L.5 Measures of variation

<u>Coefficient of quartile</u>: is a measure of dispersion, which is used to describe the spread or distribution of data. It is calculated as follows:

Coefficient of quartile deviation = (Q75 – Q25) / (Q75 + Q25)

<u>Example 1:</u>

A = {2, 4, 6, 8, 10, 12, 14}

n = 7, range = 12, mean = 8, median = 8, Q_1 = 4, Q_3 = 12,

Quartile coefficient of dispersion = 0.5

B = {1.8, 2, 2.1, 2.4, 2.6, 2.9, 3}

n = 7, range = 1.2, mean = 2.4, median = 2.4, Q_1 = 2, Q_3 = 2.9,

Quartile coefficient of dispersion = 0.18

The quartile coefficient of dispersion of data set A is 2.7 times as great (0.5 / 0.18) as that of data set B.

The probability

Probability means possibility, deals with the occurrence of a random event. The value is expressed from zero to one. It used to predict how likely events are to happen.

Probability can range from 0 to 1, where 0 means the event to be an impossible one and 1 indicates a certain event.

For example, when we toss a coin, either we get Head OR Tail, only two possible outcomes are possible (H, T). The probability of a head in any single flip of the coin

equals 1/2 . But when two coins are tossed then there will be four possible outcomes, i.e {(H, H), (H, T), (T, H), (T, T)}.

Probability of event to happen P(E) = Number of favourable outcomes/Total Number of outcomes

ex- There are 6 pillows in a bed, 3 are red, 2 are yellow and 1 is blue. What is the probability of picking a yellow pillow?

Ans: The probability is equal to the number of yellow pillows in the bed divided by the total number of pillows, i.e. 2/6 = 1/3.

Probability of an Event

Assume an event E can occur in **r** ways out of a sum of **n** probable or possible **equally likely ways**. Then the probability of happening of the event or its success is expressed as;

P(E) = r/n

The probability that the event will not occur or known as its failure is expressed as:

P(E') = (n-r)/n = **1-(r/n)**

E' represents that the event will not occur.

Therefore, now we can say;

P(E) + P(E') = 1

This means that the total of all the probabilities in any random test or experiment is equal to 1.

ex: A vessel contains 4 blue balls, 5 red balls and 11 white balls. If three balls are drawn from the vessel at random, what is the probability that the first ball is red, the second ball is blue, and the third ball is white?

Solution:

The probability to get the first ball is red or the first event is 5/20.

Since we have drawn a ball for the first event to occur, then the number of possibilities left for the second event to occur is 20 – 1 = 19.

Hence, the probability of getting the second ball as blue or the second event is 4/19.

Again with the first and second event occurring, the number of possibilities left for the third event to occur is 19 – 1 = 18.

And the probability of the third ball is white or the third event is 11/18.

Therefore, the probability is 5/20 x 4/19 x 11/18 = 44/1368 = 0.032.

Or we can express it as: P = 3.2%.

ex: Two dice are rolled, find the probability that the sum is:

1. **equal to 1**
2. **equal to 4**
3. **less than 13**

Solution:

To find the probability that the sum is equal to 1 we have to first determine the sample space S of two dice as shown below.

S = { (1,1),(1,2),(1,3),(1,4),(1,5),(1,6)

(2,1),(2,2),(2,3),(2,4),(2,5),(2,6)

(3,1),(3,2),(3,3),(3,4),(3,5),(3,6)

(4,1),(4,2),(4,3),(4,4),(4,5),(4,6)

(5,1),(5,2),(5,3),(5,4),(5,5),(5,6)

(6,1),(6,2),(6,3),(6,4),(6,5),(6,6) }

So, n(S) = 36

1) Let E be the event "sum equal to 1". Since, there are no outcomes which where a sum is equal to 1, hence,

P(E) = n(E) / n(S) = 0 / 36 = 0

2) Let A be the event of getting the sum of numbers on dice equal to 4.

