Statistics & Probability with the TI-89

Sample Activity: Exploration 7

Brendan Kelly

The students in Mr. Santos' class were asked to record the number of hours spent studying for their mathematics test. For each student, Mr. Santos wrote an ordered pair (*x*, *y*). The value of *x* represented the number of hours of study and the value of *y* was the student's mark on the test. Marsha's ordered pair was (3, 82) because she spent 3 hours studying and received a final mark of 82 out of 100. The set of ordered pairs which Mr. Santos recorded are shown on the right.

Ordered Pairs

(3.0, 82)	(5.5, 78)	(1.0, 60)	(4.9, 93)	(5.1, 86)
(2.5, 71)	(4.2, 90)	(0.5, 40)	(3.5, 88)	(7.0, 96)
(1.5, 73)	(2.4, 82)	(2.0, 53)	(6.2, 87)	(8.4, 100)
(2.6, 75)	(3.7, 85)	(5.4, 70)	(9.3, 89)	(7.6, 87)
(1.4, 48)	(0.5, 56)	(6.5, 85)	(2.3, 61)	(5.2, 74)
(1.0, 47)	(3.5, 87)	(8.2, 94)	(3.0, 86)	(5.4, 92)

WORKED EXAMPLE 1

Use the data given above to determine whether there is a relationship between the number of hours of study and the test mark.

SOLUTION

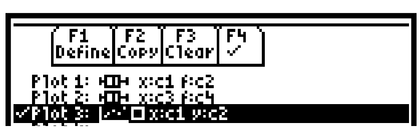
❶. Access the Data/Matrix editor and create the data variable **santos**.

Enter in c1 the first coordinates of all the ordered pairs and in c2, all the second coordinates, taking care to check that opposite each value in c1 is the corresponding second coordinate in c2.

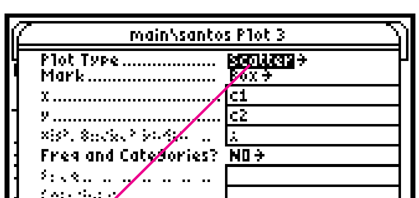
The display shows the first 7 rows of the table.

DATA		
	c1	c2
1	3	82
2	5.5	78
3	1	60
4	4.9	93
5	5.1	86
6	2.5	71
7	4.2	90

❷. Press **F2** to define the plot, and use the function key **F4** to deselect all other plots (shown by ✓). Then select the plot you wish to define.



Then press: **F1** and complete the dialog box as shown in the display.

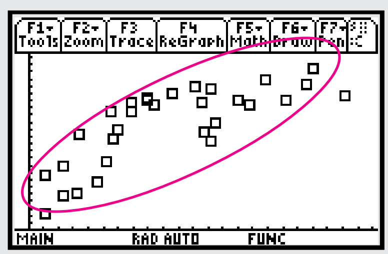


Note: We choose Plot Type, Scatter, to obtain a plot called a *scatter plot.*

❸. To obtain the scatter plot, we press:

◆ [GRAPH] followed by

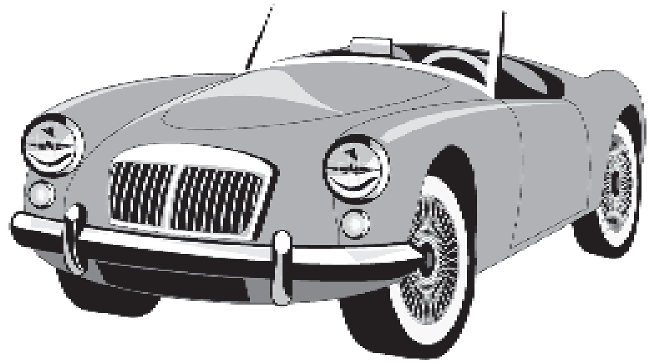
F2 **9** and we obtain:



This display shows that almost all the data points in the scatter plot can be contained in an oval which is inclined along an axis with positive slope. In such a case, the variables represented along the axes are said to be *positively correlated.* This scatter plot shows that there is a positive correlation between preparation time for a test and the test mark. When the data points cluster close to the axis of the oval, we say that the variables have a *strong positive correlation.* This scatter plot shows a fairly strong positive correlation.

