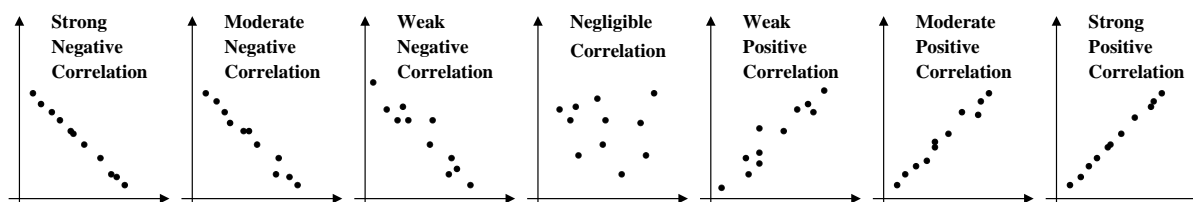### 2.1 The Product Moment Correlation Coefficient, PMCC

Given a sample a bivariate data, the Product Moment Correlation Coefficient, the PMCC, is a measure of how strong the correlation is in the sample.

Denoted r, it can have any value between -1 and $+1$ and, from the Year 1 course, there is a sense that if r is close to unity this indicates strong correlation in the parent population from which the sample was taken. On the other hand, if r is close to zero this suggests negligible correlation in the parent population.

The correlation in the parent population is denoted ρ (the Greek letter rho) and what is sought is information about ρ from the calculated value of r of the sample.

As is often the case in statistics, having only a few items of data gives rise to weird results, obviously wrong. Taking this to an extreme, consider a sample of just two pieces of bivariate data. With only two points on the scatter graph and a line of best fit between them, that line passes perfectly through the two points. This gives perfect correlation (either -1 or $+1$) but is clearly not yielding anything useful. A sample of size two can never have anything other than perfect correlation !

Intuitively, for every sample size there must be a value of r, below which there is, say, a 10% likelihood of there being negligible correlation for ρ. And a value of r below which there is a 5% likelihood of there being negligible correlation for ρ. And a value of r below which there is a 1% likelihood of negligible correlation for ρ.

For many sample sizes, these values of r at the various significance levels are known and are presented in the tables, "Critical Values for Correlation Coefficients".

From the tables for a sample size of 8 (and a one-tailed situation)
• r needs to exceed 0.7887 to be over 99 % confident there is some correlation, ρ
• r needs to exceed 0.6215 to be over 95 % confident there is some correlation, ρ
• r needs to exceed 0.5067 to be over 90 % confident there is some correlation, ρ

All of these ideas are formalised using previous knowledge of Hypothesis Testing.

2.2 Example

A researcher is investigating the effect of high altitude on mental capacity.
The researcher believes that mental capacity falls as height is gained.
As a mountaineer ascends a mountain, they periodically stop and perform
a psychometric test.

height, x, metres	3142	3453	4023	5062	4568	3712	2048
score, y, %	83	78	87	62	53	78	71

(**i**) Use your statistics calculator to determine the Product Moment
Correlation Coefficient for the data.

[2 marks]

(**ii**) Test, at the 5 % level of significance, whether there is evidence of
a negative correlation between altitude and mental capacity.
State your hypotheses clearly.

[4 marks]

2.3 Exercise

*Any solution based entirely on graphical
or numerical methods is not acceptable*

Marks Available: 42

Question 1

An examination board sets two mathematics examinations each summer.
In Paper 1 calculators are allowed but for the second paper they are not.
The examinations board believes that students scores in the two papers
are positively correlated, so that when a student misses one of the two
examinations an estimated score can be derived from their performance
in the other.

To investigate this alleged positive correlation 10 students are selected
and their scores in the two papers are presented in the following table;

Paper 1, x, %	57	69	87	66	42	93	35	46	77	68
Paper 2, y, %	63	72	92	75	54	81	42	28	65	73

(**i**) Use your statistics calculator to determine the Product Moment
Correlation Coefficient for the data and also the equation of the
line of best fit.

[2 marks]

(**ii**) Test, at the 1 % level of significance, whether there is evidence of
a positive correlation between the performances in the two papers.
State your hypotheses clearly.

[4 marks]

(**iii**) For a student who has missed Paper 2 but scored 61 % in Paper 1,
what score would be predicted for the missing paper using the
part (**i**) regression line.

[1 mark]

(**iv**) Mavis has scored 18 % in Paper 1 but was absent for Paper 2.
What do you advise the examination board do ?
Give a reason for your answer.

[2 marks]

Question 2

As part of a survey in a particular profession, age, x years, and yearly salary, £y thousands, were recorded.
The values of x and y for a randomly selected sample of ten members of the profession are as follows;

x, years	30	52	38	48	56	44	41	25	32	27
y, (£1000s)	22	38	40	34	35	32	28	27	29	41

(**i**)   Calculate to 3 decimal places, the product moment correlation coefficient between age and salary.

[2 marks]

It is suggested that there is negligible correlation between age and salary.

(**ii**)   Test this suggestion at the 5% significance level, stating your null and alternative hypotheses clearly.

HINT : Be careful - this is a TWO tailed test

[4 marks]

Question 3

It is known that there is a high positive correlation between being in bed an dying; most people die in bed.
Does this mean you can avoid dying by never being in a bed ?
Give a reason for your answer.

