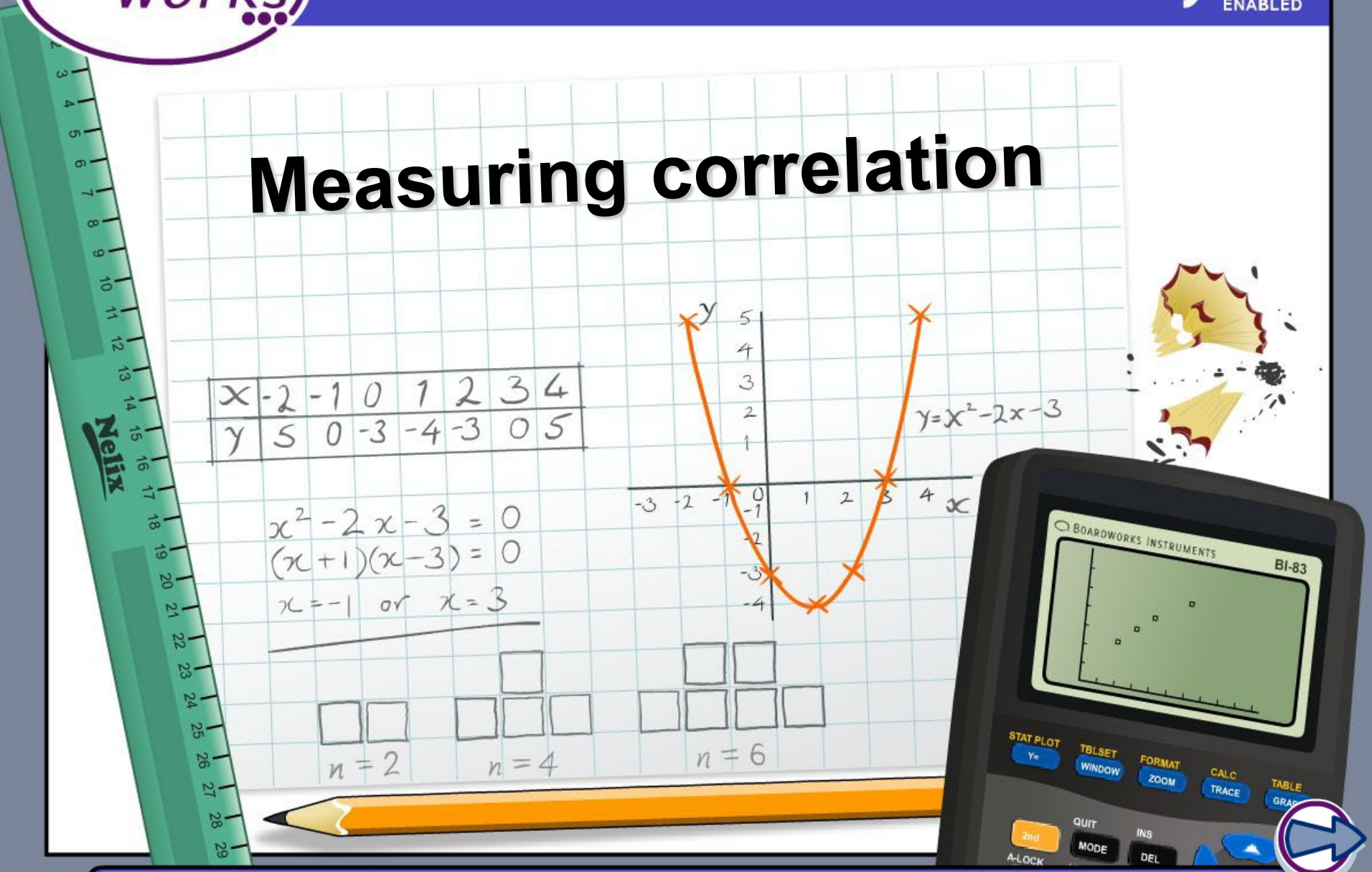
Measuring correlation



Information

Common core icons

This icon indicates a slide where the Standards for Mathematical Practice are being developed. Details of these are given in the Notes field.

