Solutions

1. Two variables have a linear correlation of –0.94. Which of the following is true?

   A The variables share a strong, positive correlation.

   B The variables share a moderate, positive correlation.

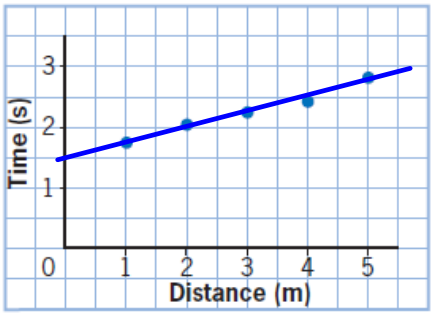
   C The variables share a strong, negative correlation.

   D The variables share a moderate, negative correlation.

C

For a correlation of r = -0.94, it is a negative correlation (r is negative) and strong (r value close to ±1)

2. The scatter plot shows the distance, d, in metres, and time, t, in seconds, for a student walking in front of a motion sensor.



Which equation represents the line of best fit?

A $d = -0.25t + 1.5$

B $d = 0.25t + 1.5$

C $d = 0.5t + 1.5$

D $d = 1.5t + 0.25$

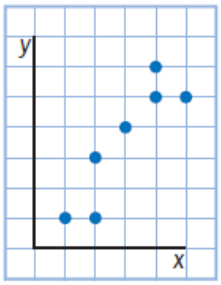
B

The line is sloping up to the right so it can't be A. The y-intercept looks like it is going to be at about 1.5 so it can't be D. We can draw in an approximate line of best fit and calculate the slope.

Use points (1,1.5) and (5, 2.75)

$$\text{Slope} = \frac{2.75 - 1.5}{5 - 1} = 1.25 \div 4 = 0.3125$$

3. The scatter plot shows the relationship between two variables, x and y. Which of the scenarios is most likely to have this relationship?



A student's math score versus student's height

B height above ground of a skydiver versus time

C computer boot-up time versus unemployment rate

D number of automobile accidents versus amount of snowfall

D

Most likely to be D because when there is more snowfall, there are more accidents. As x increase, y also increases.

A and C are unrelated. B would give a negative trend. As time increases, height above the ground would decrease.

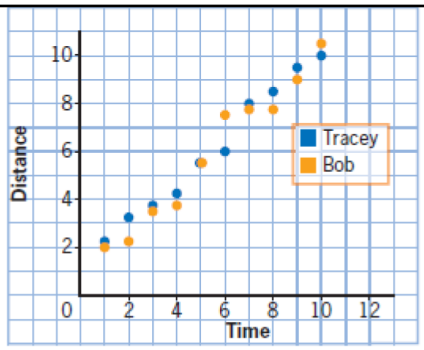
4. Compare the linear regression data for two students walking in front of a motion sensor:

Bob's walk	Tracey's walk
$d = 0.75t + 2$	$d = 0.75t + 2$
$r = 0.70$	$r = 0.95$

a) How are the movements of these walkers similar? different?

b) Sketch possible distance versus time graphs for Bob and Tracey.

b)



a) **Similar** - Both walkers are starting from a distance of 2 from the motion sensor and will have the same line of best fit for their data.

Different - As Tracey's r-value is closer to 1 it means her data is closer to the line of best fit compared to Bob's. Her speed is more consistent.

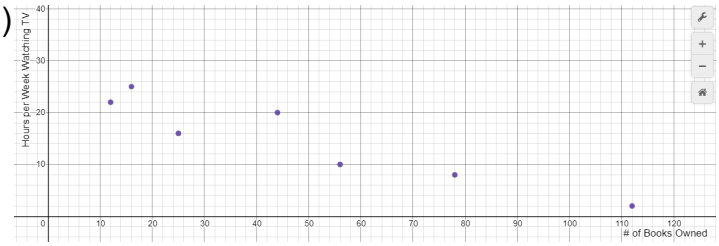
5. a) Create a scatter plot of time spent watching television versus number of books owned.

b) Characterize the correlation.

c) Perform a linear regression and record the correlation coefficient. Does this support your answer to part b)? Explain.

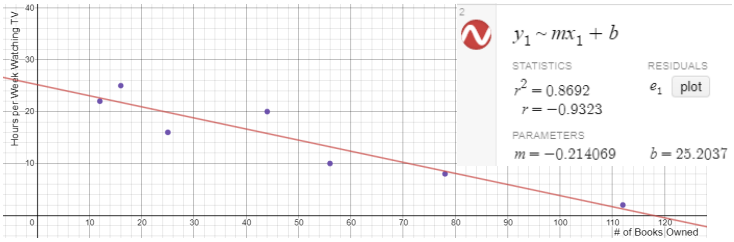
Number of Books Owned	Hours per Week Watching TV	Hours per Week Using Internet
25	16	8
44	20	12
12	22	16
16	25	13
78	8	20
112	2	15
56	10	11

a)



b) There is a strong, negative linear correlation. As the number of books owned increases, the number of hours spent watching TV decreases.

c)



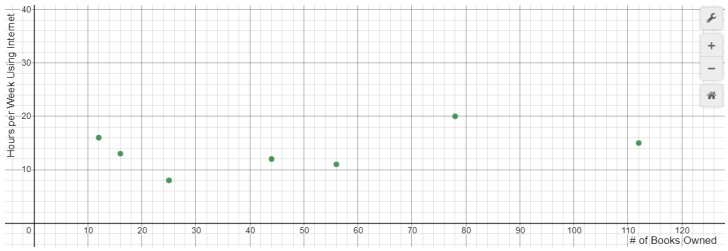
The r-value is -0.93 which reinforces our answer from (b) that there is a strong, negative linear correlation.

