Regression Line

$$Y = a + bX + e$$

$$\hat{Y} = a + bX$$

$$Y = \hat{Y} + e$$

Residuals

$$e = Y - \hat{Y}$$

OLS Estimation

- Using calculus, we obtain

$$b = \frac{\sum[(X - M_X)(Y - M_Y)]}{\sum(X - M_X)^2} = \frac{SS_{XY}}{SS_X}$$

$$a = M_Y - bM_X$$

OLS Estimation

$$b = r\left(\frac{s_Y}{s_X}\right)$$

The Standardized Model

$$\hat{z}_Y = \beta z_X$$

Standard Error of the Estimate

$$s_e = \sqrt{\frac{SSE}{df}}$$

- where $df = n - 2$ for simple linear regression

- Recall from ANOVA: $MSE = \frac{SSE}{df}$
  - Thus, s_e is sometimes called root MSE

Coefficient of Determination

- Coefficient of determination: r^2 or R^2
  - Proportion of variance in Y that is accounted for by X

Hypothesis Test for ρ^2

F-statistic:

TABLE 16.4 The *F* Table for an Analysis of Regression

Formulas for Completing the Analysis of Regression				
Source of Variation	***SS***	***df***	***MS***	F_{obt}
Regression	$r^2 SS_Y$	1	$\frac{SS_{regression}}{df_{regression}}$	$\frac{MS_{regression}}{MS_{residual}}$
Residual (error)	$(1 - r^2)SS_Y$	$n - 2$	$\frac{SS_{residual}}{df_{residual}}$	
Total	$SS_{regression} + SS_{residual}$	$n - 1$		

Hypothesis Test for α^* and β^*

t-statistic

$$t = \frac{a}{s_a} \qquad\qquad t = \frac{b}{s_b}$$

- where
  - a is the parameter estimate for the intercept and s_a is the estimated standard error for the slope
  - b is the parameter estimate for the intercept and s_b is the estimated standard error for the slope

Confidence Intervals

$$a \pm t_{cv}(s_a)$$

$$b \pm t_{cv}(s_b)$$

- If 0 is contained in the CI, then fail to reject the null hypothesis

Hypothesis Testing

$$F = t^2$$

*For SLR only