Correlation

Contents

- Correlation
- Methods of studying correlation
- Types of correlation
- Spurious Correlation
- Factors affecting Correlation Coefficient
- Data Analysis
- Video

Correlation

- Statistical technique for measuring the relationship between variables

- If the quantities(X,Y) vary in such a way that change in one variable corresponds to change in the other variable then the variables X and Y are correlated

Methods of Studying Correlation

- Scatter diagram method
- Karl Pearson's Correlation Coefficient
- Spearman's Rank Correlation Coefficient

Scatter Plot

- Graph that is used to plot the data points for two variables
- X axis --Independent variables
- Y axis -- Dependent Variables
- Idea about the kind of relationship between the two variables

Types of Correlation

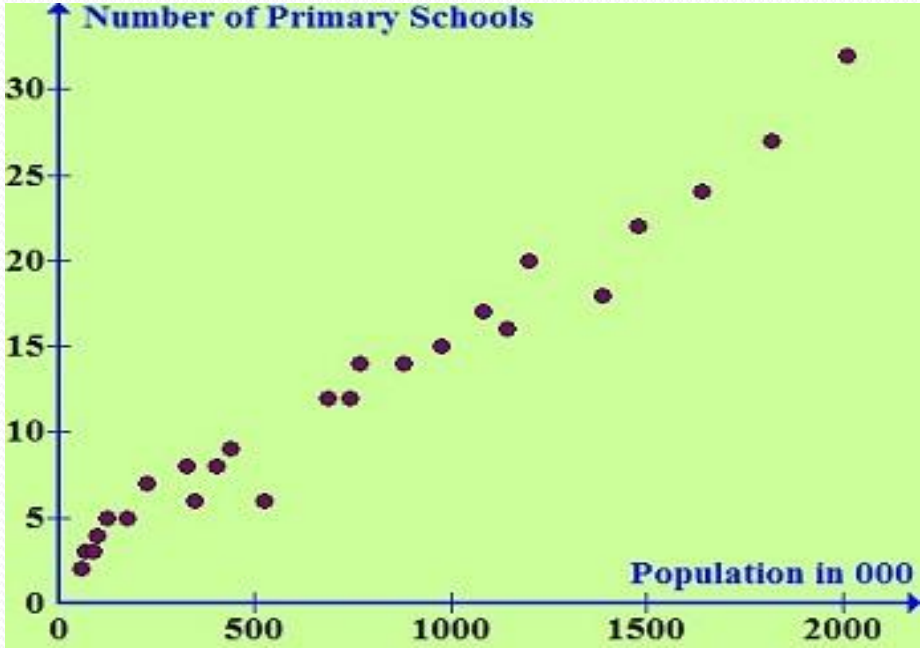
- Positive Correlation
- Negative Correlation
- Strong Correlation
- Weak correlation
- Zero correlation

Positive correlation

- X ↑, Y ↑ (increase) or X ↓, Y ↓ (decrease)

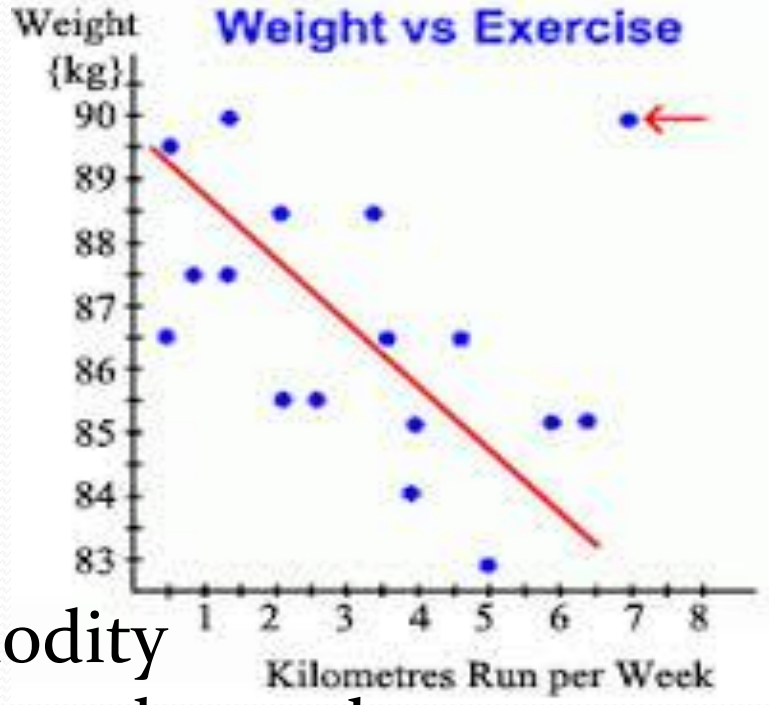
Examples

- Family income and family expenditure
- The more time you spend running on a treadmill, the more calories you will burn
- Smoking and lung diseases
- Rainfall and yield of rice
- Training and performance



Negative correlation

- As X ⬆, Y ⬇
- As X ⬇, Y ⬆



Examples

- Demand and Price of a commodity
- Length of candle and no of hours burned
- Vaccination and illness
- School achievement and days absent from school

Zero Correlation

- When one factor increases/decreases we have no idea what other would be

Examples

Height of a person and income received

Correlation Coefficient

- defined as a numerical representation of the **strength** and direction of a relationship.
- Represented by “r” and unitfree
- –1 <= r <= +1
- The closer to 0, the weaker linear relationship
- No relationship if exactly equal to 0
- The closer to –1, the stronger the negative linear relationship
- The closer to 1, the stronger the positive linear relationship
- Exactly equal to +1 or -1 is perfect relationship

