**Analysing Bivariate Data – Using iNZight**

The example below will go through using the diamonds dataset.

|  |  |
| --- | --- |
| 1. Open InZight and Import Data as for the Time Series topic. Once you have imported the ‘Diamonds’ dataset it should look like this -> |  |
| 1. Get an overview of the whole dataset.   To do this go to ‘Advanced’ and ‘Scatter Plot Matrix’. This will give the window shown on the right where I have selected all of the variables.  When you click ‘Plot’ you will get the graphs on the right.  Look for scatter graphs with a linear association (in this case Carat and Price). |  |
| 1. Decide which variable you think is more likely to explain the variation in the other one. In this case Carat is likely to be an explanatory variable for Price.   Drag the title of each variable onto ‘Variable 1’ and ‘Variable 2’. Your explanatory variable goes onto variable 1 and your response variable goes onto variable 2.  ‘Snip’ / Copy just the graph and add it to your report.  You can now examine any association between the variables to decide on possible model type.  NB: The axis titles given on iNZight may not be sufficient, so once inserted into your document you might need to put a text box over the top of them to include a proper title and axis labels (including units). |  |
| 1. You also need to add a trend curve to the plot (we normally try a linear trend first). You do this by clicking add to plot, and selecting add trend curve. Tick the box next to linear as shown to the right and you should get the graph on the right.   You can now close the window. |  |
| 1. The final thing you need to do is get the summary, you get this by pressing the ‘Get Summary’ button. |  |
| \* Explanation of correlation coefficients or r-values are given on the right ->  \*\* Only use for linear trends!! |  |
| 1. Separating Variables   For Merit and Excellence one of the things you can do is separate out the data by a categorical variable to see if this is affecting the trend. You do this by dropping a variable onto the ‘Subset By’ section on iNZight as shown below  There are three possible subsets for this data – by clarity, by lab and by colour.  Get summaries for each. |  |
| 7. Now it is your turn. For each dataset you need to produce:  - The scatter plot.  - The equation of the trend (summary)  - Then write a description using the template we have been practicing |  |
| 1. Use the SportsScience data set to practice these skills – identify at least three different relationships in the data set and write a description for each. 2. Try to identify subsets of data for each that may better explain the relationships. | |

Explanation of variables in SportsScience data set. This data set provides information about 102 male athletes and 100 female athletes at the Australian Institute of Sport.

* RCC – Red blood cell count
* WCC – White blood cell count
* HC – Haematocrit
* Hg – Haemaglobin
* Ferr – Plasma ferritin concentration
* BMI – Body mass index (weight/height2
* SSF – Sum of skin folds
* X.Bfat - % body fat
* LBM – Lean body mass (kg)
* Ht – Height (cm)
* Wt – Weight (kg)