Let's learn about

1.Central Limit Theorem

2.Probability

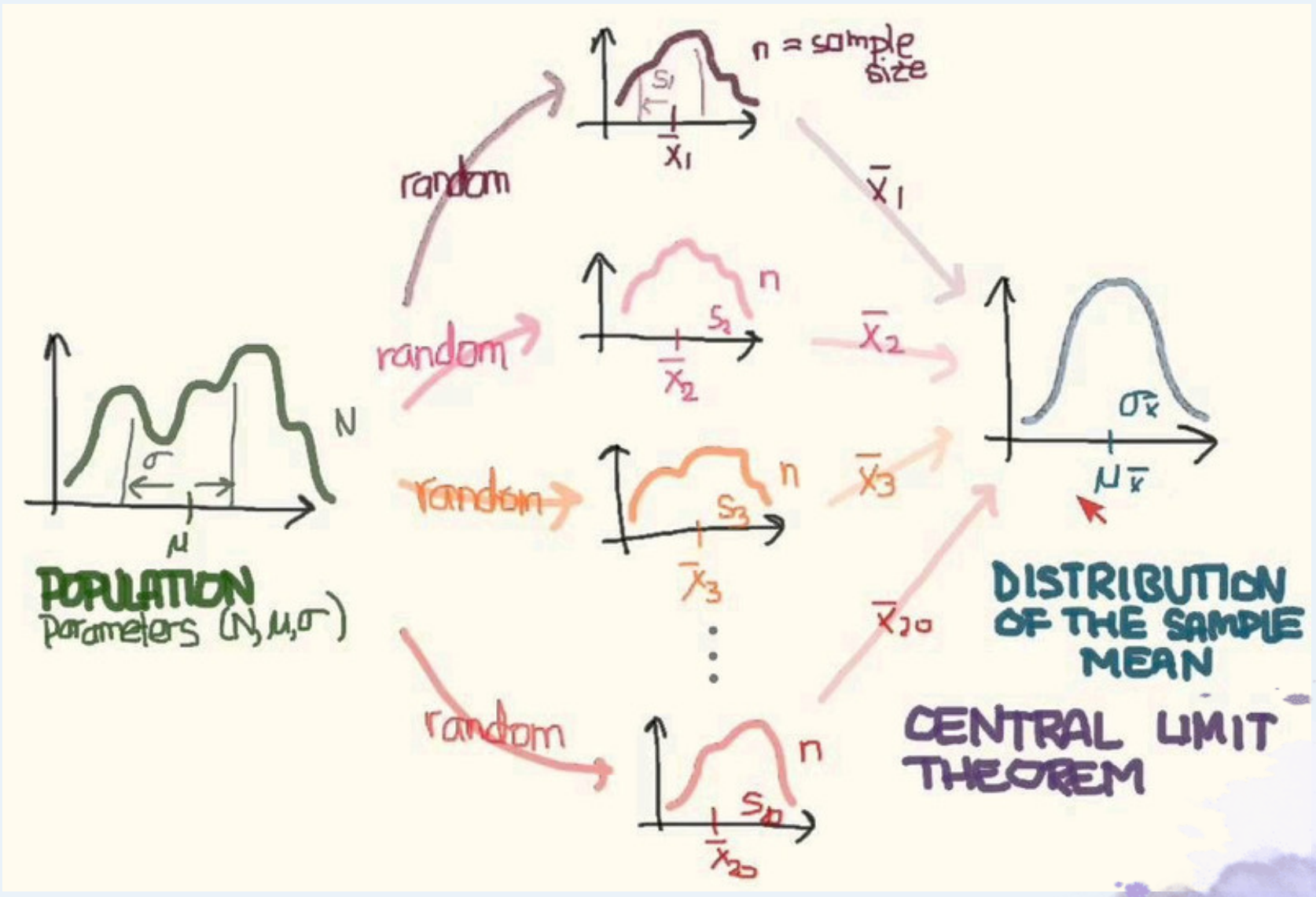
3.Permutation and Combination

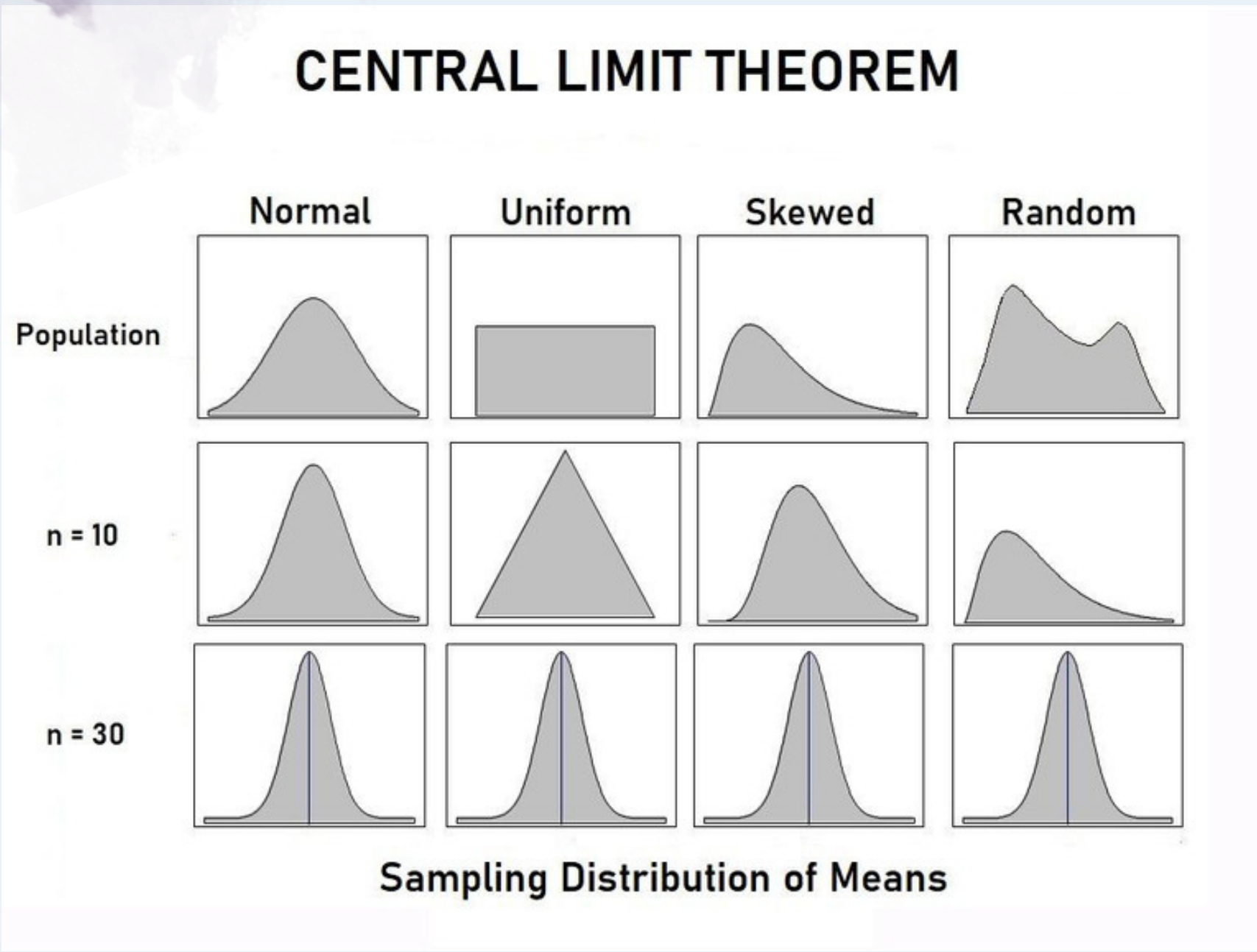
4.Covariance

- Pearson correlation coefficient
- spearman rank correlation coefficient

Central Limit Theorem

Central Limit theorm states that if we have a population data with mean **μ** and std **σ** of any distribution and take significantly large random samples from the population with replacement , then the distribution of the sample means will be approximately normally distributed.





Here we can observe that any distribution with sample size 30 follows normal distibution

Probability

Probability is a measure of the likely hood of an event.

Tossing a fair coin