Slides containing examples of mathematical modeling are marked with this stamp.

This icon indicates an opportunity for discussion or group work.

The **Standards for Mathematical Practice** outlined in the Common Core State Standards for Mathematics describe varieties of expertise that mathematics educators at all levels should seek to develop in their students.

They are:

1) Make sense of problems and persevere in solving them.

2) Reason abstractly and quantitatively.

3) Construct viable arguments and critique the reasoning of others.

4) Model with mathematics.

5) Use appropriate tools strategically.

6) Attend to precision.

7) Look for and make use of structure.

8) Look for and express regularity in repeated reasoning.

This icon indicates that the slide contains activities created in Flash. These activities are not editable.

This icon indicates teacher's notes in the Notes field.

Correlation

A **scatter plot** shows the relationship between two variables.

If there is a relationship between these variables, we say that there is **correlation**.

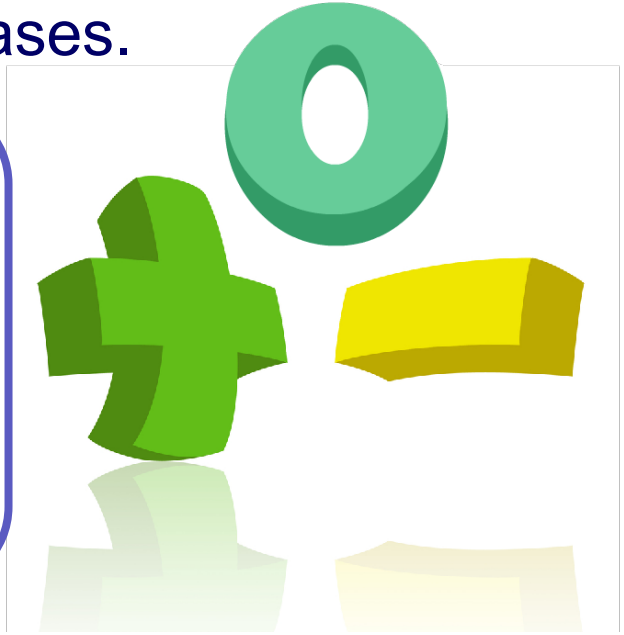
- Correlation is a general **trend**.
- Some data points may not fit this trend, these are **outliers**.
- A **line of best fit** can be drawn in order to estimate values or even predict values in some cases.

Scatter plots can show:

