**Simple Linear Regression: Review and Examples**

## Simple linear regression and correlation analysis

The statistical analysis that we have covered in the previous chapters has been concerned with the characteristics of a single variable.

Business decisions are often based upon the relationships between the two variables. If a relationship exists, then it may be possible to predict the number of sales of a product from its price or the sales of a product from the amount spent on its advertising.

Simple linear regression and correlation analysis examine a possible relationship between the variables. Simple linear regression describes the relationship between the variables. To measure the closeness of relationship between variables we use correlation analysis. The closer the actual data points to the regression line, the close the relationship between the variables.

The two measures of the strength of relationship are:

1. the correlation coefficient, $r$;
2. the coefficient of determination, $r^2$ or $R^2$.

By the end of this chapter you will be able to solve problems, such as:

The manager of Pizza Hut would like to establish a timetable to give the customers an idea of how long it will take to deliver a pizza. Seven randomly selected deliveries were used to record the number of kilometers to the delivery site from Pizza Hut and the times from the order to the delivery.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x$ (distance, km) | 4.6 | 13.4 | 15 | 6 | 9.2 | 7.8 | 14.6 |
| $y$ (delivery time, minutes) | 5 | 13 | 10 | 5 | 9 | 8 | 12 |

(a) Draw a scatter graph. Does the scatter graph indicate that a linear relationship exist between the distance and the delivery time.

(b) Find the correlation coefficient *r* and comment on its value.

(c)Find the coefficient of determination $r^2$ and comment on its value.

(d) Use the least squares method to find the regression equation for a line of best fit for the data. Plot the regression line on the scatter diagram.

(e) Use your result from (d) to predict the delivery time for a distance of 11 km. Is the prediction reliable? Give a reason for your answer.

**Scatter Diagrams**

A good first step in a study of the relationship between two variables is to make a *scatter diagram*. A scatter diagram is a picture of the ordered pairs of observations. Scatter diagrams will often show at a glance whether a relationship exists between the two variables.

We plot the **independent variable**, *x*, on the horizontal axis and the **dependent variable**, *y*, on the vertical axis. The variable we are interested in predicting is usually called the dependent variable (*y*). The variable used to predict *y* is usually called the independent variable (*x*). If the price of a certain product is used to predict the number of units sold, then the price is the independent variable *x* and the number of units sold is the dependent variable *y*.

**Example 1**

Ahmad wants to know the relationship between his monthly sales and the amount of monthly commission he gets. The following table shows the figures for the last six months:

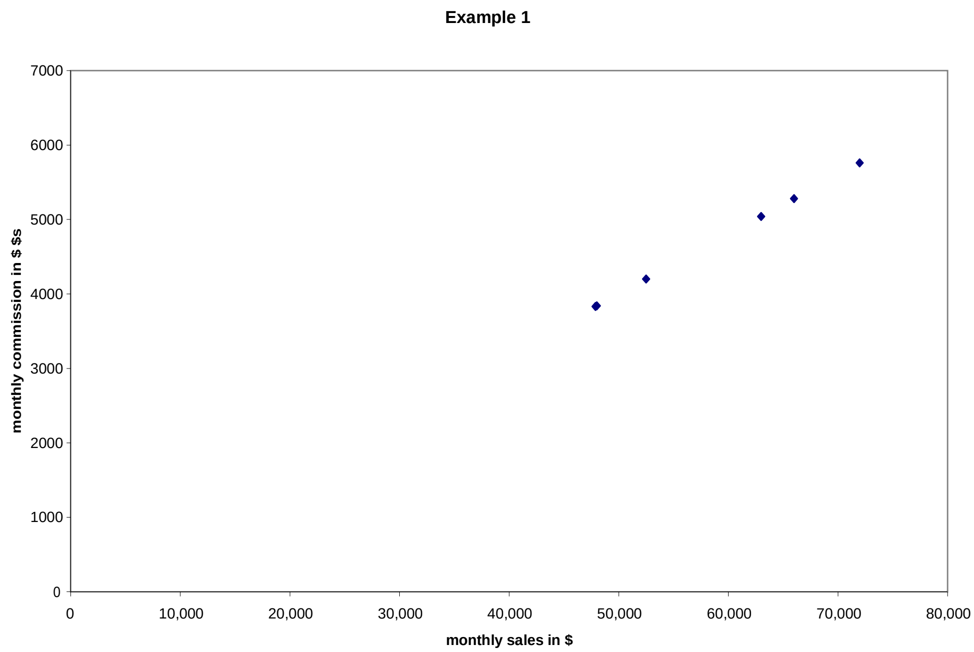
| | | | | | | |
|---|---|---|---|---|---|---|
| monthly sales in $ | 48,000 | 52,500 | 63,000 | 47,900 | 72,00 | 66,000 |
| monthly commission in $ | 3840 | 4200 | 5040 | 3832 | 5760 | 5280 |