Probability of head p(H) = 0.5

Probability of Tail p(T) = 0.5

Probability of event to happen

P(E) = Number of favourable outcomes/Total Number of outcomes

Probability

Probability is a measure of the likely hood of an event.

Rolling a dice

There are six outcomes

$$\{1,2,3,4,5,6\}$$

what is the probability of rolling a one = p(1) = (1/6)

$$p(A \text{ or } B) = p(A) + p(B).$$

probability of rolling 1 or 6 = p(1 or 6) = $\frac{1}{6} + \frac{1}{6}$

Probability

probability of rolling 1 or 6 = p(1 or 6) = $\frac{1}{6} + \frac{1}{6}$

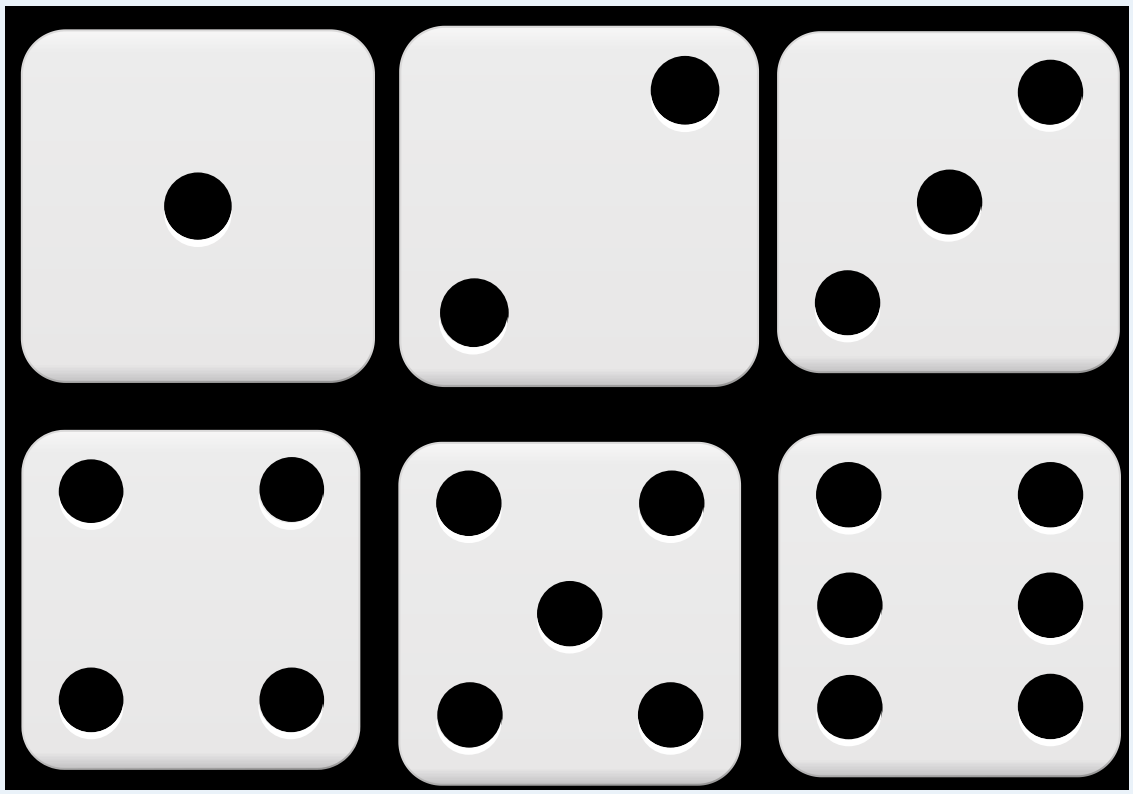
$$p(1 \text{ or } 6) = \frac{2}{6} = \frac{1}{3}$$

mutually exclusive events

Two events are mutually exclusive if they cannot occur at the same time.