positive correlation

negative correlation

zero correlation



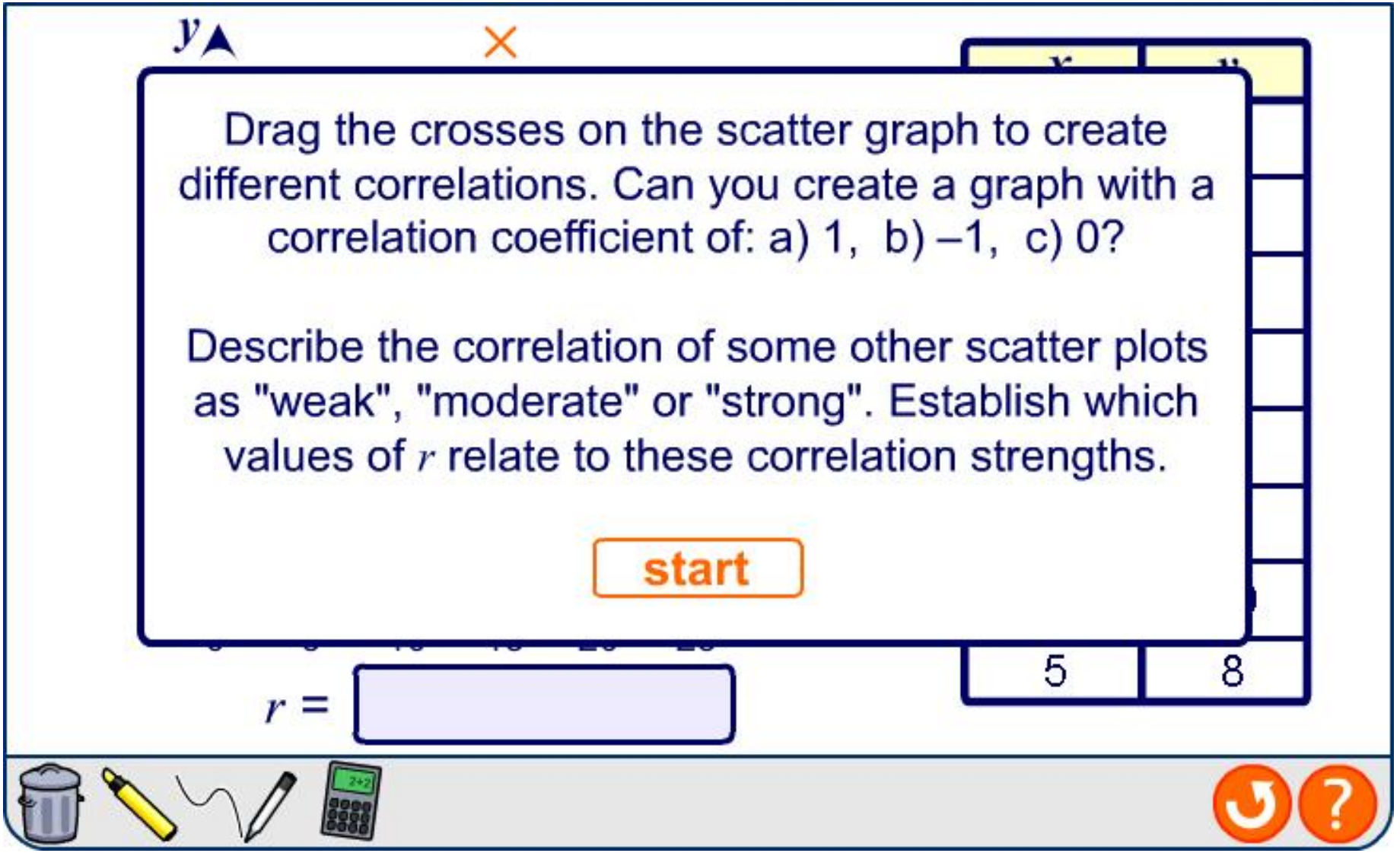
Types of correlation

Positive correlation | Negative correlation | No correlation

The **correlation coefficient**, r, quantifies the degree of linear correlation.
The value of r can be between –1 and 1.

Press on each of the tabs above to see what kinds of scatter plots show each type of correlation.

Correlation coefficients



Test marks

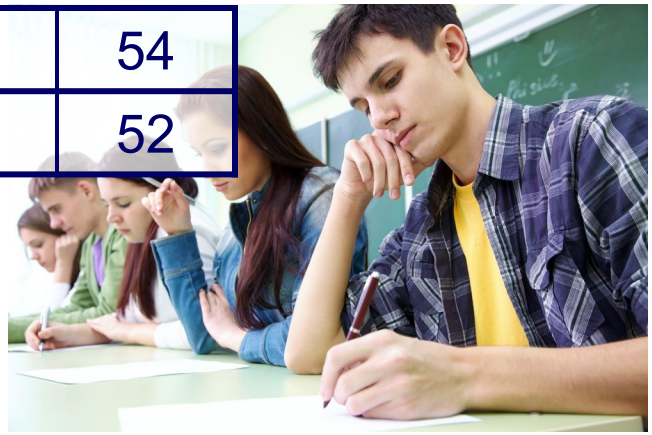
Algebra and Physics points in the end of year test

Algebra %	24	37	43	52	65	71
Physics %	19	51	62	64	71	80

How do the points in algebra and physics relate to each other?