Three Sets of Data Showing Different Directions and Degrees of Correlation

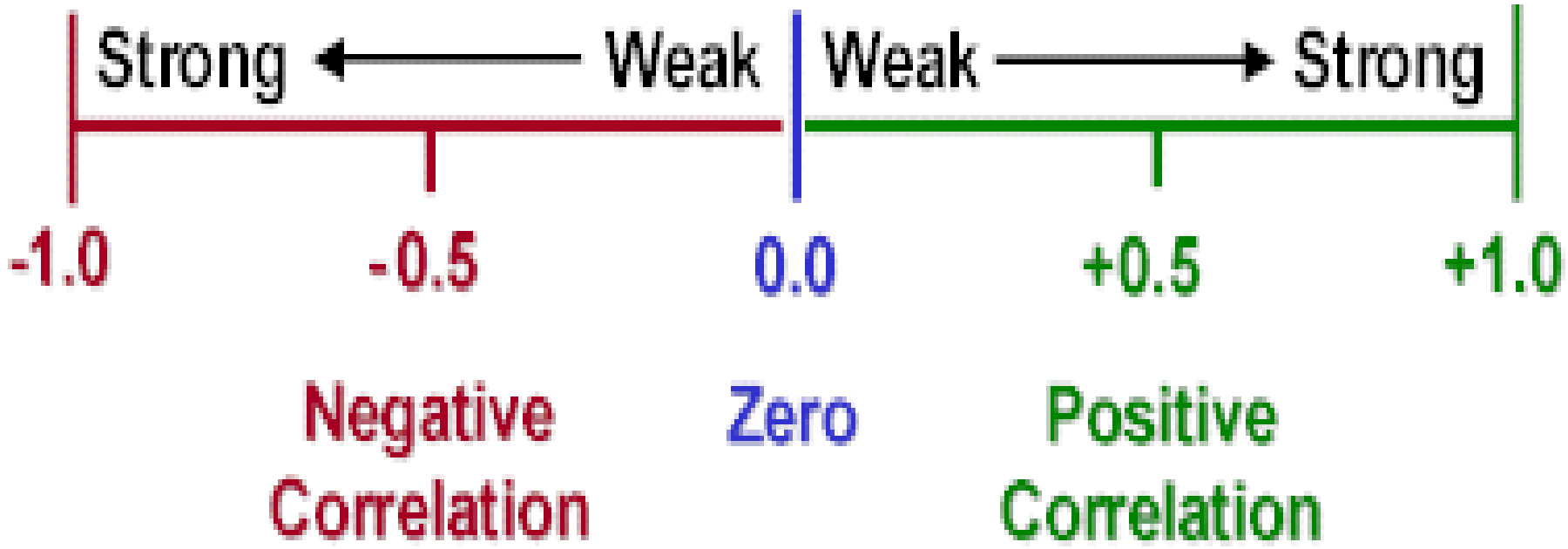
(A) $r = +1.00$		(B) $r = -1.00$		(C) $r = 0$	
X	Y	X	Y	X	Y
5	5	5	1	2	1
4	4	4	2	5	4
3	3	3	3	3	3
2	2	2	4	1	5
1	1	1	5	4	2

Interpreting Corr. Coefficient, r

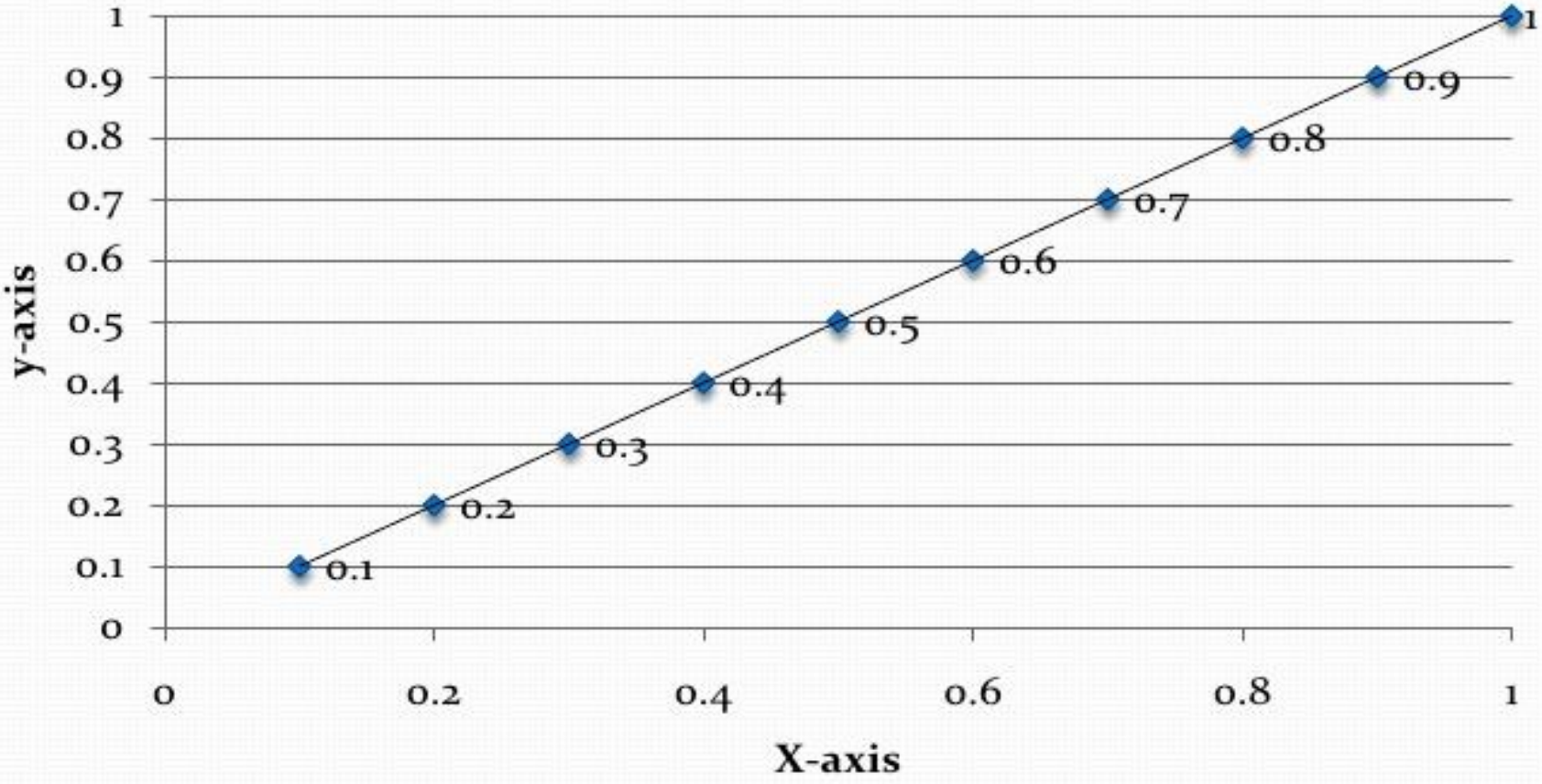
Value of Corr. Coefficient	Strength of Corr.
r equals +or- **1**	Perfect
r > **.70** or r < **–.70**	Strong
r is between **.30** & **.70** or r is between **–.30** and **–.70**	Moderate
r is between 0 and **.30** or r is between **0** and **–.30**	Weak
Equal to 0	No correlation

