Name:
Date:

Correlation Examples

- Correlation is a measure of the strength of the 'relationship' or 'pattern' between two sets of data
- Usually, the sets of data will be called $x = \{x_1, x_2, \ldots, x_n\}$ and $y = \{y_1, y_2, \ldots, y_n\}$ and may be represented in a table as so:

x	x_1	x_2	$\ldots$	x_n
y	y_1	y_2	$\ldots$	y_n

- A *data point* is a coordinate (x_i, y_i) taken from the data and the data points may be plotted on a graph to produce a scatter diagram for the data
- The strength of the correlation is known as the *product moment correlation coefficient* and is usually called r. It is always true that $-1 \leq r \leq 1$. The following is a guide to the meaning of the values of r:

Strong positive correlation:	$0.6 < r \leq 1$
Moderate positive correlation:	$0.3 < r \leq 0.6$
Weak positive correlation:	$0.1 < r \leq 0.3$
No correlation:	$-0.1 \leq r \leq 0.1$
Weak negative correlation:	$-0.3 \leq r < -0.1$
Moderate negative correlation:	$-0.6 \leq r < -0.3$
Strong negative correlation:	$-1 \leq r < -0.6$

- Importantly, r is only a test for *linear correlation*.There might be a relationship between x and y which is not a straight line!
- If you are not given all the raw data, you may have to calculate r using the formula

$$r = \frac{S_{xy}}{\sqrt{S_{xx}S_{yy}}}$$

  The formula for calculating S_{xy}, S_{xx} and S_{yy} is given in the formula book, should you ever need it.
- It is important to understand that positive correlation does not imply causation: two things can be related without one causing the other. Famously, the number of ice-creams sold each month is positively correlated with the number of deaths by drowning in that month. Can you think why?

Example (Calculating r from raw data) The table below gives data on the number of ice-creams sold each month by a newsagent in London (x) and the number of deaths by drowning in that month around the UK (y).

x	181	186	197	207	208	230
y	54	57	62	65	66	73

1. Calculate the size of the product moment correlation coefficient
2. Interpret your value for r
3. Comment on the claim that 'Eating ice-cream causes people to drown'

Example (Outliers) It is found that in the next month, the newsagent only sold 12 ice-creams but that there were 77 deaths by drowning.

1. Express this information as a data point
2. Recalculate r
3. Give a new interpretation of r
4. Comment on the claim that 'ice-cream sales and drowning are actually less correlated than previously believed'

Example (Interpreting r) Comment on the likelihood of the following:

1. The length of time spent studying and the grade achieved were found to have $r = -0.41$
2. The width of a dog's paw and its weight were found to have $r = 0.67$
3. The speed of typing and amount of coffee consumed was found to have $r = 1.05$

Example (Calculating r by formula) The cost of a car (x) and the maximum speed of the car (y) are found to have the following summary data:

$$\begin{aligned} S_{xx} &= 2417.85 \\ S_{yy} &= 47.9 \\ S_{xy} &= 30.62 \end{aligned}$$

1. Calculate the product moment correlation coefficient for this data.
2. Comment on this value of r
3. Comment on the claim that 'there is no relationship between the cost of a car and the car's maximum speed'