# Chapter 7: Outline Solutions

Minitab and SPSS have been was used in the calculations except where noted.

7.1

a. The regression equation is

Sales = 8.67 + 2.03 Spots

Predictor Coef SE Coef T P

Constant 8.671 3.662 2.37 0.056

Spots 2.0286 0.3084 6.58 0.001

S = 2.58014 R-Sq = 87.8% R-Sq(adj) = 85.8%

Analysis of Variance

Source DF SS MS F P

Regression 1 288.06 288.06 43.27 0.001

Residual Error 6 39.94 6.66

Total 7 328.00

b. The P-value for the slope is 0.001, so we can reject in favor of the alternative  In this case, the slope is clearly positive so that an increase in the number of spots increases the expected level of sales.

c. S = 2.58 so that there is still a fair amount of uncertainty (see the prediction interval below). The value R2 = 0.878 indicates that the model has accounted for about 88% of the original variance.

d. The ANOVA also has a P-value of 0.001, so the same decision is reached, as it should be: reject in favor of the alternative  Note that F = 43.27 = (6.58)2 = t2 apart from a slight rounding error.

e. Given 20 spots, the point forecast for sales is 49.24. The 90% prediction interval is

 (41.88, 56.61) indicating that there is still considerable uncertainty in the outcome.

7.2

Sales clearly decrease with Price, as expected. Revenue increases with Price but seems to be flattening out. The non-linearity in the revenue relationship leads to a lower value of R2 in that case. Indeed, the price coefficient is not significantly different from zero (p=0.124).



**Regression Analysis: Sales versus Price**

The regression equation is

Sales = 44.0 - 2.00 Price

Predictor Coef SE Coef T P

Constant 44.000 5.000 8.80 0.000

Price -2.0000 0.6124 -3.27 0.008

S = 3.46410 R-Sq = 51.6% R-Sq(adj) = 46.8%

Analysis of Variance

Source DF SS MS F P

Regression 1 128.00 128.00 10.67 0.008

Residual Error 10 120.00 12.00

Total 11 248.00

**Regression Analysis: Revenue versus Price**

The regression equation is

Revenue = 133 + 12.0 Price

Predictor Coef SE Coef T P

Constant 133.33 58.39 2.28 0.046

Price 12.000 7.151 1.68 0.124

S = 40.4508 R-Sq = 22.0% R-Sq(adj) = 14.2%

Analysis of Variance

Source DF SS MS F P

Regression 1 4608 4608 2.82 0.124

Residual Error 10 16363 1636

Total 11 20971

d. There is a large error in predicting revenue for period 5

7.3

The relationship between Sales and Week is non-existent. There is no evidence of a time trend. Clearly the model relating Sales to Spots is more useful and captures a useful marketing relationship.

**Regression Analysis: Sales versus Week**

The regression equation is

Sales = 29.3 + 0.60 Week

Predictor Coef SE Coef T P

Constant 29.321 5.629 5.21 0.002

Week 0.595 1.115 0.53 0.613

S = 7.22402 R-Sq = 4.5% R-Sq(adj) = 0.0%

Analysis of Variance

Source DF SS MS F P

Regression 1 14.88 14.88 0.29 0.613

Residual Error 6 313.12 52.19

Total 7 328.00

7.4



a. There is some curvature in the relationship for Sales vs time

b. For Sales the regression line is:

**Time Series Plot of Revenue**

**Regression Analysis: Revenue versus time**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 1 3667380 3667380 405.93 0.000

 time 1 3667380 3667380 405.93 0.000

Error 50 451721 9034

Total 51 4119101

Model Summary

 S R-sq R-sq(adj) R-sq(pred)

95.0496 89.03% 88.81% 87.82%

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant -162.5 26.7 -6.08 0.000

time 17.695 0.878 20.15 0.000 1.00

Regression Equation

Revenue = -162.5 + 17.695 time

Fits and Diagnostics for Unusual Observations

 Std

Obs Revenue Fit Resid Resid

 48 875.6 686.8 188.7 2.05 R

 52 945.2 757.6 187.6 2.05 R

R Large residual

c. The slope coefficient is clearly highly significant (p=0.000 to 3 DP).

d. 89 percent (R2) of the variation is captured by this trend model but substantial variation remains as *S* =95.0, indicating that forecasts could be high or low by as much as 190 (2 times *S*).

e. Predicted Values for New Observations are clearly on the low side. The residuals plots reveal why: the time trend model fails to capture the curvature.

