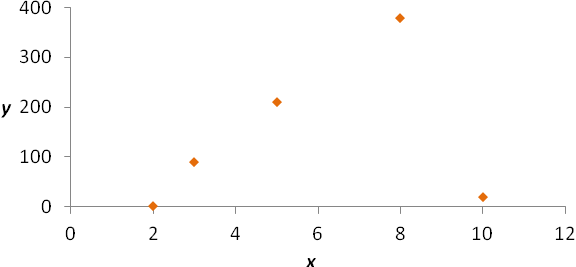
Chapter 14 Regression Analysis

Solutions



The scatterplot shows that *x* and y appear to have a positive relationship, with the exception of one possible outlier at (10, 20).



|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

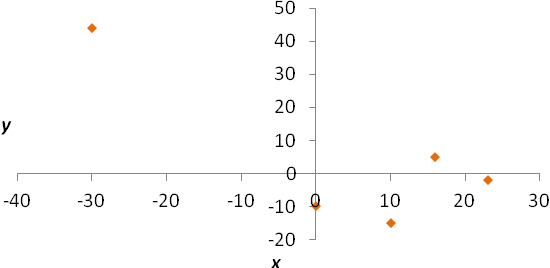
|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |



The covariance of 158.2 indicates that *x* and *y* have a positive linear relationship.



The correlation coefficient of 0.30 indicates that the strength of the positive linear relationship is moderate.



The scatterplot shows that *x* and *y* appear to have a negative linear relationship.



|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |



The covariance of ‒379.9 indicates that *x* and y have a negative linear relationship.



The correlation coefficient of ‒0.78 indicates that the strength of the negative linear relationship is strong.



* 1. The value of the test statistic, .



* 1. With Thus, we reject *H*0 if or



* 1. Since , we do not reject *H*0. At the 5% significance level, we cannot conclude that the population correlation coefficient between *x* and *y* is significantly different from zero. This means we cannot comment on the direction of the relationship between *x* and *y* based on .



* 1. The value of the test statistic,



* 1. For , the *p*-value = < 0.005.



* 1. At *p*-value < , so we reject *H*­0. At the 5% significance level, we conclude that the population correlation coefficient between *x* and *y* is significantly different from zero. Therefore, we can say that there is a significantly negative linear relationship between *x* and *y*.



* 1. . The correlation coefficient indicates a positive linear relationship.



The value of the test statistic,



The *p*-value, is between 0.05 and 0.10.



* 1. Since with the *p*-value > , we do not reject *H*0. At the 5% significance level, we cannot conclude that the population correlation coefficient between *x* and *y* is positive.



1. 1. . The correlation coefficient indicates a slightly negative linear relationship.







The value of the test statistic,



With Thus, we reject HO if > 2.069 or <‒2.069Since 2.069 < 0.85 < 2.069, we do not reject *H*0. At the 5% significance level, we cannot conclude that the population correlation coefficient is different from zero. This means we cannot comment on the direction of the relationship between *x* and *y* based on .







|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
| .  .  . | .  .  . | .  .  . |
|  |  |  |
|  |  |  |



The covariance of ‒35.15 indicates that weather and crime have a negative linear relationship. (As the number of rainy days increases, the number of shootings tends to decrease).



The correlation coefficient of ‒0.49 indicates that the strength of the negative linear relationship is moderate.



The value of the test statistic,



Here the *p*-value, , is more than 0.20.



* 1. Since *p*-value > α, we do not reject *H*­0. At the 5% significance level, we cannot conclude that the population correlation coefficient is negative. Therefore, based on the sample information, we cannot say that dreary weather and crime are negatively correlated.

1. 1. Let M = Microsoft, C = Coca Cola, B = Bank of America, and G = General Electric;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | M | C | B | G |
| Average | 26.32 | 52.81 | 14.39 | 16.17 |
| Standard deviation | 2.26 | 3.54 | 2.29 | 1.38 |

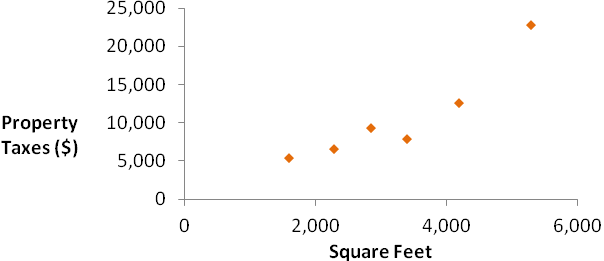
Finding the covariance between each pair, and then the correlation:

|  |  |  |  |
| --- | --- | --- | --- |
| Pair: | MC | MB | MG |
| Covariance | 4.52 | 3.24 | 2.56 |
| Correlation |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| Pair: | CB | CG | BG |
| Covariance | 0.36 | 3.06 | 1.45 |

|  |  |  |  |
| --- | --- | --- | --- |
| Correlation |  |  |  |

* 1. The correlation between Microsoft and Coca Cola is the lowest between each pair with Microsoft, with a correlation of 0.57. Therefore, the Coca Cola asset will give the maximum benefit of diversification with Microsoft.
  2. The lowest correlation between any pair of assets is with Coca Cola and Bank of America, with a correlation of 0.04. Therefore, Coca Cola and Bank of America will give the maximum benefit of diversification between any of the above pairs.



