Dimensionless Technologies Private Limited
Visit us at: www.dimensionless.in
- info@dimensionless.in
- 9923170071, 8108094992

Notes for Students – Lesson 14

Two variables

→ Different design

- How we collect our data
- How we visualize our data

→ Analysis is similar

Two Variables that are probably related

Height and weight

Time spent studying and your grade

Temperature of the weather outside and ankle injuries

Dimensionless Technologies Private Limited
Visit us at: www.dimensionless.in
- info@dimensionless.in
- 9923170071, 8108094992

DIMENSIONLESS TECHNOLOGY

Q1.

x: Miles on a car	y: Value of the car
60,000	$ 12,000
80,000	$ 10,000
90,000	$ 9,000
100,000	$ 7,500
120,000	$ 6,000

X: predictor, explanatory, independent variable

Y:
- ☐ outcome
- ☐ determiner
- ☐ response
- ☐ stand-alone
- ☐ dependent

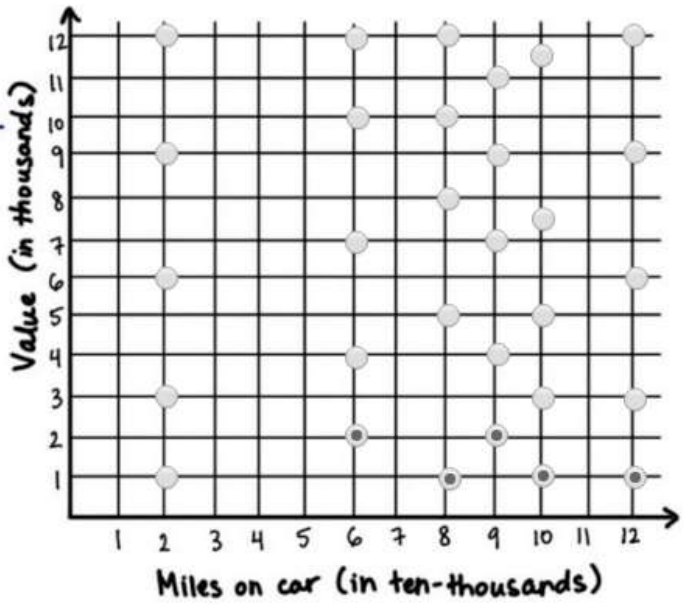
Q2.

Miles on a car	Value of the car
60,000	$ 12,000
80,000	$ 10,000
90,000	$ 9,000
100,000	$ 7,500
120,000	$ 6,000

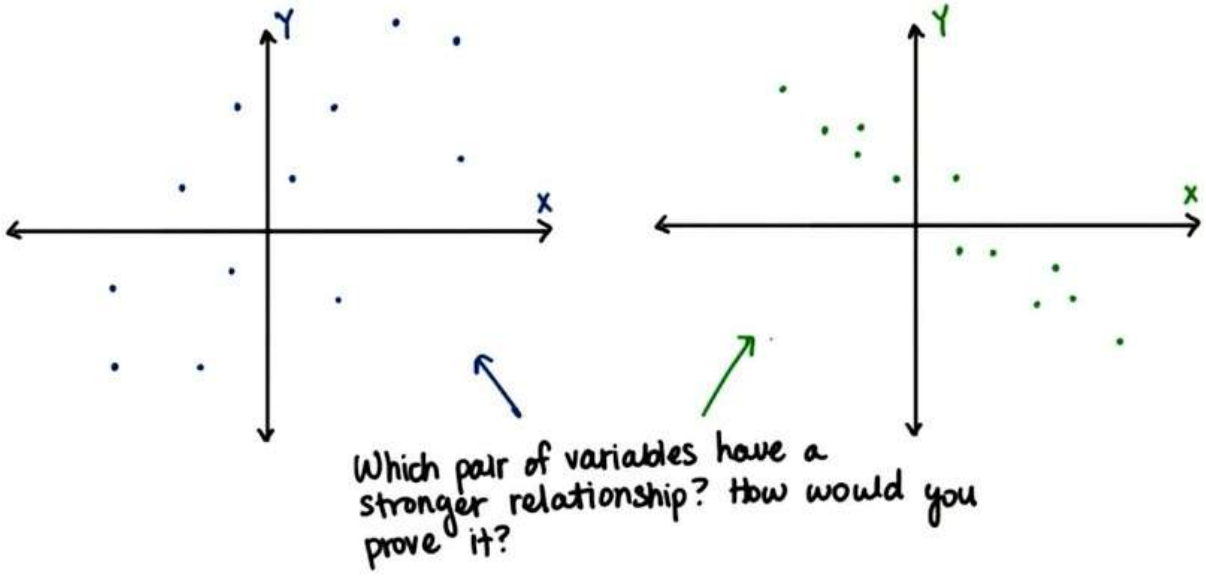
How can we show that these variables have a relationship?

