30 a. $E(Y) = \dfrac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$

b. It is an estimate of the probability that a customer that does not have a Simmons credit card will make a purchase.

c. A portion of the Minitab binary logistic regression output follows:

```
Logistic Regression Table

                                                 Odds           95% CI
Predictor      Coef    SE Coef       z       P   Ratio   Lower   Upper
Constant    -0.9445     0.3150   -3.00   0.003
Card         1.0245     0.4235    2.42   0.016    2.79    1.21    6.39

Log-Likelihood =  64.265
Test that all slopes are zero: G = 6.072, DF = 1, P-Value = 0.014
```

Thus, the estimated logit is $\hat{g}(x) = -0.9445 + 1.0245x$.

d. For customers that do not have a Simmons credit card ($x = 0$)

$\hat{g}(0) = -0.9445 + 1.245(0) = 0.9445$

and

$$\hat{y} = \frac{e^{\hat{g}(0)}}{1 + e^{\hat{g}(0)}} = \frac{e^{0.9445}}{1 + e^{0.9445}} = \frac{0.3889}{1 + 0.3889} = 0.28$$

$$t_{0.025} = 2.776$$

($n - p - 2 = 8 - 2 - 2 = 4$ degrees of freedom)

Since none of the studentized deleted residuals is less than -2.776 or greater than 2.776, we conclude that there are no outliers in the data.

For customers that have a Simmons credit card ($x = 1$)

$\hat{g}(1) = -0.9445 + 1.245(1) = 0.0800$

and

$$\hat{y} = \frac{e^{\hat{g}(1)}}{1 + e^{\hat{g}(1)}} = \frac{e^{0.08}}{1 + e^{0.08}} = \frac{1.0833}{1 + 1.0833}$$

$= 0.52$

e. Using the Minitab output shown in part (c), the estimated odds ratio is 2.79. We can conclude that the estimated odds of making a purchase for customers who have a Simmons credit card are 2.79 times greater than the estimated odds of making a purchase for customers that do not have a Simmons credit card.

32 a. $E(Y) = \dfrac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$

b. A portion of the Minitab binary logistic regression output follows:

```
Logistic Regression Table

                                                 Odds           95% CI
Predictor      Coef    SE Coef       z       P   Ratio   Lower   UPPer
Constant    -2.6335     0.7985   -3.30   0.001
Balance     0.22018    0.09002    2.45   0.014    1.25    1.04    1.49

Log-Likelihood =  25.813
Test that all slop
pes are zero: G = 9.460, DF = 1, P-Value = 0.002
```

Thus, the estimated logistic regression equation is

$$E(y) = \frac{e^{2.6355 + 0.22018x}}{1 + e^{2.6355 + 0.22018x}}$$

c. Significant result: the p-value corresponding to the G test statistic is 0.0002.

d. For an average monthly balance of €1000, $x = 10$

$$E(y) = \frac{e^{2.6355 + 0.22018x}}{1 + e^{2.6355 + 0.22018x}} = \frac{e^{2.6355 + 0.22018(10)}}{1 + e^{2.6355 + 0.22018(10}}$$

$$= \frac{e^{0.4317}}{1 + e^{0.4317}} = \frac{0.6494}{1.6494} = 0.39$$

Thus, an estimate of the probability that customers with an average monthly balance of €1000 will sign up for direct payroll deposit is 0.39.

e. Repeating the calculations in part (d) using various values for x, a value of $x = 12$ or an average monthly balance of approximately €1200 is required to achieve this level of probability.

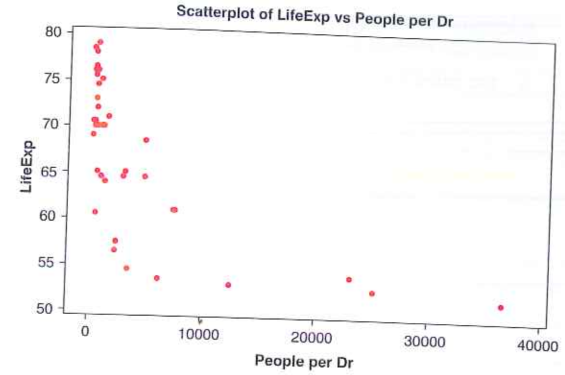
f. Using the Minitab output shown in part (b), the estimated odds ratio is 1.25. Because values of x are measured in hundreds of euros, the estimated odds of signing up for payroll direct deposit for customers that have an average monthly balance of €600 is 1.25 times greater than the estimated odds of signing up for payroll direct deposit for customers that have an average monthly balance of €500. Moreover, this interpretation is true for any €100 increment in the average monthly balance.

