IB Math Studies 2 Bell Work

1-variable stats on the graphing calculator

Enter the data lists:
L1 - 5, 3, 3, 6, 4, 5, 3, 7, 5, 7, 1, 8, 9, 5
L2 – 9, 6, 5, 5, 7, 5, 6, 7, 6, 4, 5, 8

Using the calculator:

Make: A frequency histogram
A box-and-whisker plot
Parallel box-and-whisker plots

Find: The 5-number summary of L1
The mean of L2

Chapter 11

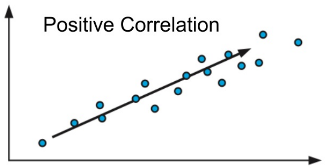
Two variable statistics

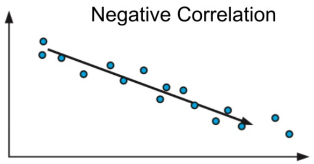
A Correlation
B Measuring correlation

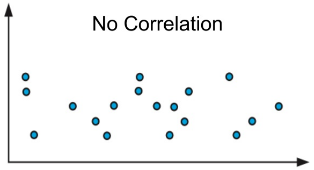
Syllabus reference: 4.2, 4.3, 4.4

A CORRELATION

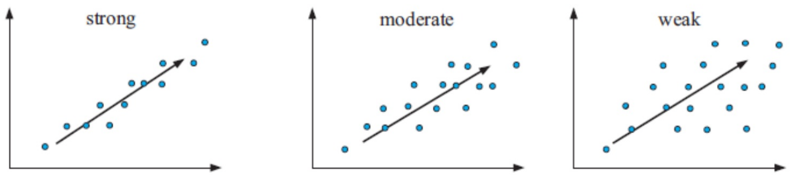
Correlation refers to the relationship or association between two variables.







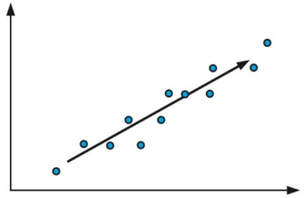
Look at the spread of points to make a judgement about the **strength** of the correlation. For **positive relationships** we would classify the following scatterplots as:



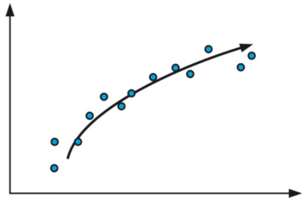
slope ≠ strength of correlation

LINEARITY

These points are roughly linear.

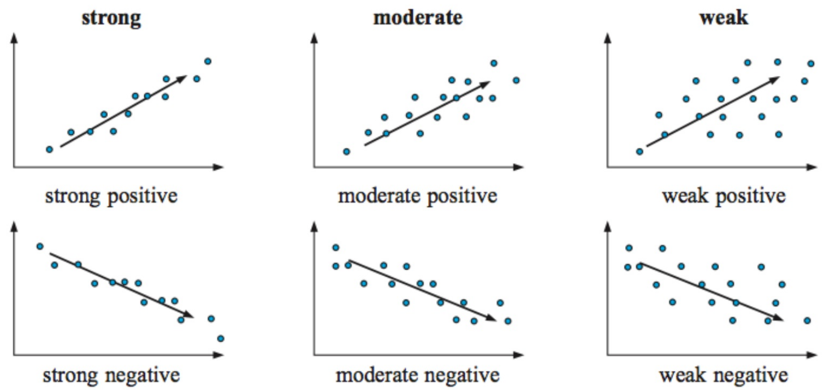


These points do not follow a linear trend.

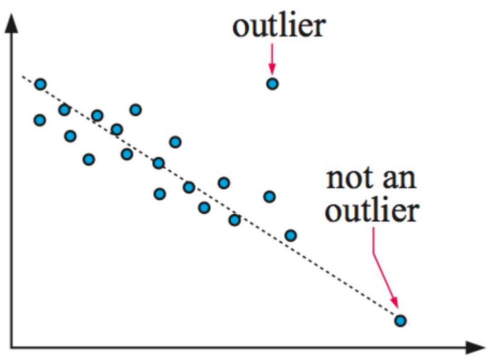


STRENGTH of the correlation:

How tightly does the data fit to a line or curve?



