

Key

**Notes 12.6 Scatter Plots, Correlation, and Regression Lines**

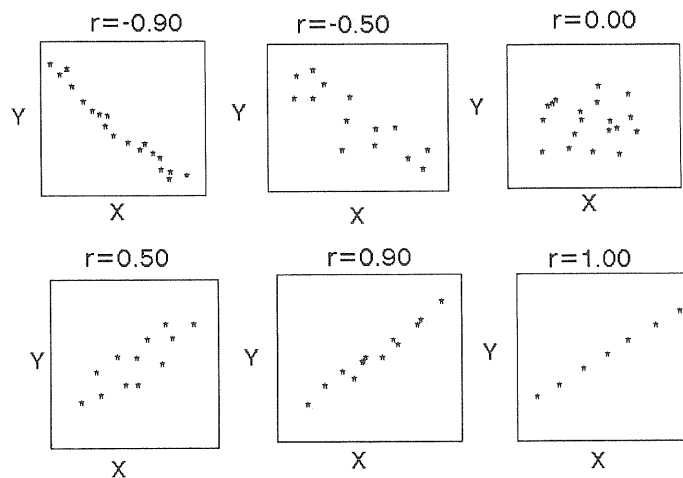
A **correlation** exists between two variables when the higher values of one variable consistently go with the higher values of another variable, or when the higher values of one variable consistently go with the lower values of another variable. A **scatter diagram** (or scatterplot) is a graph in which each point represents the values of the two variables.

A **correlation coefficient** ( $r$ ) is a number between -1 and 1. It describes the strength of the correlation between two variables. The strongest linear relationship is indicated with a correlation coefficient ( $r$ ) equal to -1 or 1. In this case, the points would all lie on a line. The weakest linear relationship is indicated with an  $r=0$ . In this case, the points would all be scattered.

**Positive correlation:** Both variables tend to increase (or decrease) together. ( $r$  is positive.)

**Negative correlation:** One variable increases while the other decreases. ( $r$  is negative)

**No Correlation:** There is no apparent relationship between the two variables. ( $r=0$ )



- 1-3 a) State whether the diagram shows a positive correlation, a negative correlation, or no correlation. Is the correlation weak or strong?
- b) Summarize any conclusions that you can draw from the diagram.

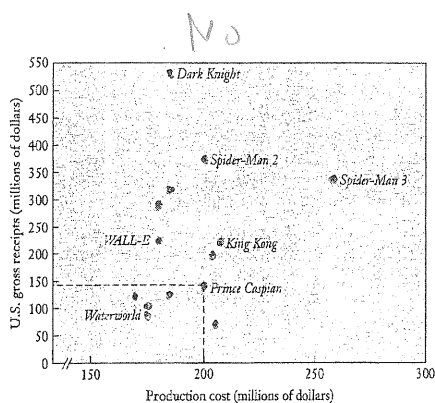


FIGURE 5.44 Scatter diagram for the data in Table 5.6.

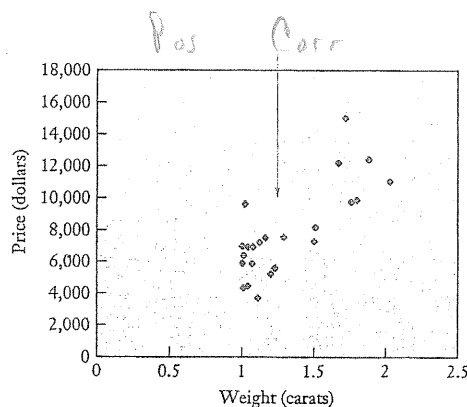


FIGURE 5.45 A scatter diagram for diamond weights and prices

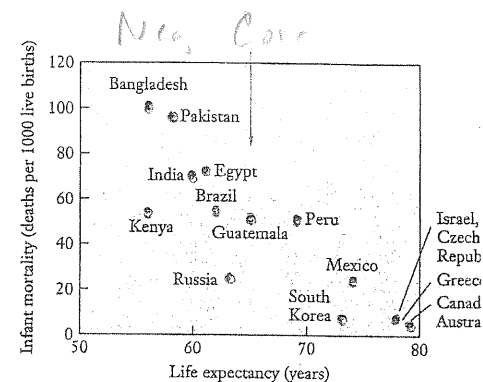


FIGURE 5.46 A scatter diagram for life expectancy and infant mortality

4-6 Are the following pairs of variables correlated? If so, is there a positive or negative correlation?

- 4) Amount of snow in Flagstaff and the number of people at Snow Bowl. *Correlated, positive*
- 5) The color of a person's hair and the amount of money a person earns. *not correlated*
- 6) The time spent on Facebook and a student's GPA. *Correlated, negative*

7-8 Think of two variables that would have a positive correlation and two that would have a negative correlation.

- + 7) *Time spent lifting weights / muscle mass* *HW / test grade*
- 8) *Amount of alcohol drink / score on driving obstacle course*  
*sleep in class / test score*

### Possible Explanations for a Correlation

- 1) The correlation is coincidence or casual connection.
- 2) Both variables might be directly influenced by some common underlying cause.
- 3) One variable may be a direct cause of the other.

9-11 For each statement below, state whether there is a positive or negative correlation. Then state if the correlation is due to coincidence, a common underlying cause, or a direct cause.

