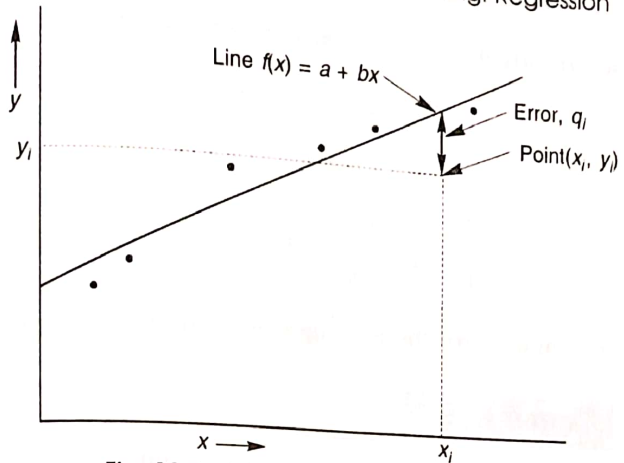

Fig. 10.2 Scatter diagram

It can be easily verified that the first two strategies do not yield a unique line for a given set of data. The third strategy overcomes this problem and guarantees a unique line. The technique of minimising the sum of squares of errors is known as *least squares regression*. In this section we consider the least-squares fit of a straight line.

Least Squares Regression

Let the sum of squares of individual errors be expressed as

$$Q = \sum_{i=1}^{n} q_i^2 = \sum_{i=1}^{n} [y_i - f(x_i)]^2$$

$$= \sum_{i=1}^{n} (y_i - a - bx_i)^2 \qquad (10.5)$$

In the method of least squares, we choose a and b such that Q is minimum. Since Q depends on a and b, a necessary condition for Q to be minimum is

$$\frac{\partial Q}{\partial a} = 0 \quad \text{and} \quad \frac{\partial Q}{\partial b} = 0$$

Then

$$\frac{\partial Q}{\partial a} = -2 \sum_{i=1}^{n} (y_i - a - bx_i) = 0$$

$$\frac{\partial Q}{\partial b} = -2 \sum_{i=1}^{n} x_i (y_i - a - bx_i) = 0 \qquad (10.6)$$

Thus

$$\sum y_i = na + b \sum x_i \qquad (10.7)$$

$$\sum x_i y_i = a \sum x_i + b \sum x_i^2$$