# Investigating Relationships between 2 numerical variables

We use scatter plots when both variables are numerical

First, you need to identify the independent and dependent variables (IV and DV)

IV always plotted on the $x$ axis (horizontal)
DV plotted on the vertical axis.

You should be able to plot points manually on the coordinate plane

| (IV, DV) |
| --- |
| $(x, y)$ |

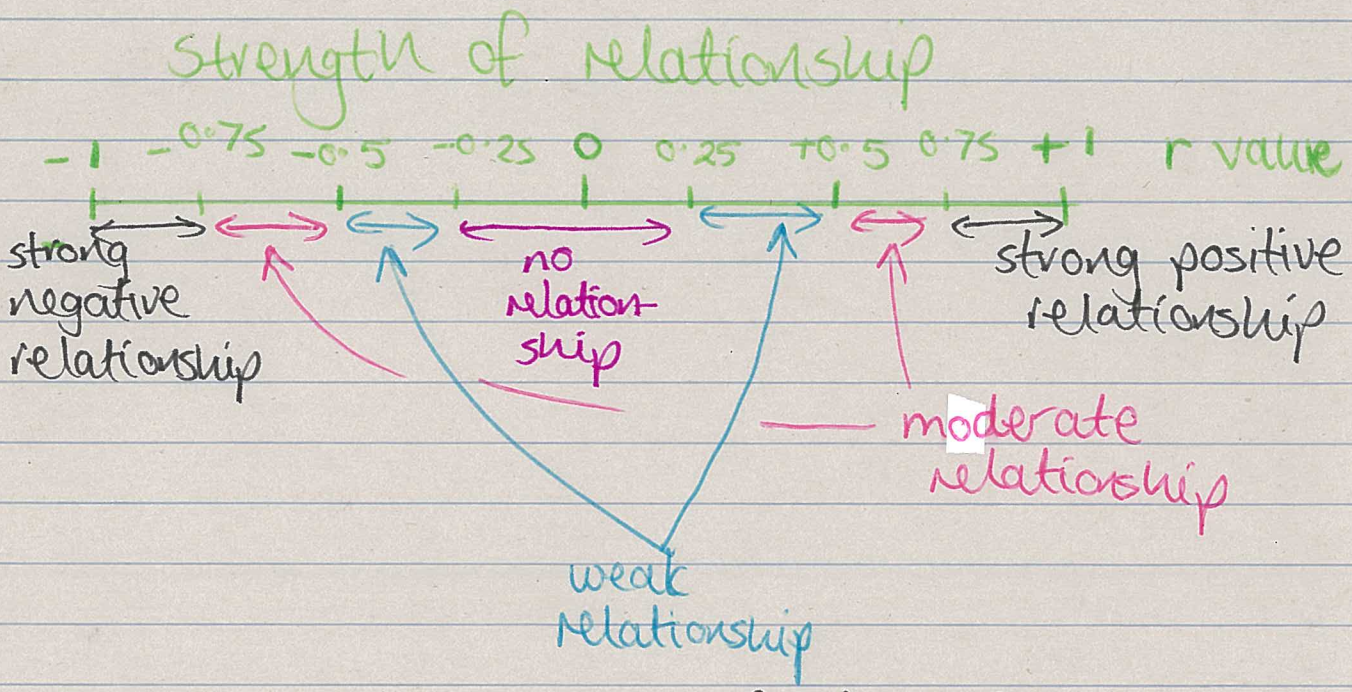
and using your CAS

## Interpreting the Scatter Plot

* Need to look for a clear pattern in the dots. If there is a pattern, then there is a relationship.

To describe the relationship look at:
- direction (and outliers) positive / negative
- form linear / non-linear
- strength correlation coefficient $r$
  varies between $-1$ to $+1$
  → or just visual observation will give a subjective measure.

# Relationship Between Two Numerical Variables




* If use r as a measure of strength assuming that
  - numerical data
  - linear relationship
  - NO outliers

To calculate r $r = \frac{\sum(x-\bar{x})(y-\bar{y})}{(n-1)s_x s_y}$

Don't worry about this too much you can use your CAS always !! YAY!

1. Enter the data in lists/spreadsheets
2. ctrl I calculator page
3. MENU → 6 → 1 → 4 linear regression
4. select x and y lists

⑤ On your screen will appear all the required information for the LEAST SQUARES REGRESSION LINE $y = a + bx$

$a$ → "y intercept"
$b$ → "gradient"
* $r$ → correlation coefficient
** $r^2$ → coefficient of determination

* Remember $r$ tells you strength of linear relationship.

** $r^2$ tells you the degree to which one variable can be predicted by another. Usually we use $r^2 \times 100$ to create a %.

So if $r = 0.83$, $r^2 = (0.83)^2$
$= 0.69$ (to 2 dec. places)

This means that:
69% of the variation in the DV is explained by the variation in the IV

Always use this wording !! OK !!

and 31% is explained by other factors

(100 − 69)