Chapter 16

Solutions

2 a. The MINITAB output is shown below:

```
The regression equation is
Y = 9.32 + 0.424 X

Predictor     Coef     SE Coef      T       P
Constant     9.315      4.196     2.22    0.113
X            0.4242     0.1944    2.18    0.117

S = 3.531   R-sq = 61.4%   R-sq(adj) = 48.5%

Analysis of Variance

SOURCE           DF     SS     MS     F      P
Regression        1   59.39  59.39  4.76  0.117
Residual Error    3   37.41  12.47
Total             4   96.80
```



The high p-value (0.117) indicates a weak relationship; note that 61.4 per cent of the variability in y has been explained by x.

b. The MINITAB output is shown below:

```
The regression equation is
Y = -8.10 + 2.41 X - 0.0480 XSQ

Predictor     Coef      SE Coef      T       P

Constant    -8.101      4.104     -1.97    0.187
X            2.4127     0.4409     5.47    0.032
XSQ         -0.04797    0.01050   -4.57    0.045

S = 1.279   R-sq = 96.6%   R-sq(adj) = 93.2%

Analysis of Variance

SOURCE           DF     SS      MS      F      P
Regression        2   93.529  46.765  28.60  0.034
Residual Error    2    3.271   1.635
Total             4   96.800
```

At the 0.05 level of significance, the relationship is significant; the fit is excellent.

c. $\hat{y} = -8.101 + 2.4127(20) - 0.04797(20)^2$

$= 20.965$

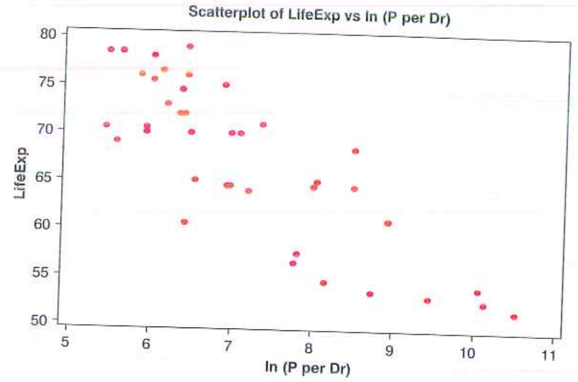
4 a. The MINITAB output is shown below:

```
The regression equation is
Y = 943 + 8.71 X

Predictor     Coef     SE Coef      T       P
Constant     943.05     59.38     15.88   0.000
X              8.714     1.544     5.64   0.005

S = 32.29   R-sq = 88.8%   R-sq(adj) = 86.1%

Analysis of Variance

SOURCE           DF     SS      MS       F       P
Regression        1   33 223  33 223  31.86   0.005
Residual Error    4    4172    1043
Total             5   37 395
```

b. The p-value of $0.005 < \alpha = 0.01$; reject H_0

6 a. The scatter diagram for LifeExp against People per Dr suggests the existence of a possible nonlinear relationship between the two variables:

b. However when the People per Dr variable is replaced by its logarithm in the scatter diagram, a linear model now seems plausible:



c. The situation is exactly analogous for the scatter diagram of LifeExp with People per TV variables.

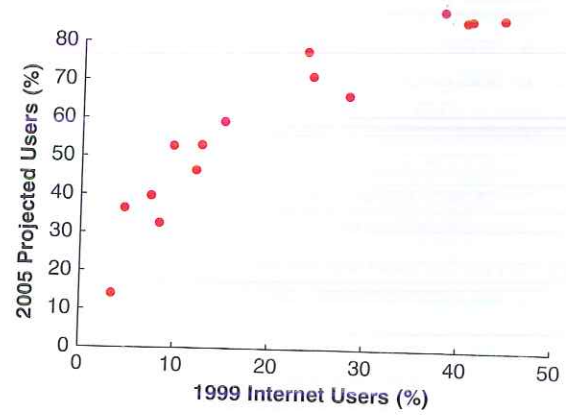
Correspondingly we have the two simple regression models:

$\widehat{\text{LifeExp}} = 77.887 - 4.26 \ln(\text{P per TV})$ $R^2 = 0.731$

$\widehat{\text{LifeExp}} = 102.873 - 4.974 \ln(\text{P per Dr})$ $R^2 = 0.693$

Neither of these relationships is causal but the first with the ln(P per TV) predictor has a slightly better R^2 value which might favour it in this instance.

8 a. The scatter diagram is shown below:



STUDENTS-HUB.com Uploaded By: anonymous