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

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Year | Quarter | Actual | Forecast | LCL | UCL |
| 2013 | 1 | 1023.96 | 775.32 | 577.00 | 973.65 |
| 2013 | 2 | 1069.37 | 793.02 | 594.27 | 991.77 |
| 2013 | 3 | 1106.00 | 810.71 | 611.52 | 1009.90 |
| 2013 | 4 | 1175.23 | 828.41 | 628.77 | 1028.05 |
| 2014 | 1 | 1270.09 | 846.10 | 645.99 | 1046.21 |
| 2014 | 2 | 1340.41 | 863.80 | 663.21 | 1064.39 |
| 2014 | 3 | 1409.43 | 881.49 | 680.41 | 1082.58 |
| 2014 | 4 | 1484.73 | 899.19 | 697.59 | 1100.78 |
| 2015 | 1 | 1573.13 | 916.88 | 714.76 | 1119.00 |
| 2015 | 2 | 1644.69 | 934.58 | 731.92 | 1137.23 |
| 2015 | 3 | 1738.36 | 952.27 | 749.06 | 1155.48 |
| 2015 | 4 | 1823.33 | 969.97 | 766.19 | 1173.74 |

 |  |  |  |  |  |
|  |  |  |  |  |  |



 The plot shows that the pattern of sales deviates systematically from the straight line regression that we have fitted. Despite the continued growth of the company, the point forecast is well below the figure for the previous quarter.

7.5

a. See exercise 2.2 for preliminary analysis.



b. **Regression Analysis: Injuries versus Train-miles**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 1 1336158 1336158 58.71 0.000

 Train-miles 1 1336158 1336158 58.71 0.000

Error 18 409639 22758

 Lack-of-Fit 13 401630 30895 19.29 0.002

 Pure Error 5 8009 1602

Total 19 1745797

Model Summary

 S R-sq R-sq(adj) R-sq(pred)

150.857 76.54% 75.23% 68.90%

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant -1719 321 -5.36 0.000

Train-miles 29.13 3.80 7.66 0.000 1.00

Regression Equation

Injuries = -1719 + 29.13 Train-miles

Fits and Diagnostics for Unusual Observations

 Std

Obs Injuries Fit Resid Resid

 18 1510.0 1048.4 461.6 3.28 R

R Large residual

c. The P-value for the slope is 0.000 (not zero but rounded to zero to 3 decimal place accuracy), so we can reject in favor of the alternative  In this case, the slope is clearly positive so that the expected level of injuries is increasing with the number of train-miles travelled.

The ANOVA also has a P-value of 0.000 to 3 decimal places, so the same decision is reached, as it should be: reject in favor of the alternative  Note that F = 58.71 = (7.66)2 = t2 apart from a slight rounding error.

d. *S* = 150.86 so that there is still a fair amount of uncertainty (see the prediction interval below). The value *R2* = 0.799 indicates that the model has accounted for about 76 percent of the original variance.

e. If Train-Miles =100, the point forecast for Injuries is 1194. The 95% prediction interval is ((844.894, 1543.18) indicating that there is still considerable uncertainty in the outcome.

f. The plot does not show any systematic deviations from a straight line. The intercept has no interpretation in this case, but the fact that it is negative means that Injuries are increasing more than in proportion to Train-Miles. i.e. if your look at injuries per train mile, in the last few years they have been higher than the early years with 2005 showing a serious jump.. This is a worrying feature that should be investigated further.

7.6





The regression output and residual plots show a high R2 but a clearly deficient model. The forecasts are clearly deficient. 16 out of 36 lie inside the 90 percent prediction limits; the rest all lie above the upper value.

The first step towards improvement would be to consider a logarithmic transform. We might also examine changes rather than absolute values. Seasonal and other relevant factors might be introduced, but such extensions require the developments in Chapters 8 and 9.



The log model is somewhat worse and 13 of the 36 actual values lie inside the 90% prediction intervals. Nevertheless, the residuals plots clearly demonstrate that major problems remain.

**Regression Analysis: LN\_median versus Index**

**Regression Analysis: ln(Price\_2013) versus ln(time\_2013)**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 1 48.30 48.2970 1808.48 0.000

 ln(time\_2013) 1 48.30 48.2970 1808.48 0.000

Error 382 10.20 0.0267

Total 383 58.50

Model Summary

 S R-sq R-sq(adj) R-sq(pred)

0.163419 82.56% 82.52% 81.84%

Coefficients

Term Coef SE Coef T-Value P-Value

Constant 10.0616 0.0435 231.47 0.000

ln(time\_2013) 0.36572 0.00860 42.53 0.000

Regression Equation

ln(Price\_2013) = 10.0616 + 0.36572 ln(time\_2013)



The results for the Mean Price are very similar, although the R2 values are slightly lower.

7.7

The large sample size ensures that the coefficient is significant but the relationship is weak (R2 =5.1%). As expected the sign is negative: an increase in the mortgage rate slows down construction. However, there are clearly other factors that need to be considered.