Three possible outcomes give a sum equal to 4 they are:

A = {(1,3),(2,2),(3,1)}

n(A) = 3

Hence, P(A) = n(A) / n(S) = 3 / 36 = 1 / 12

3) Let B be the event of getting the sum of numbers on dice is less than 13.

From the sample space, we can see all possible outcomes for the event B, which gives a sum less than B. Like:

(1,1) or (1,6) or (2,6) or (6,6).

So you can see the limit of an event to occur is when both dies have number 6, i.e. (6,6).

Thus, n(B) = 36

Hence,

P(B) = n(B) / n(S) = 36 / 36 = 1

Probability questions

1. Two dice are thrown together. Find the probability that the product of the numbers on the top of the dice is:
   (i) 6 (ii) 12 (iii) 7
2. A bag contains 10 red, 5 blue and 7 green balls. A ball is drawn at random. Find the probability of this ball being a
   (i) red ball (ii) green ball (iii) not a blue ball

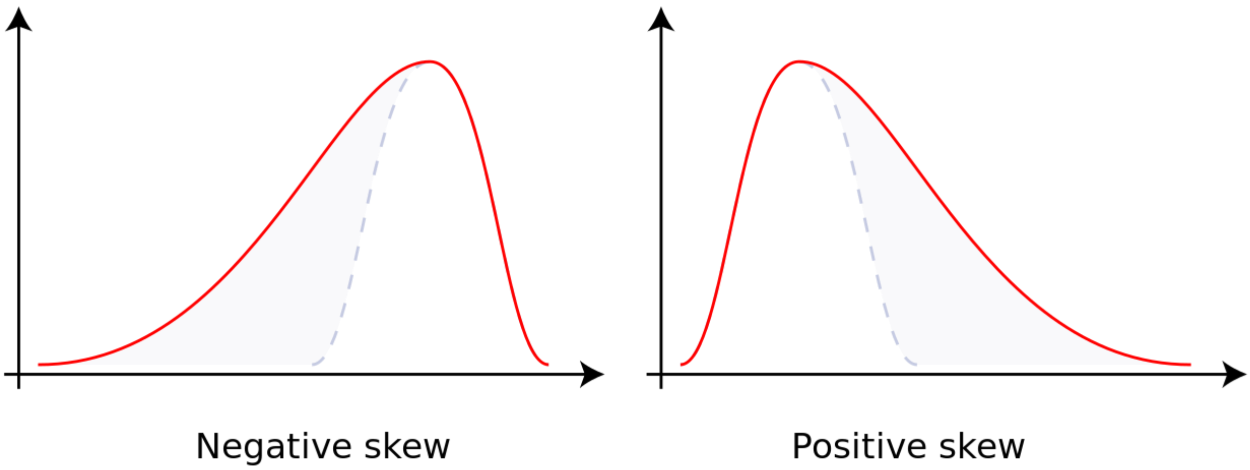
Skewness

It is a measure of the asymmetry of the probability distribution of a real-valued random variable about its mean. The skewness value can be positive, zero, negative, or undefined.

1- **Negative skew:** The left tail is longer; the mass of the distribution is concentrated on the right of the figure. The distribution is said to be left-skewed, left-tailed, or skewed to the left, despite the fact that the curve itself appears to be skewed or leaning to the right; left instead refers to the left tail being drawn out and, often, the mean being skewed to the left of a typical center of the data. A left-skewed distribution usually appears as a right-leaning curve. Mean < Median < Mode

2- **Positive skew:** The right tail is longer; the mass of the distribution is concentrated on the left of the figure. The distribution is said to be right-skewed, right-tailed, or skewed to the right, despite the fact that the curve itself appears to be skewed or leaning to the left; right instead refers to the right tail being drawn out and, often, the mean being skewed to the right of a typical center of the data. A right-skewed distribution usually appears as a left-leaning curve. Mode < Median < Mean

- **zero value** in skewness means that the tails on both sides of the mean balance out overall; this is the case for a symmetric distribution. Mean= median=mode are equal then the distribution is a **normal distribution** and the coefficient of skewness will be 0.
- asymmetric distribution where one tail is long and thin, and the other is short but fat. Thus, the judgement on the symmetry of a given distribution by using only its skewness is risky; the distribution shape must be taken into account.