Correlation Coefficient

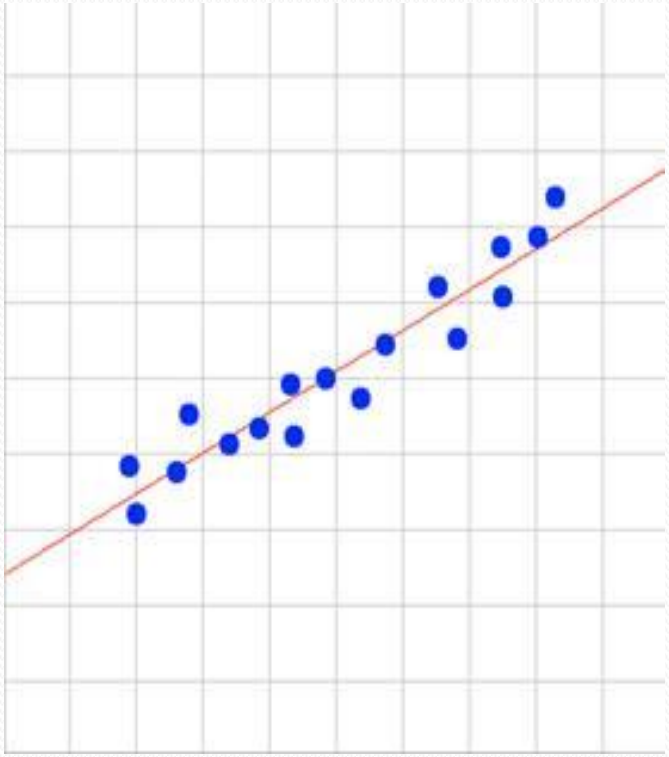
Shows Strength & Direction of Correlation



