8 a. $R^2 = \frac{SSR}{SST} = \frac{14\,052.2}{15\,182.9} = 0.926$

b. $\text{adj } R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$

$= 1 - (1 - 0.926)\frac{10-1}{10-2-1} = 0.905$

c. Yes: after adjusting for the number of independent variables in the model, we see that 90.5 per cent of the variability in y has been accounted for.

10 a. $R^2 = \frac{SSR}{SST} = \frac{12\,000}{16\,000} = 0.75$

b. $\text{adj } R^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1}$

$= 1 - (0.25)\frac{9}{7} = 0.68$

c. The adjusted coefficient of determination shows that 68 per cent of the variability has been explained by the two independent variables: thus, we conclude that the model does not explain a large proportion of variability.

12 a. $MSR = SSR/p = 6216.375/2 = 3108.188$

$MSE = \frac{SSE}{n-p-1} = \frac{507.75}{10-2-1} = 72.536$

b. $F = MSR/MSE = 3108.188/72.536 = 42.85$

Using F table (2 degrees of freedom numerator and 7 denominator), p-value is less than 0.01

Because p-value $\leq \alpha = 0.05$, the overall model is significant.

c. $t = 0.5906/0.0813 = 7.26$

Using t table (7 degrees of freedom), area in tail is less than 0.005: p-value is less than 0.01.

Because p-value $\leq \alpha$, β_1 is significant.

d. $t = 0.4980/0.0567 = 8.78$

Using t table (7 degrees of freedom), area in tail is less than 0.005: p-value is less than 0.01.

Because p-value $\leq \alpha$, β_2 is significant.

14 a. In the two independent variable case the coefficient of x_1 represents the expected change in y corresponding to a one unit increase in x_1 when x_2 is held constant. In the single independent variable case the coefficient of x_1 represents the expected change in y corresponding to a one unit increase in x_1.

b. Yes. If x_1 and x_2 are correlated one would expect a change in x_1 to be accompanied by a change in x_2.

16 a. $F = 28.38$

Using F table (2 degrees of freedom numerator And 7 denominator), p-value is less than 0.01.

Actual p-value $= 0.002$

Because p-value $\leq \alpha$, there is a significant relationship.

b. $t = 7.53$

Using t table (7 degrees of freedom), area in tail is less than 0.005: p-value is less than 0.01.

Actual p-value $= 0.001$

Because p-value $\leq \alpha$, β_1 is significant and x_1 should not be dropped from the model.

c. $t = 4.06$

Actual p-value $= 0.010$

Because p-value $\leq \alpha$, β_2 is significant and x_2 should not be dropped from the model.

18 a. Using Minitab, the 95 per cent confidence interval is 132.16 to 154.16.

b. Using Minitab, the 95 per cent prediction interval is 111.13 to 175.18.

20 a. $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2$ where

$x_2 = 0$ if level 1 and 1 if level 2

b. $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2(0) = \beta_0 + \beta_1 x_1$

c. $E(Y) = \beta_0 + \beta_1 x_1 + \beta_2(1) = \beta_0 + \beta_1 x_1 + \beta_2$

d. $\beta_2 = E(Y \mid \text{level } 2) - E(y \mid \text{level } 1)$

β_1 is the change in $E(Y)$ for a 1 unit change in x_1 holding x_2 constant.

22 a. €15 300

b. Estimate of sales $= 10.1 - 4.2(2) + 6.8(8) + 15.3(0) = 56.1$ or €56 100

c. Estimate of sales $= 10.1 - 4.2(1) + 6.8(3) + 15.3(1) = 41.6$ or €41 600

24 a. Relevant correlation details are as follows:

```
             Tar     Nicotine
Nicotine     0.977
             0.000

CO           0.957   0.926
             0.000   0.000

Cell Contents: Pearson correlation
               P-Value
```

Clearly the independent variables Tar and Nicotine are highly correlated. Not surprisingly when both predictors are fitted in the model a VIF of 21.6 is observed for each variable indicating the presence of serious multicollinearity.

b. Because of its slightly higher correlation with CO only the Tar variable is therefore considered as a predictor in the relevant regression model. Selected Minitab output is shown below:

```
The regression equation is
CO = 2.74 + 0.80 Tar

Predictor    Coef     SE Coef       T       P
Constant   2.7433     0.6752     4.06   0.000
Tar        0.8010    0.08360    15.92   0.000

S = 1.40   R-Sq = 91.7%   R-Sq(adj) = 91.3%
```

c. The p-value corresponding to $t = 15.92$ is $0.000 < \alpha = 0.05$: thus, Tar is a statistically significant predictor.

26 a. The Minitab output is shown below:

```
The regression equation is
Y - 0.20 + 2.60 X
Predictor     Coef    SE Coef       T       P
Constant     0.200      2.132    0.09   0.931
X           2.6000     0.6429    4.04   0.027
S = 2.033  R-Sq = 84.5%  R-Sq(adj) = 79.3%
Analysis of Variance
Source           DF      SS      MS      F      P
Regression        1  67.600  67.600  16.35  0.027
Residual Error    3  12.400   4.133
Total             4  80.000
```

b. Using Minitab we obtained the following values:

x_i	y_i	$\hat{y}_i$	Standardized residual
1.	3	2.8	0.16
2	7	5.4	0.94
3	5	8.0	−1.65
4	11	10.6	0.24
5	14	13.2	0.62

The point (3, 5) does not appear to follow the trend of remaining data: however, the value of the standardized residual for this point, −1.65, is not large enough for us to conclude that (3, 5) is an outlier.

c. Using Minitab, we obtained the following values:

x_i	y_i	Studentized deleted residual
1	3	0.13
2	7	0.91
3	5	−4.42
4	11	0.19
5	14	0.54

$t_{0.025} = 4.303$

$(n - p - 2 = 5 - 1 - 2 = 2$ degrees of freedom$)$

Since the studentized deleted residual for (3, 5) is $-4.42 < -4.303$, we conclude that the third observation is an outlier.

28 a. The Minitab output appears in the solution to part (b) of exercise 5; the estimated regression equation is:

Revenue = 83.2 + 2.29 TVAdv + 1.30 NewsAdv

b. Using Minitab we obtained the following values:

$\hat{y}_i$	Standardized residual
96.63	−1.62
90.41	−1.08
94.34	1.22
92.21	−0.37
94.39	1.10
94.24	−0.40
94.42	−1.12
93.35	1.08

With the relatively few observations, it is difficult to determine if any of the assumptions regarding the error term have been violated. For instance, an argument could be made that there does not appear to be any pattern in the plot; alternatively an argument could be made that there is a curvilinear pattern in the plot.

c. The values of the standardized residuals are greater than −2 and less than +2; thus, using test, there are no outliers. As a further check for outliers, we used Minitab to compute the following studentized deleted residuals:

Observation	Studentized deleted residual
1	−2.11
2	−1.10
3	1.31
4	−.33
5	1.13
6	−.36
7	−1.16
8	1.10

d. Using Minitab we obtained the following values:

Observation	h_i	D_i
1	0.63	1.52
2	0.65	0.70
3	0.30	0.22
4	0.23	0.01
5	0.26	0.14
6	0.14	0.01
7	0.66	0.81
8	0.13	0.06

The critical average value is

$\frac{3(p+1)}{n} = \frac{3(2+1)}{8} = 1.125$

Since none of the values exceed 1.125, we conclude that there are no influential observations. However, using Cook's distance measure, we see that $D_1 > 1$ (rule of thumb critical value); thus, we conclude the first observation is influential. Final conclusion: observation 1 is an influential observation.

STUDENTS-HUB.com Uploaded By: anonymous