CHAPTER 12: Multiple Regression

12.7 [LO 1]

a. The plots show a linear (or somewhat linear) relationship between Price & Demand, IndPrice & Demand, PriceDiff & Demand, and AdvExp & Demand.

b. The mean demand for the large size bottle of Fresh when the price of Fresh is $3.70, the average industry price of competitors' similar detergents is $3.90 and the advertising expenditure to promote Fresh is 6.50 ($650,000).

c. β_0 = meaningless in practical terms

β_1 = the mean change in demand for each additional dollar in the price of Fresh holding all other predictor variables constant.

β_2 = the mean change in demand for each additional dollar in the average price of competitors' detergents holding all other predictor variables constant.

β_3 = the mean change in demand for each additional $100,000 spent on advertising Fresh holding all other predictor variables constant.

ε = all other factors that influence the demand for Fresh detergent

d. The plots for Demand vs. AdvExp and Demand vs. PriceDif appear to be more linear than the other two plots.

12.15 [LO 4, 5]

a. SSE = 1.4318, s^2 = 1.4317/(30-4) = .0551

b. Total variation = 13.4586

Explained variation = 12.0268

c. R^2 = .894 = 12.0268/13.4586

Adjusted R^2 = .881 = (.894 – (3/29))(29/26)

Approximately 89% of the variance in demand is predicted by price, average price, and advertising, which drops to 88% when adjusted for the number of predictors.

d. F = MSexplained/MSE = 72.80

e. at <.05, model is significant, F-critical = 2.98

f. at <.01, model is significant, F-critical = 4.64

g. output p = .000000000000888

12.19 [LO 3]

a. b_0 = 1946.8020, b_1 = 0.0386, b_2 = 1.0394, b_0 = –413.7578

b_0 = labour hours when x-ray = 0, bed days = 0, and length of stay = 0 which is probably meaningless as the hospital has no patients staying there.

b_1 = implies that labour hours increases 0.04 for each unit increase in x-rays, when bed days and length of stay remain constant (predicted change).

b_2 = implies that labour hours increases by 1.04 for each unit of increase in bed days when x-rays and length of stay remain constant (predicted change).

b_3 = implies that labour hours decreases by 413.76 when length of stay decreases by one unit and both x-rays and bed days remain constant (predicted change).

b. $\hat{y}$ = 1946.802 + .0386 (56194) + 1.0394 (14077.88) – 413.7578 (6.89) = 15897.65

c. Therefore, actual hours were 17207.31 – 15896.25 = 1311.06 hours greater than predicted.

12.29 [LO 6]

y = 17207.31 is above the upper limit of the interval [14906.2, 16886.3]; this y-value is unusually high.

12.33 [S 12.8]

The shorter interval is from the model using x_4. This model is better.

12.35 [S 12.9]

Multiply: $x_1 x_2$

12.37 [S 12.9]

y-hat = –2.3497 + 2.3611x_1 + 4.1831x_2 – 0.3489x_1x_2

x_1 = radio / TV x_2 = print

a. x_2 = 1, slope = 2.0122; x_2 = 2, slope = 1.6633; x_2 = 3, slope = 1.3144; x_2 = 4, slope = 0.9655; x_2 = 5, slope = 0.6166.

These slopes are the estimated average sales volume increase (in units of $10,000) for every $1,000 increase in radio and tv ads.

b. x_1 = 1, slope = 3.8342; x_1 = 2, slope 3.4853; x_1 = 3, slope = 3.1364, x_1 = 4, 2.7875; x_1 = 5, slope = 2.4386.

These slopes are the estimated average sales volume increase (in units of $10,000) for every $1,000 increase in print ads.

c. The smallest print slope is bigger than the largest radio/tv slope.

12.39 [S 12.10]

An independent variable, the levels of which are defined by describing them.

12.41 [S 12.10]

The effect of the qualitative independent variable on the dependent variable.

12.45 [S 12.10]

a. No interaction since *p*-values are so large.

b. $\hat{y}$ = 8.61178 (861,178 bottles)

95% prediction interval = [8.27089,8.95266]—slightly bigger

12.47 [S 12.11]

$(k - g)$ denotes the number of regression parameters we have set equal to zero in H_0.
$[n - (k + 1)]$ denotes the denominator degrees of freedom.

12.49 [S 12.11]

Model 3—complete
Model 1—reduced
$H_0 : \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$

$$F = \frac{\frac{1.4318 - .5347}{4}}{\frac{.5347}{22}} = 9.228$$

$F_{.05} = 2.82$ based on 4 and 22 degrees of freedom.
$F_{.01} = 4.31$ based on 4 and 22 degrees of freedom.
Since 9.228 > 4.31, reject H_0 at $\alpha = .05$ and .01; Because the null hypothesis was that the equations have the same slope and intercept, rejecting the Ho means that at least one of these claims is false.