WORKED EXAMPLES

AN INVESTMENT OR AN EXPENDITURE?



When an item which you purchase tends to increase in value, we usually refer to it as an *investment*. Conversely, an item which decreases in value, is called an *expenditure*. Understanding the difference between an investment and an expenditure is an important key to managing your personal finances. To determine whether a car is an expenditure or an investment, we must find out whether its value increases or decreases with its age.

The table displayed here shows for a popular model car, the resale value in dollars for each age between 1 and 10 years.

Age	Value	Age	Value
1	$15,975	6	$3,448
2	$ 9,285	7	$2,755
3	$ 8,000	8	$1,995
4	$ 6,790	9	$1,400
5	$ 3,150	10	$1,268

WORKED EXAMPLE 2

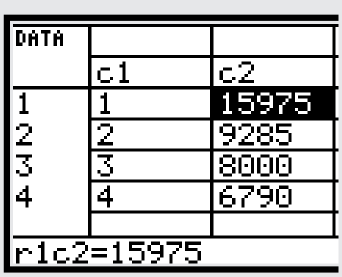
Use the data given in the table to determine whether there is a correlation between the age of a car and its resale value.

SOLUTION

By scanning the table, we see that the resale value of the car tends to decrease with increasing age, so there seems to be a relationship; i.e., a *correlation*, between these two variables. To determine the strength of the correlation, we may proceed as in worked example 1.

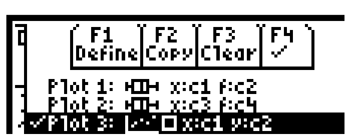
❶. Access the Data/Matrix editor and create the data variable **car**.

Enter in c1 the ages from 1 through 10, and in c2, the corresponding values, taking care to check that opposite each value in c1 is the correct value c2.



❷. Press **F2** to define the plot, and use the function key **F4** to deselect all other plots (shown by ✓). Then select the plot you wish to define.

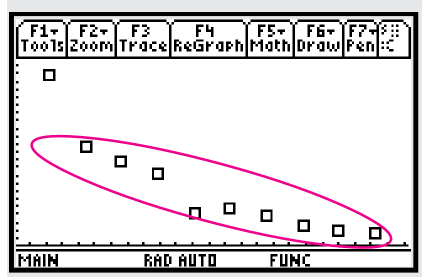
Then press: **F1** and define the plot as shown in Plot 3 of the display below.



❸. To obtain the scatterplot, press:

◆ [GRAPH] followed by

F2 **9** and we obtain:



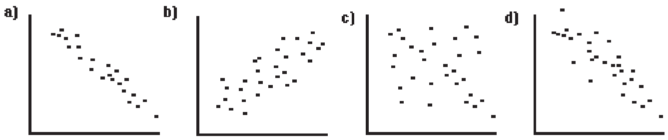
The display shows that almost all the data points in the scatter plot can be contained in an oval which is inclined along an axis with negative slope. In such a case we say that the variables plotted on both axes are *negatively correlated.* The scatter plot shows that there is a negative correlation between the age of a car and its resale value. This scatter plot shows a very strong negative correlation.

Exercises & Investigations

❶. Write in your own words the meanings of the following phrases:

- positive correlation
- strong positive correlation
- negative correlation
- strong negative correlation
- no correlation

❷. Use the phrases above to identify the kind of correlation which is characterized by each of the following scatter plots.



❸. The table below shows 30 ordered pairs. The first coordinate in each ordered pair is the age of a randomly selected married woman. The second component is the age of her spouse. Use the procedure in the worked examples to create a scatter plot which will reveal whether there is a correlation between the age of a woman and her spouse.