Probability

mutually exclusive events



rolling a dice is also a mutually exclusive events.

Probability

non mutual exclusive events

Two events cannot occur at the same time.

picking randomly a card from a deck of cards two events "heart" and "king" can be selected.

Probability

mutual exclusive events example

what is the probability of coin landing on heads or tails?

addition rule for mutual exclusive events

$$p(A \text{ or } B) = p(A) + p(B).$$

$$p(\text{H or T}) = \frac{1}{2} + \frac{1}{2}$$

$$p(\text{H or T}) = 1$$

Probability

mutual exclusive events example

what is the probability of getting 1 or 6 or 3

addition rule for mutual exclusive events

$$p(A \text{ or } B \text{ or } C) = p(A) + p(B) + p(C)$$

$$p(1 \text{ or } 6 \text{ or } 3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6}$$

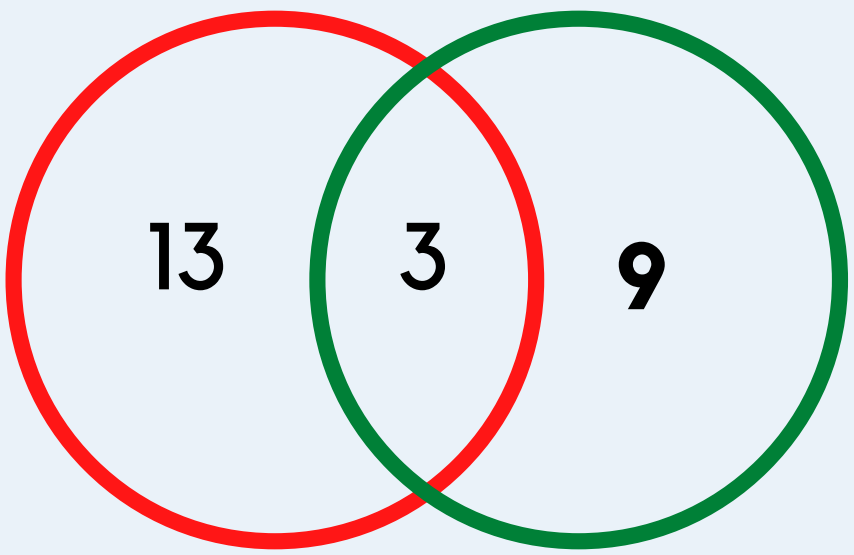
$$p(1 \text{ or } 6 \text{ or } 3) = 0.5$$

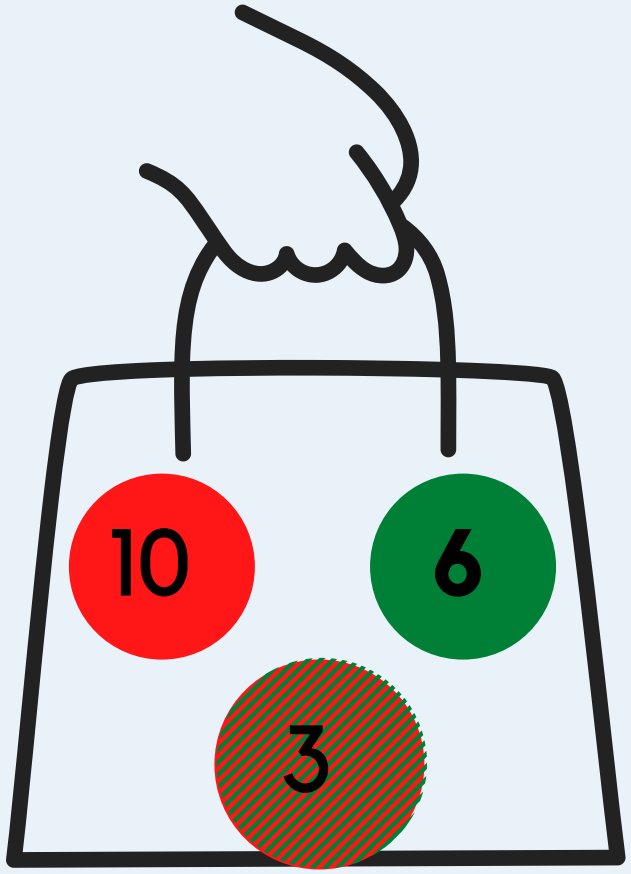
Probability

non mutual exclusive events example

Bag of Marbles: 10 red, 6 green , 3 red and green

when picking randomly from a bag of marbles what is the probability of choosing a marble that is red or green?