Q3. Let's draw a scatter plot !!

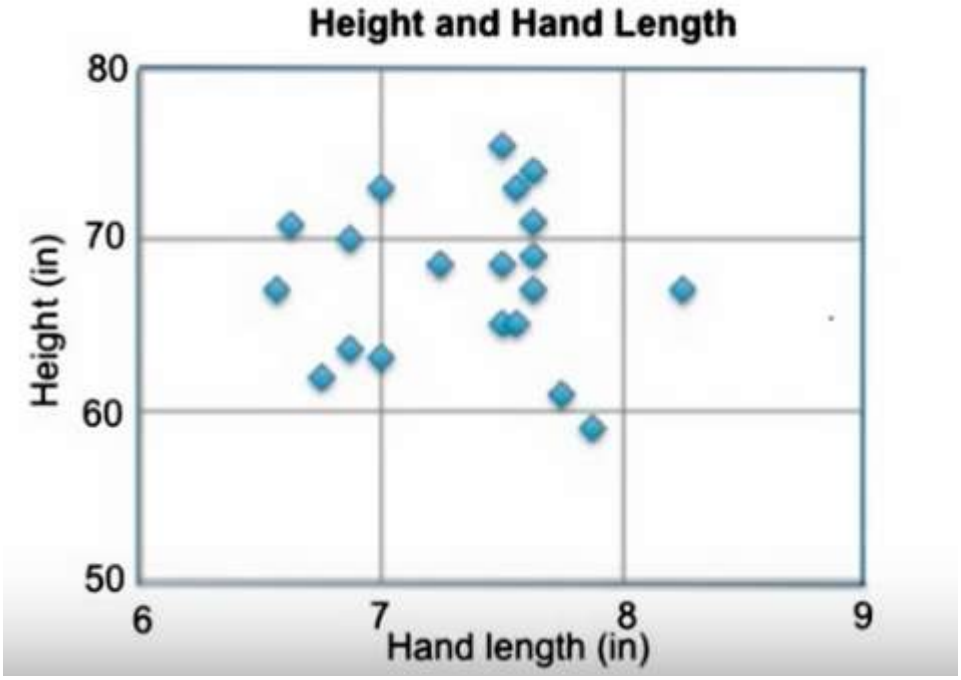
Miles on a car	Value of the car
60,000	$ 12,000
80,000	$ 10,000
90,000	$ 9,000
100,000	$ 7,500
120,000	$ 6,000