- 9) The more murders a city has, the more library books it has. *pos common underlying cause*
- 10) Stock prices go down when the number of scorpions go up. *neg coincidence* *higher pop.*
- 11) A man's weight increases as the number of calories he consumes increases. *pos / direct cause*

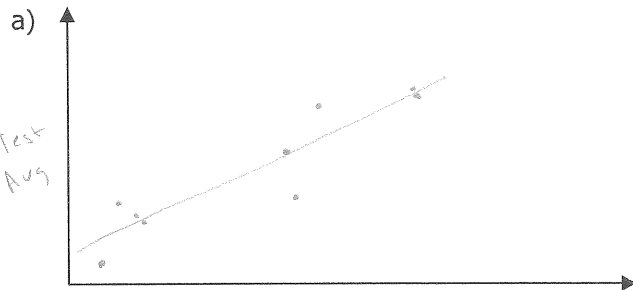
12-13 For each set of data:

- a) Draw a scatterplot using a graphing calculator.
- b) Find  $r$ .
- c) Find the equation of the line of best fit.
- d) Draw a conclusion.

**\*These are also the directions to your homework on problems (1 – 4) All \***

12) Below are the test averages and the homework averages of ten Year 4 students.

Student	Homework Average (x)	Test Average (y)
Rene	70	61
Thomas	100	95
Mark	10	44
Wanda	100	93
Judy	30	63
Haydn	68	80
Barbara	32	62
Karen	100	95
Mary	14	65
Phil	76	88



b)  $r = .895 = .9$  hw avg

c) Equation of Regression line:  $y = .45x + 47.71$

d) Conclusion: Pos strong correlation between hw average and test average.

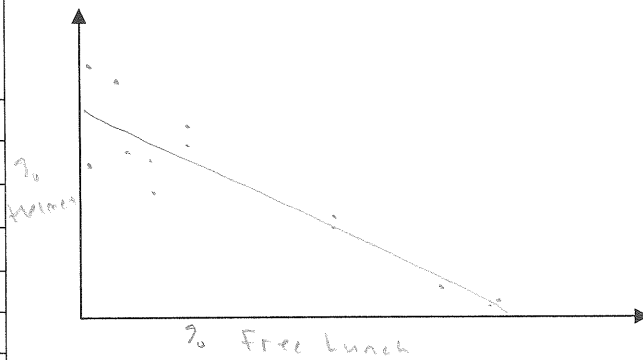
e) What would be the test average of a student whose homework average is 83%?

$84.9\%$

13) The table below shows the percentage of children in a neighborhood who receive reduced fee lunches and the percent of bike riders in that neighborhood that wear a bicycle helmet.

Neighborhood	% Receiving Reduced Fee Lunches (x)	% Wearing Bicycle Helmets (y)
Fair Oaks	50	22.1
Strandwood	11	35.9
Walnut Acres	2	57.9
Daisy Bay	19	22.2
Belshaw	26	42.4
Toyon	19	32.4
Kennedy	73	5.8
Cassell	81	3.6
Miner	51	21.4
Sedgewick	11	55.2
Sakamoto	2	33.3
Lietz	25	38.4

a) Sketch scatter plot from calculator



b)  $r = -.85$

c) Equation of Regression line:  $y = -.54x + 47.49$

d) Conclusion: Free lunch is neg correlated with wearing helmets

e) What percentage of people that wear helmets can we anticipate for a neighborhood where 65% of the people receive reduced fee lunches?

$y = 12.48$  wear helmets

## Scatter Plots with a Graphing Calculator

### How to Input Data

- 1) Press **Stat** then **Edit** (1)
- 2) Type in all values of  $x$  under  $L_1$  then type in all values of  $y$  under  $L_2$

### How to Graph a Scatterplot

(Steps 3-7 are done only once to prepare your GC to make scatter plots. Once is set, go from step 2 to step 8).

- 3) Press **Stat Plot** ( $2^{\text{nd}}$   $y=$ ) and select Plot 1
- 4) Arrow to **On** and press enter
- 5) Arrow Down to **Type** then arrow to the 1<sup>st</sup> graph and press enter
- 6) Arrow down to **XList** and type in  $L_1$  then arrow down to **YList** and type in  $L_2$
- 7) Arrow down to **Mark** then arrow to the 1<sup>st</sup> one (small box) and press enter
- 8) Now press **Zoom** then **9:ZoomStat**      Presto! There's your scatterplot!

### How to Find the Equation of the Line of Best Fit (regression line) and the Correlation Coefficient

\*First make sure that your calculator's diagnostic is activated. To do this press **Catalog** ( $2^{\text{nd}}$  0). Arrow down to **DiagnosticOn** then press enter twice.

- 9) Press **Stat** then arrow over to **Calc**
- 10) Arrow down to **4:LinReg (ax+b)** and press enter

You did it!

### How to Graph your Line of Best Fit

- 11) Press **y=** (clear anything that is there)
- 12) Press **Vars**
- 13) Arrow down to **5:Statistics** press enter
- 14) Arrow over to **EQ** then select **1:RegEQ**

You should now have the equation on the  $y=$  screen.

- 15) Press **graph** and you'll be done!