MATHS 911

MATHS Exam Paper 3

Correlation Coefficient

LIBERTY LIFE

Assessment Standard: 12.4.1 (e)
Calculate the Correlation Coefficient

Use available technology to calculate the correlation coefficient of a set of bivariate numerical data to make relevant deductions.

In this module we will be using examples and data used in the module on The Regression Function (published last Tuesday).

Correlation analysis measures the strength of a linear association between two variables x and y. At school level we work with a measure called the Pearson's correlation coefficient, (r), which lies between -1 and 1. That is

$$-1 \leq r \leq 1$$

Collect your Paper 3 Lessons every week!!

Guys, both NSC and IEB examinations candidates have the option of writing Paper 3 at the end of the year! Paper 3 covers additional mathematics material and is out of 100 marks. Maths Paper 3 will really set you apart in the job market, and make studying technical subjects at tertiary level easier. We have hooked you up with these lessons - written by IEB Maths Paper 3 examiner Heather Frankiskos. Though the lessons apply to both IEB and NSC candidates, where there are differences, we will point them out! The lesson this week applies to candidates from both examining bodies. **Give it a go!**

We interpret the value of r as follows:

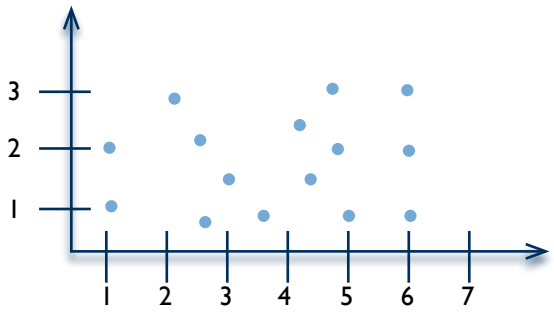
	value of r	Interpretation
1)	r = 0	No relation between x and y. Data is randomly scattered (Figure 1)
2)	r = -1	Perfect negative linear correlation. All data points lie on a straight line. As x increases y decreases (Figure 2).
3)	r = 1	Perfect positive linear correlation. All data points lie on a straight line. As x increases y increases. (Figure 3).
4)	-1 < r < 0 eg: -0,89 = r -0,56 = r	Negative linear correlation . A strong negative linear relationship. A moderate negative linear relationship.
5)	0 < r < 1 eg: 0,32 = r	Positive linear correlation. A weak positive linear relationship. (Figure 4 - Points not clustered so closed to the trend line).

PLEASE NOTE: PAPER 3 IS AN OPTIONAL EXAMINATION PAPER, YOU DO NOT HAVE TO WRITE IT. WE HIGHLY ENCOURAGE YOU TO ATTEMPT PAPER 3, IF IT IS NOT AVAILABLE AT YOUR SCHOOL, ASK YOUR TEACHERS ABOUT THE POSSIBILITY OF WRITING IT.

Bubblegum ©2009

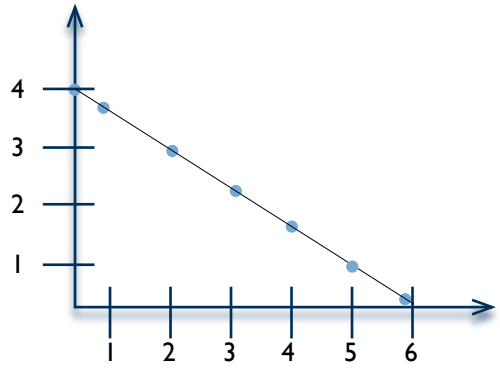
Correlation Coefficient

Figure 1



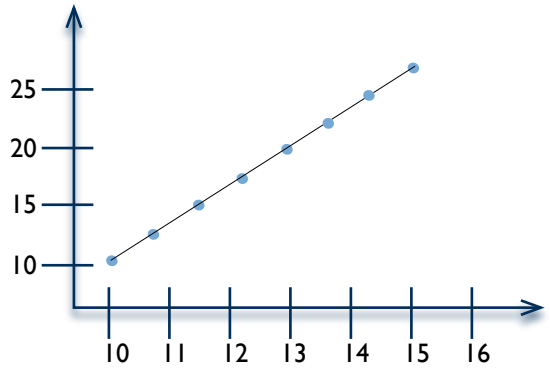
r = 0

Figure 2



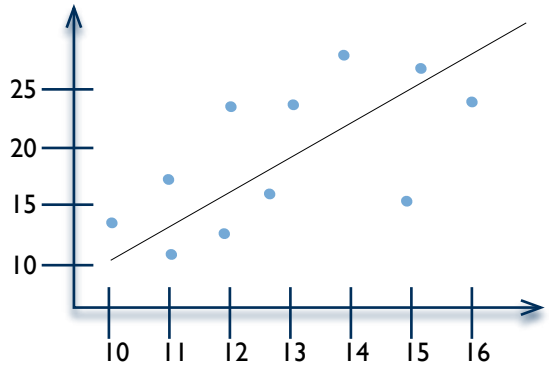
r = -1

Figure 3



r = 1

Figure 4



r = 0,32

PLEASE NOTE: PAPER 3 IS AN OPTIONAL EXAMINATION PAPER, YOU DO NOT HAVE TO WRITE IT. WE HIGHLY ENCOURAGE YOU TO ATTEMPT PAPER 3, IF IT IS NOT AVAILABLE AT YOUR SCHOOL, ASK YOUR TEACHERS ABOUT THE POSSIBILITY OF WRITING IT.

