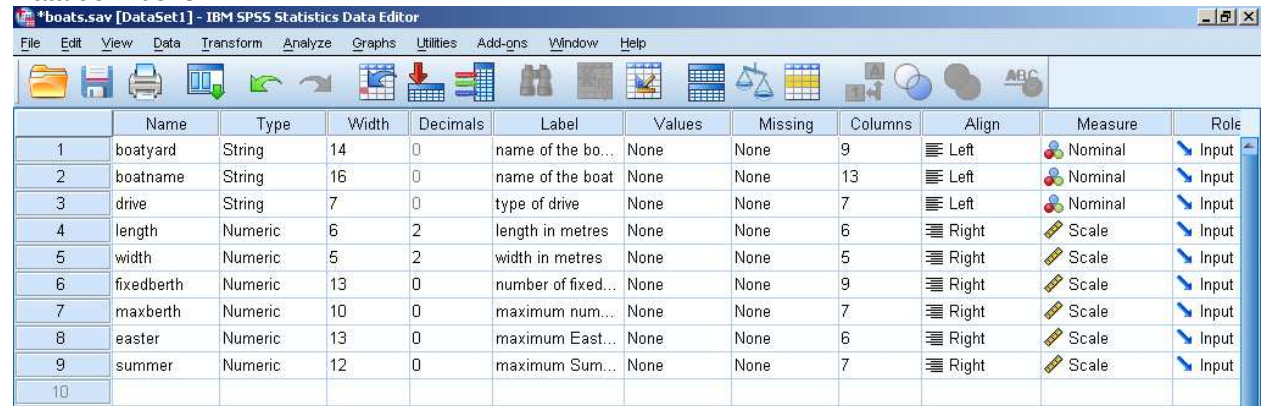


# Interpreting Regression Output

## Using the boats data for multiple regression in SPSS

Data was collected on the prices charged for weekly boat hire at Easter and in the Summer from a number of boatyards on the Norfolk Broads. For each boat, the fields shown below were recorded.

Data definitions:



|    | Name       | Type    | Width | Decimals | Label              | Values | Missing | Columns | Align | Measure | Role  |
|----|------------|---------|-------|----------|--------------------|--------|---------|---------|-------|---------|-------|
| 1  | boatyard   | String  | 14    | 0        | name of the bo...  | None   | None    | 9       | Left  | Nominal | Input |
| 2  | boatname   | String  | 16    | 0        | name of the boat   | None   | None    | 13      | Left  | Nominal | Input |
| 3  | drive      | String  | 7     | 0        | type of drive      | None   | None    | 7       | Left  | Nominal | Input |
| 4  | length     | Numeric | 6     | 2        | length in metres   | None   | None    | 6       | Right | Scale   | Input |
| 5  | width      | Numeric | 5     | 2        | width in metres    | None   | None    | 5       | Right | Scale   | Input |
| 6  | fixedberth | Numeric | 13    | 0        | number of fixed... | None   | None    | 9       | Right | Scale   | Input |
| 7  | maxberth   | Numeric | 10    | 0        | maximum num...     | None   | None    | 7       | Right | Scale   | Input |
| 8  | easter     | Numeric | 13    | 0        | maximum East...    | None   | None    | 6       | Right | Scale   | Input |
| 9  | summer     | Numeric | 12    | 0        | maximum Sum...     | None   | None    | 7       | Right | Scale   | Input |
| 10 |            |         |       |          |                    |        |         |         |       |         |       |

We want to see how the maximum price charged during Easter for a week's hire is related to the attributes of the boat (length, width, number of fixed berths, maximum number of berths )

