CHAPTER 12: Multiple Regression and Model Building

12.6 **a.** The plot of price versus size has a straight line appearance and thus the model $y = \beta_0 + \beta_1 x_1 + \varepsilon$ is appropriate. The plot of price versus rating has a straight line appearance. The model $y = \beta_0 + \beta_1 x_1 + \varepsilon$ is appropriate. Combining these two models, we obtain the model $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$.

b. $\mu_{y|x_1=20, x_2=9}$ is the mean (or average) of the sales prices of all houses having 20 hundred (that is, 2000) square feet and a niceness rating of 9.

c. β_0 = mean sales price of all houses having 0 square feet and 0 niceness rating—meaningless.
β_1 = change in mean sales price associated with each increase in house size of 100 square feet, when niceness rating stays constant.
β_2 = change in mean sales price associated with each increase in niceness rating of 1, when house size remains constant.

d. The error term represents all factors other than the square footage and the niceness rating. One such factor is the ability and effort of the real estate agent listing the house.

12.11 **a.** b_0 = 29.347, b_1 = 5.6128, b_2 = 3.8344

b_0 = meaningless

b_1 = 5.6128 implies that we estimate that mean sales price increases by \$5,612.80 for each increase of 100 square feet in house size, when the niceness rating stays constant.

b_2 = 3.8344 implies that we estimate that mean sales price increases by \$3,834.40 for each increase in niceness rating of 1, when the square footage remains constant.

b. 172.28. From $\hat{y}$ = 29.347 + 5.6128(20) + 3.8344(8)

12.17 **a.** $SSE = 73.6; s^2 = \dfrac{SSE}{n-(k+1)} = \dfrac{73.6}{10-(2+1)} = \dfrac{73.6}{7} = 10.5; s = \sqrt{10.5} = 3.242$

b. Total variation = 7447.5
Unexplained variation = 73.6
Explained variation = 7374

c. $R^2 = \dfrac{7374}{7447.5} = .99 \quad \bar{R}^2 = \left(R^2 - \dfrac{k}{n-1}\right)\left(\dfrac{n-1}{n-(k+1)}\right)$

$$= \left(.99 - \frac{2}{10-1}\right)\left(\frac{10-1}{10-(2+1)}\right)$$

$$= .987$$

R^2 and $\bar{R}^2$ close together and close to 1.

d. F(model) $= \dfrac{(\text{Explained variation})/k}{(\text{Unexplained variation})/(n-(k+1))}$

$$= \frac{7374/2}{73.6/(10-(2+1))} = \frac{7374/2}{73.6/7} = 350.87$$

e. Based on 2 and 7 degrees of freedom, $F_{.05}$ = 4.74. Since F(model) = 350.87 > 4.74, we reject $H_0 : \beta_1 = \beta_2 = 0$ by setting $\alpha = .05$.

f. Based on 2 and 7 degrees of freedom, $F_{.01}$ = 9.55. Since F(model) = 350.87 > 9.55, we reject $H_0 : \beta_1 = \beta_2 = 0$ by setting $\alpha = .01$.

g. p-value = 0.00 (which means less than .001). Since this p-value is less than α = .10, .05, .01, and .001, we have extremely strong evidence that $H_0 : \beta_1 = \beta_2 = 0$ is false. That is, we have extremely strong evidence that at least one of x_1 and x_2 is significantly related to y.

12.22 We first consider the intercept β_0

a. b_0 = 29.347, s_{b_0} = 4.891, t = 6.00

where t = b_0/s_{b_0} = 29.347/4.891 = 6.00

b. We reject H_0: $\beta_0 = 0$ (and conclude that the intercept is significant) with α = .05

if $|t| > t_{.05/2} = t_{.025}$

Since $t_{.025}$ = 2.365 (with n – (k + 1) = 10 – (2 + 1) = 7 degrees of freedom), we have t = 6.00 > $t_{.025}$ = 2.365.

We reject H_0 : $\beta_0 = 0$ with α = .05 and conclude that the intercept is significant at the .05 level.

c. We reject H_0 : $\beta_0 = 0$ with α = .01 if $|t| > t_{.01/2} = t_{.005}$

Since $t_{.005}$ = 3.499 (with 7 degrees of freedom), we have t = 6.00 > $t_{.005}$ = 3.499.