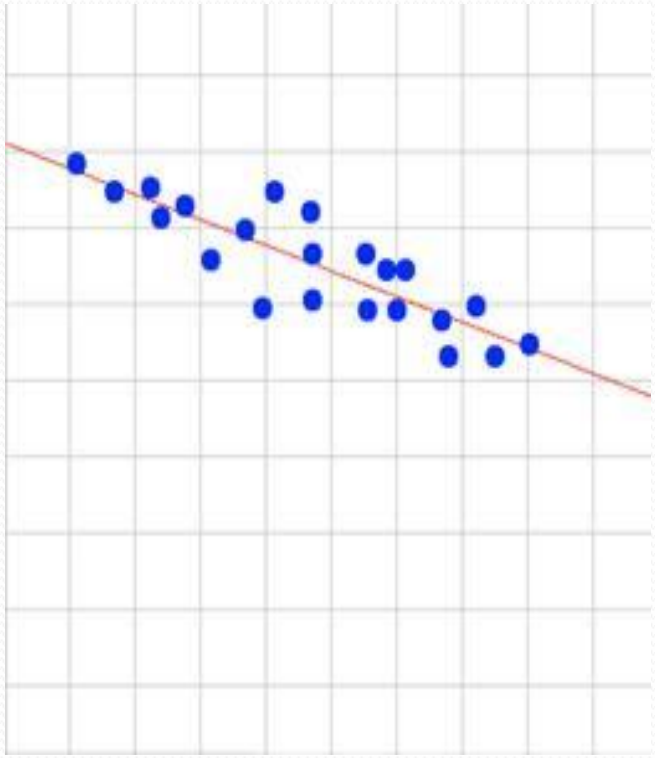
Perfect Correlation



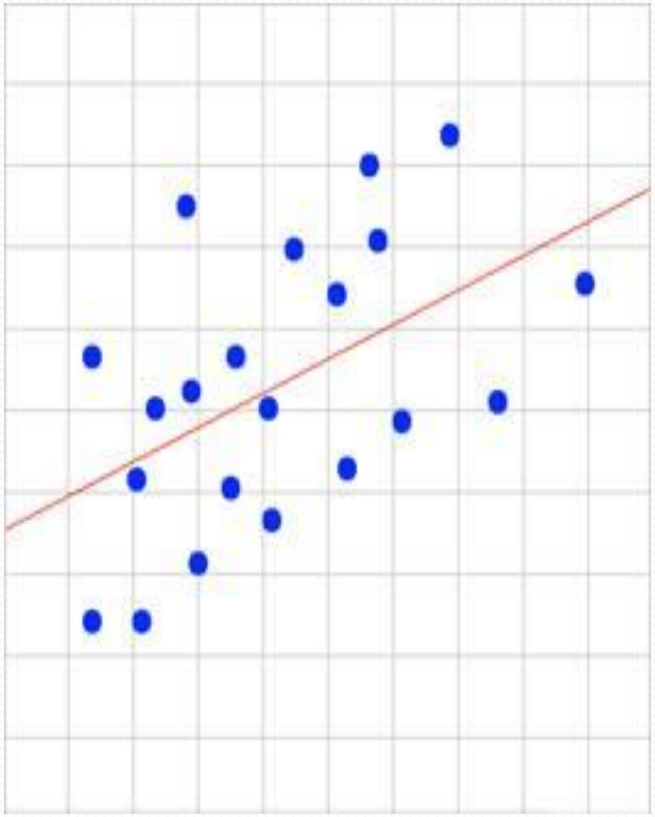
Strong positive correlation
r=0.95



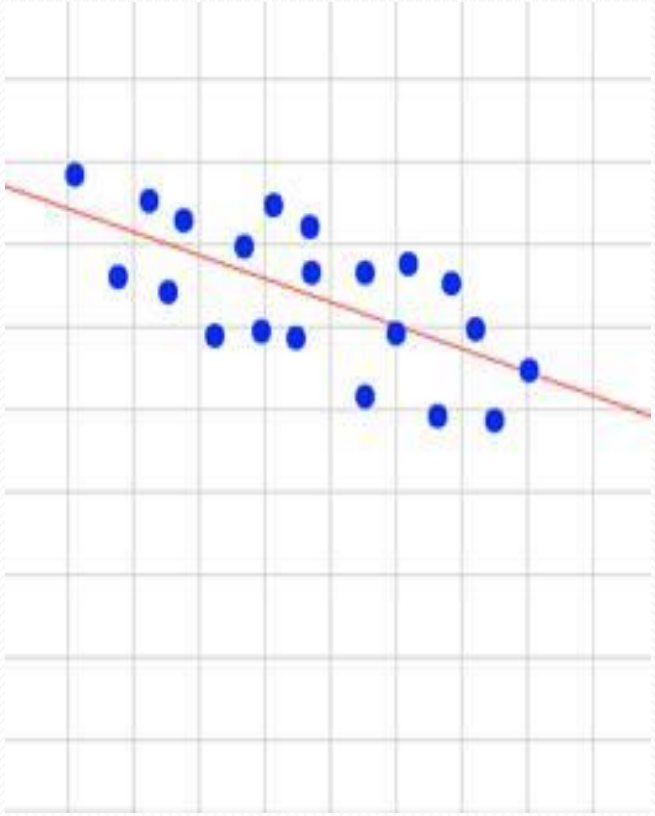
Strong negative correlation
r=-0.85



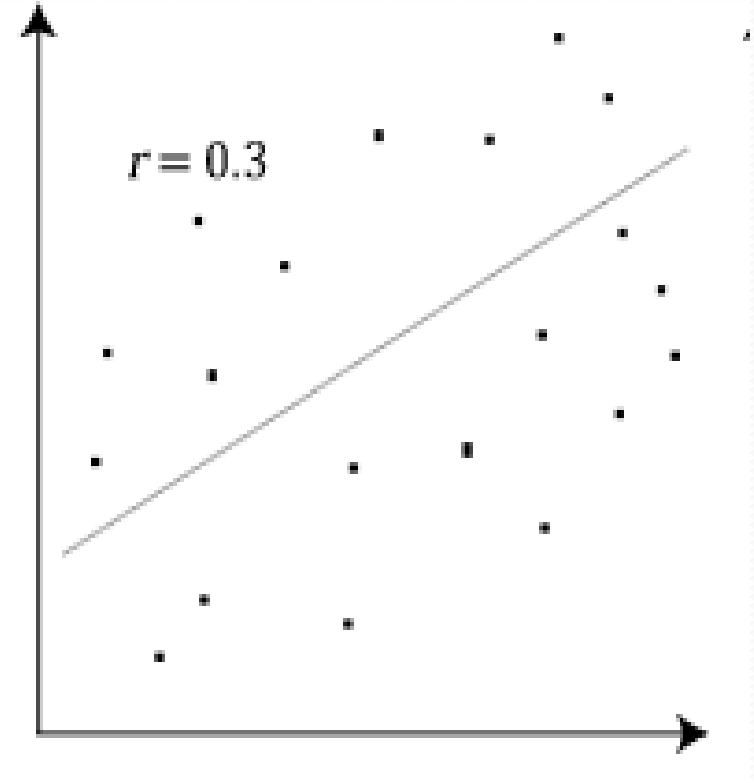
moderate positive correlation
r=0.5



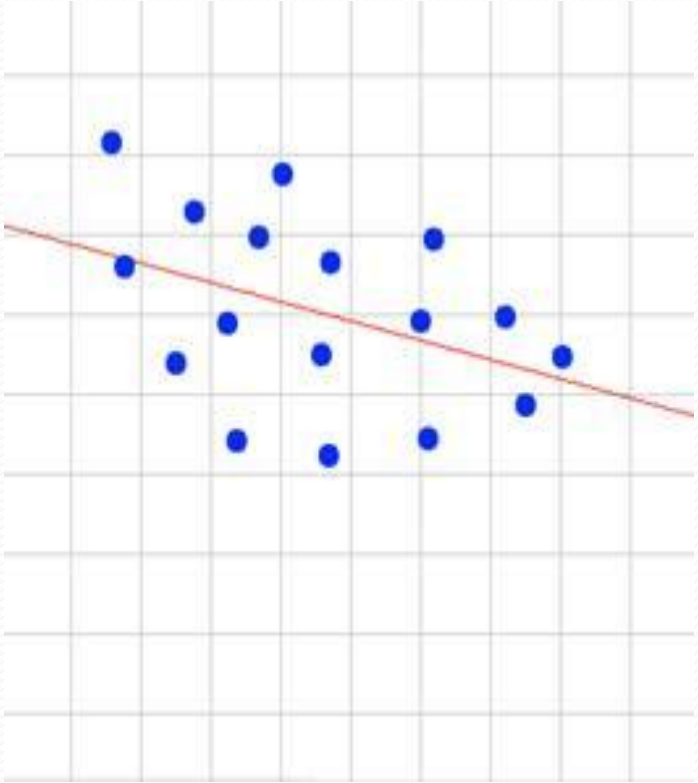
moderate negative correlation
r=-0.75



Weak Positive Correlation, r=0.3

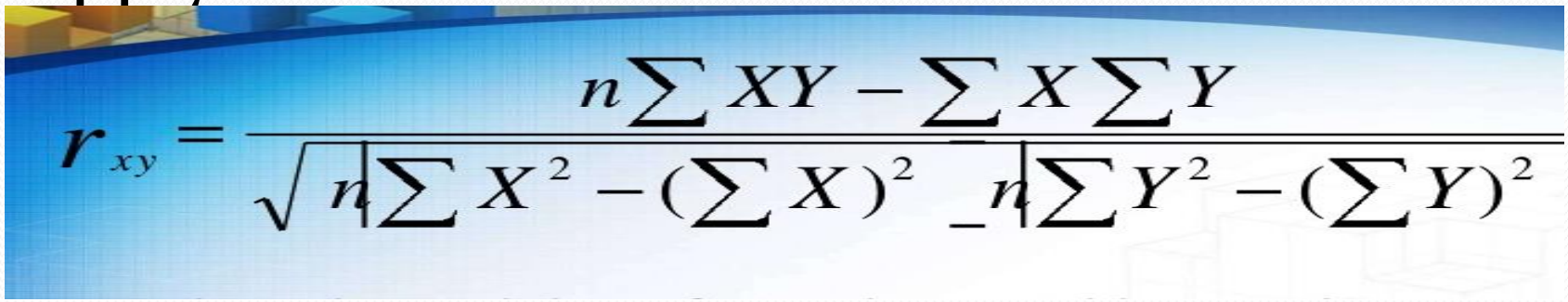


Weak negative correlation r= -0.40



Pearson's Correlation coefficient

- An agricultural research organisation tested a particular chemical fertiliser to try to find out whether an increase in the fertiliser used would lead to corresponding increase in food supply



n- sample size
X- value of Independent variable
Y-value of dependent variable

Illustration

Sl.Nos	Fertiliser (X)	Bushels of Beans(Y)	XY	X^2	Y^2
1	2	4	8	4	16
2	1	3	3	1	9
3	3	4	12	9	16
4	2	3	6	4	9
5	4	6	24	16	36
6	5	5	25	25	25
7	3	5	15	9	25

Σ	ΣX= 20	ΣY= 30	ΣXY= 93	Σ X^2 =68	Σ Y^2 =136

r=0.81

Spearman's rank coefficient

- When the values of the two variables are converted to their **ranks**, and there from the correlation is obtained, the correlations known as rank correlation
- Example: Intelligence and Beauty

$$r = 1 - \left(\frac{6\Sigma d^2}{n(n^2-1)}\right)$$

n=No of observations

D=difference between the ranks

Spurious Correlation

- A situation in which measures of two or more variables are statistically related (they cover) but are not in fact causally linked—usually because the statistical relation is caused by a third variable.