The correlation matrix gives:

|                          |                     | Correlations     |                 |                          |                        |                      |
|--------------------------|---------------------|------------------|-----------------|--------------------------|------------------------|----------------------|
|                          |                     | length in metres | width in metres | maximum number of berths | number of fixed berths | maximum Easter price |
| length in metres         | Pearson Correlation | 1                | .688**          | .769**                   | .828**                 | .857**               |
|                          | Sig. (2-tailed)     |                  | .000            | .000                     | .000                   | .000                 |
|                          | N                   | 112              | 112             | 112                      | 112                    | 112                  |
| width in metres          | Pearson Correlation | .688**           | 1               | .498**                   | .492**                 | .537**               |
|                          | Sig. (2-tailed)     | .000             |                 | .000                     | .000                   | .000                 |
|                          | N                   | 112              | 112             | 112                      | 112                    | 112                  |
| maximum number of berths | Pearson Correlation | .769**           | .498**          | 1                        | .891**                 | .748**               |
|                          | Sig. (2-tailed)     | .000             | .000            |                          | .000                   | .000                 |
|                          | N                   | 112              | 112             | 112                      | 112                    | 112                  |
| number of fixed berths   | Pearson Correlation | .828**           | .492**          | .891**                   | 1                      | .863**               |
|                          | Sig. (2-tailed)     | .000             | .000            | .000                     |                        | .000                 |
|                          | N                   | 112              | 112             | 112                      | 112                    | 112                  |
| maximum Easter price     | Pearson Correlation | .857**           | .537**          | .748**                   | .863**                 | 1                    |
|                          | Sig. (2-tailed)     | .000             | .000            | .000                     | .000                   |                      |
|                          | N                   | 112              | 112             | 112                      | 112                    | 112                  |

\*\* . Correlation is significant at the 0.01 level (2-tailed).

There appears to be some multi-collinearity here as there are strong correlations between some of the "independent" variables. In particular, "number of fixed berths" is strongly correlated with "maximum number of berths". Thus we would not expect both of these variables to be present in a "good" model.

**Using the step-wise approach:**

**Variables Entered/Removed<sup>a</sup>**

| Model | Variables Entered      | Variables Removed | Method                                                                                      |
|-------|------------------------|-------------------|---------------------------------------------------------------------------------------------|
| 1     | number of fixed berths |                   | Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100). |
| 2     | length in metres       |                   | Stepwise (Criteria: Probability-of-F-to-enter <= .050, Probability-of-F-to-remove >= .100). |

a. Dependent Variable: maximum Easter price

*The first model just used "number of fixed berths" as this has the highest correlation with maximum Easter price. To this model, the variable "length in metres" was added. No further additions were made as the other variables correlated strongly with these two.*

**Model Summary<sup>c</sup>**

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics |          |     |     |               |
|-------|-------------------|----------|-------------------|----------------------------|-------------------|----------|-----|-----|---------------|
|       |                   |          |                   |                            | R Square Change   | F Change | df1 | df2 | Sig. F Change |
| 1     | .863 <sup>a</sup> | .745     | .743              | 67.792                     | .745              | 322.149  | 1   | 110 | .000          |
| 2     | .900 <sup>b</sup> | .810     | .806              | 58.845                     | .064              | 36.988   | 1   | 109 | .000          |

a. Predictors: (Constant), number of fixed berths

b. Predictors: (Constant), number of fixed berths, length in metres

c. Dependent Variable: maximum Easter price

*The value of adjusted  $r^2$  was 0.743 for model 1 (number of fixed berths). This rose to  $r^2 = 0.806$  when the "length in metres" was added to model 1 to form model 2.*

