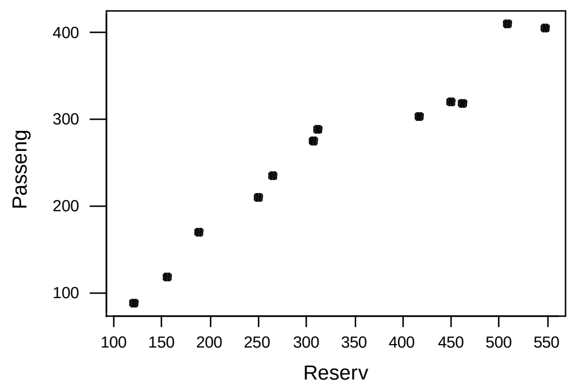
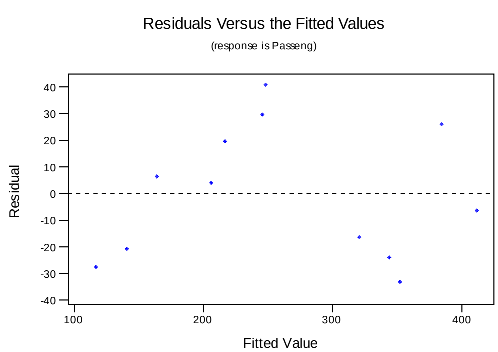
Outline Solutions to Tutorial Sheet 11

1. The corresponding Minitab plots and regression output are below. The relevant parts of the output are highlighted.

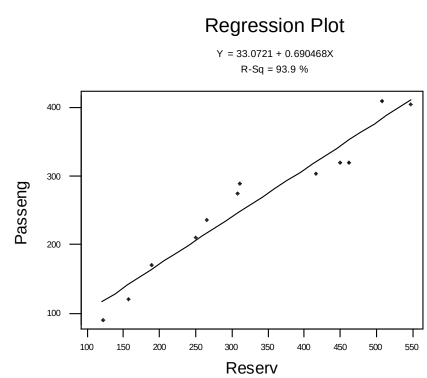


Regression Analysis

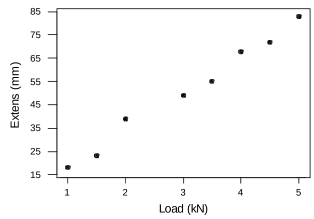
```
The regression equation is
Passeng = 33.1 + 0.690 Reserv

Predictor          Coef        StDev          T        P
Constant          33.07        19.93       1.66    0.128
Reserv          0.69047      0.05553      12.43    0.000

S = 26.26       R-Sq = 93.9%      R-Sq(adj) = 93.3%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         1      106666      106666    154.62     0.000
Residual Error    10        6898         690
Total             11      113564
```



Fitted Line Plot



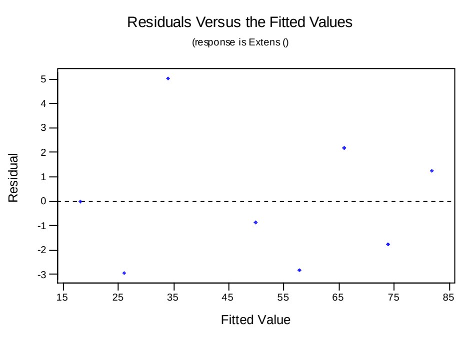
2. (i) The scatter plot shows a strong positive linear relationship.



(ii) **Regression Analysis**

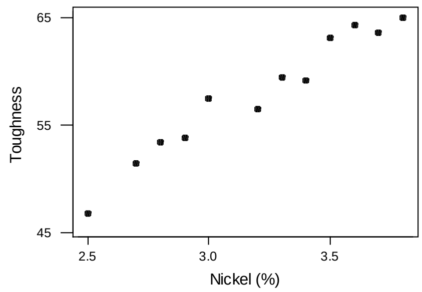
```
The regression equation is
Extens (mm) = 2.07 + 15.9 Load (kN)

Predictor          Coef        StDev           T        P
Constant          2.070        2.590        0.80    0.455
Load (kN        15.9363       0.7733       20.61    0.000

S = 2.967       R-Sq = 98.6%     R-Sq(adj) = 98.4%

Analysis of Variance

Source             DF           SS          MS         F         P
Regression          1       3738.1      3738.1    424.66     0.000
Residual Error      6         52.8         8.8
Total               7       3790.9
```

The regression is highly significant (*P=0.000*) and the R^2 value tells us that 98.6% of the observed variation in extension can be attributed to the different loads used. The slope (15.9) tells us that if load is increased by 1kN, we would expect the extension to increase by 15.9mm. The residual plot below looks random about zero so the model is reasonable.