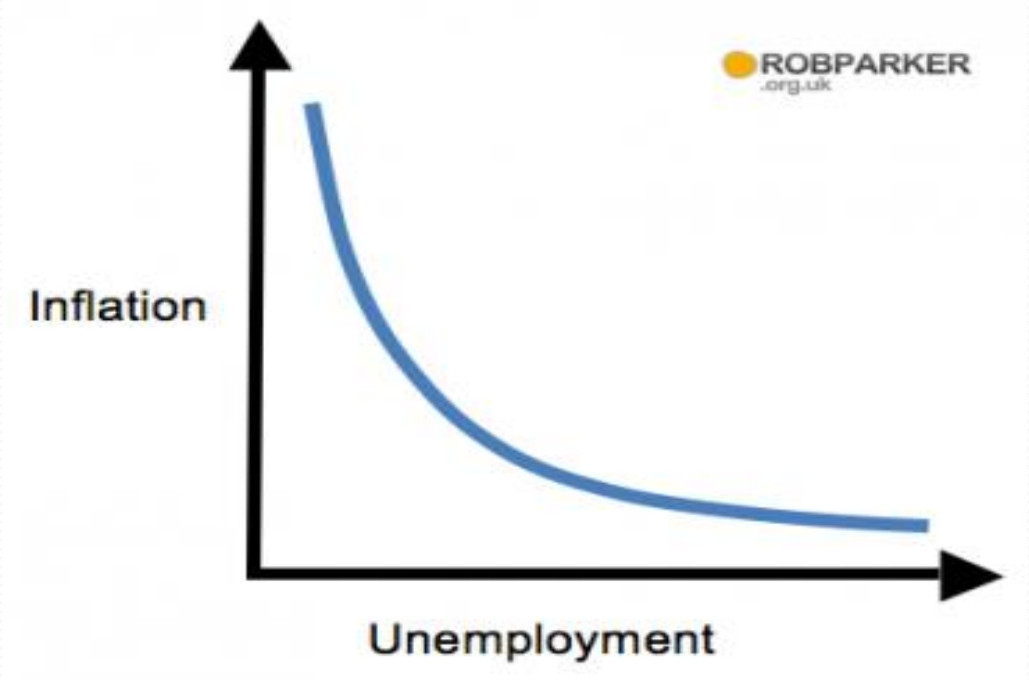
Spurious Correlation

- correlation is not present in the original observations but is produced by the way the data are handled

