3A – Fitting a least squares regression line

**DEFINITIONS**

The process of modelling an association with a straight line is called \_\_\_\_\_\_\_\_\_\_\_\_ regression.

The line of best fit through the data is called the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ line.

The equation of a line relating two variables x and y is:

x is the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ variable and goes on the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ axis

y is the \_\_\_\_\_\_\_\_\_\_\_\_\_\_ variable and goes on the \_\_\_\_\_\_\_\_\_\_\_\_\_ axis

a is the \_\_\_ \_\_\_\_\_\_\_\_\_\_\_\_\_\_

b is the \_\_\_\_\_\_\_\_ (also called the gradient)

When a regression line goes through points on a scatterplot, the vertical distances

between each point and the line are called the \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_:

A graph of a line

Description automatically generated with medium confidence

**THE LEAST SQUARES REGRESSION LINE**

The least squares regression line minimises the sum of the \_\_\_\_\_\_\_\_\_\_\_\_ of the residuals.

For a least squares regression line to be used, the data must be numerical, with a

\_\_\_\_\_\_\_\_\_\_\_\_\_\_ association and no \_\_\_\_\_\_\_\_\_\_\_\_\_.

The slope of the least squares regression line is given by:

\_\_\_\_ is the Pearson’s correlation coefficient

\_\_\_\_ is the standard deviation of the y values

\_\_\_\_ is the standard deviation of the x values

The y-intercept of the least squares regression line is given by:

\_\_\_\_ is the mean of the y values

\_\_\_\_ is the mean of the x values

\_\_\_\_ is the slope of the line

If you have to work out the correlation coefficient from the slope, the first formula can be rearranged to give:

*EG1: Find the equation of the least squares regression line from the data below:*

*(used to predict weight from height)*

A table with numbers and symbols

Description automatically generated

*EG2: Find the value of the Pearson’s correlation coefficient from the data below:*

*(used to predict exam scores from hours studied)*

A table with numbers and a few words

Description automatically generated with medium confidence

*EG3: The following bivariate data is used to predict weight from height:*



*Use your Chapter 3 Mathematica notebook to:*

1. *Find the Pearson’s correlation coefficient*
2. *Find the equation of the least squares regression line and write it below with in-context variables:*
3. *Produce a plot of this regression line and see if it fits the scatterplot*