Nonparametric skew: defined as where is the mean, is the median,

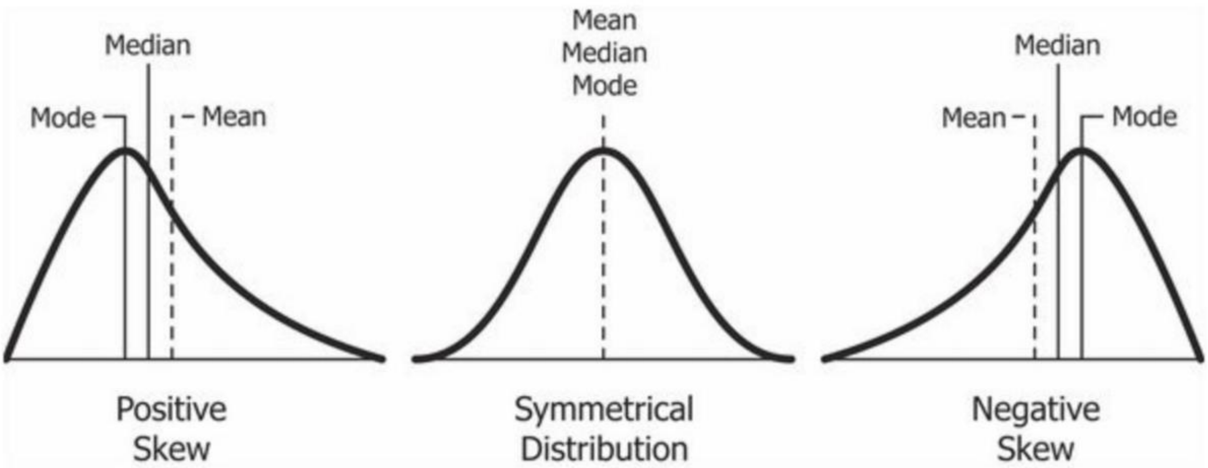
and is the standard deviation.

The skewness is defined in terms of this relationship:

- positive/right nonparametric skew means the mean is greater than (to the right of) the median,
- while negative/left nonparametric skew means the mean is less than (to the left of) the median.

If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness.

If the distribution is both symmetric and unimodal, then the mean = median = mode. This is the case of a coin toss or the series 1,2,3,4,...



Coefficient of Skewness Formula

Coefficient of Skewness

Using Mode: $\frac{\bar{x} - \text{Mode}}{s}$

Using Median: $\frac{3(\bar{x} - \text{Median})}{s}$

How to Calculate Coefficient of Skewness?

Depending upon the data available either of the two formulas can be used to calculate the coefficient of skewness.

EX: the mean of a data set is 60.5, the mode is 75, the median is 70 and the standard deviation is 10. The steps to calculate the coefficient of skewness are as follows:

Using Mode

- Step 1: Subtract the mode from the mean. mean -mode = 60.5 - 75 = -14.5
- Step 2: Divide this value by the standard deviation to get the coefficient of skewness. Thus, sk_1 = -14.5 / 10 = -1.45.

Using Median : the most popular called (Karl Pearson Coefficient of Skewness formula)

- Step 1: Subtract the median from the mean. mean- mode = 60.5 - 70 = -9.5
- Step 2: Multiply this value by 3. This gives -28.5.
- Step 3: Divide the value from step 2 by the standard deviation to obtain the coefficient of skewness. Thus, sk_2 = -28.5 / 10 = -2.85