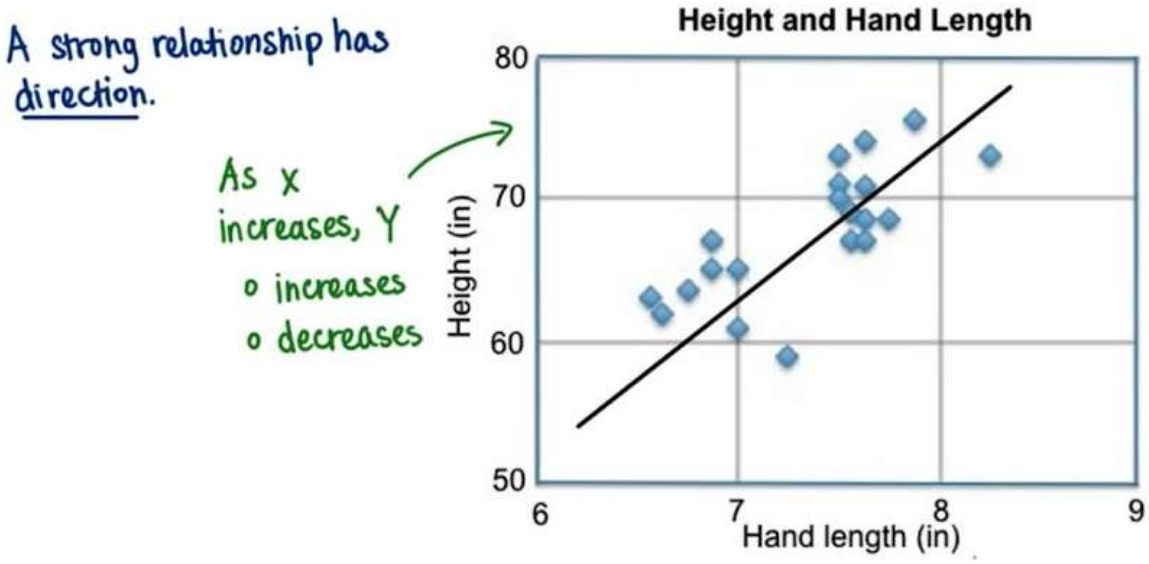


Q4.



Q5.

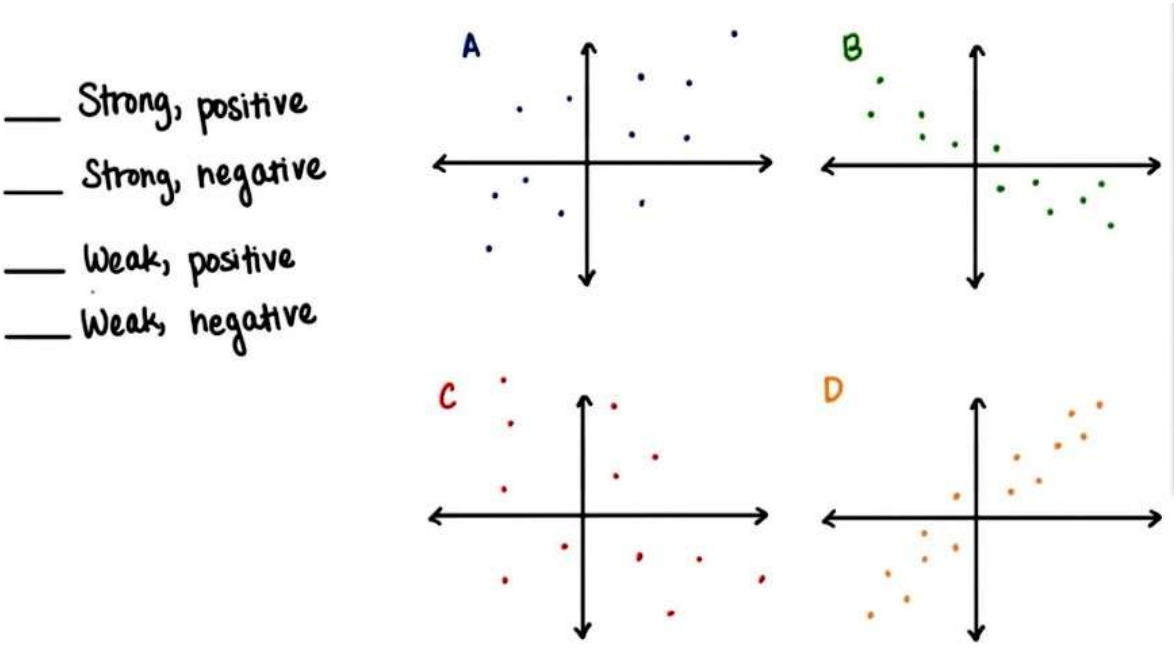

Height and Hand Length
80
70
60
50
Height (in)
6
7
8
9
Hand length (in)


A strong relationship has direction.
As x increases, Y
o increases
o decreases
Height and Hand Length
80
70
60
50
Height (in)
6
7
8
9
Hand length (in)

Q6.

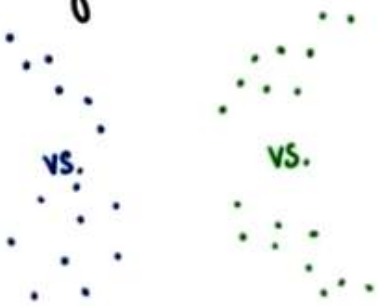
We often refer to the relationship between two variables as the correlation. If two variables have a strong relationship, then they have a high correlation. We can also say strong correlation.

Which of these has a strong positive correlation? Or strong in negative? Which shows a weak and positive correlation and finally which is a weak but negative correlation. Put the appropriate letter in the boxes here on the left.