Ages of Women and their Husbands

(29, 34)	(37, 38)	(19, 20)	(57, 57)	(34, 32)
(23, 25)	(51, 54)	(72, 81)	(29, 23)	(70, 70)
(45, 54)	(39, 37)	(58, 56)	(64, 71)	(35, 35)
(42, 50)	(25, 24)	(37, 48)	(36, 36)	(27, 32)
(56, 42)	(17, 24)	(28, 28)	(57, 26)	(41, 39)
(47, 47)	(16, 18)	(24, 27)	(84, 87)	(55, 59)

If you discover a correlation, indicate whether it is positive or negative and describe it as strong or weak.

❹. A biologist who was studying crickets, hypothesized that the number of chirps per minute made by a cricket is strongly correlated to the outside temperature increases. He recorded the data shown in the table. Create a scatter plot to test his hypothesis.

Temperature in ˚ C	Chirps per min
17	105
18	110
19	110
20	126
21	126
22	130
23	130
24	152
24	156
25	160
26	170
27	171
28	175
29	196
30	212

❺. For each of the following pairs of variables, indicate what kind of correlation (if any) you would expect to exist. Explain why.

a) height and weight
b) age and personal savings
c) level of education and income
d) latitude of a U. S. city and its mean temperature in winter
e) travel time and speed of travel for a particular distance
f) A golfer's age and golf score

❻. In order to quantify the "degree" of correlation between two variables, Karl Pearson, one of the pioneers of statistics, defined the *correlation coefficient*, r between two variables x and y by the equation:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

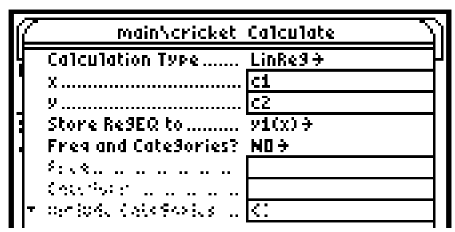
where y_i is the value of y corresponding to $x = x_i$. $i = 1 \ldots n$

a) Suppose that x and y are so strongly correlated that y is a linear function of x; i.e., $y = ax + b$ for some constants a, b. Prove algebraically that $|r| = 1$. How are x and y related when $r = 1$? How are they related when $r = -1$?

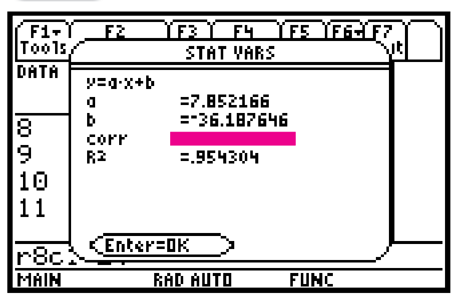
b) What range of values of r would indicate:
a strong correlation? a weak correlation? no correlation?

c) Create a data variable, cricket, and enter into your Data/Matrix editor, the temperatures and corresponding chirps per minute given in exercise 4.

Press **F5** and complete the **TwoVars** dialog box as shown below by selecting from the pop-up menus.



Pressing **ENTER** yields the following display.



Follow the above procedures to find the value of the correlation coefficient, r, (denoted by **corr**), between cricket chirps per minute and temperature.

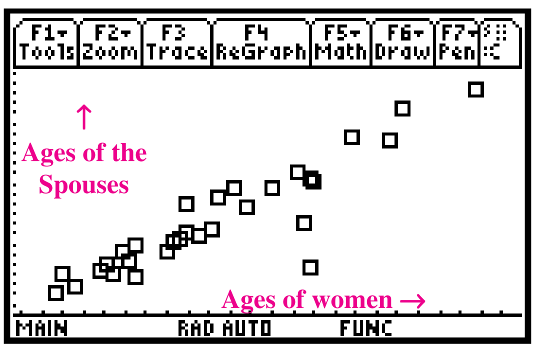
❼. a) Calculate the correlation coefficient between the ages of a woman and her spouse (See exercises 3 and 6.)
b) Calculate the correlation coefficient between hours of study and test mark presented in *Worked Example* 1.