**Coefficients<sup>a</sup>**

| Model |                        | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. | Collinearity Statistics |       |
|-------|------------------------|-----------------------------|------------|---------------------------|--------|------|-------------------------|-------|
|       |                        | B                           | Std. Error | Beta                      |        |      | Tolerance               | VIF   |
| 1     | (Constant)             | 263.311                     | 15.076     |                           | 17.466 | .000 |                         |       |
|       | number of fixed berths | 62.209                      | 3.466      | .863                      | 17.949 | .000 | 1.000                   | 1.000 |
| 2     | (Constant)             | 12.597                      | 43.251     |                           | .291   | .771 |                         |       |
|       | number of fixed berths | 35.206                      | 5.363      | .489                      | 6.564  | .000 | .315                    | 3.178 |
|       | length in metres       | 33.623                      | 5.528      | .453                      | 6.082  | .000 | .315                    | 3.178 |

a. Dependent Variable: maximum Easter price

*In model 1 . The coefficient of number of fixed berths is significantly different from zero ( $p=0.000<0.05$ )  
In model 2, both coefficients are significantly different from zero ( $p=0.000$  in both cases )*

**Casewise Diagnostics<sup>a</sup>**

| Case Number | Std. Residual | maximum Easter price | Predicted Value | Residual |
|-------------|---------------|----------------------|-----------------|----------|
| 15          | 3.397         | 945                  | 745.12          | 199.877  |

### Casewise Diagnostics<sup>a</sup>

| Case Number | Std. Residual | maximum Easter price | Predicted Value | Residual |
|-------------|---------------|----------------------|-----------------|----------|
| 15          | 3.397         | 945                  | 745.12          | 199.877  |

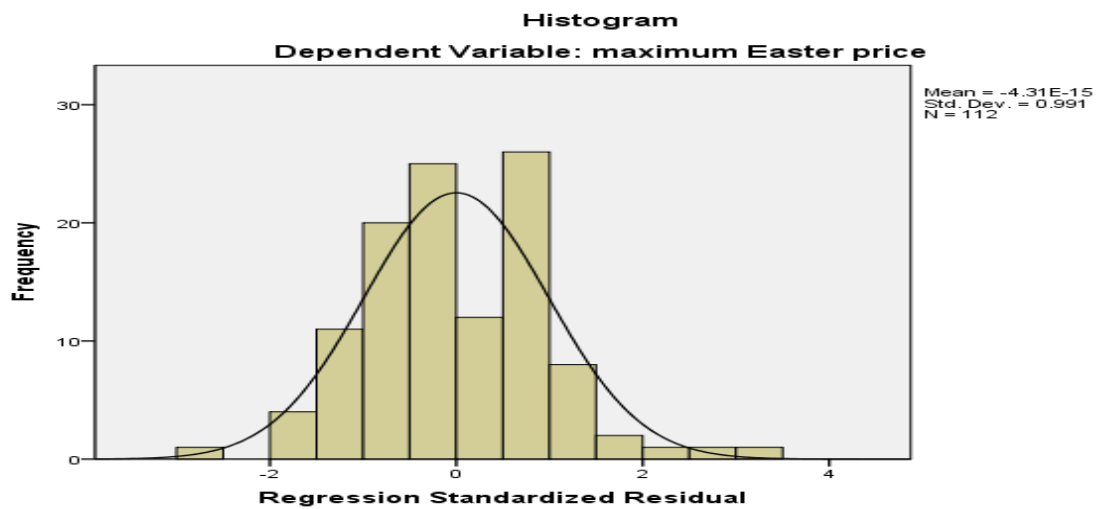
a. Dependent Variable: maximum Easter price

*The only real outlier is case number 15 as it is the only point with a standardised residual outside the range -3 to + 3.*