Correlation Coefficient

Strength & Direction of Relationships



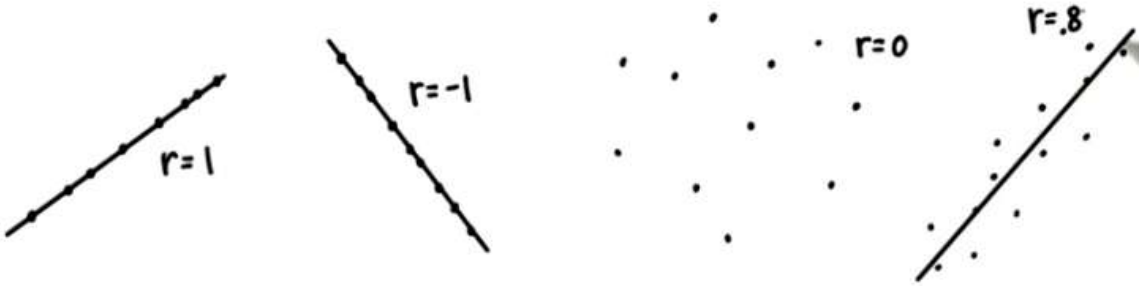
Pearson's r

Quantify relationship with correlation coefficient (r).

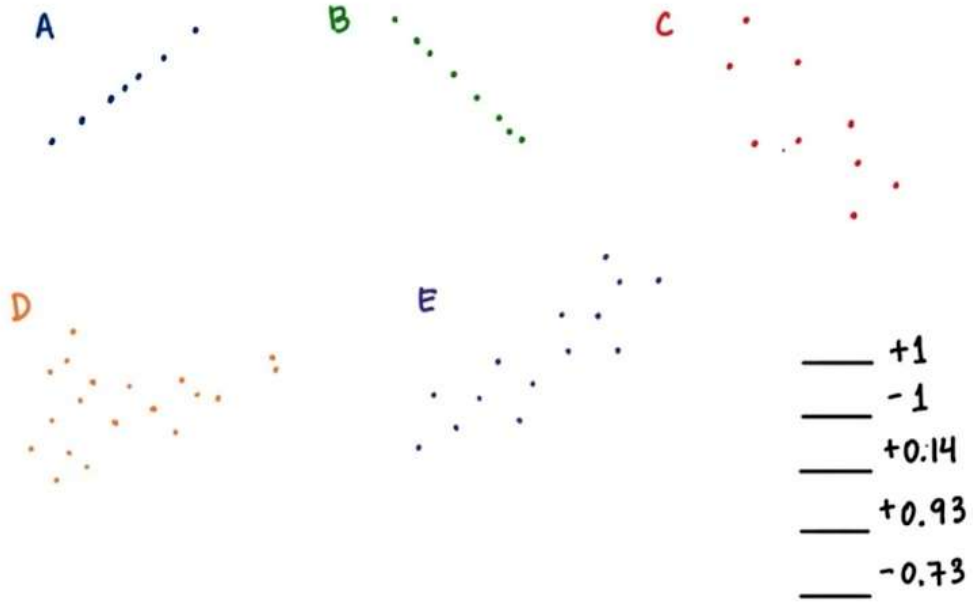
$$r = \frac{cov(x,Y)}{S_x \cdot S_y} = \frac{cov_{x,Y}}{S_x \cdot S_y}$$ (not a percentage)

r^2 = % of the variation in Y explained by the variation in x.

$$r = \frac{1}{n-1}\left(\frac{\sum_x \sum_y (x-\bar{x})(y-\bar{y})}{s_x s_y}\right)$$



Q7. Match with r.



Q8.

What are the correlation coefficients?

X = Age in years

Y = Age in months

r=

Q9.

What are the correlation coefficients?

X = Hours awake
Y = Hours asleep

r =

Calculating r in Excel

Q10.

= PEARSON(Start cell for variable x : end cell for variable x,
Start cell for variable y : end cell for variable y)

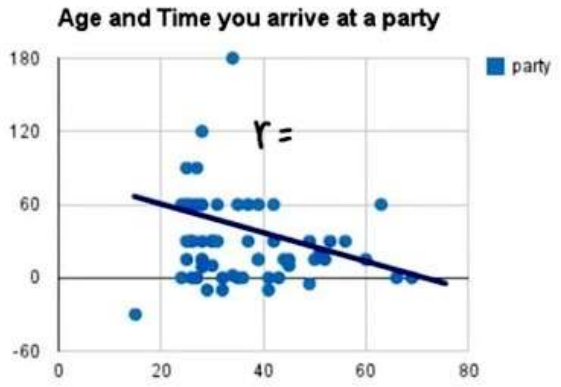
Click link below to see poll results.