The scatterplot indicates that there is a positive relationship between size of house and property taxes.

b.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
| .  .  . | .  .  . | .  .  . |
|  |  |  |

|  |  |  |
| --- | --- | --- |
|  |  |  |



The covariance indicates that house size and property taxes have a positive linear relationship.



The correlation coefficient of 0.92 indicates that the strength of the positive linear relationship is very strong.



The value of the test statistic,



For , the *p*-value for the two-tailed test is 0.005, the *p*-value is less than 2 0.005, that is, less than 0.01.



* 1. At α = 0.05, the *p*-value < α, so we reject *H*0. At the 5% significance level, we conclude that the population correlation between house size and property taxes is significantly different from zero.



|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |

|  |  |  |
| --- | --- | --- |
| .  .  . | .  .  . | .  .  . |
|  |  |  |
|  |  |  |



The correlation coefficient of 0.57 indicates a moderate relationship between age and happiness.



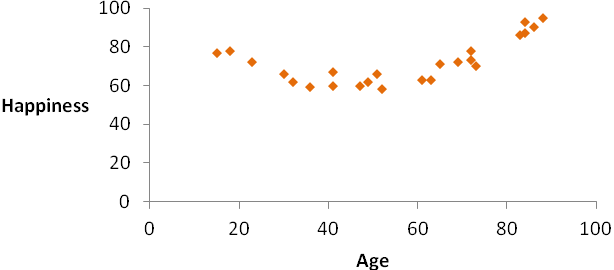
The value of the test statistic,



For , the *p*-value for a two-tailed test is Thus, the *p*-value is less than 2 0.005, that is, less than 0.01. At α = 0.01, the *p*-value < α, so we reject *H*0­. At the 1% significance level, the correlation coefficient is statistically significant.







Although there does appear to be a relationship between age and happiness, the relationship is not linear. This conflicts with the above conclusion that the linear relationship between age and happiness is significant.



|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
| .  .  . | .  .  . | .  .  . |
|  |  |  |
|  |  |  |



The correlation coefficient of 0.23 indicates a positive relationship between PPG and MPG.







The value of the test statistic,



For *,* the *p*-value =is more than 0.20.



* 1. At α = 0.05, the *p*-value > α, so we do not reject *H*0­. At the 5% significance level, we cannot conclude that the correlation between PPG and MPG is significantly positive. This is surprising, as one typically thinks that the longer a player plays in the game, the more points he is likely to score.



* 1. With



If equals 40, the predicted -value is:







* 1. With



If equals ―20, the predicted -value is:







* 1. With



. If equals 25, the predicted -value is:







* 1. . If equals ―5, the predicted -value is:







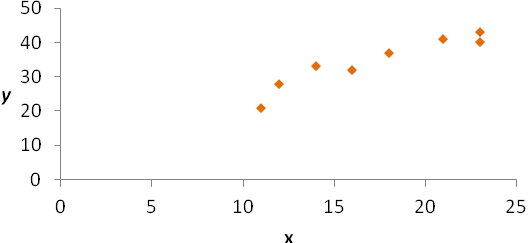
* 1. If doubles, When doubled from 10 to 20, increased by 25 (65 ‒ 40).



1. 1. The estimated slope coefficient equals ‒17. Thus, as increases by 1 unit, is predicted to *decrease* by 17 units.







There appears to be a linear relationship between x and y; a linear regression model is appropriate.



|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| .  .  . | .  .  . | .  .  . | .  .  . |
|  |  |  |  |
|  |  |  |  |



The sample regression equation is,



* 1. If = 10,



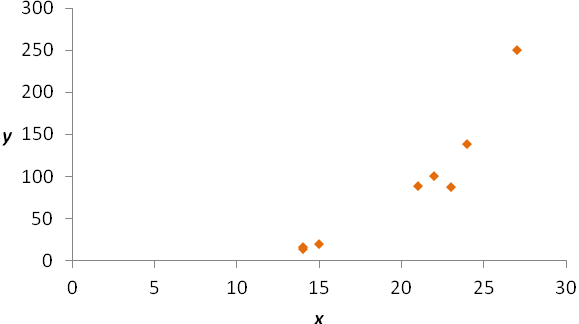
If = 15,



If = 20,







There appears to be a positive relationship between *x* and *y*, so estimating a simple linear regression is appropriate.



|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| .  .  . | .  .  . | .  .  . | .  .  . |
|  |  |  |  |
|  |  |  |  |



The sample regression equation is,



* 1. If = 20,



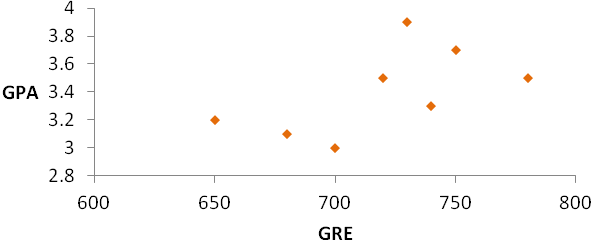
If = 100,



If = 200,











|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| .  .  . | .  .  . | .  .  . | .  .  . |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |



The sample regression equation is,



* 1. The predicted GPA for a GRE score of 710 is .







|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| .  .  . | .  .  . | .  .  . | .  .  . |
|  |  |  |  |
|  |  |  |  |



The sample regression equation is,



* 1. As Education increases by 1 unit (1 additional year of higher education), an individual’s annual salary (measured in thousands of dollars) is predicted to increase by 9.38, or by $9,380.
  2. , or $88,510.