Probability

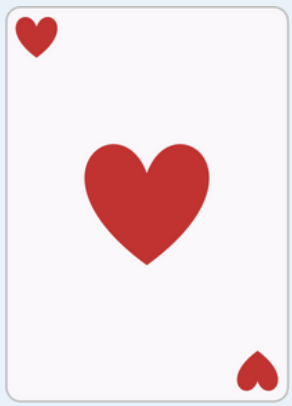
non mutual exclusive events example

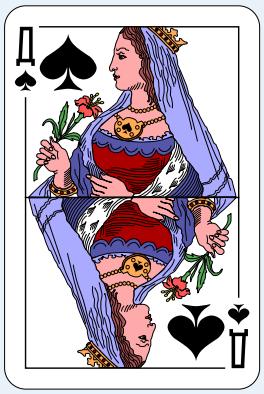
addition rule for mutual exclusive events

(A or B) = p(A) + p(B) − p(A and B).

$$p(\text{red or green}) = \frac{13}{19} + \frac{9}{19} - \frac{3}{19}$$

deck of cards - what is the probability of choosing heart or queen





Probability

non mutual exclusive events example

deck of cards - what is the probability of choosing heart or queen

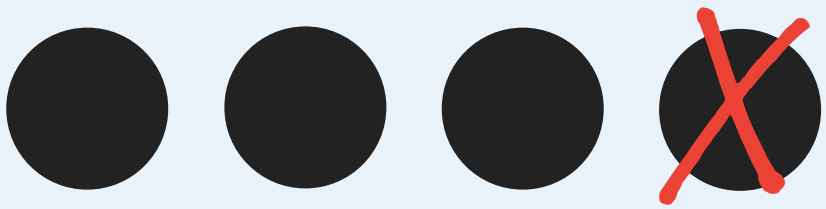
$$(A \text{ or } B) = p(A) + p(B) - p(A \text{ and } B).$$

$$p(\text{heart or queen}) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52}$$

Probability

multiplication rule

dependents events: two events are dependent if they affect one another.

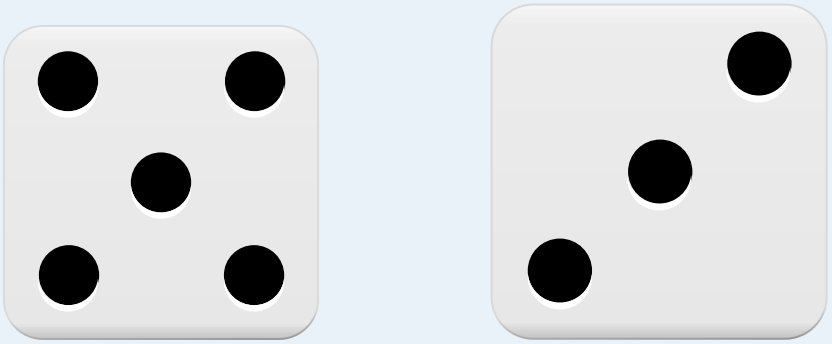


$$p(balck) = \frac{4}{7} \longrightarrow p(yellow) = \frac{3}{6}$$

Probability

multiplication rule

what is the probability of rolling "5" and rolling "3" with a normal 6 sided dice.



*p(A and B) = p(A) * p(B).*

$$\frac{1}{6} + \frac{1}{6} = \frac{1}{36}$$

Probability

P(A or B)

Mutual exclusive

Non Mutual exclusive

p(A or B) = p(A) + p(B). Mutual exclusive

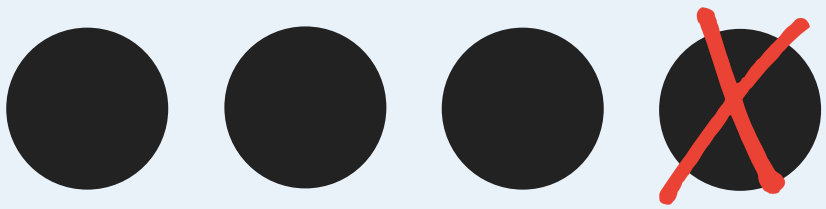
p(A or B) = p(A) + p(B) − p(A and B).

Non Mutual exclusive

Probability

dependent events

dependents events: two events are dependent if they affect one another.



$$p(A \text{ and } B) = p(A) * p(B|A)$$

$$p(\text{balck}) = \frac{4}{7} \longrightarrow p(\text{yellow}) = \frac{3}{6}$$

$$p(\text{balck and yellow}) = \frac{4}{7} * \frac{3}{6} = \frac{2}{7}$$

Permutations

A permutation is an arrangement in a definite order of several objects taken, some or all at a time, with permutations, every tiny detail matters. It means the order in which elements are arranged is significant.

$$ {}^{n}P_{r} = \frac{n!}{(n - r)!} $$

An example of permutations is the number of 2 letter words that can be formed by using the letters in a word say, **GREAT**

$$ {}^{5}P_{2} = \frac{5!}{(5 - 2)!} = \frac{5!}{3!} $$

$$ \frac{5*4*3*2*1}{3*2*1} = 20 $$

Combination

The combination is a way of selecting elements from a set so that the order of selection doesn't matter. With the combination, only choosing elements matters. It means the order in which elements are chosen is not essential.

$$ {}^{n}C_{r} = \frac{n!}{(n-r)!\ r!} $$

Find the number combinations if n = 12 and r = 2.

$$ {}^{12}C_{2} = \frac{12!}{(12-2)!\ 2!} = \frac{12!}{10!\ \ 2!} $$

$$ \frac{12 * 11 * 10!}{2! * 10!} = 66 $$

Covariance

Covariance measures the direction of the relationship between two variables. A positive covariance means that both variables tend to be high or low at the same time. A negative covariance means that when one variable is high, the other tends to be low.

Covariance is mainly used for feature selection with covariance we can find is it positively co-related or negatively co-related.

