3C – Conducting a regression analysis using data

**THE COEFFICIENT OF DETERMINATION**

The value of r2 can be considered a measure of the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ power of a regression equation.

As a guide, any relationship with a coefficient of determination greater than \_\_\_\_\_% can be regarded as having good predictive power.

**THE RESIDUAL PLOT**

The residual for a data point is the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ between the actual value of the response variable and the predicted value using the regression equation:

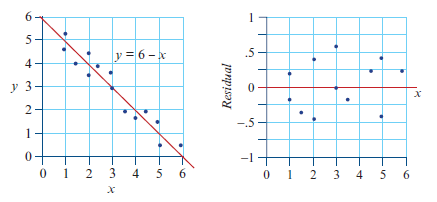
**Residual value = actual RV value – predicted RV value**

If all of the residuals for a data set are plotted on the same graph, we can look to see

if there is a \_\_\_\_\_\_\_\_\_\_\_\_ in them.

A pattern involving positive and negative residuals indicates that the trend in the

data may be \_\_\_\_\_ – \_\_\_\_\_\_\_\_\_\_\_\_ in nature. Random residuals indicate linearity.



A graph of a function

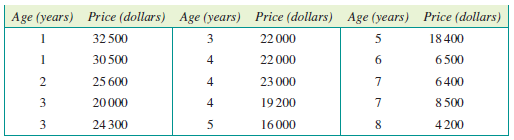
Description automatically generated

3C – Conducting a regression analysis using data

**CONDUCTING A REGRESSION ANALYSIS AND WRITING A REPORT**

1. Construct a scatterplot
2. Calculate the correlation coefficient and determine the strength, direction and form of the association
3. Determine the equation of the regression line
4. Interpret the y-intercept (a) and slope (b) of the regression line in context
5. Calculate and interpret the coefficient of determination
6. Use the regression line to make predictions if required
7. Calculate residuals and use a residual plot to test the assumption of linearity

*The age in years and price of a selection of second hand cars is shown below:*



Use your Chapter 3 Mathematica notebook to perform a regression analysis, make the predictions below, and write a report:

*Predict the price of a car that is 5.5 years old:*

*Predict the number of years until a car is worth $9000:*

From the scatterplot, we see that there is a \_\_\_\_\_\_\_\_\_\_\_\_ , \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ , \_\_\_\_\_\_\_\_\_\_\_\_\_

association between the price of a second hand car and its age, with r = \_\_\_\_\_\_\_\_. There are no

obvious outliers.

The equation of the least squares regression line is:

The slope of the regression line predicts that, on average, the price of these cars decreased by

\_\_\_\_\_\_\_\_\_\_\_\_ each year.

The intercept predicts that, on average, the price of these cars when new was \_\_\_\_\_\_\_\_\_\_\_\_\_\_.

The coefficient of determination indicates that \_\_\_\_\_\_% of the variation in the price of these cars

is explained by the variation in their age.

The lack of a clear pattern in the residual plot confirms the assumption of a \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ association between the price and the age of these cars.