Biology and English points for the same students in the end of year tests

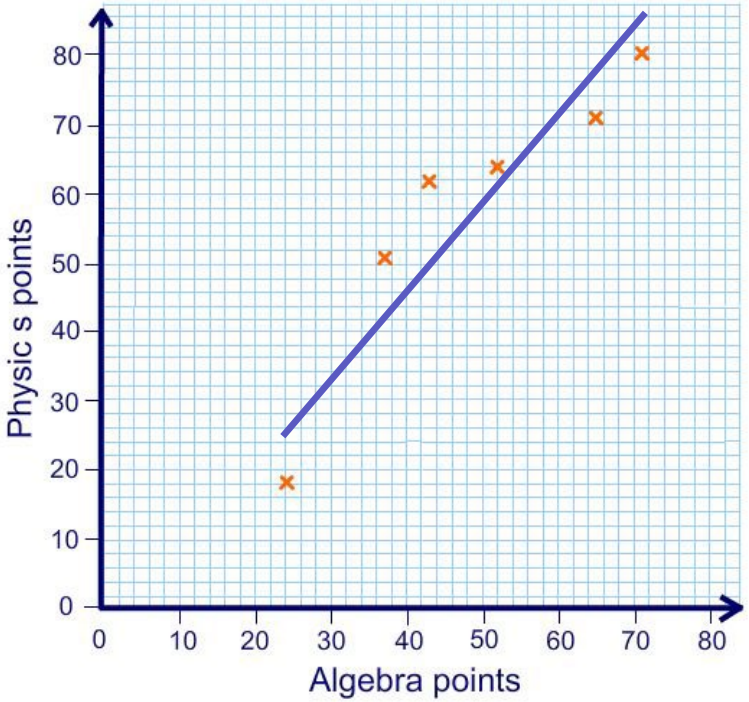
Biology %	14	27	34	39	42	54
English %	76	72	68	64	55	52

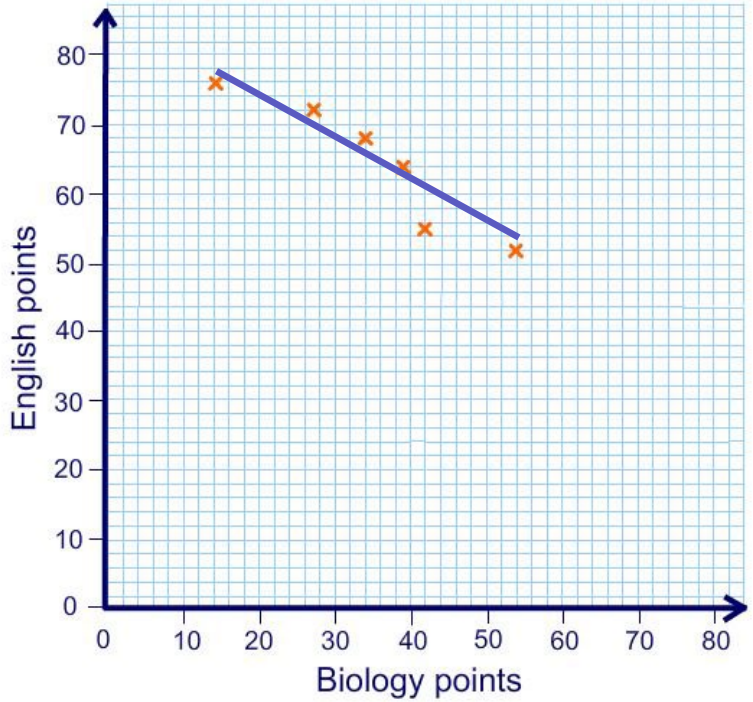


How do the points in chemistry and English relate to each other?

Scatter graph results

Here is a scatter graph of each set of test results:





Compare and discuss the graphs. Is the correlation weak or strong? Where would you draw a line of best fit? Estimate the correlation coefficient for each graph.

Is there a correlation?