Population Covariance Formula

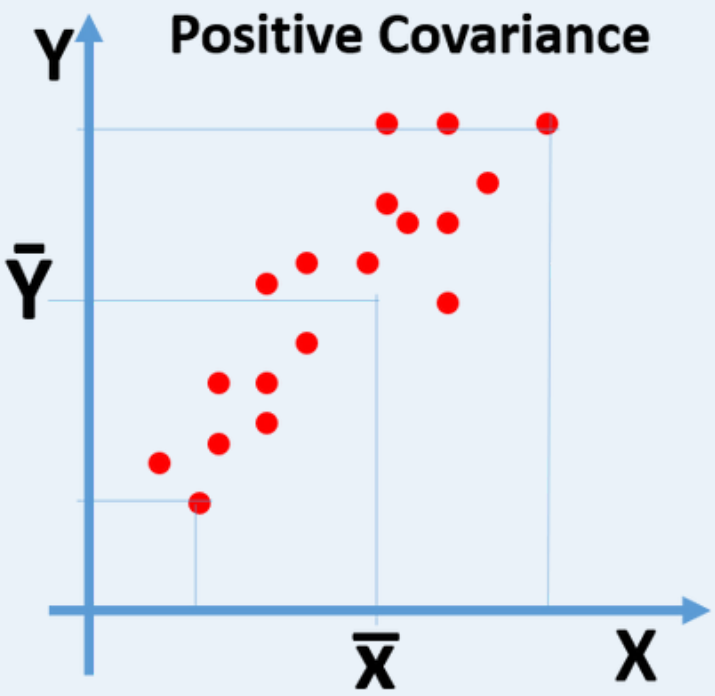
$$Cov(x,y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{N}$$

Sample Covariance

$$Cov(x,y) = \frac{\sum(x_i - \bar{x})(y_i - y)}{N - 1}$$

Positive Covariance

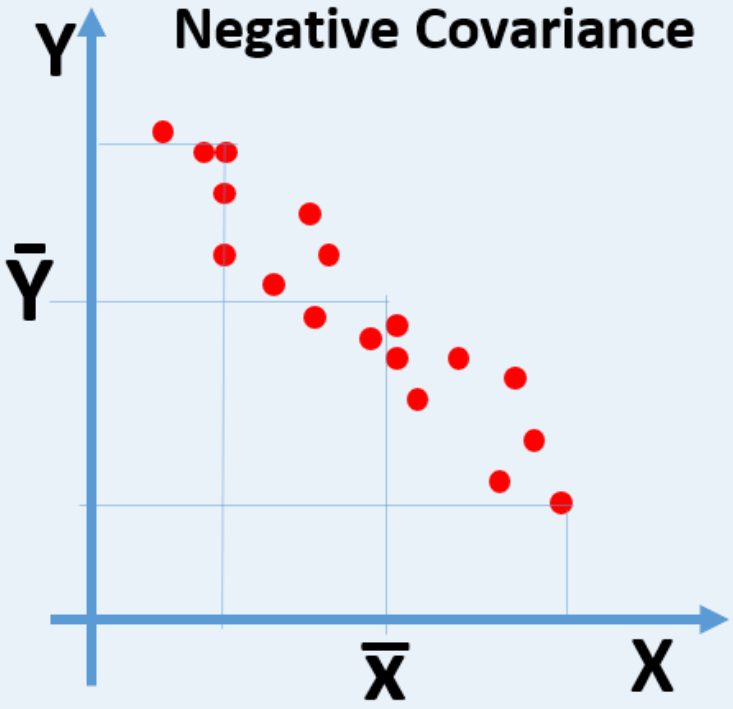
A positive covariance between two variables indicates that these variables tend to be higher or lower at the same time. In other words, a positive covariance between variables x and y indicates that x is higher than average at the same times that y is higher than average, and vice versa. When charted on a two-dimensional graph, the data points will tend to slope upwards.



Here you can observe
X increases then Y increases
or
X decreases then Y decreases

Negative Covariance

When the calculated covariance is less than zero, this indicates that the two variables have an inverse relationship. In other words, an x value that is lower than average tends to be paired with a y that is greater than average, and vice versa.



Here you can observe
X increases then Y decreases
or
X decreases then Y increases

Example of Covariance Calculation

Assume an analyst in a company has a five-quarter data set that shows quarterly gross domestic product GDP growth in percentages (x) and a company's new product line growth in percentages (y). The data set may look like:

- Q1: x = 2, y = 10
- Q2: x = 3, y = 14
- Q3: x = 2.7, y = 12
- Q4: x = 3.2, y = 15
- Q5: x = 4.1, y = 20

The average x value equals 3, and the average y value equals 14.2. To calculate the covariance, the sum of the products of the xi values minus the average x value, multiplied by the yi values minus the average y values would be divided by (n-1), as follows:

Cov(x,y) = ((2 - 3) x (10 - 14.2) + (3 - 3) x (14 - 14.2) + ... (4.1 - 3) x (20 - 14.2)) / 4 = (4.2 + 0 + 0.66 + 0.16 + 6.38) / 4 = 2.85

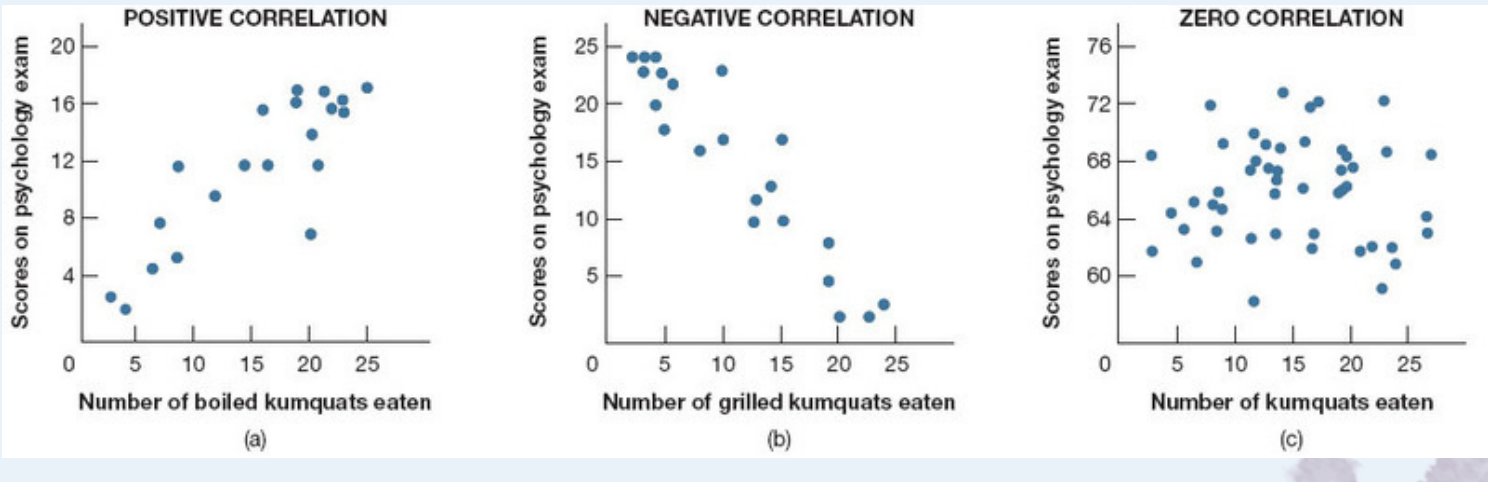
Having calculated a positive covariance here, the analyst can say that the growth of the company's new product line has a positive relationship with quarterly GDP growth.

