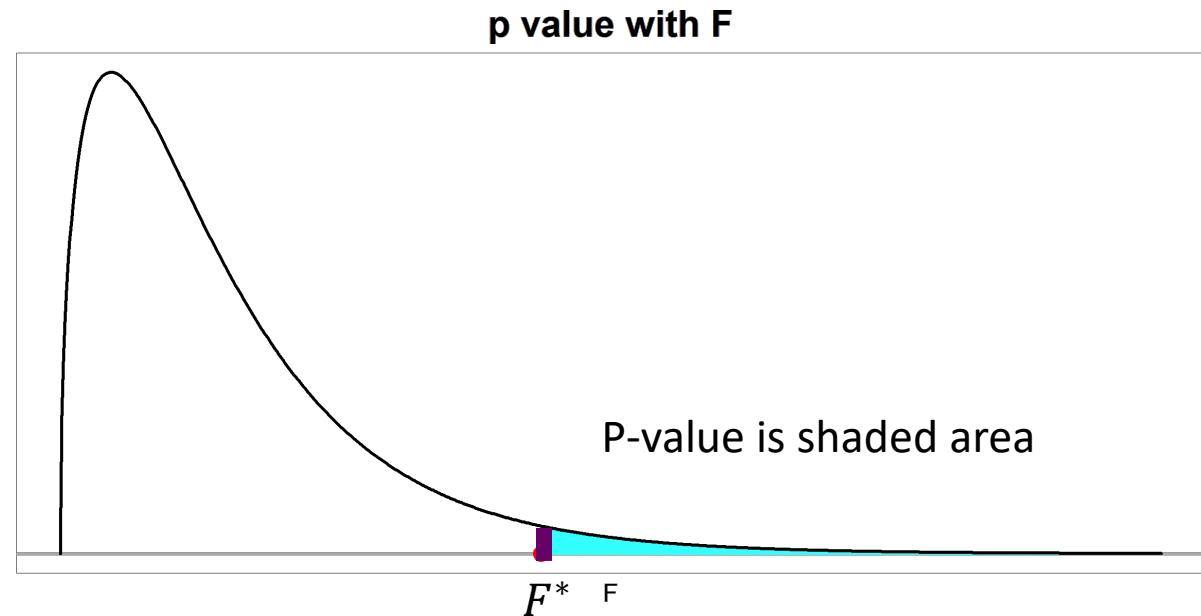
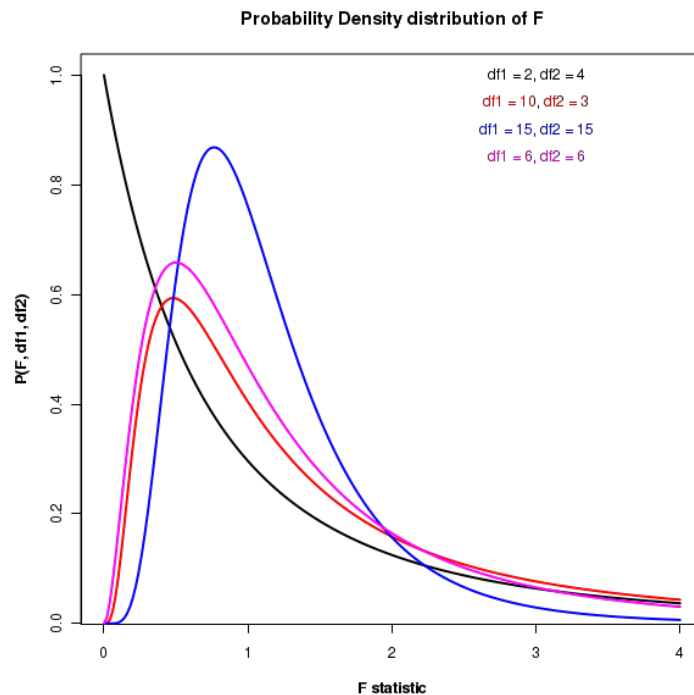
- Equivalent to the t-Test for the Slope
 - Identical Hypotheses (in Simple Linear Regression)
 - Perform Simple Linear Regression and Obtain Table
 - Different Test Statistic

$$F^* = \frac{SSModel/1}{SSE/(n-2)} = \frac{MSModel}{MSE}$$

- Compute p-value Based off F-Distribution and Not t-Distribution
 - F-distribution based off numerator degrees of freedom and denominator degrees of freedom
- Decision and Interpretation is the Same (in SLR)

F-Distribution

- Typically, Right Skewed



- Used Typically When Test Statistic is a Ratio of Variances

Example: Fatalities

- Question: *Is there a linear relationship between the proportion of young drivers in a state and the number of vehicle fatalities in a state?*
- Data for States from 1982 Through 1988
 - From AER Package
 - *fatal* = Number of Vehicle Fatalities
 - *youngdrivers* = Percent of Drivers Aged 15 – 24 (inclusive)
- Question: *What is the problem with fitting a linear regression for the relationship **fatal** versus **youngdrivers**?*

Supplement for Lecture 10

- Conducting t-Test for Slope
- Confidence Interval for Parameters
- Getting ANOVA Table
- Conducting F-test and Comparing Results

Make Reasonable Decisions