How are these variables correlated?

How much sleep you have and performance in a memory test.

negative correlation

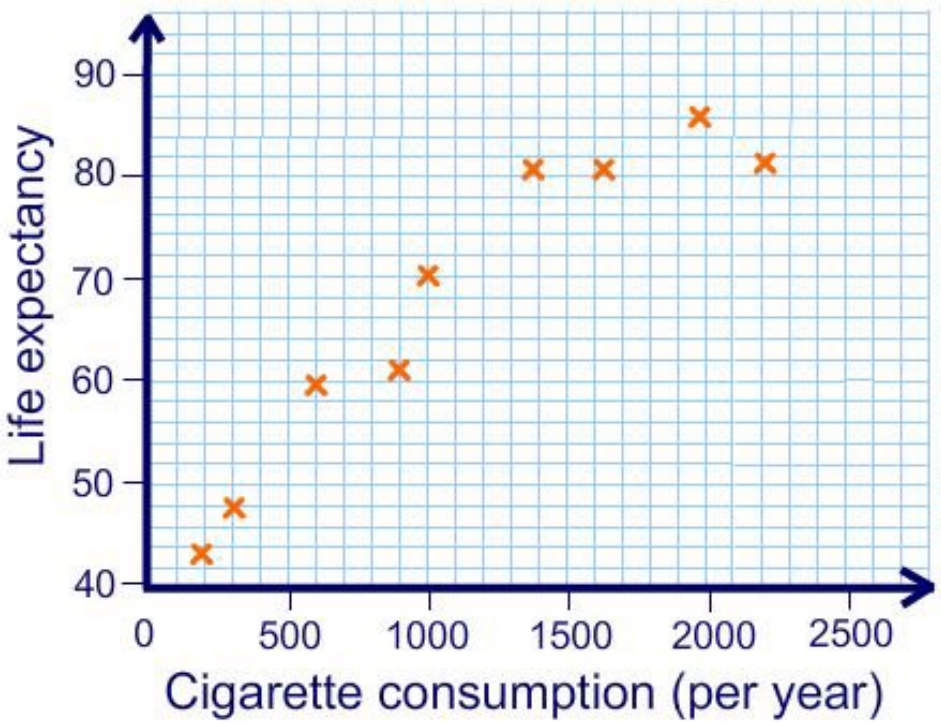
positive correlation

no correlation

Correlation vs. causation

The scatter plot below shows life expectancy at birth vs. annual cigarette consumption for a sample of 9 countries.

What trend does the scatter plot show?



How does this demonstrate the importance of interpreting scatter plots with caution?

Correlation vs. causation

The scatter plot for life expectancy vs. cigarette consumption shows a positive correlation. However, it would be **wrong** to conclude that consuming more cigarettes causes people to live longer.

This type of correlation is sometimes referred to as **nonsense correlation** or **causation**.

The relationship could be explained by the fact that both life expectancy and cigarette consumption for a country are correlated with a **third variable**: the country's wealth.

Cause and effect

A study finds a positive correlation between the number of cars in a town and the number of babies born.

The local newspaper reports:

"Buying a new car can help you get pregnant!"

Does the study support this conclusion?

Correlation does not necessarily imply that there is a causal relationship between the two variables. There may be some other cause.

What might be the cause in the example above?

Cause and effect

A study finds a negative correlation between the number of sleds sold and the temperature.

The local newspaper reports:

"If you want it to snow, go out and buy a sled!"

Does the study support this conclusion? Explain your answer.

Misleading headlines

Discuss these headlines. Press one for an explanation.

"Taller students do better in new math test!"

"The more coffee you drink, the more stressed you are."

"Chocolate causes lower grades at college."

"Counseling can make you depressed."

"New hospital exercise regime causes rise in injuries."

Wealth vs. depression

This table contains data for 8 countries, showing:

- the percentage of the population diagnosed as depressed
- the GDP (i.e. wealth) of the country, in international dollars.