We reject H_0 : β_0 = 0 with α = .01 and conclude that the intercept is significant at the .01 level.

d. The Minitab output tells us that the p-value for testing H_0 : $\beta_0 = 0$ is 0.000. Since this p-value is less than each given value of α, we reject H_0 : $\beta_0 = 0$ at each of these values of α. We can conclude that the intercept β_0 is significant at the .10, .05, .01, and .001 levels of significance.

e. A 95% confidence interval for β_0 is

$[b_0 \pm t_{\alpha/2}\, s_{b_0}] = [b_0 \pm t_{.025}\, s_{b_0}]$

$= [29.347 \pm 2.365(4.891)]$

$= [17.780, 40.914]$

This interval has no practical interpretation since β_0 is meaningless.

f. A 99% confidence interval for β_0 is

$[b_0 \pm t_{.005}\, s_{b_0}] = [29.347 \pm 3.499(4.891)]$

$= [12.233, 46.461]$

We next consider β_1.

(a) $b_1 = 5.6128$, $s_{b_1} = .2285$, $t = 24.56$

where $t = b_1 / s_{b_1} = 5.6128/.2285 = 24.56$

(b), (c), and (d):

We reject $H_0: \beta_1 = 0$ (and conclude that the independent variable x_1 is significant) at level of significance α if $|t| > t_{\alpha/2}$. Here $t_{\alpha/2}$ is based on $n - (k + 1) = 10 - 3 = 7$ d.f.

For $\alpha = .05$, $t_{\alpha/2} = t_{.025} = 2.365$, and for $\alpha = .01$, $t_{\alpha/2} = t_{.005} = 3.499$.

Since $t = 24.56 > t_{.025} = 2.365$, we reject $H_0: \beta_1 = 0$ with $\alpha = .05$.

Since $t = 24.56 > t_{.005} = 3.499$, we reject $H_0: \beta_1 = 0$ with $\alpha = .01$.

Further, the Minitab output tells us that the *p*-value related to testing $H_0: \beta_1 = 0$ is 0.000. Since this *p*-value is less than each given value of α, we reject H_0 at each of these values of α (.10, .05, .01, and .001).

The rejection points and *p*-values tell us to reject $H_0: \beta_1 = 0$ with $\alpha = .10$, $\alpha = .05$, $\alpha = .01$, and $\alpha = .001$. We conclude that the independent variable x_1 (home size) is significant at the .10, .05, .01, and .001 levels of significance.

(e) and (f):

95% interval for β_1:

$[b_1 \pm t_{.025}\, s_{b_1}] = [5.6128 \pm 2.365(.2285)]$

$= [5.072, 6.153]$

99% interval for β_1:

$[b_1 \pm t_{.005}\, s_{b_1}] = [5.6128 \pm 3.499(.2285)]$

$= [4.813, 6.412]$

For instance, we are 95% confident that the mean sales price increases by between \$5072 and \$6153 for each increase of 100 square feet in home size, when the rating stays constant.

Last, we consider β_2.

(a) $b_2 = 3.8344$, $s_{b_2} = .4332$, $t = 8.85$

where $t = b_2/s_{b_2} = 3.8344/.4332 = 8.85$

(b), (c), and (d):

We reject $H_0: \beta_2 = 0$ (and conclude that the independent variable x_2 is significant) at level of significance α if $|t| > t_{\alpha/2}$. Here, $t_{\alpha/2}$ is based on $n - (k+1) = 10 - 3 = 7$ d.f.

For $\alpha = .05$, $t_{\alpha/2} = t_{.025} = 2.365$, and for $\alpha = .01$, $t_{\alpha/2} = t_{.005} = 3.499$.

Since $t = 8.85 > t_{.025} = 2.365$, we reject $H_0: \beta_2 = 0$ with $\alpha = .05$.

Since $t = 8.85 > t_{.005} = 3.499$, we reject $H_0: \beta_2 = 0$ with $\alpha = .01$.

Further, the Minitab output tells us that the *p*-value related to testing $H_0: \beta_2 = 0$ is 0.000. Since this *p*-value is less than each given value of α, we reject H_0 at each of these values of α (.10, .05, .01, and .001).

The rejection points and *p*-values tell us to reject $H_0: \beta_2 = 0$ with $\alpha = .10$, $\alpha = .05$, $\alpha = .01$, and $\alpha = .001$.

We conclude that the independent variable x_2 (niceness rating) is significant at the .10, .05, .01, and .001 levels of significance.

(e) and (f):

95% interval for β_2:

$[b_2 \pm t_{.025} s_{b_2}] = [3.8344 \pm 2.365(.4332)]$

$= [2.810, 4.860]$

99% interval for β_2:

$[b_2 \pm t_{.005} s_{b_2}] = [3.8344 \pm 3.499(.4332)]$

$= [2.319, 5.350]$

For instance, we are 95% confident that the mean sales price increases by between \$2810 and \$4860 for each increase of one rating point, when the home size remains constant.

12.70 The estimates of the coefficients indicate that at a specified square footage, adding rooms increases selling price while adding bedrooms reduces selling price. Thus building both a family room and a living room (while maintaining square footage) should increase sales price. In addition, adding a bedroom at the cost of another room will tend to decrease selling price.