- Example-

1. Correlation between Long Hair and test score

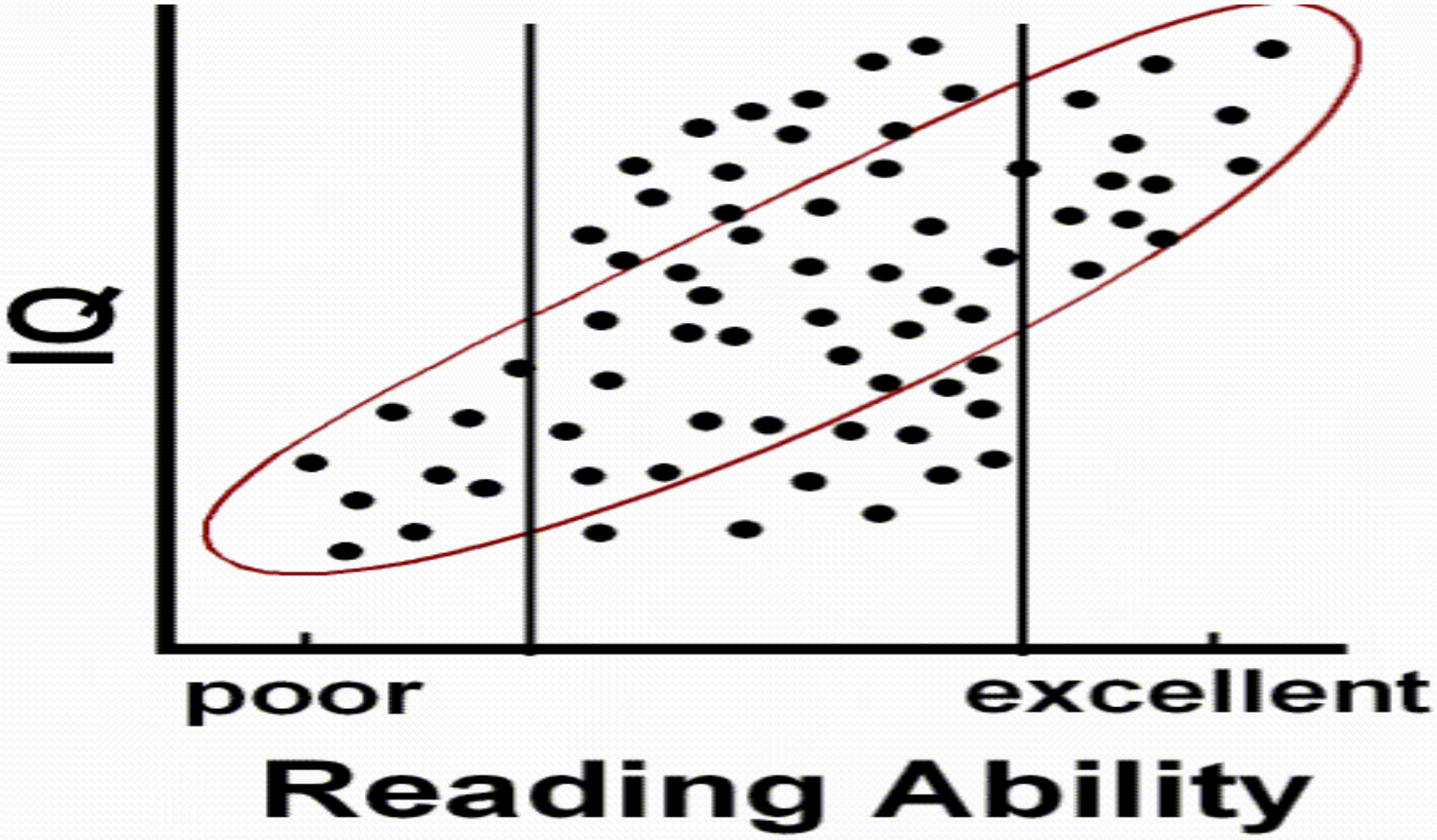
2. From economics is the Phillips curve



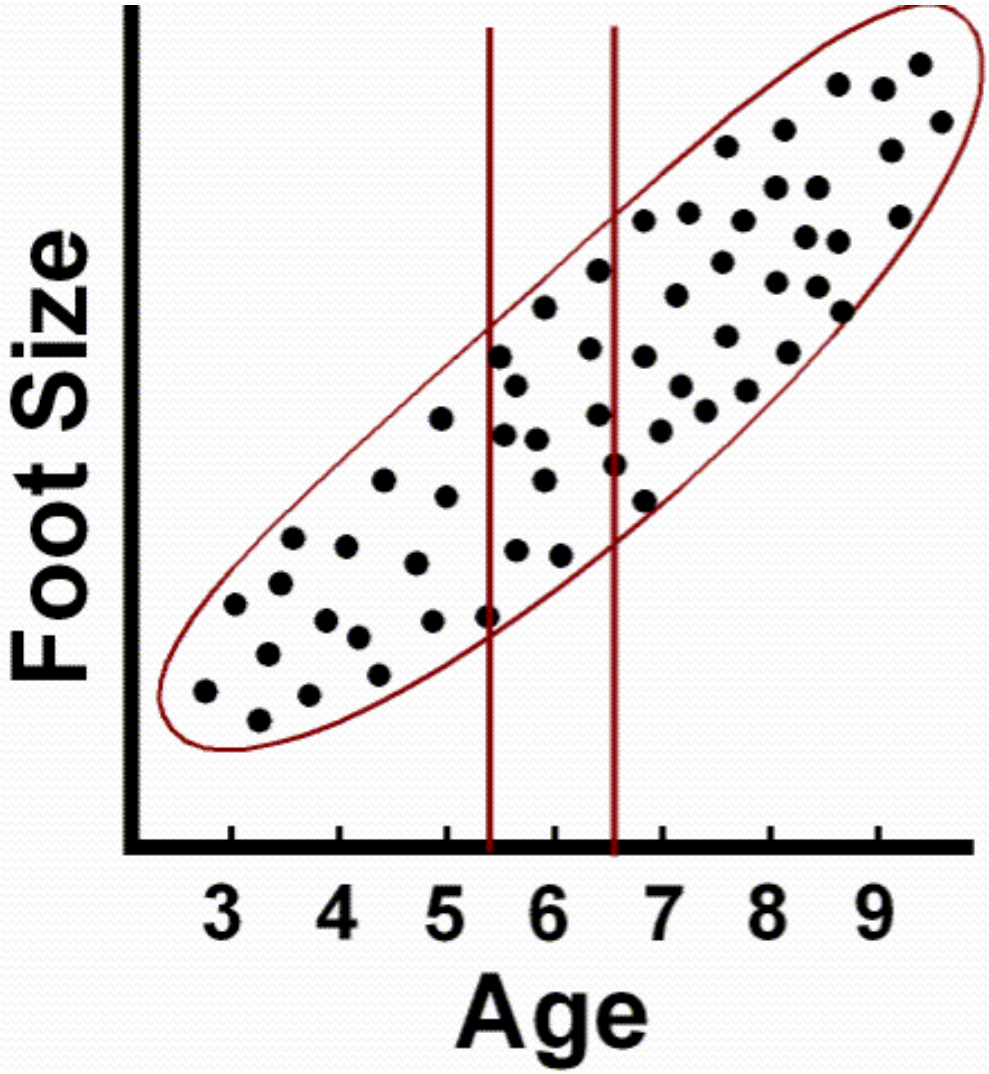
But US entered a period of Stagflation (high unemployment *and* high inflation) caused by shocks to the oil supply and government wage/price controls. (1970's)