* 1. Choose Data>Data Analysis>Regression from the menu. Select the Consumption data as your Input *y* range, and Disposable Income as your Input *x* range. Click OK.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.9949 |  |  |  |  |  |
| R Square | 0.9899 |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Adj. R Square | 0.9894 |  |  |  |  |  |
| Std. Error | 742.859 |  |  |  |  |  |
| Observations | 22 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *Df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 1 | 1.08E+09 | 1.08E+09 | 1965.385 | 1.9E‒21 |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Residual | 20 | 11036796 | 551839.8 |  |  |  |
| Total | 21 | 1.1E+09 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *p-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 8550.675 | 593.92 | 14.3969 | 5.12E‒12 | 7311.77 | 9789.58 |
| disposable income | 0.6860 | 0.0155 | 44.3327 | 1.9E‒21 | 0.6538 | 0.7183 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |



* 1. As disposable income increases by $1, consumption is predicted to increase by $0.686. In other words, the marginal propensity to consume is $0.686 per $1 of disposable income.
  2. .



1. 1. Excel Output:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.1530 |  |  |  |  |  |
| R Square | 0.0234 |  |  |  |  |  |
| Adj. R Square | ‒0.0987 |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Std. Error | 5.4443 |  |  |  |  |  |
| Observations | 10.0 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 1 | 5.69 | 5.69 | 0.19 | 0.67 |  |
| Residual | 8 | 237.12 | 29.64 |  |  |  |
| Total | 9 | 242.81 |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 18.433989 | 24.5467 | 0.7510 | 0.4742 | ‒38.1707 | 75.0387 |
| ERA | ‒3.821807 | 8.7264 | ‒0.4380 | 0.6730 | ‒23.9449 | 16.3012 |
|  |  |  |  |  |  |  |



For a one-unit increase in ERA, predicted salary *decreases* by $3.82 million.



|  |  |  |
| --- | --- | --- |
|  | ERA | Predicted Salary (in $ millions) |
| J. Santana | 2.53 |  |
| C. Lee | 2.54 |  |
| T. Lincecum | 2.62 |  |
| C. Sabathia | 2.70 |  |
| R. Halladay | 2.78 |  |
| J. Peavy | 2.85 |  |
| D. Matsuzaka | 2.90 |  |
| R. Dempster | 2.96 |  |
| B. Sheets | 3.09 |  |
| C. Hamels | 3.09 |  |



|  |  |  |  |
| --- | --- | --- | --- |
|  | Predicted Salary | Actual Salary | Residual |
| J. Santana | 8.76 | 17.0 | 8.24 |
| C. Lee | 8.73 | 4.0 | ‒4.73 |
| T. Lincecum | 8.42 | 0.4 | ‒8.02 |
| C. Sabathia | 8.12 | 11.0 | 2.88 |
| R. Halladay | 7.81 | 10.0 | 2.19 |
| J. Peavy | 7.54 | 6.5 | ‒1.04 |
| D. Matsuzaka | 7.35 | 8.3 | 0.95 |
| R. Dempster | 7.12 | 7.3 | 0.18 |
| B. Sheets | 6.63 | 12.1 | 5.48 |

|  |  |  |  |
| --- | --- | --- | --- |
| C. Hamels | 6.63 | 0.5 | ‒6.12 |

There might be factors other than ERA that contribute to the pitchers salaries. This model only considers ERA and, given high residuals, is not a very good predictor of salary.



Excel Output:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.5716 |  |  |  |  |  |
| R Square | 0.3267 |  |  |  |  |  |
| Adj. R Square | 0.2961 |  |  |  |  |  |
| Std. Error | 9.4696 |  |  |  |  |  |
| Observations | 24 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 1 | 957.19 | 957.19 | 10.67 | 0.0035 |  |
| Residual | 22 | 1972.81 | 89.67 |  |  |  |
| Total | 23 | 2930 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 56.1772 | 5.2145 | 10.7732 | 3.06E‒10 | 45.3630 | 66.9914 |
| Age | 0.2844 | 0.0871 | 3.2671 | 0.0035 | 0.1039 | 0.4650 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |

The estimated model is



* 1. When



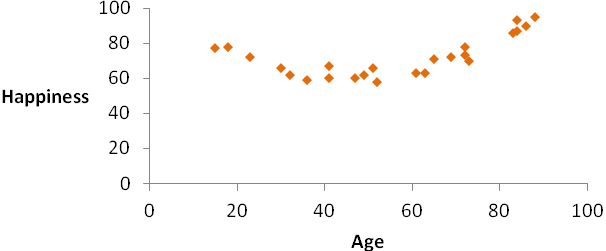
When



When







According to the scatterplot, happiness seems to start higher, be lowest between ages 40 – 50, and then increase. In other words, the relationship between Age and Happiness in not linear. However, our linear model predicts a person aged 50 to be happier than a person aged 25.Therefore, our predictions based on our linear model are not accurate.