covariance is nothing but variance

$$\mathbf{Cov(X,X)= \Sigma (x i - x)(x i - x)/ N}$$

What Is Covariance vs. Variance?

Covariance and variance are both used to measure the distribution of points in a data set. However, variance is typically used in data sets with only one variable, and indicates how closely those data points are clustered around the average. Covariance measures the direction of the relationship between two variables. A positive covariance means that both variables tend to be high or low at the same time. A negative covariance means that when one variable is high, the other tends to be low.



What Is the Difference Between Covariance and Correlation?

Covariance measures the direction of a relationship between two variables, while correlation measures the strength of that relationship. Both correlation and covariance are positive when the variables move in the same direction, and negative when they move in opposite directions. However, a correlation coefficient must always be between -1 and +1, with the extreme values indicating a strong relationship.

Pearson correlation coefficient

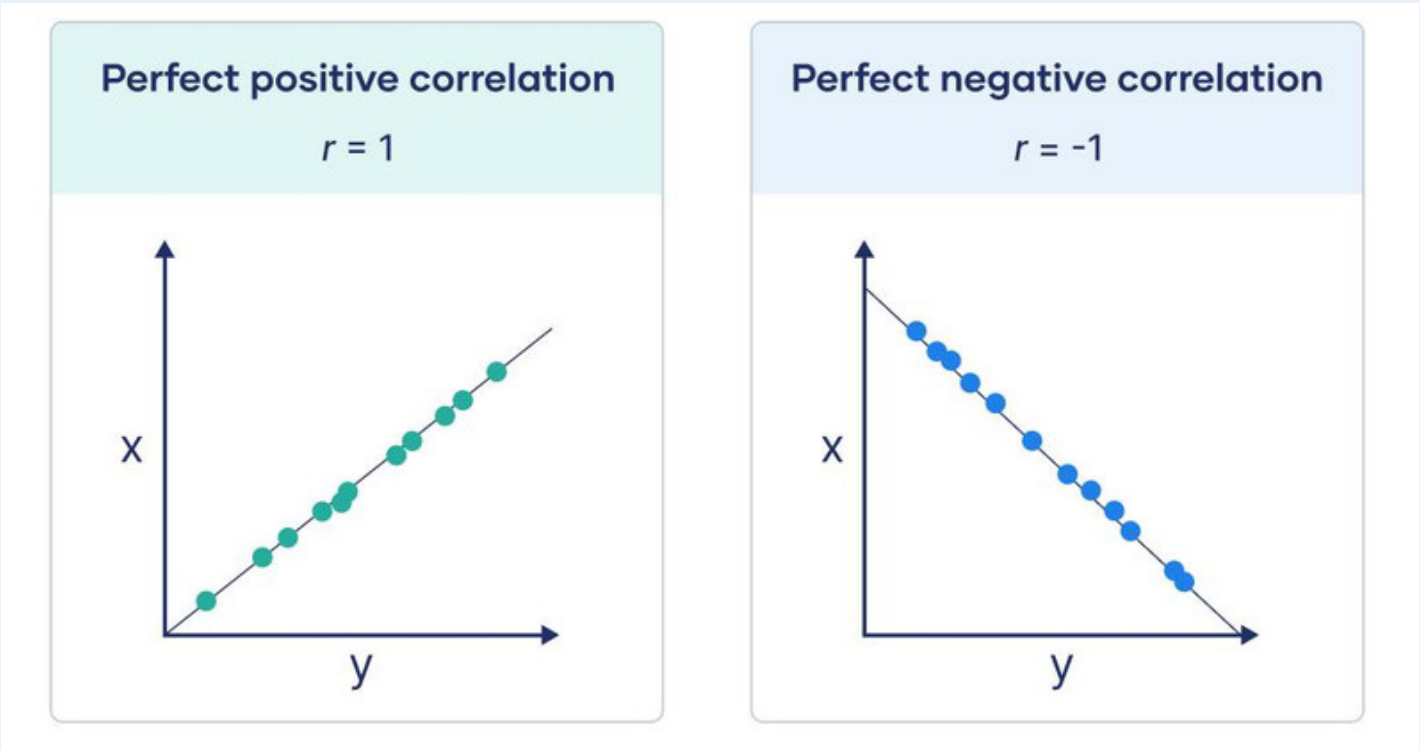
The Pearson correlation coefficient (r) is the most common way of measuring a linear correlation. It is a number between –1 and 1 that measures the strength and direction of the relationship between two variables

