TUTORIAL SHEET 11

Regression and Correlation

1. Use the data for the worked example in section 5.5.2 (page 157) of the lecture notes which was analysed in Excel. Re-do the analysis in Minitab and make sure you can identify the appropriate parts of the output for interpretation. These are the equation of the best fitting line, the overall significance level and the R^2 value. For the residual plot use the *Residuals versus Fits* option under *Graphs...* on the Regression window.

   For the following questions use either Minitab or Excel as you prefer. (In Minitab, a useful summary plot is provided by a *Fitted Line Plot*. Try it for the data for question 1.)

2. The data below show the extension, y, in mm of a steel rod under various loads, x, in kN.

x (kN)	1	1.5	2	3	3.5	4	4.5	5
y (mm)	18	23	39	49	55	68	72	83

   (i) Plot the data on a scatter diagram and interpret the plot.
   (ii) Find the equation of the regression line for predicting extension for a given load and interpret the relevant parts of the output. Include an appropriate residual plot.
   (iii) Use the estimated regression equation to predict the extension of the rod under loads of 2.5kN, 6kN and 10kN. Comment on these predictions.

3. The data below gives the toughness of 12 laboratory specimens of a certain type of alloy steel, in which the percentage nickel content was varied.

Nickel (%)	Toughness
3.5	63.2
2.5	46.8
3.3	59.5
2.7	51.5
3.8	65.1
3.2	56.5
2.8	53.4
3.6	64.4
3.4	59.2
2.9	53.8
3.7	63.7
3.0	57.5

(i) Plot a scatter diagram to confirm the relationship is linear.
(ii) Calculate the correlation coefficient and test it for significance.
(iii) Use the regression line to estimate how you would expect toughness to change, on average, if nickel content is increased by 1%. Also estimate toughness of alloy steel with 2.2% nickel content.

4. The following are coded measurements of two quality characteristics for a sample of 10 manufactured items. It is required to see whether the easily measured X characteristic can be used to estimate the more difficult to measure Y characteristic.

X	10	13	21	24	29	36	41	45	47	50
Y	10	16	11	13	16	21	23	35	44	59

Plot a scatter diagram and comment on the shape of the relationship.
Form new columns containing log transformed data values *ln(X)* and *ln(Y)*. (Use *Calc→Calculator...* in Minitab or *Insert→Function...* in Excel.) Plot scatter diagrams and calculate correlation coefficients for Y against *ln(X)*, *ln(Y)* against X and *ln(Y)* against *ln(X)*. Hence decide which of these transformations to use for a regression analysis.

Using your selected transformation, perform the regression analysis. Interpret the slope and estimate Y for an item that has an X-value of 30 coded units.

5. The *plywood.mtw* worksheet (which you should have stored) contains 24 observations of 4 variables. Full details of the data are on the **Some Data Sets** sheet . The problem involved inserting a chuck into a log, which is then spun. It is required to investigate factors affecting the amount of torque applied before the chuck slips out. These factors are the diameter of the log (in inches), the temperature of the log (°F) and penetration of the chuck into the log (in inches).

Perform three simple regression analyses of torque using each of the three variables separately and, by looking at the R^2 values, decide which of the three is the best predictor of torque.

Actually, it is possible to perform a regression analysis on all three variables *at the same time*. This is called **multiple regression analysis**. In Minitab, just enter further columns in the *Predictors* box of the regression window. Do it with this data. Try to interpret the output and use the regression equation to predict the torque which can be applied when a log of 6 inches diameter at 100°F is used with a chuck inserted 2 inches into the log.