[2 marks]

Question 4

In the town of Shrewsbury a weekly local newspaper called *The Shrewsbury AdMag* is delivered free to around 41,000 homes. A new edition is published each Thursday. Over a six week period the manager of a local business experimented with placing a differently sized advertisement each week.

He believes that the bigger the advertisement, the greater the sales.

The manager recorded the sales between Thursdays to see if there was the expected positive correlation between the size of the advertisement and the resulting sales.

The following table gives the data set;

size, x, percentage of a page	60	25	12.5	10	40	20
sales, y, in £1000s	15.4	12.7	9.1	11.4	12.8	8.5

(**i**) Use your statistics calculator to determine the Product Moment Correlation Coefficient for the data and also the equation of the regression line.

[2 marks]

(**ii**) Test, at the 5 % level of significance, whether there is evidence of positive correlation between the size of the advertisement and the resulting sales. State your hypotheses clearly.

[4 marks]

(**iii**)   If a half page advertisement were to be placed, what would the resulting sales be predicted to be ?

[1 mark]

(**iv**) The manager is considering running a two page advertisement to boost sales. Why might a statistician advise caution ?

[2 marks]

Question 5

A-Level Examination Question from 22nd May 2014, S3, Q8 (a) (b) (Edexcel)

The heights, in metres, and weights, in kilograms, of a random sample of 9 men are shown in the table below;

Man	A	B	C	D	E	F	G	H	I
Height (x)	1.68	1.74	1.75	1.76	1.78	1.82	1.84	1.88	1.98
Weight (y)	75	76	100	77	90	95	110	96	120

(a) Calculate, to 3 decimal places, the product moment correlation coefficient between height and weight for these men.

[2 marks]

(b) Use your value of the product moment correlation coefficient to test whether or not there is evidence of a positive correlation between the height and weight of men. Use a 5 % significance level.
State your hypotheses clearly.

[4 marks]

Question 6

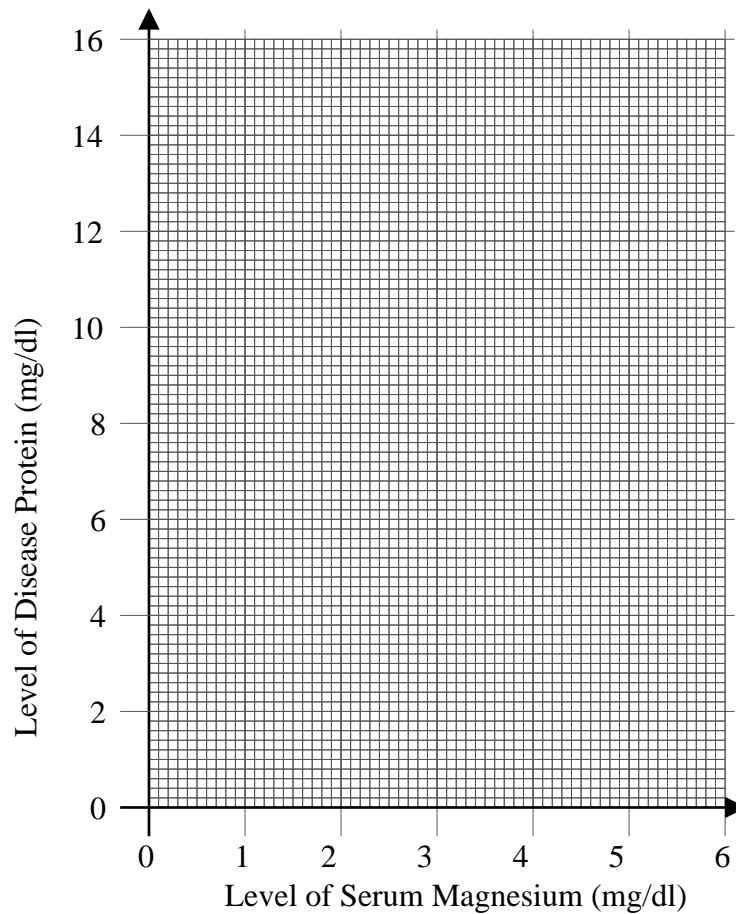*A-Level Examination Question from 9th June 2005, S3, Q4 (edited)*

Over a period of time, researchers took 10 blood samples from one patient with a blood disease. For each sample, they measured the levels of serum magnesium, s, mg/dl, in the blood and the corresponding level of the disease protein, d mg/dl. The results are shown in the table;

s	1.2	1.9	3.2	3.9	2.5	4.5	5.7	4.0	1.1	5.9
d	3.8	7.0	11.0	12.0	9.0	12.0	13.5	12.2	2.0	13.9

(a) State what type of correlation is measured by the product moment correlation coefficient, the PMCC.

[1 mark]

(b) On the adjacent page, draw a scatter diagram to represent these data.

[3 marks]

(**c**) Calculate the value of the PMCC between *s* and *d*

[2 marks]

(**d**) Stating your hypotheses clearly, test, at the 1 % significance level, whether or not the correlation coefficient is greater than zero.

[3 marks]

(**e**) With reference to your scatter diagram, comment on your result in part (d).

[1 mark]