Bivariate Data Computer room Session 3 Groups

1. Use the data set **Fisher Iris**. In the 1930s Dr. Edgar Anderson collected data on 150 iris specimens. This data set was published in 1936 by R. A. Fisher, the well-known British statistician.

Question: What sort of relationship is there between the petal length and petal width for iris flowers?

Explanatory variable is ‘Petal width in mm’. Response variable is ‘Petal length in mm’

|  |  |
| --- | --- |
| 1 Draw a scatter plot using the whole data set. Describe what you see. | 7. Colour code the different species of iris on the one graph. Do they form groups of points?  |
| 2 State the variables with units. Does it matter which variable goes on the x-axis? | 8. Draw separate scatter plots for each of the 3 species. What is the effect of splitting the data into the separate species groups? Comment on the strength of the relationship, the steepness of the trend, the value of a prediction |
| 3 Fit a linear regression model. Write down the equation of this line. |
| 4 Explain the trend in context. |
| 5 Find the correlation coefficient, r. Comment on what it tells you. |
| 6. Make a prediction. |

1. Use the data set **Fisher Iris**. ‘Sepal width’. and ‘Sepal length’

Write a purpose statement:

|  |  |
| --- | --- |
| 1 Draw a scatter plot using the whole data set. Describe what you see. | 4. Draw separate scatter plots for each of the 3 species. What is the effect of splitting the data into the separate species groups? Comment on the strength of the relationship, the steepness of the trend, the value of a prediction for each type of iris. |
| 2 State the variables with units. Does it matter which variable goes on the x-axis? Explain what you will do and why. |
| 3. Colour code the different species if iris on the one graph. Do they form groups of points? |