* 1. If



* 1. As increases by one unit, is predicted to increase by 12.9 units, holding constant.



* 1. If



* 1. As increases by one unit, is predicted to decrease by 47.2 units, holding constant.



1. 1. Excel output:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.9291 |  |  |  |  |  |
| R Square | 0.8632 |  |  |  |  |  |
| Adjusted R Square | 0.8085 |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Standard Error | 4.6868 |  |  |  |  |  |
| Observations | 8.0 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 2 | 693.05 | 346.52 | 15.78 | 0.01 |  |
| Residual | 5 | 109.83 | 21.97 |  |  |  |
| Total | 7 | 802.88 |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 22.8074 | 6.8384 | 3.3352 | 0.0207 | 5.2288 | 40.3861 |
| x1 | 0.8460 | 0.1523 | 5.5546 | 0.0026 | 0.4545 | 1.2376 |
| x2 | ‒0.7053 | 0.2451 | ‒2.8773 | 0.0347 | ‒1.3353 | ‒0.0752 |
|  |  |  |  |  |  |  |

Using the estimates, we derive the sample regression as .



The regression coefficient of is 0.85, suggesting that as increases by one unit, is predicted to increase by 0.85 units, holding constant. Likewise, the regression coefficient of is ‒0.71, suggesting that as increases by one unit, is predicted to decrease by 0.71 units, holding constant.



* 1. If



1. 1. Excel output:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.9432 |  |  |  |  |  |
| R Square | 0.8897 |  |  |  |  |  |
| Adj. R Square | 0.8455 |  |  |  |  |  |
| Standard Error | 2.6816 |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Observations | 8 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *Df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 2 | 289.92 | 144.96 | 20.16 | 0.00 |  |
| Residual | 5 | 35.96 | 7.19 |  |  |  |
| Total | 7 | 325.88 |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 13.8256 | 6.5880 | 2.0986 | 0.0899 | ‒3.1095 | 30.7606 |
| x1 | 2.5263 | 0.4013 | 6.2951 | 0.0015 | 1.4947 | 3.5580 |
| x2 | 0.2946 | 0.1752 | 1.6820 | 0.1534 | ‒0.1556 | 0.7449 |
|  |  |  |  |  |  |  |

Using the estimates, we derive the sample regression as .



The regression coefficient of is 0.29, suggesting that as increases by one unit, is predicted to increase by 0.29 units, holding constant.



* 1. If



1. 1. Using the estimated coefficients, the sample regression is .



* 1. The regression coefficient of 8.47 suggests that as the number of women over the age of 60 increases by 1 million, sales of StrongBones is predicted to increase by $8.47 million, holding income constant. Similarly, the slope coefficient of 7.62 suggests that as the average income of women over the age of 60 increases by $1000, sales of StrongBones is predicted to increase by $7.62 million, holding population constant.
  2. If population = 1.5 million and income = 44 thousand, then ($ million).



* 1. The coefficient for Income is not as expected, since according to the sociologist’s hypothesis we would expect it to have a negative sign. The positive sign for the Poverty coefficient is as expected.
  2. The coefficient of 53.16 suggests that as poverty increases by one percent, the crime rate is predicted to rise by 53.16 crimes per 100,000 residents, holding income constant.
  3. If Poverty = 20% and Income = $50 thousand, then (crimes per 100,000 residents).



* 1. Choose Data>Data Analysis>Regression from the menu. Select the Price data as your Input Y range, and simultaneously choose the Sqft, Beds, *and* Baths columns as your Input X range. Click OK. Excel Output:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.8507 |  |  |  |  |  |
| R Square | 0.7237 |  |  |  |  |  |
| Adj. R Square | 0.6978 |  |  |  |  |  |
| Std. Error | 74984.98 |  |  |  |  |  |
| Observations | 36 |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |  |
|  | *Df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 3 | 5E+11 | 2E+11 | 3E+01 | 5E‒09 |  |
| Residual | 32 | 2E+11 | 6E+09 |  |  |  |
| Total | 35 | 7E+11 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 153348.27 | 57141.79 | 2.6836 | 0.0114 | 36954.2414 | 269742.2914 |
| Sqft | 95.86 | 35.40 | 2.7078 | 0.0108 | 23.7490 | 167.9629 |
| Beds | 556.89 | 20280.31 | 0.0275 | 0.9783 | 40752.7546 | 41866.5360 |
| Baths | 92022.91 | 25012.30 | 3.6791 | 0.0009 | 41074.5297 | 142971.2955 |
|  |  |  |  |  |  |  |

* 1. The slope coefficient of 95.86 suggests that for every additional square foot, the predicted price of a home increases by $95.86, holding number of bedrooms and bathrooms constant.

The slope coefficient of 556.89 suggests that for every additional bedroom, the predicted price of a home increases by $556.89, holding square footage and number of baths constant.

The slope coefficient of 92022.91 suggests that for every additional bathroom, the predicted price of a home increases by $92,022.91, holding square footage and number of bedrooms constant.

