CORRELATION & REGRESSION

KEY WORDS & DEFINITIONS

1 Correlation A description of the linear relationship between two variables.

2 Bivariate data Pairs of values for two variables

3 Causal relationship Where a change in a variable causes a change in another. Not always true.

4 Least squares regression line
A type of line of best fit which is a straight line in the form $y = a + bx$

5 'b' of a regression line
The gradient of the line; indicating positive correlation if it is positive and negative correlation if it is negative.

6 Independent or Explanatory variable
The variable which occurs regardless of the other variable (e.g. time passing). Plotted on the x axis.

7 Dependent or Response variable
The variable whose value depends on the independent variable's data points.

8 Interpolation Estimating a value within the range of the data. Reliable.

9 Extrapolation Estimating a value outside of the range of the data. NOT reliable.

10 Product Moment Correlation Coefficient
A measure of the strength and type of correlation.

WHAT DO I NEED TO KNOW

Interpreting 'b' of a regression line:
Refer to the change in the variable y for each unit change of the variable x IN CONTEXT

PMCC, r is the PMCC for a population sample

PMCC, ρ is the PMCC for the entire population

Range of PMCC, r: $-1 \leq r \leq 1$

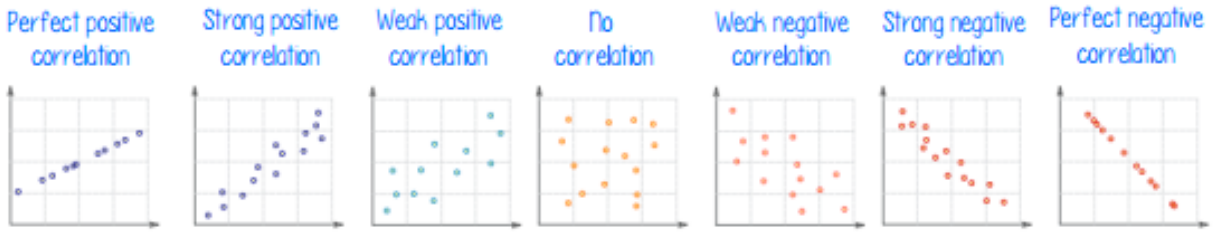
Hypotheses for one tailed test on PMCC:
$H_0: \rho = 0$
$H_1: \rho > 0$ or $H_1: \rho < 0$

Hypotheses for two tailed test on PMCC:
$H_0: \rho = 0$
$H_1: \rho \neq 0$

Check sample size is big enough to draw a valid conclusion and comment on it if not.

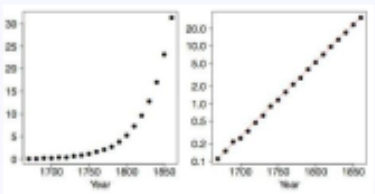
A regression line is only a valid model when the data shows linear correlation.

Only make predictions for the dependent variable using the regression line of y on x within the range of the original data



EXPONENTIAL MODELS

You can use logarithms and coding to transform graphs and examine trends in non-linear data



If $y = ax^n$ then $\log y = \log a + n \log x$

If $y = kb^x$ then $\log y = \log k + x \log b$