	USA	Spain	Netherlands	Belgium	Italy	France	Mexico	Germany
% population depressed	9.6	4.9	6.5	11.7	3.8	8.5	4.8	3.6
GDP (intl. $)	48,387	30,626	42,183	37,737	30,464	35,153	14,610	37,897

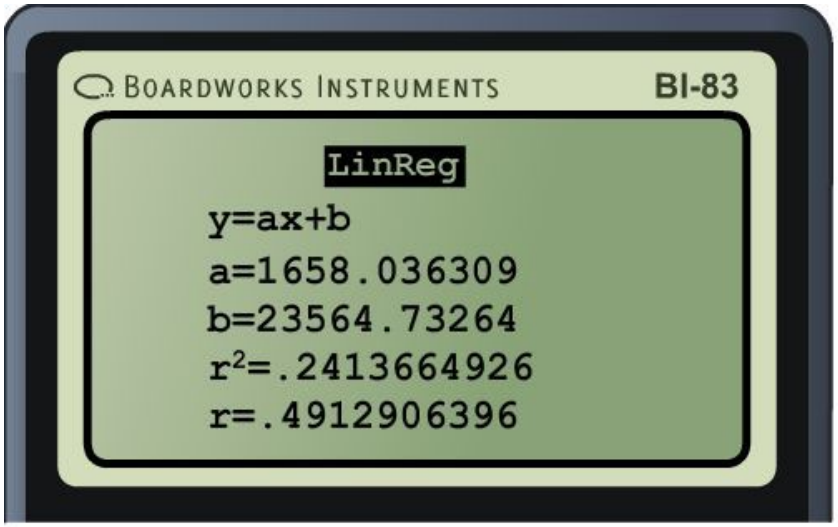
How would you expect these two variables to be related? Graph the data. Is the relationship linear?

We can find the value of the correlation coefficient using a graphing calculator.

Using a graphing calculator

Using the STAT feature of your graphing calculator, enter the depression figures in one list and the corresponding GDP figures in the second list.

Press STAT again and go to the “CALC” menu. Select linear regression, “LinReg”, and “Calculate”.



The correlation coefficient, r, is **0.491** (to nearest thousandth).

This shows that there is a moderate **positive** correlation between a country’s wealth (GDP) and the percentage of its population that is depressed. This seems the opposite of what might be expected.

What other factors might affect this result?

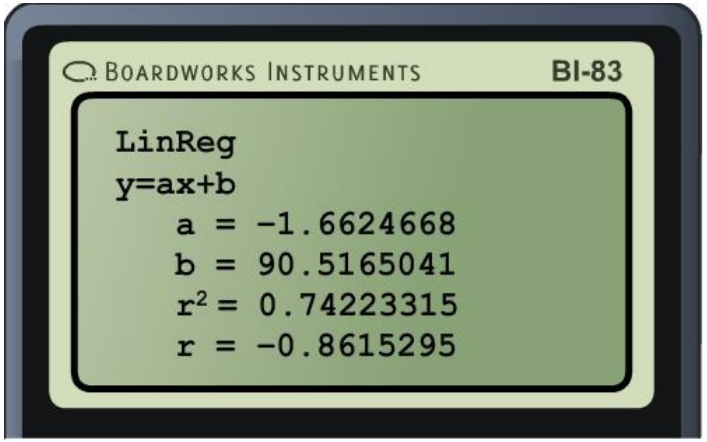
Children's eating habits

	USA	Spain	UK	Russia	Italy	France	Canada	Germany
% of children overweight	25.1	16.9	15.8	7.6	15.2	11.2	19.5	16.0
% eating breakfast	47.2	72.2	56.1	78.0	62.4	71.4	58.2	67.0

Liz says: "There's a fairly strong negative correlation between the two variables in this table."

Is she correct? Explain.

Discuss the result as a class.



Use the STAT feature of your graphing calculator and select LinReg.

Liz is correct; the value of r, is **–0.862**, to the nearest thousandth, giving a fairly strong negative correlation.