* 1. If sqft = 2500, beds = 3, and baths = 2;



* 1. As the student to teacher ratio (STR) increases, the teacher is not able to spend as much time with each student, so this will likely have a negative effect on score.

As the teacher’s salary (TSAL) increases, the teacher is likely to be more motivated (and experienced) and thus may do a better job teaching the students, to this is likely to have a positive effect on score.

As the median household income (INC) increases, parents are likely to be well educated who instill the value of education on their children. This is likely to have a positive effect on score.

As the number of single family households (SGL) increases, children tend to have less stable home lives, more chores, or parents who are busier and cannot help with homework as much. This is likely to have a negative effect on score.

* 1. Choose Data>Data Analysis>Regression from the menu. Select the Score data as your Input Y range, and simultaneously choose the STR, TSAL, INC, *and* SGL columns as your Input X range. Click OK. Excel Output:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.7858 |  |  |  |  |  |
| R Square | 0.6174 |  |  |  |  |  |
| Adj. R Square | 0.6105 |  |  |  |  |  |
| Std. Error | 4.3042 |  |  |  |  |  |
| Observations | 224 |  |  |  |  |  |
|  |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 4 | 6548.49 | 1637.12 | 88.37 | 0.00 |  |
| Residual | 219 | 4057.28 | 18.53 |  |  |  |
| Total | 223 | 10605.77 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Intercept | 231.89 | 3.37 | 68.8201 | 0.0000 | 225.2535 | 238.5354 |
| STR | ‒0.50 | 0.13 | ‒3.7944 | 0.0002 | ‒0.7538 | ‒0.2384 |
| TSAL | ‒0.02 | 0.07 | ‒0.3129 | 0.7547 | ‒0.1704 | 0.1237 |
| INC | 0.29 | 0.03 | 8.5093 | 0.0000 | 0.2250 | 0.3607 |
| SGL | ‒0.88 | 0.17 | ‒5.0556 | 0.0000 | ‒1.2203 | ‒0.5357 |
|  |  |  |  |  |  |  |

Using the estimates, , we derive the sample regression as:



.



The signs on the coefficients are all as expected except for TSAL, which has a slightly negative effect on score, holding all other variables constant.

* 1. .



* 1. When income increases by 20, holding everything else the same, the predicted score will increase by points to 240.69.



* 1. Choose Data>Data Analysis>Regression from the menu. Select the Salary data as your Input Y range, and simultaneously choose the PCT, TD, and Age columns as your Input X range. Click OK. Excel Output:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.6270 |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| R Square | 0.3931 |  |  |  |  |  |
| Adj. R Square | 0.3280 |  |  |  |  |  |
| Std. Error | 5.4989 |  |  |  |  |  |
| Observations | 32 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 3 | 548.35 | 182.78 | 6.04 | 0.00 |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Residual | 28 | 846.68 | 30.24 |  |  |  |
| Total | 31 | 1395.02 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 32.7867 | 20.73 | 1.5815 | 0.1250 | ‒9.6788 | 75.2522 |
| PCT | ‒0.8265 | 0.44 | 1.8855 | 0.0698 | ‒1.7244 | 0.0714 |
| TD | 0.7850 | 0.25 | 3.1491 | 0.0039 | 0.2744 | 1.2956 |
| Age | 0.3862 | 0.25 | 1.5248 | 0.1385 | ‒0.1326 | 0.9050 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |

Using the estimates, we derive the sample regression as:



.



* 1. Yes, the estimated coefficient of ‒0.83 is surprising. This coefficient suggests that as a quarterback’s pass completion percentage increases by 1 percent, his salary decreases by $0.83 million. This is surprising.



* 1. Drew Brees residual =



Tom Brady residual =



The estimated model implies that Drew Brees is slightly overpaid whereas Tom Brady is grossly underpaid. Perhaps there are other important factors, that impact salary, which were not included in the model.



.



* 1. Given SSE = 1250 and SST = 1500, we compute .



.



* 1. Given SSE = 2540 and SST = 13,870, we compute .



* 1. .



* 1. Given SSE = 35 and SST = 90, we compute .



* 1. where Thus, .



* 1. (Note, we can also compute it as which is the same answer as above.)



* 1. .



* 1. In order to determine if the model seems promising, we take Since the ratio is less than 0.20, the model seems promising.



* 1. Therefore, 46.18% of the sample variation in *y* is explained by the estimated regression model.



* 1. From the ANOVA table we get *k* = 2 and *n* = 29+1 = 30. Thus,



1. 1. In order to determine if the model seems promising, we take Since the ratio is less than 0.20, the model seems promising.



* 1. Therefore, 80.55% of the sample variation in *y* is explained by the estimated regression model.



* 1. From the ANOVA table we get *k* = 2 and *n* = 19+1 = 20. Thus, a



1. 1. Model 1:



Model 2:



* 1. Model 1:



Model 2:



* 1. Model 1 provides a better fit. It has a lower standard error of the estimate and a higher coefficient of determination.



Since the ratio is greater than 0.20, this model is not very promising.



* 1. The proportion that is explained is The proportion that is unexplained is







Since the ratio is much greater than 0.20, the model is highly unpromising.