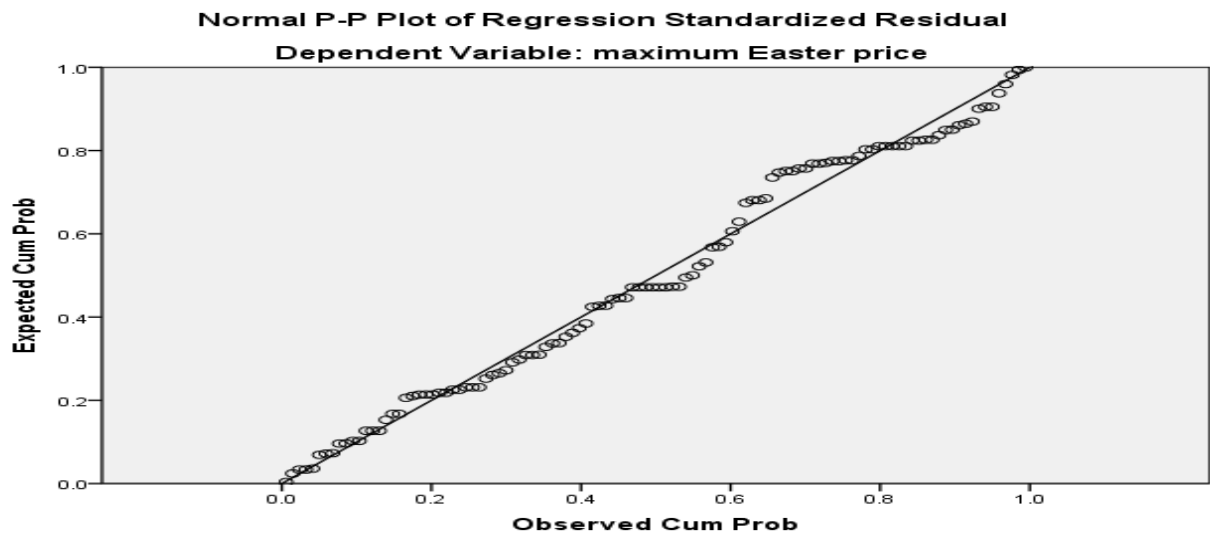
### Residuals Statistics<sup>a</sup>

|                      | Minimum  | Maximum | Mean   | Std. Deviation | N   |
|----------------------|----------|---------|--------|----------------|-----|
| Predicted Value      | 328.79   | 806.65  | 508.26 | 120.382        | 112 |
| Residual             | -158.095 | 199.877 | .000   | 58.313         | 112 |
| Std. Predicted Value | -1.491   | 2.479   | .000   | 1.000          | 112 |
| Std. Residual        | -2.687   | 3.397   | .000   | .991           | 112 |

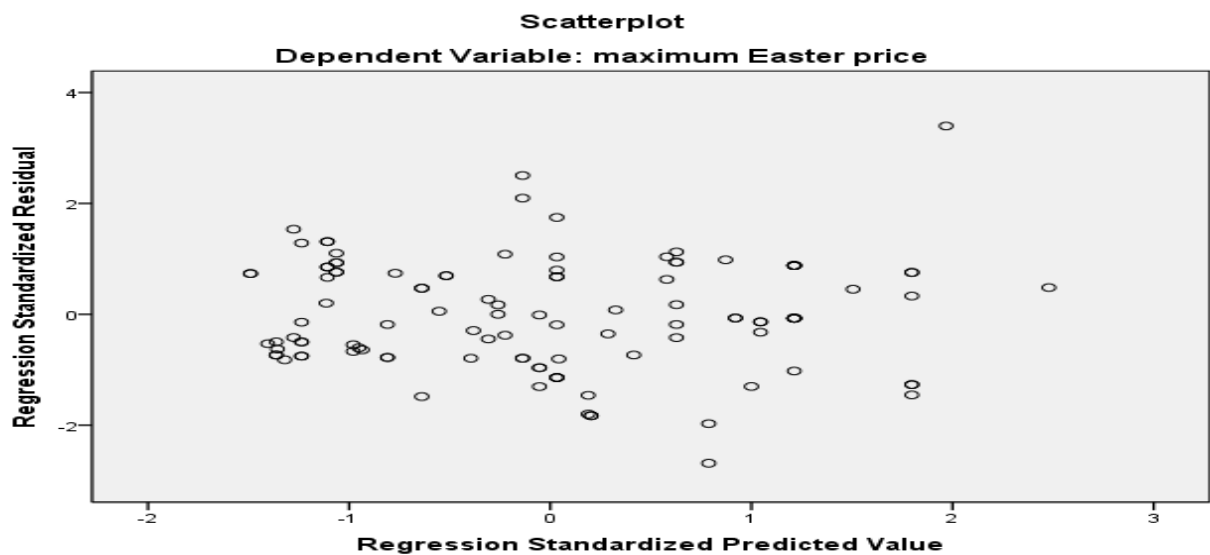
a. Dependent Variable: maximum Easter price