"Monthly commission" is the dependent variable, *y*, and "monthly sales" is the independent variable, *x*. Monthly commission changes as the volume of sales changes.

The value of *y* depends on the value of *x*.

We use the horizontal axis for the *x*-values and the vertical axis for the *y*-values. We represent each (*x*, *y*) pair by a dot on the rectangular coordinate system.

The following figure shows the scatter diagram for this data:




The pattern made by the points on the scatter diagram suggests a perfect linear relationship between monthly sales and monthly commission. The scatter diagram indicates that higher sales will result in higher commission.

**Example 2**

We want to analyze the relationship between production and manufacturing expenses. A sample of nine companies gives the following data.

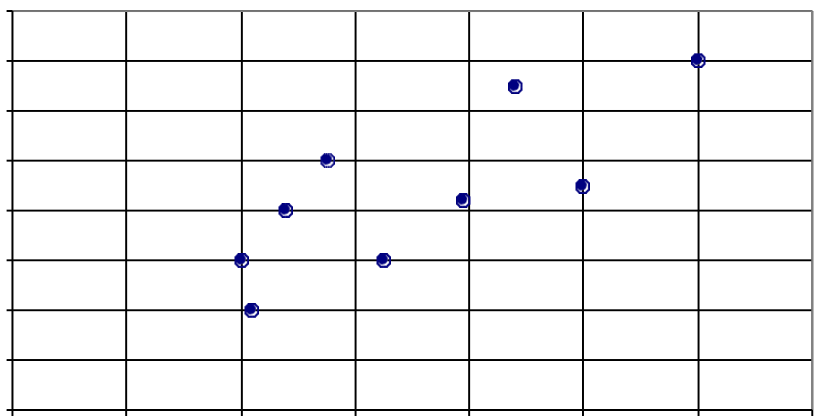
"Manufacturing expenses" is the dependent variable, *y*, and "production" is the independent variable, *x*. Manufacturing expenses change as the volume of production changes.

The value of *y* depends on the value of *x*.

| Production, *x* (in thousands of units) | 40 | 42 | 48 | 55 | 65 | 79 | 88 | 100 | 120 |
|---|---|---|---|---|---|---|---|---|---|
| Manufacturing expenses, *y* (in thousands of $) | 150 | 140 | 160 | 170 | 150 | 162 | 185 | 165 | 190 |

We use the horizontal axis for the *x*-values and the vertical axis for the *y*-values.
We represent each (*x*, *y*) pair by a dot on the rectangular coordinate system.

The following figure shows the scatter diagram for this data.



The pattern made by the points on the scatter diagram suggests a linear relationship between production and manufacturing expenses. The scatter diagram indicates that companies with higher production will have higher manufacturing costs.