* 1. From the ANOVA table we get *k* = 2 and *n* = 29+1 = 30. Thus, a



1. This table reports the regression results from each model:

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Model 1 | Model 2 | Model 3 |
| Intercept | ‒79.0342 | 151.0092 | ‒30.6198 |
| Yards Made | 0.3860 | NA | 0.3501 |
| Yards Allowed | NA | ‒0.3003 | ‒0.1085 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
|  | 13.0472 | 18.0975 | 12.8867 |
|  | 0.5931 | 0.2171 | 0.6163 |
|  | 0.5795 | 0.1910 | 0.5898 |

Notes: Parameter estimates are in the top half of the table. The lower part of the table contains goodness-of-fit measures

* 1. Model 1 has a lower standard error of the estimate and a higher ; thus Model 1 appears to be the better model for prediction.



* 1. Since the number of explanatory variables, is not the same between Model 3 and Models 1 & 2, we must use Adjusted to compare these models. Model 3 has an Adjusted of 0.5898, compared to 0.5795 and 0.1910 for Model 1 and Model 2, respectively. Therefore, with the highest Adjusted Model 3 is an improvement over the other two models. Model 3 also has the lowest standard error.







|  |  |  |  |
| --- | --- | --- | --- |
| Explanatory Variables | Adj ROA  (Model 1) | Adj Return  (Model 2) | Total Assets  (Model 3) |
|  | 9.79 | 9.78 | 9.00 |
|  | 0.0004 | 0.0024 | 0.1560 |

Since Model 3 has the lowest and the highest , Model 3 is the best model.







|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Explanatory Variables | Adj. ROA & Adj Return | Adj. ROA & Total Assets | Adj. Return & Total Assets | All 3 Variables |
|  | 9.79 | 8.99 | 8.99 | 8.98 |
| Adjusted | ‒0.0018 | 0.1567 | 0.1567 | 0.1585 |

Based on the highest Adjusted , and lowest the model using all 3 explanatory variables is the best.



* 1. . The correlation coefficient of 0.94 indicates a strong positive relationship.







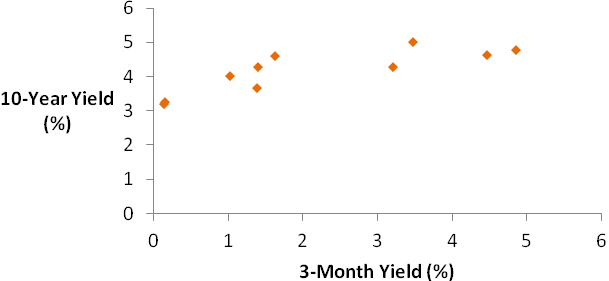
The value of the test statistic,



Thus, we reject *H*0 if Since 4.77 > 3.182, we reject *H*0. At the 5% significance level, we conclude that the returns on the mutual funds are correlated.







The scatterplot suggests a positive correlation between the two yields.



|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
| .  .  . | .  .  . | .  .  . | .  .  . |
|  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
|  |  |  |  |



(slightly different from Excel's estimate of 0.83, due to rounding)



To test if the population correlation is statistically significant, set up the hypotheses:



The value of the test statistic



With Thus, we reject *H*0 if >2.306 or Since 3.91 > 2.306, we reject *H*0. At the 5% significance level, we can conclude that the correlation is significantly different from zero.







The sample regression equation is







|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.248689 |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| R Square | 0.061846 |  |  |  |  |  |
| Adjusted R Square | 0.0427 |  |  |  |  |  |
| Standard Error | 5.768515 |  |  |  |  |  |
| Observations | 51 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 1 | 107.4888 | 107.4888 | 3.230243 | 0.078454 |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Residual | 49 | 1630.513 | 33.27577 |  |  |  |
| Total | 50 | 1738.002 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 78.9791 | 5.4855 | 14.3977 | 0.0000 | 67.9555 | 90.0027 |
| Income | ‒0.0002 | 0.0001 | ‒1.7973 | 0.0785 | ‒0.0004 | 0.0000 |

Using the excel output:

,



The sample regression equation is



For a $1,000 increase in income, homeownership rate is predicted to decrease by 2%. This negative relationship is surprising. Note, however, that this model ignores home affordability. For example, homes are very affordable in Alabama and as a result its homeownership rate is 74.1% despite a low median household

income of $39,980. By contrast, California has only a 57% homeownership rate even thought its median household income is $56,134.

* 1. (from above)



Since the ratio is less than 0.20, this model is promising.



* 1. This means that 6.18% of the sample variation in is explained by the estimated regression equation.







|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.6346 |  |  |  |  |  |
| R Square | 0.4028 |  |  |  |  |  |
| Adj. R Square | 0.3585 |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Standard Error | 13.6377 |  |  |  |  |  |
| Observations | 30 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *Df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 2 | 3386.46 | 1693.23 | 9.10 | 0.00 |  |
| Residual | 27 | 5021.63 | 185.99 |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total | 29 | 8408.09 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | ‒33.3966 | 12.7798 | ‒2.6132 | 0.0145 | ‒59.6185 | ‒7.1747 |
| P/E | 3.9674 | 0.9587 | 4.1381 | 0.0003 | 2.0002 | 5.9346 |
| P/S | ‒3.3681 | 2.6294 | ‒1.2809 | 0.2111 | ‒8.7632 | 2.0271 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |

The signs are as expected. As P/E increases, the predicted returns increase, and as P/S decreases, the predicted returns increase.