*The graph of the residuals does roughly resemble a normal distribution curve. Thus the assumption that the residuals are normally distributed seems reasonable.*



The points are close to the straight line and there is no real pattern in the points , thus the assumption of normality seems reasonable.



The residuals are fairly evenly spread about the horizontal line through zero. The difference from the zero line does not change much as the standardised predictions increase.

Conclusion : This is a good model.

Compare this output with one that includes all possible variables.

| Variables Entered/Removed <sup>b</sup> |                                                                                                       |                   |        |
|----------------------------------------|-------------------------------------------------------------------------------------------------------|-------------------|--------|
| Model                                  | Variables Entered                                                                                     | Variables Removed | Method |
| 1                                      | maximum<br>number of<br>berths, width in<br>metres, length<br>in metres,<br>number of fixed<br>berths | .                 | Enter  |

a. All requested variables entered.

b. Dependent Variable: maximum Easter price

**Model Summary<sup>b</sup>**

| Model | R                 | R Square | Adjusted R Square | Std. Error of the Estimate | Change Statistics |          |     |     |               |
|-------|-------------------|----------|-------------------|----------------------------|-------------------|----------|-----|-----|---------------|
|       |                   |          |                   |                            | R Square Change   | F Change | df1 | df2 | Sig. F Change |
| 1     | .903 <sup>a</sup> | .816     | .809              | 58.396                     | .816              | 118.850  | 4   | 107 | .000          |

a. Predictors: (Constant), maximum number of berths, width in metres, length in metres, number of fixed berths

b. Dependent Variable: maximum Easter price

**ANOVA<sup>b</sup>**

| Model |            | Sum of Squares | df  | Mean Square | F       | Sig.              |
|-------|------------|----------------|-----|-------------|---------|-------------------|
| 1     | Regression | 1621146.826    | 4   | 405286.707  | 118.850 | .000 <sup>a</sup> |
|       | Residual   | 364878.665     | 107 | 3410.081    |         |                   |
|       | Total      | 1986025.491    | 111 |             |         |                   |

a. Predictors: (Constant), maximum number of berths, width in metres, length in metres, number of fixed berths

b. Dependent Variable: maximum Easter price

*The F statistics is significant and thus indicates that there is a linear relationship between maximum Easter price and at least one of the other variables*

**Coefficients<sup>a</sup>**

| Model |                          | Unstandardized Coefficients |            | Standardized Coefficients | t      | Sig. | Collinearity Statistics |       |
|-------|--------------------------|-----------------------------|------------|---------------------------|--------|------|-------------------------|-------|
|       |                          | B                           | Std. Error | Beta                      |        |      | Tolerance               | VIF   |
| 1     | (Constant)               | 34.804                      | 73.494     |                           | .474   | .637 |                         |       |
|       | length in metres         | 36.055                      | 6.704      | .485                      | 5.378  | .000 | .211                    | 4.746 |
|       | width in metres          | -8.502                      | 27.532     | -.018                     | -.309  | .758 | .503                    | 1.987 |
|       | number of fixed berths   | 44.915                      | 7.726      | .623                      | 5.814  | .000 | .149                    | 6.696 |
|       | maximum number of berths | -10.983                     | 5.921      | -.172                     | -1.855 | .066 | .201                    | 4.985 |

a. Dependent Variable: maximum Easter price

(If we were building a model ourselves, we would eliminate “width in metres” and “maximum number of berths” as the coefficients are not significantly different from zero and the signs are not what we would expect. We would expect wider boats with more berths to have an increase in price.)