**Regression Analysis: Starts\_2013 versus Rate\_2013 (Sample to Dec 2012)**

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 1 1518752 1518752 9.05 0.003

 Rate\_2013 1 1518752 1518752 9.05 0.003

Error 382 64113936 167838

 Lack-of-Fit 312 61068983 195734 4.50 0.000

 Pure Error 70 3044953 43499

Total 383 65632688

Model Summary

 S R-sq R-sq(adj) R-sq(pred)

409.680 2.31% 2.06% 0.96%

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant 1232.8 58.9 20.92 0.000

Rate\_2013 19.63 6.52 3.01 0.003 1.00

Regression Equation

Starts\_2013 = 1232.8 + 19.63 Rate\_2013

Fits and Diagnostics for Unusual Observations

Obs Starts\_2013 Fit Resid Std Resid

 5 1140.0 1554.7 -414.7 -1.02 X

 6 1045.0 1560.6 -515.6 -1.27 X

 7 1041.0 1563.1 -522.1 -1.29 X

 8 940.0 1572.1 -632.1 -1.56 X

 9 911.0 1589.2 -678.2 -1.68 X

 10 873.0 1594.9 -721.9 -1.79 X

 11 837.0 1582.7 -745.7 -1.84 X

 12 910.0 1564.9 -654.9 -1.62 X

 13 843.0 1574.3 -731.3 -1.81 X

 14 866.0 1578.2 -712.2 -1.76 X

 15 931.0 1569.6 -638.6 -1.58 X

 16 917.0 1564.3 -647.3 -1.60 X

 17 1025.0 1560.2 -535.2 -1.32 X

 18 902.0 1560.6 -658.6 -1.62 X

 19 1166.0 1562.9 -396.9 -0.98 X

 20 1046.0 1552.1 -506.1 -1.25 X

290 2207.0 1343.3 863.7 2.11 R

301 2273.0 1353.5 919.5 2.25 R

337 490.0 1332.1 -842.1 -2.06 R

339 505.0 1330.9 -825.9 -2.02 R

340 478.0 1327.2 -849.2 -2.08 R

R Large residual

X Unusual X

The residual analysis above shows major problems with the model and this is illustrated by the residual plots below.



Note the very poor summary statistics and wide prediction intervals.



Something wrong here! The higher the interest rate the higher the starts. Every single prediction is above the actual though within the very wide upper prediction interval. Extending the data base to 2015 we get the following parameters

Model Summary

 S R-sq R-sq(adj) R-sq(pred)

400.539 5.48% 5.26% 4.32%

Coefficients

Term Coef SE Coef T-Value P-Value

Constant 1130.8 51.5 21.96 0.000

Mortgage 29.09 5.91 4.92 0.000

Whilst if we only use data to 2008) we get:

Model Summary

 S R-sq R-sq(adj) R-sq(pred)

302.241 7.56% 7.28% 6.29%

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant 1765.2 52.0 33.97 0.000

Mortgage -28.57 5.47 -5.23 0.000 1.00

Regression Equation

Housing Starts = 1765.2 - 28.57 Mortgage

 A negative effect of the interest rate as we would have expected! There’s obviously been a major change in the economy and we will need the more complicated methods of chapter 9 to understand better what’s going on.

7.8

The regression results are as shown in Section 7.6.3. The diagnostic indicate two outliers (unfortunately November and December 2008, the last two observations, suggesting problems with the model). The log model accounts somewhat better for the volatility in the later part of the series and the diagnostics are clearly superior to those for the original model.

All the forecasts lie within the prediction intervals, but we may hope for better models by taking additional factors into account.

**Regression Analysis: LN\_Unleaded versus LN\_L1\_Crude**

The regression equation is

|  |
| --- |
| **Coefficientsa** |
| Model | Unstandardized Coefficients | t | Sig. |
| B | Std. Error |
| 1 | (Constant) | 2.897 | .038 | 76.827 | .000 |
| ln\_L1\_Crude | .628 | .011 | 59.459 | .000 |
| a. Dependent Variable: ln\_Unleaded2010 |

|  |
| --- |
| **Model Summary** |
| Model | R | R Square | Adjusted R Square | Std. Error of the Estimate |
| 1 | .979a | .959 | .958 | .07630 |
| a. Predictors: (Constant), ln\_L1\_Crude |





**Regression Analysis: Unleaded2010 versus L1\_Crude\_price**

Method

Rows unused 85

Analysis of Variance

Source DF Adj SS Adj MS F-Value P-Value

Regression 1 789165 789165 2544.35 0.000

 L1\_Crude\_price 1 789165 789165 2544.35 0.000

Error 153 47455 310

 Lack-of-Fit 152 47437 312 17.34 0.189

 Pure Error 1 18 18

Total 154 836620

Model Summary

 S R-sq R-sq(adj) R-sq(pred)