Factors Affecting r Correlation

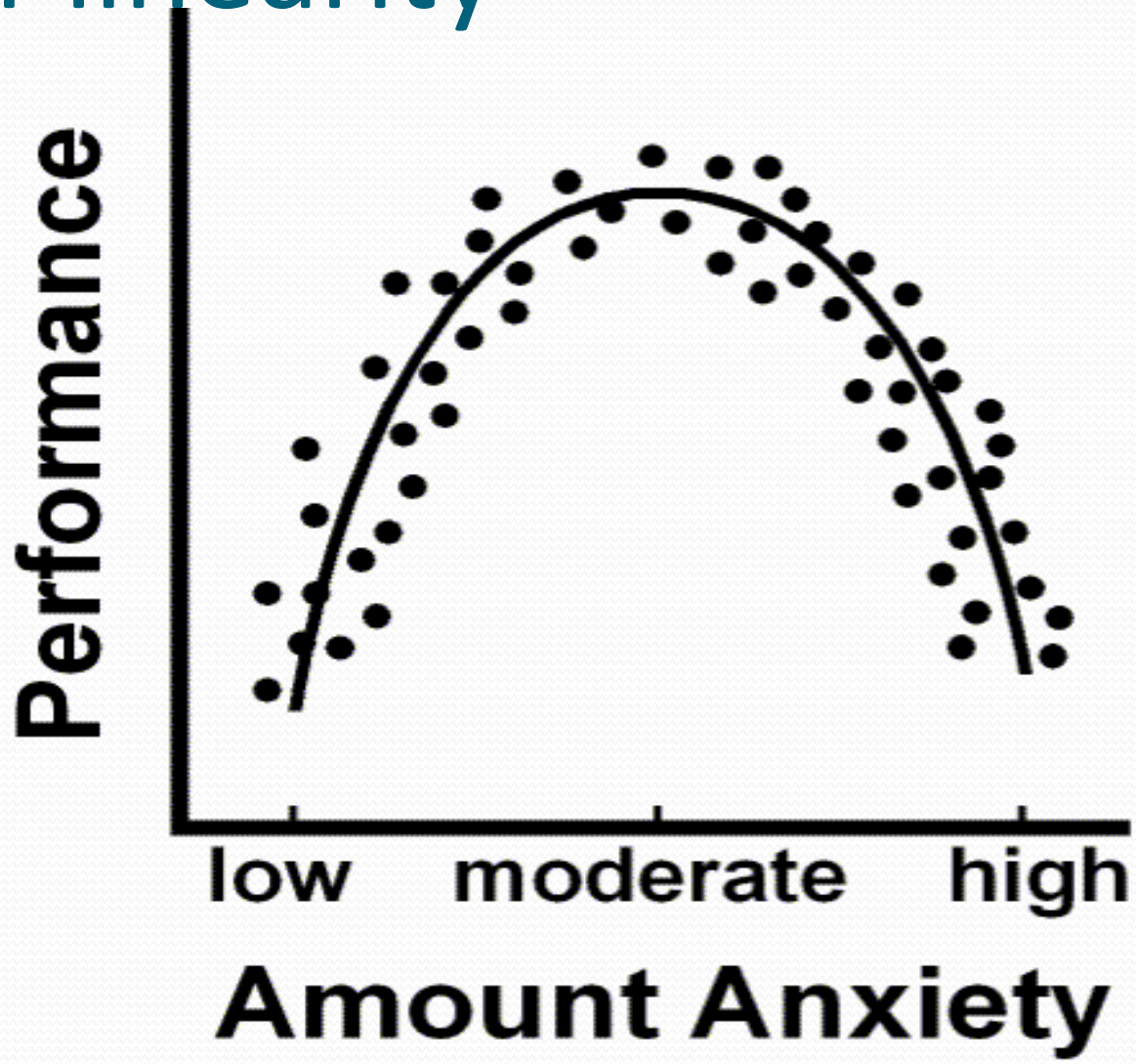
- Heterogeneous subsamples



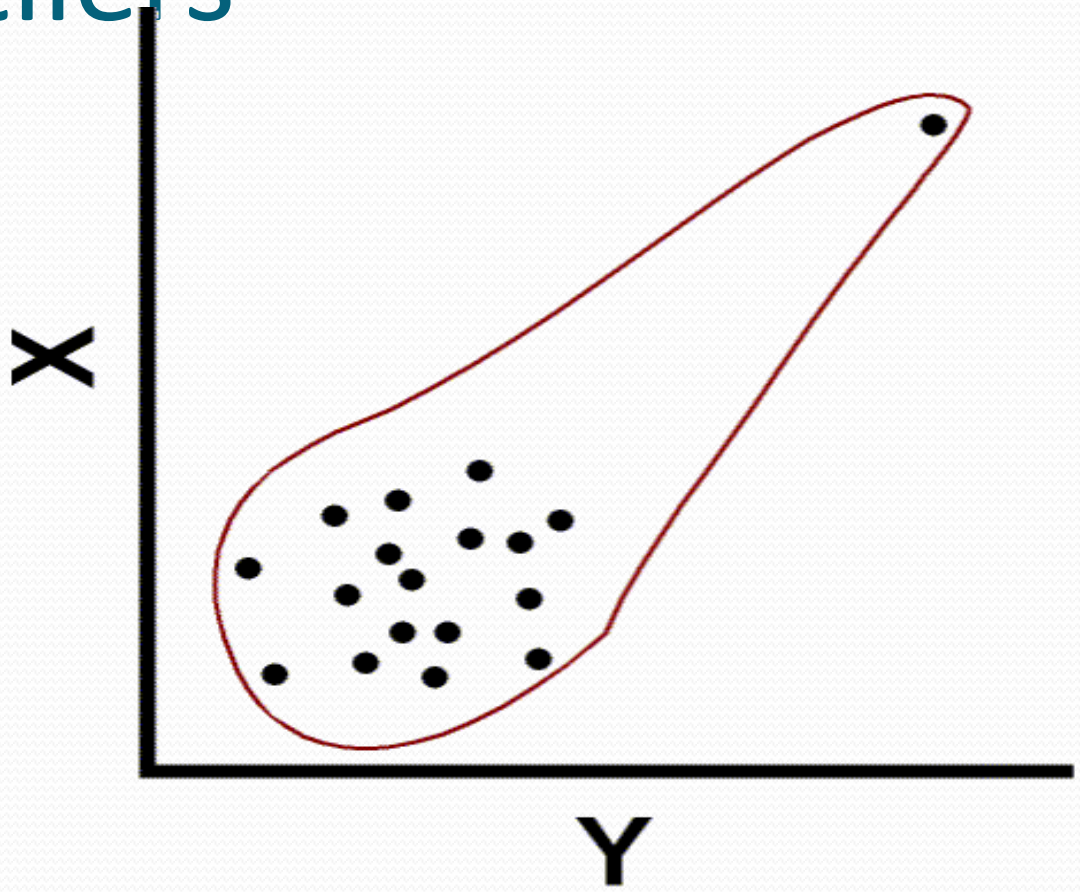
Truncation



Non-linearity



Outliers



- May be due to wrong measurement

An Example
Heart Disease and Smoking

- Does smoking cigarettes increase systolic blood pressure(Heart disease)?
- Data on heart disease and cigarette smoking in 21 developed countries (Landwehr and Watkins, 1987)
- Data have been rounded for computational convenience.
  - The results were not affected.

The Data

Surprisingly, the U.S. is the first country on the list--the country with the highest consumption and highest mortality.

Country	Cigarettes	CHD
1	11	26
2	9	21
3	9	24
4	9	21
5	8	19
6	8	13
7	8	19
8	6	11
9	6	23
10	5	15
11	5	13
12	5	4
13	5	18
14	5	12
15	5	3
16	4	11
17	4	15
18	4	6
19	3	13
20	3	4
21	3	14

What Does the Scatterplot Show?

- As smoking increases, so does coronary heart disease mortality.
- Fairly moderate relationship and positive
- Not all data points on line.
  - This gives us “residuals” or “errors of prediction”

Thank You…