The coefficient of determination

Tuesday, 26 February 2019 5:56 PM

★ By the end of the lesson I would hope that you have an understanding and be able to apply to questions the following concepts:

- Know what the Coefficient of Determination is
- Know how to find the coefficient of determination
- Know how to interpret what the coefficient of determination tells you

RECAP:

In a previous lesson we were introduced to the idea of Pearson's correlation coefficient.
This is basically a numerical value to give a measure of the linearity of a series of points on a scatter plot.

The values of r, Pearson's correlation coefficient, fall between -1 and 1.
The value of r can only be used to describe linear relationships.

We found that there were two ways of calculating Pearson's Correlation Coefficient:

- By hand (hard!), or
- Using a CAS

The formula, by hand was:

$$r = \frac{\sum(x - \bar{x})(y - \bar{y})}{(n-1)s_x s_y}$$

x: x data item
y: y data item
$\bar{x}$: Mean (or average) of the x data
$\bar{y}$: Mean (or average) of the y data
s_x: Standard Deviation of the x data
s_y: Standard Deviation of the y data

Using the CAS was much, much simpler.

This lesson is going to introduce a new term, **coefficient of determination.**

The coefficient of determination

It's really important to know what this value is, how to find it and, more importantly, how to use it in exam questions!

What is it:
The degree (given as a percentage value) to which one variable can be predicted from another related variable (linear).

How to find it:
This is really simple!
Once we know Pearson's correlation coefficient, we can use the following formula

$$coefficient\ of\ determination = r^2$$

Yup! It's that simple.
Take the value of the correlation coefficient and square it.

Note: This means the value of the coefficient of determination can **never be negative**.

Writing the right thing in a report
It's important to make sure you understand which variable is the explanatory variable and which is the response variable. Once you know this, and you have the value of r^2 then you have boilerplate sentences to complete for some pretty easy marks!

sentences to complete for some pretty easy marks!

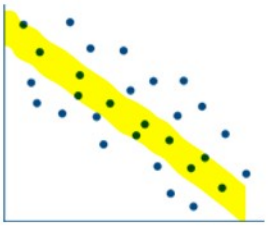
Examples

The following examples have been taken, with permission, from the Cambridge Further Maths Units 3 and 4 Textbook

Example 1

For the relationship described by this scatterplot, the coefficient of determination = 0.5210.

Determine the value of the correlation coefficient, r.



Note:
There is a trick in this question as we can have two values of r from a value of r^2.
You need to look at the graph and work out which value is the correct one.

Example 2

Carbon monoxide (CO) levels in the air and traffic volume are linearly related, with:

$r_{CO level, traffic\ volume} = +0.985$

Determine the value of the coefficient of determination, write it in percentage terms and interpret. In this relationship, traffic volume is the explanatory variable.

Note: The boiler plate response is:

The value of the coefficient of determination suggests that (**percentage**)% of the variation observed in the (**response variable**) can be explained by the variation in the (**explanatory variable**)

Report response:
*The value of the coefficient of determination suggests that **97%** of the variation observed in **carbon monoxide levels in the air** can be explained by the variation in **traffic volume**.*

Example 3

Scores on tests of verbal and mathematical ability are linearly related with:

$r_{mathematical,verbal} = +0.275$

Determine the value of the coefficient of determination, write it in percentage terms, and interpret. In this relationship, verbal ability is the explanatory variable.

$r = 0.275$

$r^2 = 0.075625 \times 100$

$r^2 = 7.56\%$



VCAA Exam Question on this concept
2016 Paper 2

app

act

c. The coefficient of determination for the association between the variables *apparent temperature* and *actual temperature* is 0.97

Interpret the coefficient of determination in terms of these variables. 1 mark



VCAA Exam Question on this concept
2017 Paper 2

$r = 0.862 \quad r^2 = 0.743$

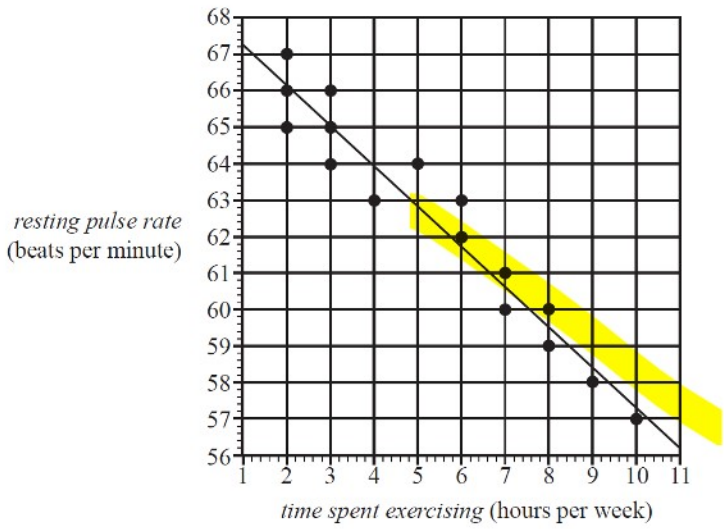
iv. The correlation coefficient is $r = 0.862$

Determine the percentage of the variation in *egg density* in the forest explained by the variation in the *number of female moths* caught in the trap.

Round your answer to one decimal place. 1 mark

74.3%

The scatterplot below displays the *resting pulse rate*, in beats per minute, and the *time spent exercising*, in hours per week, of 16 students. A least squares line has been fitted to the data.



$r^2 = 0.8339$

$r = \pm 0.913$

-0.913

VCAA Exam Question on this concept
2018 Paper 1

Question 9

The coefficient of determination is 0.8339

The correlation coefficient r is closest to

A. –0.913
B. –0.834
C. –0.695
D. 0.834
E. 0.913

VCAA Exam Question on this concept
2018 Paper 2

e. The value of the correlation coefficient r is 0.92

What percentage of the variation in the evening congestion level can be explained by the variation in the morning congestion level?

Round your answer to the nearest whole number. 1 mark

85%.

$r = 0.92$

$r^2 = 0.8464$

$r^2 = 84.64\%$