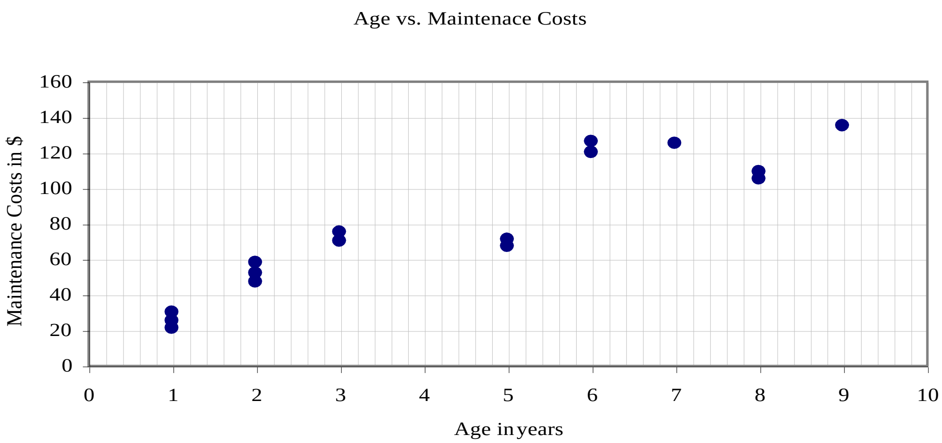
**Example 3**

Omar Suits wants to know the relationship between the age and the annual maintenance costs of sewing machines. A sample of 16 machines shows the following ages and maintenance costs during the past year.

| | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Age (years), $x$ | 8 | 3 | 1 | 9 | 5 | 7 | 5 | 2 | 1 | 3 | 6 | 2 | 1 | 2 | 6 | 8 |
| Maintenance costs, $y$ ($) | 109 | 75 | 21 | 135 | 67 | 125 | 71 | 52 | 25 | 70 | 126 | 58 | 30 | 47 | 120 | 105 |

"Maintenance costs" is the dependent variable, $y$, and "age" is the independent variable, $x$. Maintenance costs change as the age of the sewing machine changes. The value of $y$ depends on the value of $x$.

We use the horizontal axis for the $x$-values and the vertical axis for the $y$-values. We




represent each ($x$, $y$) pair by a dot on the rectangular coordinate system. The following figure shows the scatter diagram for this data.

.

The pattern made by the points on the scatter diagram suggests a linear relationship between maintenance costs and age. The scatter diagram indicates that older machines will have higher maintenance costs.

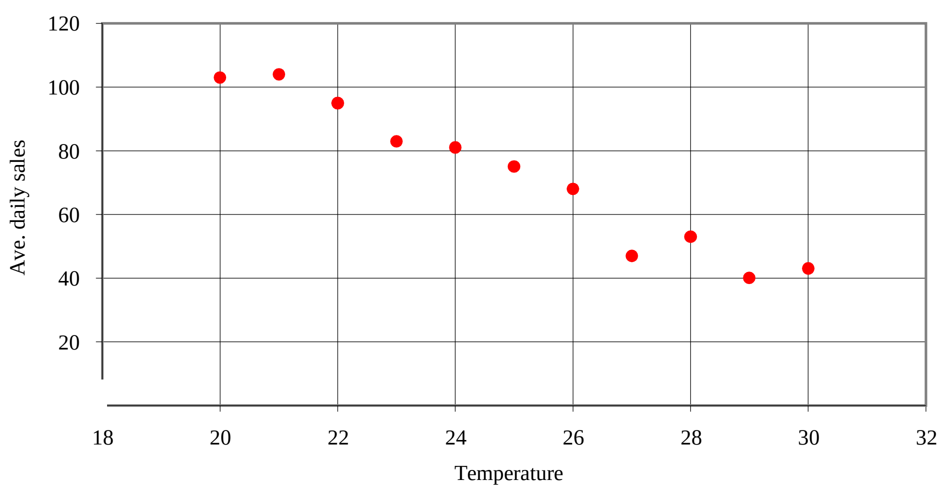
**Example 4**

Over a five-month period the average daily temperatures and the average daily sales of a certain product for each temperature are given in the following table.

| Temperature, $x$ (°C) | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Average daily sales, $y$ | 103 | 104 | 95 | 83 | 81 | 75 | 68 | 47 | 53 | 40 | 43 |

"Average daily sales" is the dependent variable, $y$, and "temperature" is the independent variable, $x$. Average daily sales of a certain product changes as the temperature changes. The value of $y$ depends on the value of $x$.