(iii) From the regression equation, predicted extensions are **41.8mm**, **97.5mm** and **161.1mm** respectively. The first is an interpolation (X = 2.5 is within the observed range of loads) so, given the strength of the relationship, it is likely to be very accurate. The second is an extrapolation (X = 6 is just outside the data range) and should be treated with caution. However it may be fairly accurate since R^2 is so large and it is only a *slight* extrapolation. The last is an extrapolation well outside the range of data and could be wildly inaccurate – we have no idea what the relationship is like for such values of load.

3. (i) The scatter plot confirms a linear relationship.



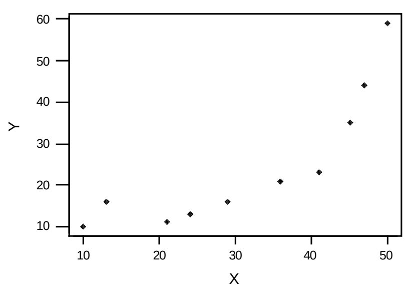
(ii) The correlation coefficient is r = **0.976** and the 5% critical value for 12 pairs of values is **0.576**. Thus since the calculated value is greater than the critical value, there really is a relationship between nickel content and toughness.

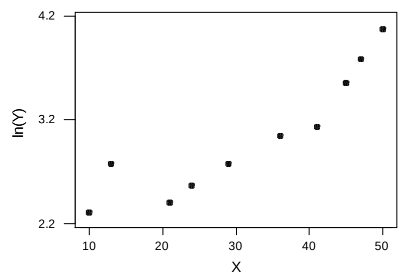
(iii) The regression equation is: `Toughness = 15.1 + 13.4 Nickel (%)`

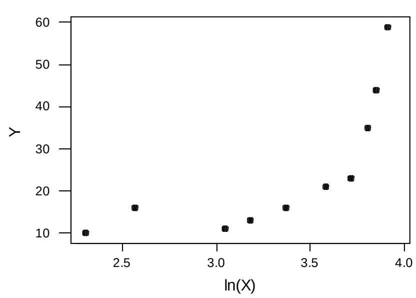
The required estimate is given by the slope (13.4). Thus we would expect toughness to increase by **13.4 units** if nickel content is increased by 1 unit (1%).

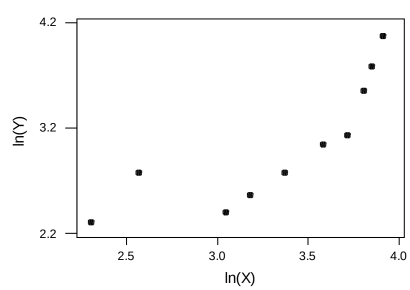
When *Nickel(%)* = 2.2, *Toughness* = 15.1 + (13.4)(2.2) = **44.58** units.

4. The various scatter diagrams are below. The relationship between *X* and *Y* is clearly non-linear. The plot of *ln(Y)* against *X* is most linear.









The correlation coefficients (which measure the extent of *linear* association) are **0.910** for *ln(Y)* against *X*, **0.774** for *Y* against *ln(X)* and **0.826** for *ln(Y)* against *ln(X)*. The fact that 0.910 is the largest confirms that the best of the three transformations is to just log the Y values. Minitab oe Excel then gives the regression line **ln(Y) = 1.84 + 0.0380 X**

The slope of 0.0380 means that if *X* is increased by 1 unit, we would expect *ln(Y)* to increase by 0.0380 units.

When $X = 30$, estimated $ln(Y)$ is 1.84 + (0.0380)(30) = 2.98
Thus, estimated Y is $e^{2.98}$ = **19.7** units.

5. R^2 values are 71.1% for diameter, 10.5% for penetration and 4.7% for temperature. Diameter has easily the highest R^2 value and is therefore the best predictor. Over 70% of variations in torque can be explained by the fact that the logs were of different diameters.

The multiple regression output is below with relevant parts highlighted.

Regression Analysis

```
The regression equation is
TORQUE = - 4.59 + 4.60 DIAMETER + 3.12 PENETRTN - 0.0474 TEMP

Predictor        Coef       StDev          T        P
Constant       -4.587       3.782      -1.21    0.239
DIAMETER       4.5958      0.4513      10.18    0.000
PENETRTN       3.1225      0.7986       3.91    0.001
TEMP         -0.04744     0.01809      -2.62    0.016

S = 3.317       R-Sq = 86.3%      R-Sq(adj) = 84.2%

Analysis of Variance

Source            DF          SS          MS         F        P
Regression         3     1384.37      461.46     41.95    0.000
Residual Error    20      220.01       11.00
Total             23     1604.38
```

Note that the overall significance (given as 0.000) means that the three variables *together* are useful predictors of torque – variations in these three factors explain 86.3% of the variation in torque. The *individual* significance values given at the top are all less than 0.05 which suggests that all three variables are contributing something useful to the model.

Using the regression equation, the predicted torque is:

$-4.59 + (4.60)(6) + (3.12)(2) - (0.0474)(100) =$ **24.51** units.