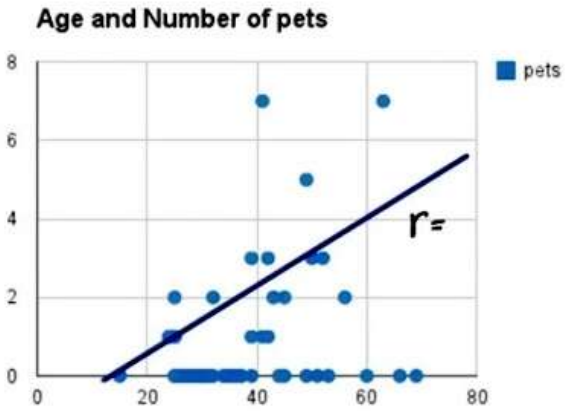
- Age, when you arrive at a party
- Age, number of pets

Create a scatterplot for each. Does it look like there is a relationship?

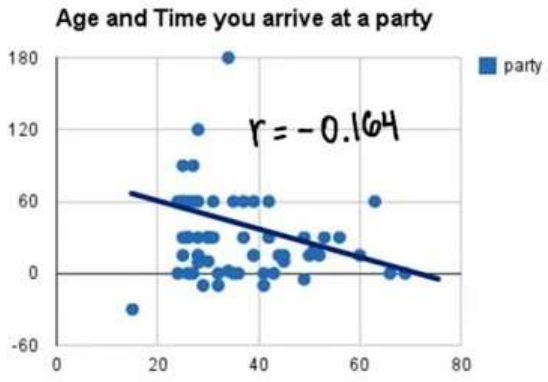
https://docs.google.com/spreadsheets/d/14Av6vXwfcC166cN6MoGyYg97KD8Tv1mGQyWgTBHr9go/edit?usp=sharing

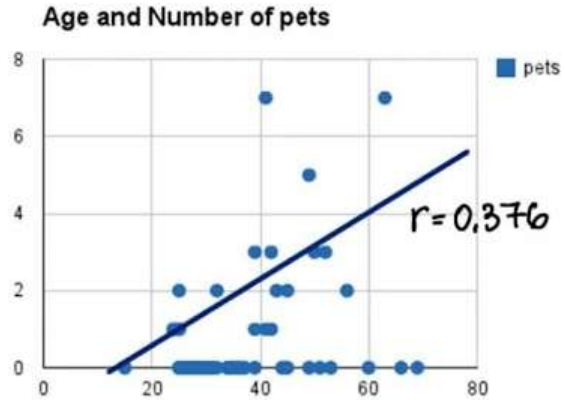
Q11.





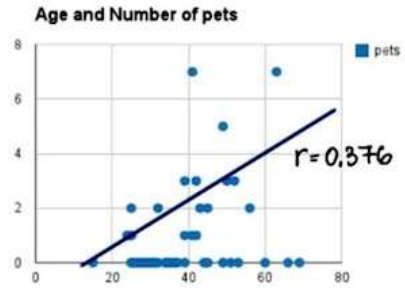
Q12. Which one is stronger ??





Q13.

Even though two variables might appear to have a relationship if we look at the sample data, this could just be due to chance. Just because our sample data works like it has a relationship, the actual population of people might not or vice versa. Our sample data might not have a relationship but the population actually does. Things like this naturally happen with sampling error as you've seen before. And we are surer as the sample size increases. The question is how much variation due to sampling error is allowable before we decide that our results are not due to chance? Again, we do Hypothesis Testing. Now, r is the correlation we get from our sample data. Rho,is the true correlation for the population. Given this what do you think the null and alternative hypothesis are?



ρ (rho) = true correlation for population.

Hypothesis Testing

$H_0: r = 0$
$H_A: r < 0$, $r > 0$, $r \neq 0$

$H_0: \rho = 0$
$H_A: \rho < 0$, $\rho > 0$, $\rho \neq 0$

$H_0: \rho < 0$, $\rho > 0$, $\rho \neq 0$
$H_A: \rho = 0$

$H_0: r < 0$, $r > 0$, $r \neq 0$
$H_A: r = 0$

Q14.