We use the horizontal axis for the $x$-values and the vertical axis for the $y$-values. We represent each $(x, y)$ pair by a dot on the rectangular coordinate system.
The following figure shows the scatter diagram for this data.




The pattern made by the points on the scatter diagram suggests a linear relationship between average daily sales and temperature. The scatter diagram indicates that the average daily sales will be less at higher temperatures.

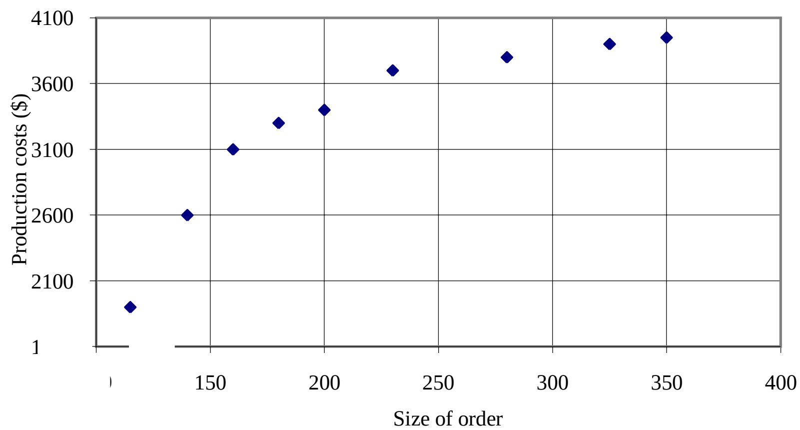
**Example 5**

A company wants to investigate the relationship between the size of order of a certain product and the cost of producing the order. The following data have been recorded for different sized orders.

| Size of order, *x* | 115 | 140 | 160 | 180 | 200 | 230 | 280 | 325 | 350 |
|---|---|---|---|---|---|---|---|---|---|
| Production costs, *y* ($) | 1900 | 2600 | 3100 | 3300 | 3400 | 3700 | 3800 | 3900 | 3950 |

“Production costs” is the dependent variable, *y*, and “size of order” is the independent variable, *x*. We use the horizontal axis for the *x*-values and the vertical axis for the *y*-values. We represent each (*x*, *y*) pair by a dot on the rectangular coordinate system.

The following figure shows the scatter diagram for this data



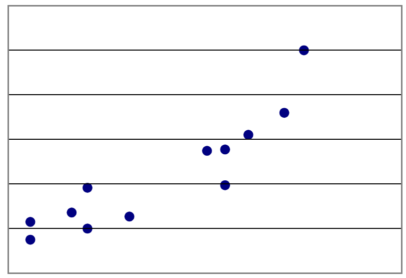

.

The pattern made by the points on the scatter diagram suggests a non-linear relationship. Production costs and the size of orders don’t increase at the same rate. As the size of the order increases the production costs per unit will increase at a lower rate.

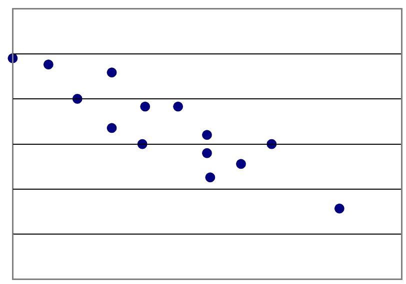
**Example 6**

Scatter diagrams will often show at a glance whether a relationship exists between two variables.

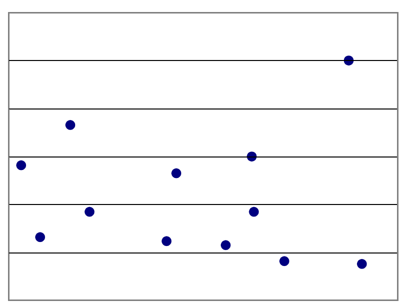
(a) Positive linear relationship



(b) Negative linear relationship



(c) No relationship



(d) Non-linear relationship