Year 2

~ Statistics ~

C O R R E L A T I O N
I I

Is there a correlation between investment and profit ?

Lesson 1

A-Level Applied Mathematics : Statistics : Year 2

Correlation II

1.1 Scatter Graph Mathematics

By way of an introductory example, suppose that a researcher thinks there is a link between the height of a person and how good they are at doing mathematics.
More specifically, her hypothesis is that;

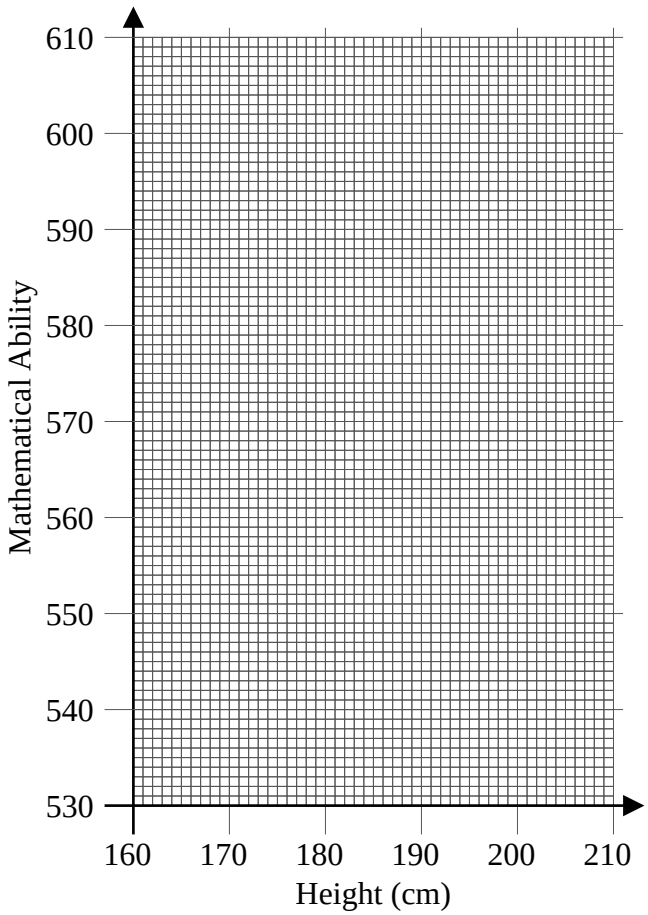
Taller people are better at mathematics than shorter people

In order to test the hypothesis she measured the height, x, to the nearest cm, of a random sample of nine people.
She also devised a mathematics test to measure their mathematical ability.

The data are shown in the table below.

height, x	179	169	187	166	162	193	161	177	168
ability, y	569	561	579	561	540	598	542	565	573

(i) Draw a scatter diagram to illustrate the bivariate data.



[4 marks]

To enter the table of bivariate data into your statistics calculator:

- Press : **SHIFT** : **SETUP** : ∇ : **3** : **2** (make sure frequencies are off)
- Press : **MENU** : **6** (Statistics)
- Press : **2** ($y = a + bx$)

- Now type in the data, using the arrow keys to move around and pressing the = button to enter each number.

- Press : **AC** (to finish)

To view the equation for the line of best fit, and the PMCC;

- Press : **OPTN 3** (to view the results)

You should find that the line of best fit, also called the least squares regression line is; $y = 317 + 1.43x$

and that the Product Moment Correlation Coefficient, the PMCC is;

$$r = 0.897$$

(ii) To add the line of best fit to the scatter graph, let $x = 160$ in the equation of the regression line to get a point (160, y_1), then let $x = 200$ to get another point (200, y_2) and then draw a straight line the two points.

[3 marks]

(iii) Give an interpretation of the 1.43 in the equation for the line of best fit.

[1 mark]

(iv) Estimate the score in the mathematics test of a person of height 190 cm.

[1 mark]

(v) State the domain of values of x for which estimates are reliable.

[1 mark]

(vi) Give an interpretation of the PMCC, $r = 0.897$.
Give a reason for your answer.

[1 mark]

(vii) How could the researcher increase the robustness of her research ?

[2 marks]