* 1. As the P/S ratio increases by 1 unit, the predicted return of the firm decreases by 3.37%, holding P/E constant.
  2. %.



* 1. (from above)



Since the ratio is greater than 0.20, this model is not very promising.



* 1. This means that 40.28% of the sample variation in is explained by the sample regression equation.



* 1. This table reports the regression results from each model:

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Model 1 | Model 2 | Model 3 |
| Intercept | 1616.3627 | 1259.6376 | 1104.2580 |
| Income | 0.0015 | NA | 0.0017 |
| GPA | NA | 141.4680 | 150.9920 |

|  |  |  |  |
| --- | --- | --- | --- |
|  | | | |
|  | 76.2217 | 56.6262 | 32.4902 |
|  | 0.2211 | 0.5701 | 0.8649 |
|  | 0.1857 | 0.5506 | 0.8520 |

Notes: Parameter estimates are in the top half of the table. The lower part of the table contains goodness-of-fit measures

Using the excel output, the three models are:

Model 1:



Model 2:



Model 3:



* 1. Since the models have different numbers of explanatory variables, we use Adjusted to compare models. Since Model 3 has the highest Adjusted , Model 3 is the best fitting model. Model 3 also has the lowest standard error of the estimate.



* 1. We compute mean value of the explanatory variables as and . Using Model 3 (the best model), .







|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.8466 |  |  |  |  |  |
| R Square | 0.7168 |  |  |  |  |  |
| Adj. R Square | 0.7106 |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Standard Error | 3.1373 |  |  |  |  |  |
| Observations | 143 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 3 | 3462.07 | 1154.02 | 117.25 | 0.00 |  |
| Residual | 139 | 1368.14 | 9.84 |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total | 142 | 4830.21 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 0.4190 | 0.5124 | 0.8178 | 0.4149 | ‒0.5940 | 1.4321 |
| Research | 0.0087 | 0.0012 | 7.1332 | 0.0000 | 0.0063 | 0.0111 |
| Patents | 0.0517 | 0.0199 | 2.5967 | 0.0104 | 0.0123 | 0.0910 |
| Duration | ‒0.0194 | 0.0236 | ‒0.8238 | 0.4115 | ‒0.0661 | 0.0272 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |

Using the excel output, the estimated model is



* 1. startups



* 1. A $1 million increase in research expenditure results in a predicted increase in the number of startups by 0.0087 holding everything else constant. Thus, approximately $114.94 million () in additional research



expenditures would be needed to have 1 additional predicted startup, everything else being the same. Note that $114.94 × 0.0087 equals (approximately) 1.



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SUMMARY OUTPUT | |  |  |  |  |  |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Multiple R | 0.5851 |  |  |  |  |  |
| R Square | 0.3424 |  |  |  |  |  |
| Adj. R Square | 0.2995 |  |  |  |  |  |
| Standard Error | 5.4126 |  |  |  |  |  |
| Observations | 50 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *df* | *SS* | *MS* | *F* | *Significance F* |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Regression | 3 | 701.58 | 233.86 | 7.98 | 0.00 |  |
| Residual | 46 | 1347.60 | 29.30 |  |  |  |
| Total | 49 | 2049.18 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 7.8730 | 4.0872 | 1.9263 | 0.0603 | ‒0.3541 | 16.1001 |
| EDUC | 1.4371 | 0.3386 | 4.2436 | 0.0001 | 0.7554 | 2.1187 |
| EXPER | 0.4483 | 0.1419 | 3.1599 | 0.0028 | 0.1627 | 0.7338 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| AGE | ‒0.0114 | 0.0834 | ‒0.1365 | 0.8920 | ‒0.1793 | 0.1565 |
|  |  |  |  |  |  |  |

Using the excel output, the estimated regression is:



* 1. The signs on the coefficients for education and experience are as expected; an increase in education and/or experience are associated with an increase in wage, holding everything else fixed. The negative coefficient for age is somewhat surprising.
  2. For a 1 year increase in higher education, wage is expected to increase by $1.44/hour, holding experience and age constant.
  3. meaning that 34.24% of the sample variation in wage is explained by the estimated regression model.



* 1. /hour.



**Case Study 14.1**

1. The sample correlations coefficients of each fund with Magellan are:

|  |  |  |  |
| --- | --- | --- | --- |
| M & Total | M & Short-Term | M & Int'l | M & Freedom |
| 0.460449 | 0.389657 | 0.85557 | 0.928237 |

The highest sample correlation coefficient is with the Freedom Income fund and the lowest correlation is with the Short-Term Bond fund.

1. To test if each correlation coefficient is statistically significant, set up the hypotheses which will be the same for each test:



Using a 5% significance level, . Thus, we reject HO if



M & Total:



Since 3.95 > 2, we reject H0 and conclude that the correlation between the Magellan and Total Bond Fund is statistically significant.

M & Short-Term:



Since 3.23 > 2, we reject H0 and conclude that the correlation between the Magellan and the Short-Term Bond Fund is statistically significant.