Answers to the Exercises in Exploration 7

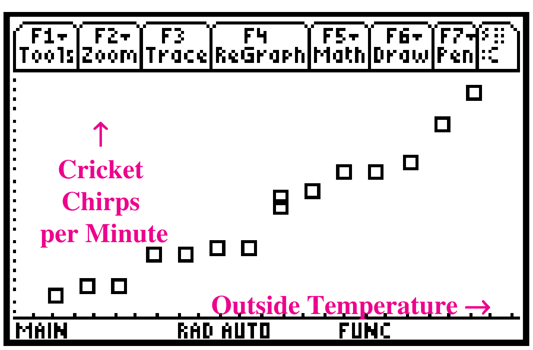
1. In general, two variables are said to be *positively correlated* if large values of each variable are associated with large values of the other variable. The degree of this correlation may be described as weak or strong depending upon how few violations of this relationship exist in the set of data points. Two variables are said to be *negatively correlated* if large values of each variable are associated with small values of the other variable. When the correlation between two variables is approximately linear, the data points in the corresponding scatter plot cluster along a straight line and most of the data points can be contained within an eccentric ellipse with the straight line as its major axis.
When two variables are not correlated, we say that they show *no correlation.*

2. a) strong negative correlation   b) positive correlation
c) no correlation   d) negative correlation

3. The scatter plot for the ages of women and their spouses is shown below. Since the data points cluster along a straight line of positive slope, we say the ages of women are positively correlated to the ages of their spouses. Since the clustering is close to a line, we would describe this relationship as a *strong positive correlation*.



4. The scatter plot for the number of cricket chirps per minute and the outside temperature is shown below.



This scatter plot shows that there is a strong positive correlation between the number of cricket chirps per minute and the outside temperature. The scientist's hypothesis seems to be valid. Some might suggest that advancing such an hypothesis without using a scatter plot wouldn't be cricket. We are inclined to chirp in agreement.

5. a) A positive correlation: Taller people will tend to be heavier than shorter people, so a large height is generally associated with a large weight.

b) A positive correlation: Substantial personal savings are usually associated with increased age since few young people have had the opportunity to accumulate wealth.

c) A positive correlation: The largest incomes are usually earned by people who are highly educated and the lowest incomes by people who have little or no formal education.

d) A negative correlation. Cities of large latitude are closer to the earth's poles than cities of smaller latitude and are therefore colder (lower mean temperature) in winter than cities of smaller latitude.

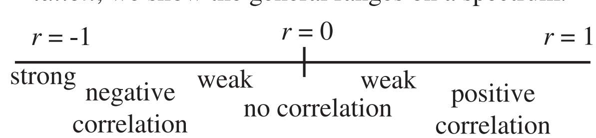
e) No correlation. For an individual, golf score may drop as the person improves during middle age. In older age this trend may reverse itself. However, if a data point consists of a randomly selected person's age and his (or her) golf score, then there would be little or no correlation expected, since golf scores would seem to be randomly distributed at every age level depending on experience and skill.

f) A strong negative correlation. Travel time for a particular trip is inversely proportional to the speed of travel. Since large travel times correspond to small travel speeds, these variables are negatively correlated. Since there are *no* deviations from this correspondence, the negative correlation is strong.

g) No correlation. Unless the olives are consumed with dry martinis, there is little reason to expect any correlation between these variables.

6. a) If $y = ax + b$, then $y_i = ax_i + b$ and $\bar{y} = a\bar{x} + b$, so $y_i - \bar{y} = a(x_i - \bar{x})$. Substituting into the formula for r, we obtain: $r = \frac{a}{|a|}$ and so, $|r| = 1$. That is, $r = 1$ when $y = ax + b$ and $a > 0$. $r = -1$ when $y = ax + b$ and $a < 0$.

b) Since there is no universally agreed range of r corresponding to the terms *strong correlation* and *weak correlation*, we show the general ranges on a spectrum.



The correlation coefficient between the number of cricket chirps per minute and the outside temperature is: $r = 0.9768\ldots$

7. a) The (linear) correlation coefficient between the age of a woman and the age of her spouse is: $r = 0.90716\ldots$.
b) The (linear) correlation coefficient between the hours of study and the test mark is $r = 0.7650\ldots$.
Note: Compare these coefficients with the correlation of 0.4 between the IQ's of parents and their children, or the correlation of 0.5 between the intelligence of siblings.