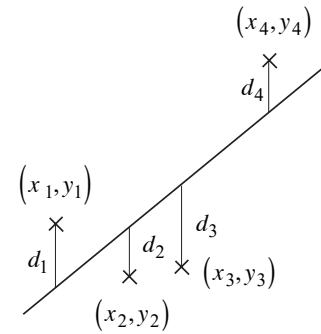


Correlation and Regression

Essential information

- The **equation of the line of regression** is based on minimising the square of the distances of the data points from the regression line,

$$\text{i.e. minimising } \sum_{i=1}^n d_i^2$$



- and the **y and x line of regression** is given by

$$y - \bar{y} = \frac{S_{xy}}{S_x^2}(x - \bar{x})$$

when
$$S_{xy} = \frac{1}{n} \sum_{i=1}^n x_i y_i - \bar{x} \bar{y}$$

and
$$S_x^2 = \frac{1}{n} \sum_{i=1}^n x_i^2 - \bar{x}^2$$

- Bivariate data** is data that has two variables