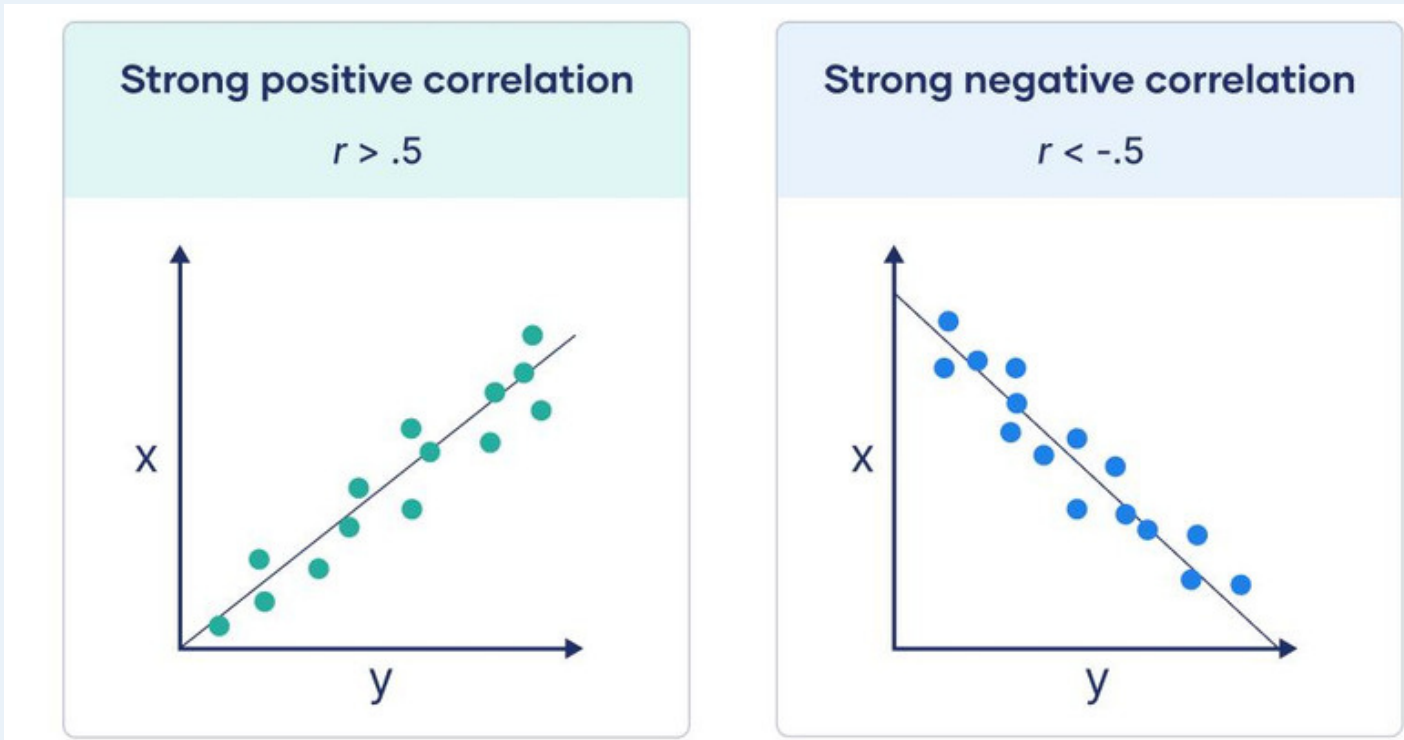
$$\rho\,(x,y) = \frac{\mathbf{cov(x,y)}}{\sigma\,\mathbf{x} * \sigma\,\mathbf{y}}$$