**Casewise Diagnostics<sup>a</sup>**

| Case Number | Std. Residual | maximum Easter price | Predicted Value | Residual |
|-------------|---------------|----------------------|-----------------|----------|
| 15          | 3.191         | 945                  | 758.64          | 186.357  |

a. Dependent Variable: maximum Easter price

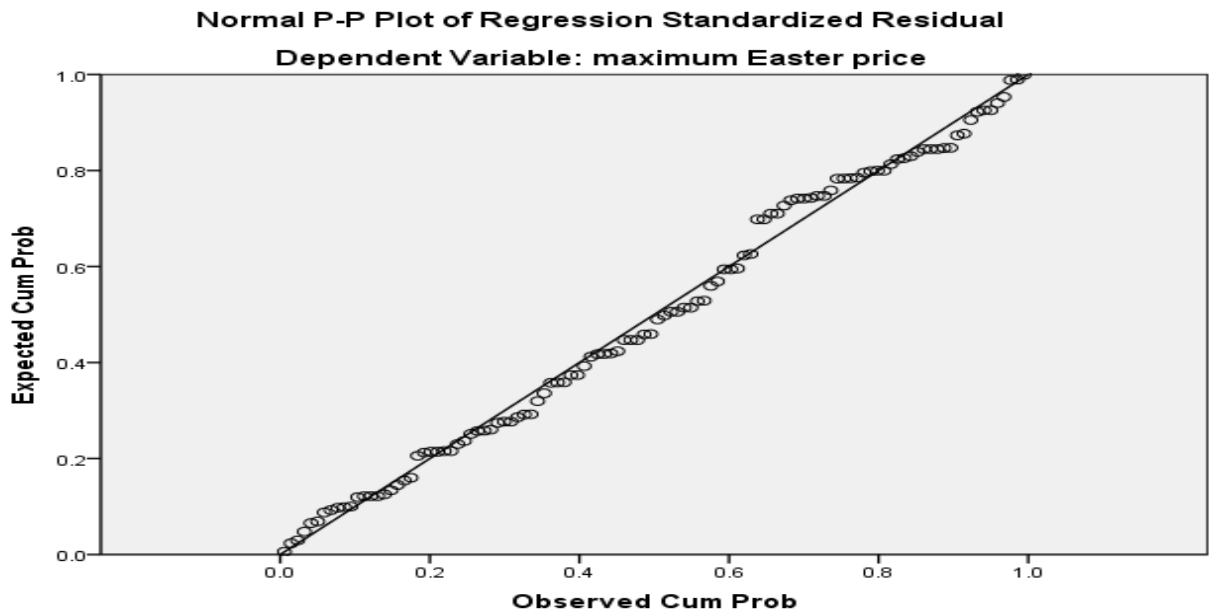
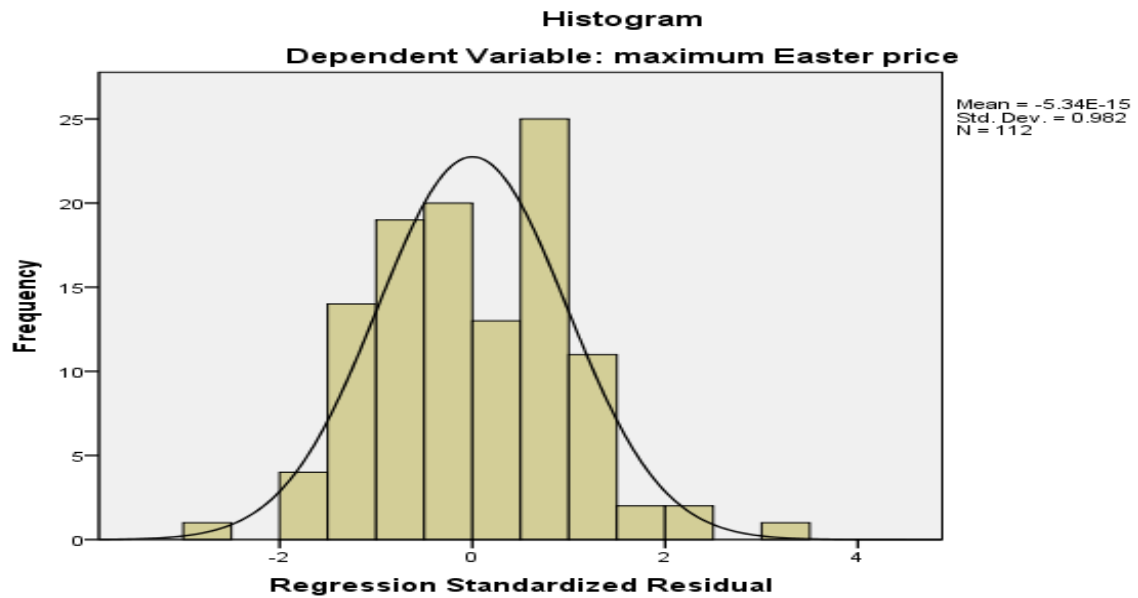
**Residuals Statistics<sup>a</sup>**