12.51 [LO 6]

$$\hat{y} = 30{,}626 + 3.893(28000) - 29{,}607(1.56) + 86.52(1821.7) \cong 251{,}056$$

12.53 [LO 5]

a. Output for all:

SUMMARY OUTPUT All

Regression Statistics	
Multiple R	0.878394
R Square	0.771577
Adjusted R Square	0.754006
Standard Error	1.319372
Observations	29

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	152.8786	76.43932	43.91193	4.61E-09
Residual	26	45.25929	1.740742		
Total	28	198.1379			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	16.94219	1.435079	11.80575	6.01E-12	13.99234	19.89204	13.99234	19.89204
Age(x1)	-0.00066	0.013029	-0.05035	0.96023	-0.02744	0.026126	-0.02744	0.026126
Price(x2)	-0.05548	0.006086	-9.11638	1.4E-09	-0.06799	-0.04297	-0.06799	-0.04297

Significant regression model. Price is the only significant predictor.

b. Outputs:

SUMMARY OUTPUT Males

Regression Statistics	
Multiple R	0.817165
R Square	0.667758
Adjusted R Square	0.607351
Standard Error	1.408836
Observations	14

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	43.88127	21.94063	11.05422	0.002333
Residual	11	21.83302	1.98482		
Total	13	65.71429			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	13.23223	2.361521	5.603267	0.00016	8.034556	18.42991	8.034556	18.42991
Age(x1)	-0.07728	0.035932	-2.15066	0.054585	-0.15636	0.001808	-0.15636	0.001808
Price(x2)	-0.01943	0.017359	-1.11942	0.286809	-0.05764	0.018774	-0.05764	0.018774

SUMMARY OUTPUT Females

Regression Statistics	
Multiple R	0.948477
R Square	0.899609
Adjusted R Square	0.882877
Standard Error	1.000841
Observations	15

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	2	107.7131	53.85657	53.7661	1.02E-06
Residual	12	12.02019	1.001683		
Total	14	119.7333			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	13.83504	4.13341	3.347125	0.005811	4.829112	22.84097	4.829112	22.84097
X Variable 1	0.028102	0.02989	0.940197	0.365658	-0.03702	0.093226	-0.03702	0.093226
X Variable 2	-0.04714	0.013665	-3.44982	0.004807	-0.07691	-0.01737	-0.07691	-0.01737

Models are significant for both men and women. For men, Age has a slight negative relationship with interest (younger more interested, $p<.10$). For women, Price is the significant predictor (greater interest with lower prices, $p<.01$).

12.55 [S 12.9]

a. Interaction term is not a significant predictor ($p>.10$).

b. Introducing the interaction term decreases the F-value but increases the Multiple R slightly.

12.57 [LO 5]

a. β_5 : b_5 = 0.2137, Confidence Interval = [0.0851,0.3423], p–value = .0022, significant at 0.01 but not 0.001 so we have very strong evidence.

β_5 : b_6 = 0.3818, Confidence Interval = [0.2551,0.5085], p–value < .001, significant at 0.001 so we have extremely strong evidence.

b. b_6 = .1681 Confidence Interval: [.0363,.29], p–value = .0147, strong evidence.

c. $\mu_{[d,a,C]} - \mu_{[d,a,A]} = [\beta_0 + \beta_1 d + \beta_2 a + \beta_3 a^2 + \beta_4 da + \beta_5(0) + \beta_6(1) + \beta_7 a(0) + \beta_8 a(1)]$

$- [\beta_0 + \beta_1 d + \beta_2 a + \beta_3 a^2 + \beta_4 da + \beta_5(0) + \beta_6(0) + \beta_7 a(0) + \beta_8 a(0)]$

$= \beta_6 + \beta_8 a$

= –.9351+.2035(6.2)=.3266

= –.9351+.2035(6.6)=.408

$\mu_{[d,a,C]} - \mu_{[d,a,B]} = [\beta_0 + \beta_1 d + \beta_2 a + \beta_3 a^2 + \beta_4 da + \beta_5(0) + \beta_6(1) + \beta_7 a(0) + \beta_8 a(1)]$

$- [\beta_0 + \beta_1 d + \beta_2 a + \beta_3 a^2 + \beta_4 da + \beta_5(1) + \beta_6(0) + \beta_7 a(1) + \beta_8 a(0)]$

$= \beta_6 + \beta_8 a - \beta_5 - \beta_7 a$

$= \beta_6 - \beta_5 + \beta_8 a - \beta_7 a$

= –.9351–(–.4807)+.2035(6.2)–.1072(6.2) = .14266

= –.9351–(–.4807)+.2035(6.6)–.1072(6.6) = .18118

Both differences increased with the larger value of *a*.

d. The prediction interval for the third model is slightly shorter.

The differences between campaign A and campaigns B & C change as volume level changes.