Hypothesis testing about the true population correlation is a type of t test. It uses the t distribution. Since we already know all about t tests, we know when to reject or fail to reject the null.

Example :-

Hypothesis Testing

$H_0: \rho = 0$
$H_A: \rho < 0$
$\rho > 0$
$\rho \neq 0$

ρ (rho) = true correlation for population.

$$t = \frac{r\sqrt{N-2}}{\sqrt{1-r^2}} \qquad df = N-2$$

	X	Y
1		
2		
⋮		
N		

25 participants (N = 25)
Non-directional test ($\alpha = 0.05$)
$t = 2.71$

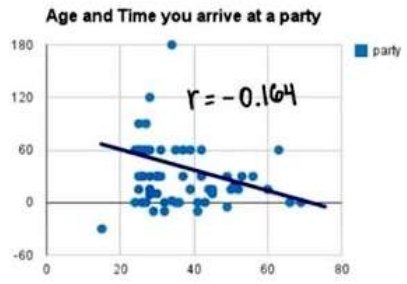
There { is / is not } evidence to reject the null;

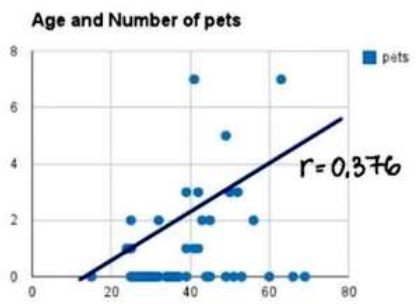
there { is / is not } a significant relationship between x and y.

Q15.

Since we know that the sample statistics often aren't exactly what the population parameters are, we might want a confidence interval for the true correlation. In other words, a range of likely values for the population correlation coefficient rho.

For the correlation between age and time you arrive at a party, the 95% confidence interval went from negative 0.3995, to positive 0.0914. And for age and number of pets, this ranged from 0.1369 to 0.5733.Based on these Confidence Intervals, what could you logically conclude? What would be your statistical decision for each?





$H_0: \rho = 0$
$H_A: \rho < 0$
$\rho > 0$
$\rho \neq 0$

ρ (rho) = true correlation for population.

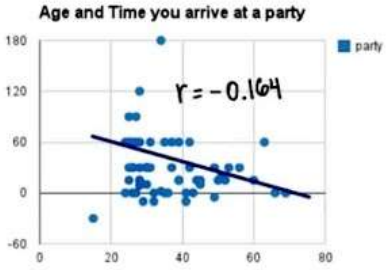
95% CI =
(-0.3995, 0.0914)

- Reject the null
- Fail to reject the null

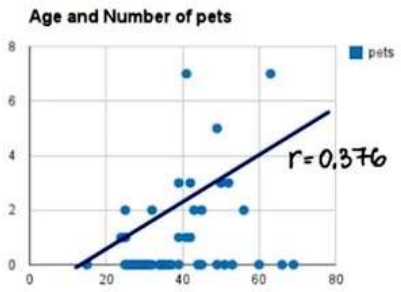
95% CI =
(0.1369, 0.5733)

- Reject the null
- Fail to reject the null

Q16.



95% CI =
(-0.3995, 0.0914)
Fail to reject the null
t = -1.2787 df = 59
p=



95% CI =
(0.1369, 0.5733)
Reject the null
t = 3.1146 df = 59
p=

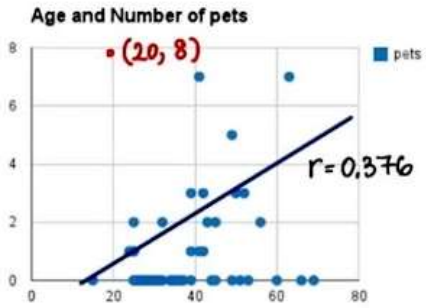
$H_0: \rho = 0$
$H_A: \rho < 0$
$\rho > 0$
$\rho \neq 0$

ρ (rho) = true correlation for population.

$\alpha = 0.05$

OUTLIER

Q17.



95% CI =
(0.1369, 0.5733)
Reject the null
t = 3.1146 df = 59
p = 0.0028

What is the new correlation coefficient when we add the point (20, 8) to the data?

Dimensionless Technologies Private Limited
Visit us at: www.dimensionless.in
- info@dimensionless.in
- 9923170071, 8108094992