|                      | Minimum  | Maximum | Mean   | Std. Deviation | N   |
|----------------------|----------|---------|--------|----------------|-----|
| Predicted Value      | 336.28   | 780.69  | 508.26 | 120.851        | 112 |
| Residual             | -147.015 | 186.357 | .000   | 57.334         | 112 |
| Std. Predicted Value | -1.423   | 2.254   | .000   | 1.000          | 112 |
| Std. Residual        | -2.518   | 3.191   | .000   | .982           | 112 |

Casewise Diagnostics<sup>a</sup>

| Case Number | Std. Residual | maximum<br>Easter price | Predicted Value | Residual |
|-------------|---------------|-------------------------|-----------------|----------|
| 15          | 3.191         | 945                     | 758.64          | 186.357  |

a. Dependent Variable: maximum Easter price



## Features of a Good Linear Regression Model [SPSS printout]

|   | Source of information  | Desirable feature                                                                                                                                                                                                                                                         |
|---|------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| 1 | Model Summary          | A high value of adjusted $r^2$<br><br><i>[A low value indicates a poor linear fit]</i>                                                                                                                                                                                    |
| 2 | Table of Correlations  | Model chosen should not contain independent variables that are highly correlated with each other.<br><br><i>[Model should not exhibit any multi-collinearity]</i>                                                                                                         |
| 3 | Coefficients table     | All coefficients significantly different from zero<br>i.e. Sig column has values below 0.025<br><br><i>[If a coefficient is close to zero it adds nothing useful to the model]</i>                                                                                        |
| 4 | ANOVA                  | The F statistic is significantly different from zero indicating that there is a relationship between the dependent variable and at least one of the independent variables.                                                                                                |
| 5 | Casewise diagnostics   | Only 5% of readings are outside the range -2 to +2 in 'Std. Residual' column<br>Any outliers are outside the range -3 to +3<br><br><i>[If there are a number of outliers than data points could just be 'odd' cases or could indicate lack of a linear relationship.]</i> |
| 6 | Histogram of residuals | Histogram fits the superimposed Normal curve (or is close)<br><br><i>[If histogram does not approximate to Normal curve, it implies residuals are not normally distributed and thus model is not a good fit. Should consider possibility of non-linear relationship]</i>  |
| 7 | Normal p-p plot        | Points are:<br>scattered about the diagonal line<br>close to the diagonal line<br>do not exhibit any pattern<br><br><i>[If points are not as stated above, the model is not a good fit. Should consider possibility of non-linear relationship]</i>                       |