M & Int’l:



Since 12.83 > 2, we reject H0 and conclude that the correlation between the Magellan and the International Small Cap Fund is statistically significant.

M & Freedom:



Since 19.27 > 2, we reject HO. At the 5% significance level, we conclude that the correlation between the Magellan and the Freedom Income Fund is statistically significant.

1. At the 5% significance level, each correlation coefficient is statistically significant. Further, since the correlation between the Magellan Fund and the Short-Term Bond Fund is the lowest, the recommendation should be to invest in the Short-Term Bond Fund.

**Case Study 14.2**

1. The Excel output for the simple and multiple regression models are given below:

Simple Linear Model:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY OUTPUT | |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.6234 |  |  |  |  |  |
| R Square | 0.3887 |  |  |  |  |  |
| Adj. R Square | 0.3479 |  |  |  |  |  |
| Standard Error | 2.3078 |  |  |  |  |  |
| Observations | 17 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | *Df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 1 | 50.79 | 50.79 | 9.54 | 0.01 |  |
| Residual | 15 | 79.89 | 5.33 |  |  |  |
| Total | 16 | 130.68 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 15.2607 | 4.3746 | 3.4885 | 0.0033 | 5.9365 | 24.5849 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| AdsCost | 0.0230 | 0.0075 | 3.0881 | 0.0075 | 0.0071 | 0.0389 |
|  |  |  |  |  |  |  |

Multiple Regression Model:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY OUTPUT | |  |  |  |  |  |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Multiple R | 0.7455 |  |  |  |  |  |
| R Square | 0.5558 |  |  |  |  |  |
| Adj. R Square | 0.4924 |  |  |  |  |  |
| Standard Error | 2.0362 |  |  |  |  |  |
| Observations | 17 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *Df* | *SS* | *MS* | *F* | *Significance F* |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Regression | 2 | 72.64 | 36.32 | 8.76 | 0.00 |  |
| Residual | 14 | 58.04 | 4.15 |  |  |  |
| Total | 16 | 130.68 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 17.5060 | 3.9817 | 4.3966 | 0.0006 | 8.9661 | 26.0458 |
| AdsCost | 0.0266 | 0.0068 | 3.9323 | 0.0015 | 0.0121 | 0.0410 |
| Unemp | ‒0.6879 | 0.2997 | ‒2.2955 | 0.0377 | ‒1.3306 | ‒0.0452 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |

Using the Excel outputs, the estimated models are:



1. Since the multiple regression model has a higher adjusted R2 than the simple linear model (0.4924 > 0.3479), the multiple regression model is more appropriate. Note that the multiple regression model also has a lower standard error of the estimate (2.0362 < 2.3078).
2. Predicted sales with 6% unemployment and $400 advertising costs is:

thousand, or $24,019.



With the same unemployment rate and $600 in advertising costs:

thousand, or $29,339.



Note that the difference in the predicted value is $29,339 ― $24,019 = $$5,320, which is 0.0266 ×200 .

**Case Study 14.3**

1. The mean and the standard deviation of each of the variable is shown below.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Price ($1000) | SQFT (1000) | Bed | Bath | LTSZ |
| Average | 208.79 | 2.60 | 3.56 | 2.49 | 0.82 |
| Standard Dev. | 79.64 | 1.02 | 0.67 | 0.75 | 1.37 |

1. The Excel regression output is shown below.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |
| SUMMARY OUTPUT | |  |  |  |  |  |
|  |  |  |  |  |  |  |
| *Regression Statistics* | |  |  |  |  |  |
| Multiple R | 0.9194 |  |  |  |  |  |
| R Square | 0.8453 |  |  |  |  |  |
| Adj. R Square | 0.8315 |  |  |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Standard Error | 32.6850 |  |  |  |  |  |
| Observations | 50 |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
|  | *Df* | *SS* | *MS* | *F* | *Significance F* |  |
| Regression | 4 | 2.63E+11 | 6.57E+10 | 61.47143 | 1.16E‒17 |  |
| Residual | 45 | 4.81E+10 | 1.07E+09 |  |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Total | 49 | 3.11E+11 |  |  |  |  |
|  |  |  |  |  |  |  |
|  | *Coefficients* | *Standard Error* | *t Stat* | *P-value* | *Lower 95%* | *Upper 95%* |
| Intercept | 23714.63 | 25435.27 | 0.932352 | 0.35613 | ‒27514.6 | 74943.88 |
| SQFT | 44.97168 | 6.262365 | 7.18126 | 5.49E‒09 | 32.35863 | 57.58473 |
| Bed | ‒5028.72 | 7921.085 | ‒0.63485 | 0.52874 | ‒20982.6 | 10925.17 |
| Bath | 26142.43 | 8917.572 | 2.931564 | 0.005285 | 8181.52 | 44103.34 |
| LTSZ | 25725.12 | 3437.085 | 7.484576 | 1.96E‒09 | 18802.48 | 32647.77 |

The estimated model is:



1. The *R2* of 0.8453 means that 84.53% of the sample variation in asking price of houses is explained by the estimated regression model. The remaining 15.47% of the sample variation is not explained by the estimated model.