OUTLIERS



Be careful!

Correlation ≠ Causation

Consider the following:

1 The *arm length* and *running speed* of a sample of young children were measured, and a strong, positive correlation was found to exist between the variables.

Does this mean that short arms cause a reduction in running speed or that a high running speed causes your arms to grow long? This would clearly be nonsense.

Rather, the strong, positive correlation between the variables is attributed to the fact that both *arm length* and *running speed* are closely related to a third variable, *age*. Up to a certain age, both *arm length* and *running speed* increase with *age*.

B MEASURING CORRELATION

To say that a correlation is strong, moderate, or weak is not a very accurate measure. To measure correlation then, we calculate **Pearson's Correlation Coefficient.**

For a set of n data given as ordered pairs (x_1, y_1), (x_2, y_2), (x_3, y_3), ..., (x_n, y_n),

Pearson's correlation coefficient is $$r = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2\sum(y-\overline{y})^2}}$$

where $\overline{x}$ and $\overline{y}$ are the means of the x and y data respectively, and $\sum$ means the sum over all the data values.

Calculating Pearson's Correlation Coefficient by hand:

$$r = \frac{\sum(x-\overline{x})(y-\overline{y})}{\sqrt{\sum(x-\overline{x})^2\sum(y-\overline{y})^2}}$$

X	1	2	3	4	5	6
Y	3	2	5	5	9	6

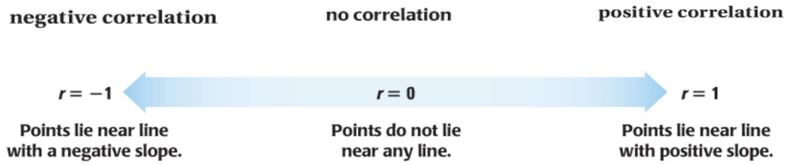
$\overline{x} = 3.5$

$\overline{y} = 5$

$\sum$ = "sum of all"

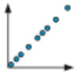
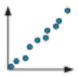
$$r = \frac{(1-3.5)(3-5)+(2-3.5)(2-5)+(3-3.5)(5-5)+(4-3.5)(5-5)+(5-3.5)(9-5)+(6-3.5)(6-5)}{\sqrt{[(1-3.5)^2+(2-3.5)^2+(3-3.5)^2+(4-3.5)^2+(5-3.5)^2+(6-3.5)^2][(3-5)^2+(2-5)^2+(5-5)^2+(5-5)^2+(9-5)^2+(6-5)^2]}}$$

PEARSON'S CORRELATION COEFFICIENT



The following table is a guide for describing the strength of linear correlation using r.

Positive correlation

$r = 1$	perfect positive correlation	
$0.95 \leqslant r < 1$	very strong positive correlation	
$0.87 \leqslant r < 0.95$	strong positive correlation	
$0.5 \leqslant r < 0.87$	moderate positive correlation	
$0.1 \leqslant r < 0.5$	weak positive correlation	
$0 \leqslant r < 0.1$	no correlation	

Negative correlation

$r = -1$	perfect negative correlation	
$-1 < r \leqslant -0.95$	very strong negative correlation	
$-0.95 < r \leqslant -0.87$	strong negative correlation	
$-0.87 < r \leqslant -0.5$	moderate negative correlation	
$-0.5 < r \leqslant -0.1$	weak negative correlation	
$-0.1 < r \leqslant 0$	no correlation	

Assignment:

Exercise **11 A #4, 5**
11 B.1 #2, 4, 6
~~**11 B.3 #3, 4**~~

Chapter 11

Two variable statistics

A Correlation
B Measuring correlation