more the values towards +1 more positively co related

more the values towards -1 more negatively co related

Pearson correlation coefficient

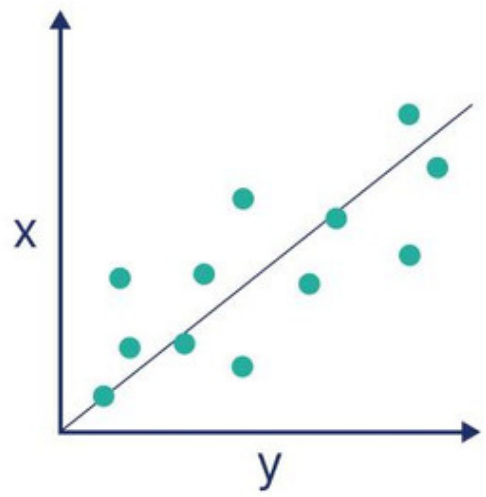




Pearson correlation coefficient

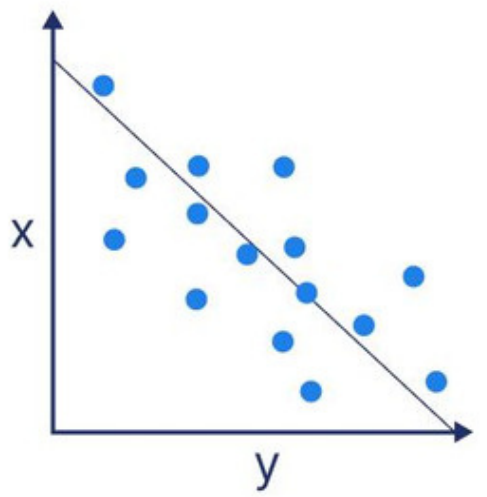
Weak positive correlation

$.3 > r > 0$



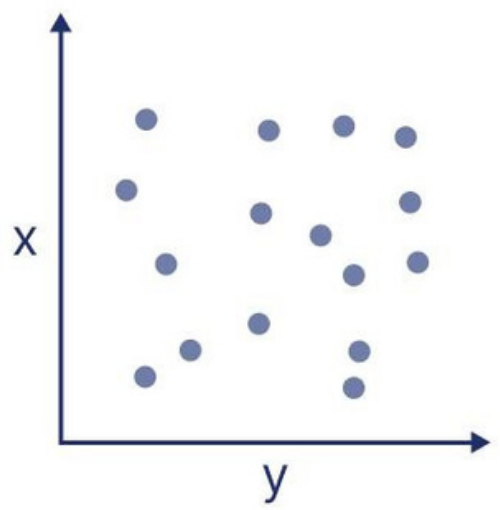
Weak negative correlation

$0 > r > -.3$



No correlation

$r = 0$



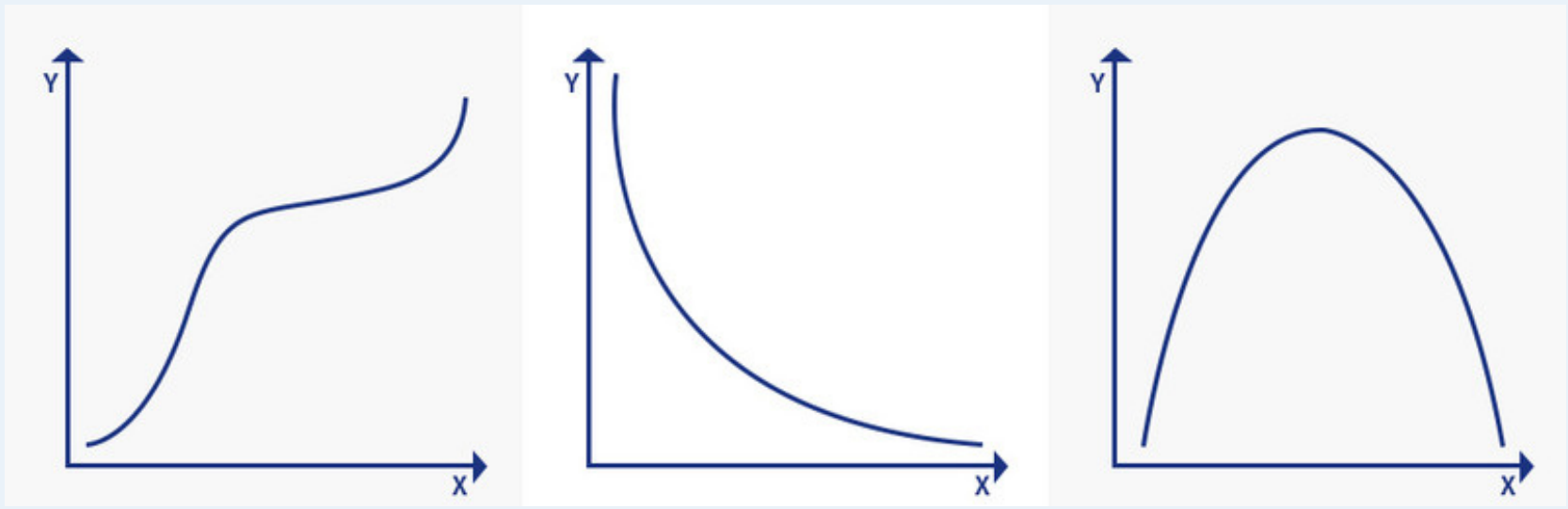
Spearman correlation coefficient

The Spearman's rank coefficient of correlation is a nonparametric measure of rank correlation (statistical dependence of ranking between two variables).
Named after Charles Spearman, it is often denoted by the Greek letter **'ρ'** (rho) and is primarily used for data analysis.
It measures the strength and direction of the association between two ranked variables. But before we talk about the Spearman correlation coefficient, it is important to understand Pearson's correlation first. A Pearson correlation is a statistical measure of the strength of a linear relationship between paired data.

$$\rho(R(x),R(y)) = \frac{\mathbf{cov(R(x),R(y))}}{\sigma \mathbf{R(x)} * \sigma \mathbf{R(y)}}$$

Spearman correlation coefficient

What Is Monotonic Function?
To understand Spearman's rank correlation, it is important to understand monotonic function. A monotonic function is one that either never increases or never decreases as its independent variable changes.
The following graph illustrates the monotonic function:



- Monotonically Increasing: As the variable X increases, the variable Y never decreases.
- Monotonically Decreasing: As the variable X increases, the variable Y never increases.
- Not Monotonic: As the X variable increases, the Y variable sometimes decreases and sometimes increases.

Spearman correlation coefficient

$$r_R = 1 - \frac{6\Sigma_i {d_i}^2}{n(n^2-1)}$$

Here,

n= number of data points of the two variables

di= difference in ranks of the "ith" element

The Spearman Coefficient, ρ, can take a value between +1 to -1 where,

- A ρ value of +1 means a perfect association of rank
- A ρ value of 0 means no association of ranks
- A ρ value of -1 means a perfect negative association between ranks.

Closer the ρ value to 0, weaker is the association between the two ranks.

Spearman correlation coefficient

Example of Spearman's Rank Correlation
Consider the score of 5 students in Maths and Science that are mentioned in the table

Students	Maths	Science
A	35	24
B	20	35
C	49	39
D	44	48
E	30	45

Step 1: Create a table for the given data.

Step 2: Rank both the data in descending order. The highest marks will get a rank of 1 and the lowest marks will get a rank of 5.

Step 3: Calculate the difference between the ranks (d) and the square value of d.
Step 4: Add all your d square values.

Spearman correlation coefficient

Students	Maths	Rank	Science	Rank	d	d square
A	35	3	24	5	2	4
B	20	5	35	4	1	1
C	49	1	39	3	2	4
D	44	2	48	1	1	1
E	30	4	45	2	2	4
						14

Step 5: Insert these values into the formula.

$$r_R = 1 - \frac{6\Sigma_i {d_i}^2}{n(n^2 - 1)}$$

= 1 - (6 * 14) / 5(25 - 1)
= 0.3
The Spearman's Rank Correlation for the given data is 0.3. The value is near 0, which means that there is a weak correlation between the two ranks.

Thank you!