6. Repeat the analysis performed in #5 for

a) time spent on the Internet versus number of books owned

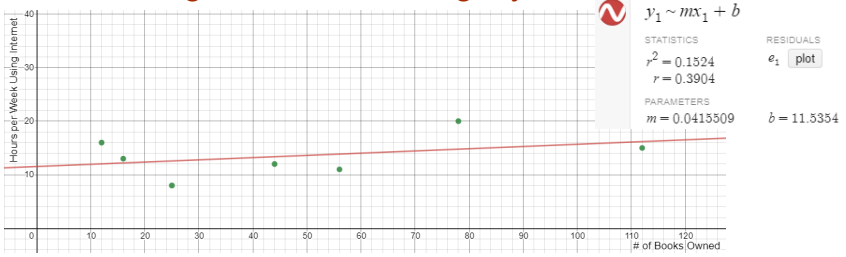
Number of Books Owned	Hours per Week Watching TV	Hours per Week Using Internet
25	16	8
44	20	12
12	22	16
16	25	13
78	8	20
112	2	15
56	10	11

a)



b) There is a weak/moderate, positive linear correlation. As the number of books owned increases, the number of hours spent watching TV increases slightly.

c)



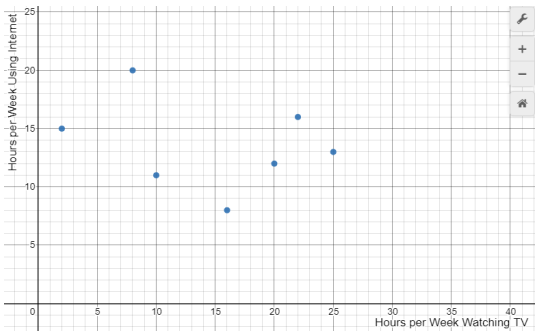
The r-value is 0.39 which reinforces our answer from (b) that there is a weak/moderate, positive linear correlation.

6. Repeat the analysis performed in #5 for

b) time spent on the Internet versus time spent watching television

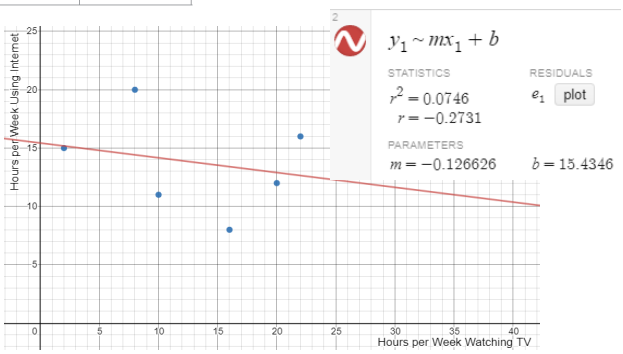
Number of Books Owned	Hours per Week Watching TV	Hours per Week Using Internet
25	16	8
44	20	12
12	22	16
16	25	13
78	8	20
112	2	15
56	10	11

a)



b) There is a weak, negative linear correlation. As the number of hours spent watching TV increases, the number of hours on the Internet decreases.

c)



The r-value is -0.27 which reinforces our answer from (b) that there is a weak, negative linear correlation.

9. Open Question The result of a linear regression between two variables is:

$y = -0.25x + 15 \qquad r^2 = 0.75$

a) Sketch what the scatter plot of y versus x could look like.

b) Explain the reasoning behind your sketch.

x_1	y_1
14	10.9
2	14.0
4	13.1
6	13.9
8	11.7
10	12.8
12	11.3

$y_1 \sim mx_1 + b$

STATISTICS

$r^2 = 0.7532$

$r = -0.8678$

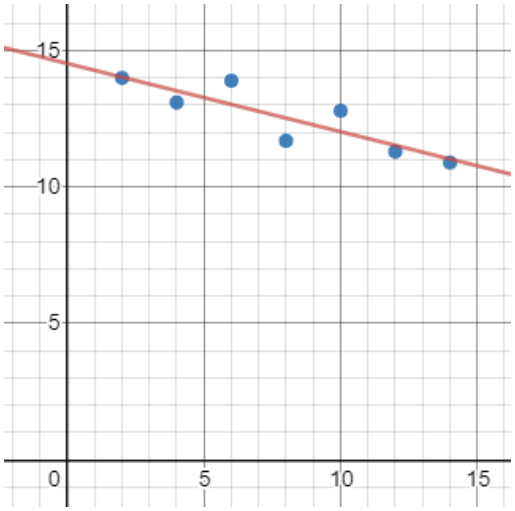
RESIDUALS

e_1 plot

PARAMETERS

$m = -0.25$ $b = 14.5286$

a)



b) This line has a slope (m) of -0.25, a y-intercept (b) of 14.5 and an r^2 value of 0.75. This implies that the start value is 14.5 and that for every increase of 1 for x, we get a decrease of 0.25 for y. The r^2 value of 0.75 means we will have an r-value of -0.866 which gives a strong, negative correlation.

13. Perform some research on the coefficient of determination, r^2.

a) How is it useful?

b) What values can it have?

c) How is it similar to the correlation coefficient? How is it different?

a) Many relationships between two variables follow patterns that are non-linear. Non-linear regression is another type of analytical technique used to find a model that provides a better fit for the given data. The coefficient of determination enables you to check how good a fit the model is for your data. The closer to 1, the better the fit.

b) The coefficient of determination can have values from 0 to 1.

c) Both of these coefficients tell us how good a fit the model is for the given data. However, the correlation coefficient (r) only applies to linear regression, where as the coefficient of determination (r^2) applies to any type of regression model (quadratic, cubic, exponential, logarithmic, sinusoidal, etc). They also each have different values where $-1 \leq r \leq 1$ and $0 \leq r^2 \leq 1$.