17.6115 94.33% 94.29% 94.10%

Coefficients

Term Coef SE Coef T-Value P-Value VIF

Constant 68.76 2.59 26.54 0.000

L1\_Crude\_price 2.7069 0.0537 50.44 0.000 1.00

Regression Equation

Unleaded2010 = 68.76 + 2.7069 L1\_Crude\_price

Fits and Diagnostics for Unusual Observations

Obs Unleaded2010 Fit Resid Std Resid

117 290.30 244.68 45.62 2.61 R

124 274.20 238.45 35.75 2.04 R

127 298.10 260.81 37.29 2.13 R

136 284.50 232.36 52.14 2.98 R

137 314.60 241.95 72.65 4.15 R

138 305.60 240.54 65.06 3.72 R

139 296.50 251.45 45.05 2.58 R

148 345.80 354.20 -8.40 -0.49 X

149 376.60 373.50 3.10 0.18 X

150 405.40 408.20 -2.80 -0.17 X

151 406.20 431.16 -24.96 -1.48 X

152 377.90 429.78 -51.88 -3.08 R X

153 370.30 384.57 -14.27 -0.84 X

154 305.10 350.57 -45.47 -2.64 R X

155 214.70 276.13 -61.43 -3.52 R

156 168.70 223.89 -55.19 -3.15 R

R Large residual

X Unusual X



## Forecast from the two models and their respective 95% prediction intervals are shown below. There is little to choose between the two models, the level (lvl) limits being marginally tighter and both containing the actual values.



## Minicase 7.1

Exercise 7.8 provides signposts on how to proceed. The main purpose behind the minicase is to become acquainted with the data sources and with the series themselves. In question 4 where the model is to be developed in constant prices, the comparison of accuracy needs to be consistent between the constant and nominal price models. With just crude\_price in the model there is no reason to believe a constant price model is more economically persuasive although some elements in the mark up over the crude price is related to real prices in the economy. So the CPI might well be a separate influence.

## Minicase 7.2

Is *Consumer Confidence* a leading indicator for *Unemployment*? Or is it the other way around? As a start, we may look at the relationship between the two for different lagged values of *Consumer Confidence.*  Lag 3 seems to be the best of these three; further lags could be explored along with other changes (e.g. differencing?). Residual plots should of course be examined.

**Regression Analysis: Unemployment versus CS\_1**

Predictor Coef SE Coef T P

Constant 8.1385 0.3830 21.25 0.000

CS\_1 -0.034498 0.004134 -8.35 0.000

S = 0.565650 R-Sq = 35.2% R-Sq(adj) = 34.7%

**Regression Analysis: Unemployment versus CS\_2**

Predictor Coef SE Coef T P

Constant 8.3609 0.3904 21.41 0.000

CS\_2 -0.036782 0.004204 -8.75 0.000

S = 0.556781 R-Sq = 37.6% R-Sq(adj) = 37.1%

**Regression Analysis: Unemployment versus CS\_3**

Predictor Coef SE Coef T P

Constant 8.5688 0.3868 22.16 0.000

CS\_3 -0.038943 0.004158 -9.37 0.000

S = 0.543024 R-Sq = 41.0% R-Sq(adj) = 40.6%

### Minicase 7.3

The original analysis of 176 cases yields:

**Regression Analysis: Salary ($000s) versus Years in Majors**

The regression equation is

Salary ($000s) = 224 + 43.6 Years in Majors

Predictor Coef SE Coef T P

Constant 223.56 40.52 5.52 0.000

Years in Majors 43.638 5.228 8.35 0.000

S = 315.239 R-Sq = 28.6% R-Sq(adj) = 28.2%

Analysis of Variance

Source DF SS MS F P

Regression 1 6922565 6922565 69.66 0.000

Residual Error 174 17291310 99375

Total 175 24213875



The revised analysis of 156 cases yields the following.

Note:

* Improved fit
* Reduced *S*
* Much higher slope
* Fewer extreme observations

**Regression Analysis: Salary ($000s) versus Years in Majors**

The regression equation is

Salary ($000s) = 1.3 + 93.8 Years in Majors

Predictor Coef SE Coef T P

Constant 1.35 42.10 0.03 0.975

Years in Majors 93.779 7.223 12.98 0.000

S = 262.399 R-Sq = 52.3% R-Sq(adj) = 51.9%

Analysis of Variance

Source DF SS MS F P

Regression 1 11605105 11605105 168.55 0.000

Residual Error 154 10603393 68853

Total 155 22208498