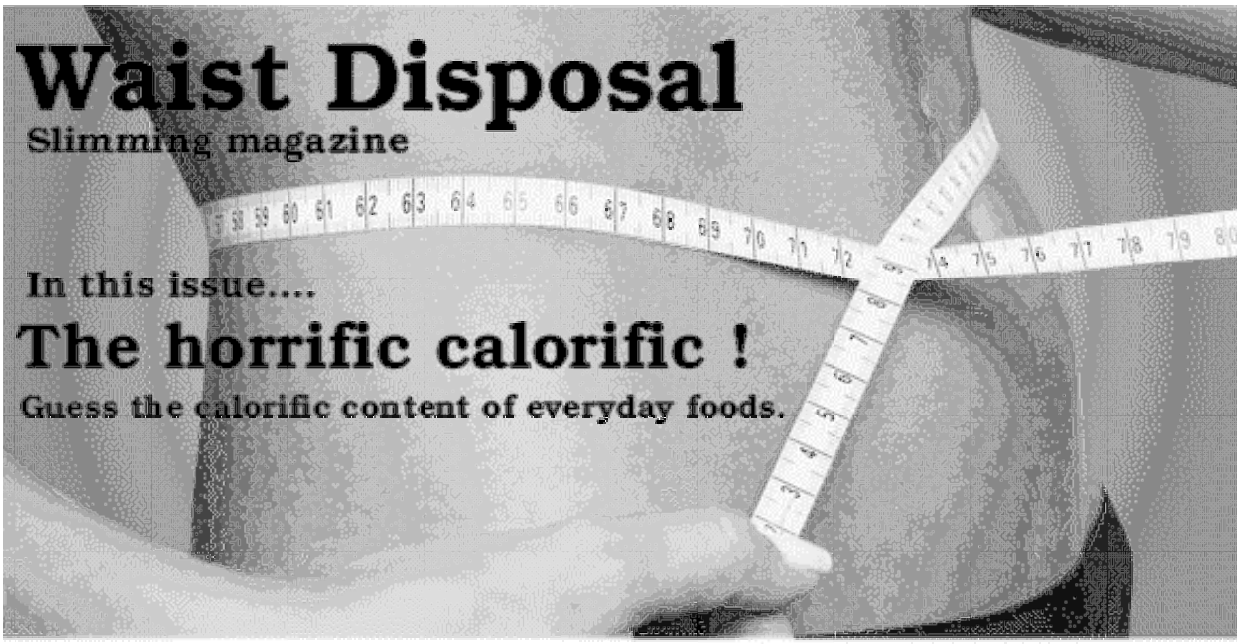
1.2 Exercise

Marks Available : 20

The slimming magazine, *Waist Disposal*, wishes to investigate how aware its readers are of the calorific content of various foods. In a recent issue readers were invited to guess the calorific content of ten everyday foods.

(i) Without drawing any diagram or making any calculations state if you are expecting there to be significant correlation.
Give a reason for your answer.

[2 marks]

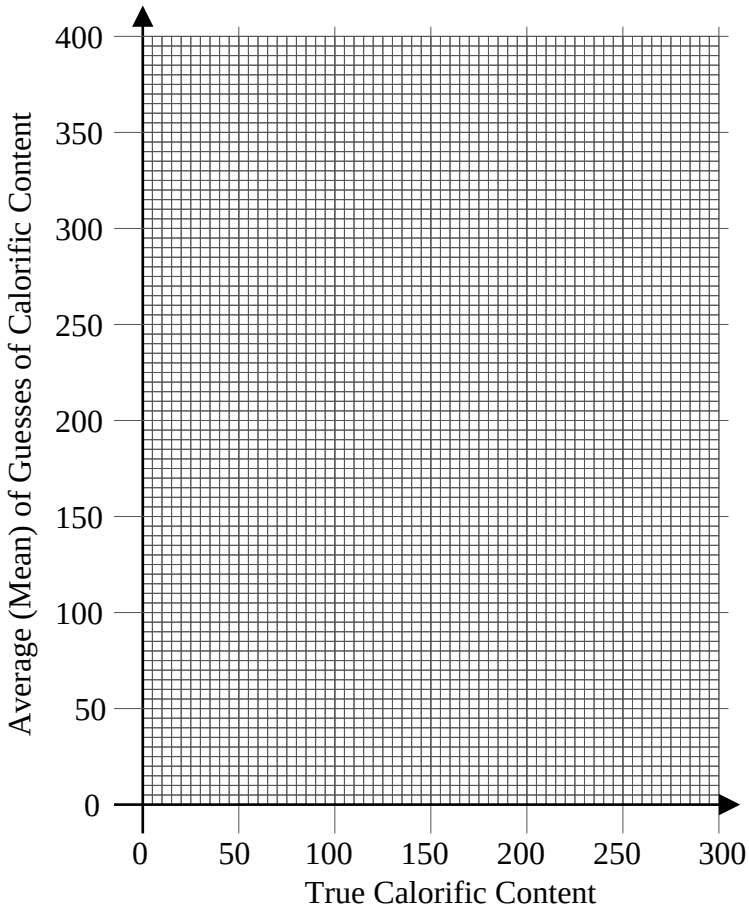


From the results sent in, the average (mean) of the guesses was calculated.

The table given below was published in the following issue of the magazine.
It gives, for each food, the true calorific content, *t*, and the average reader's guess, *g*.

Food	True, *t*	Av of Guesses, *g*
Crisps (34.5 g packet Walkers Salt and Vinegar)	180	280
Egg (Sainsbury's organic, large)	89	130
Yogurt (Activia fat free, 125 g pot)	65	110
Ryvita crisp bread (11 g slice)	35	70
Mars bar (58 g)	260	340
Pears (half a tin, Tesco's, in juice)	75	135
Orange Juice (200 ml serving, Tropicana)	94	115
Beef Burger (Sainsbury's organic, 100 g)	236	210
Sardines (half a tin, in tomato sauce, 60 g)	106	260
Baked Beans (half a tin, Heinz)	151	200

(ii) Draw a scatter diagram to illustrate the data given in the table.



[4 marks]

(iii) Enter the table of bivariate data into your statistics calculator.
Then view the equation for the line of best fit, and the PMCC.
Write down what your calculator tells you here;

[3 marks]

(iv) To add the line of best fit to the scatter graph letting $x = 0$ in the equation for the regression line to get a point (0, y_1), then let $x = 300$ to get another point (300, y_2) and then draw a straight line between the two points.

[3 marks]

(v) Give an interpretation of your correlation coefficient.
Give a reason for your answer.

[2 marks]

(vi) State, with a reason, whether or not the value of the product moment correlation coefficient (PMCC) changes if all the guesses are 50 calories higher than the values in the table.

[2 marks]

(vii) Give a practical interpretation of the slope, *b*, of your line of best fit.

[1 mark]

(viii) The published article in *Waist Disposal* concluded that all readers were much more aware of the calorific content of foods than the general population. (In which the PMCC was 0.548).
Thus by being a reader of *Waist Disposal* you were much more able to experience weight loss.
Comment on the magazine's conclusion.

[3 marks]