Bubblegum ©2009

Note that:
a) A weak correlation might suggest that a nonlinear relationship might fit better.
b) A correlation does not necessarily imply a cause and effect relationship - it is just an observation.
c) The r value will be more accurate for a bigger sample.

Using our data input for our Headmaster example in the previous module,when we get the screen from which we selected the values for $\hat{y} = A + Bx$ the screen looked like this.

1: A	2: B
3: r	4: $\hat{x}$
5: $\hat{y}$	

For the correlation coefficient (r) we choose option 3 : r. For our data we get r = 0,9090362582

This value suggests a strong positive linear relationship. It also suggests that for a better mark in Mathematics one expects a better mark in English.

Using the data for Activity 1 we calculate:
r = 0,9 371 00053 (a strong positive linear correlation)

Using the data for Activity 2 we calculate:
r = 0,92 (a strong positive linear correlation)

Using the data for Activity 3 we calculate:
r = -0,8 (a strong negative linear correlation)

Are you ready for exam type questions?
These will combine regression analysis and the correlation coefficient.

Question 1

A varsity student who waits on tables at a local restaurant recorded the cost of meals for two people (to the nearest rand) and the tip left (to the closest rand)

Meal Cost (R)	(x)	60	75	53	107	49	121	65	137
Tip (R)	(y)	5	6	6	13	5	10	7	13

a) Use your calculator to find the line of best fit /the regression function.
b) Using your answer from a) above, calculate how much the waiter should expect to be given as a tip if a meal costs R100.
c) Determine the correlation coefficient.
d) What can be deduced from your answer in c)?
e) Do you think that the regression function that you found in a) above would be a reliable predictor for a meal costing R400?

PLEASE NOTE: PAPER 3 IS AN OPTIONAL EXAMINATION PAPER, YOU DO NOT HAVE TO WRITE IT. WE HIGHLY ENCOURAGE YOU TO ATTEMPT PAPER 3, IF IT IS NOT AVAILABLE AT YOUR SCHOOL, ASK YOUR TEACHERS ABOUT THE POSSIBILITY OF WRITING IT.
Bubblegum ©2009

Solution

a) $\hat{y} = 0,422 + 0,092x$
b) $\hat{y} = 0,422 + 0,092\ (100) = R9,62 \approx R10$
c) $r = 0,91$
d) There is a high positive correlation so the predicted values found from the regression equation could be useful.
e) No - this x value seems way bigger than the data used. Most likely at this restaurant a bill for two people would not come to R400. Extrapolation is usually unreliable.

Question 2

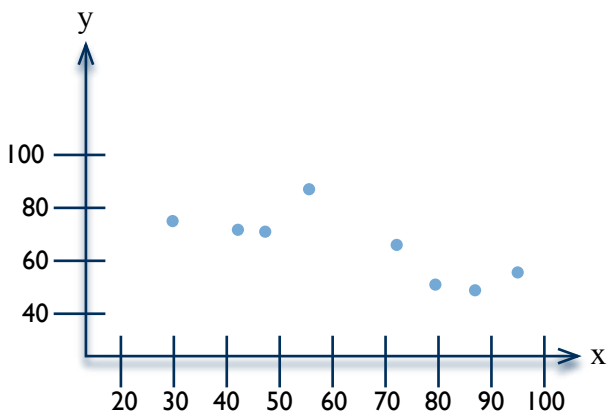
The following data for 8 learners for two tests is given.

x	80	73	95	43	56	30	48	88
y	56	61	47	72	68	78	75	54

a) Draw a scatter plot for this data.
b) Find the equation of the line of best fit using your calculator.
c) What does the sign (+/-) of the gradient tell you?
d) What mark do you predict for Richard who got 72% for test x?
e) Calculate the correlation coefficient.
f) What can be deduced from your answer in e)?
g) What mark can you expect a learner to have achieved for test x if he achieved 85 for test y?

Solution

a)



Please note:

The value for the gradient of the regression function (B) is - 0,47. This means that for every 1 mark increase in test x, we expect a 0,47 mark decrease in test y.

The value for r - (the correlation coefficient) is -0,983. This means that there is a strong negative linear correlation - the data points were clustered close to the regression function found.

b) $\hat{y} = 93,88 - 0,47x$
c) It is negative which means that it is a decreasing function: - as x increases y decreases
d) $\hat{y} = 93,88 - 0,47\ (70) = 60,04\%$
e) $r = - 0,983$
f) r value indicates a strong negative linear relationship and also most likely a reliable predictor.
g) We cannot use this regression function to calculate an x value from a given y, only a y value from a given x.

PLEASE NOTE: PAPER 3 IS AN OPTIONAL EXAMINATION PAPER, YOU DO NOT HAVE TO WRITE IT. WE HIGHLY ENCOURAGE YOU TO ATTEMPT PAPER 3, IF IT IS NOT AVAILABLE AT YOUR SCHOOL, ASK YOUR TEACHERS ABOUT THE POSSIBILITY OF WRITING IT.
Bubblegum ©2009