**Measuring the Dispersion**

**Characteristics of an ideal measure of Dispersion.**

The desired characteristics, for an ideal measure of dispersion arc the same as those for all ideal ·measure of central tendency,

1. It should he rigidly defined.
2. It should be easy to calculate and understand.
3. It should be based on all the observations
4. It should be amenable to further mathematical treatment.
5. It should not be affected by sampling methods.

The following are the various measures of dispersion

1. Range
2. Quartile deviation or semi quartile deviation
3. Mean deviation
4. Standard deviation
5. Variance

An average does not tell full story.it is hardly fully representative of a mass unless we know the way the individual items scatter around it. Further description of series is necessary if we are to gauge how representative the average is.

**Dispersion is the measure of variation of the items.---- AL Bowly**

**The degree to which numerical data tend to spread about an average value is called variation or dispersion of data.**

**Range**

It is the difference between the greatest (Maximum) and smallest(minimum) observations of the distribution.

Range = X max– X min

The range is simplest of all the measures of dispersion.

If A and B are the greatest and smallest observations respectively in a distribution, then its range is A-B.

Coefficient of the range = (X max– Xmin)/ **(**X max + X min)

This technique is used to neutralise the units

**Quasi Range:**

**Merits and demerits of range**

**Merits**

1. It is rigidly defined and easy to understand.

**Demerits**

1. Range is not based on entire set of data.
2. It is affected by fluctuations of sampling.
3. Range cannot be used if you are using open ended classes.
4. It is not suitable for mathematical treatment.

**Fractile**

In a frequency distribution the location of a value at or above a given fraction of data is called Fractile.

The Median is 0.5 fractile because half of the data set is less than or equal to this value.

Fractiles are like percentages.

25% of data lie at or below 0.25 fractiles. Similarly, 25% of data lie below the 25th **percentile**

The inter fractile range is a measure of the spread between two fractiles in frequency distribution i.e. the difference between the values of fractiles.

Fractiles have special names depending on the number of equal parts into which they divide the data.

Fractile that divide data into 10 equal parts are called **deciles**.

Percentiles divide the data into 100 equal parts.

**Quartiles**

Quartiles divide data into four equal parts.

Inter quartile range = Q3-Q1

**Quartile deviation** is obtained by dividing the inter quartile range by 2

It is also known as semi quartile range,

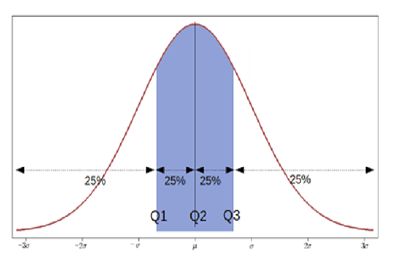
**Q.D = (Q3 -Q1)/2**

2, 11, 14, 20, 21, 25, 29, 35, 38, 46, 52, 70

Lower Quartile Median Upper Quartile

Q1(17) Q2(27) Q3(42)

**Coefficient of Q.D =**



**Examples**

1. For the following scores of students in mathematics find the 80th percentile

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 95 | 81 | 59 | 68 | 100 | 92 | 75 | 67 | 85 | 79 |
| 71 | 88 | 100 | 94 | 87 | 65 | 93 | 72 | 83 | 91 |

**Percentile range**: This is the measure of dispersion based on the difference between certain percentiles. If Pi is the ith percentile and Pj is the jth percentile. Then i-j percentile range is given by Pj-Pi (i<j).

i-j semi percentile range is given by (Pj-Pi) **/** 2.

Co-efficient of i-j percentile = Pj-Pi

Pj+Pi

1. For the following data compute ∑
2. Range
3. Inter fractile range between 20th and 80th percentiles.
4. Inter Quartile Range

25,38,36,26,22,38,22,37,24,36,21,29,32,28,34,32,23,39,33

Find 1) inter quartile range, quartile deviation, co-efficient of quartile deviation for the following distribution.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Class Interval | 0-15 | 15-30- | 30-45 | 45-60 | 60-75 | 75-90 | 90-105 |
| Frequency | 8 | 26 | 30 | 45 | 20 | 17 | 4 |

**Solution: N=** ∑ f = 150

N/4 = 37.5

3N/4 = 112.5

|  |  |  |
| --- | --- | --- |
| Class Interval | f | <cf |
| 0-15 | 8 | 8 |
| 15-30 | 26 | 34 |
| 30-45 | 30 | 64 |
| 45-60 | 45 | 109 |
| 60-75 | 20 | 129 |
| 75-90 | 17 | 146 |
| 90-105 | 4 | 150 |

Q1=l+h/f(N/4-C) Q3= l+h/f(3N/4-C)

=30+3.5/2 = 60+ (3\*3.5)/4

=30+1.75 = 60+ 2.625

=31.75 = 62.625

L = lower value of the median class, h= width of the median class, f= frequency of the median class, N= total number of observations, C cumulative frequency upto the median class

Inter quartile range = Q3-Q1

= 62.625-31.750

=30.875

Quartile deviation = (Q3-Q1) / 2

= 30.875 / 2 = 15.44

Co-efficient of quartile deviation = (Q3-Q1)/ Q3+Q1 = 30.88/94.38 = 0.33

**Mean deviation**

Mean deviation or Average Deviation is the average amount of scatter of items in a distribution from either the mean or from the median ignoring the signs of deviations.

The average often taken is mean, so the name mean deviation.

Computation of mean deviation

If x1, x2, x3….xn  are n given observations then their mean deviation is given by about an average say Mean ( M) is given by

M.D = ∑ │ X-M│ = where d = │ X-M │.

For Example: From a frequency distribution or grouped or continuous frequency distribution, mean deviation about an average A is given by

MD= 1/N ∑f│ X-A│ where f is a frequency of the class interval. N is the total frequency.

If it is a frequency distribution, X is the value of the variable.

If it is a class interval, X is the mid value of the interval.

For Example: calculate the mean deviation from the mean of the following data

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Class Interval | 2-4 | 4-6 | 6-8 | 8-10 |
| Frequency | 3 | 4 | 2 | 1 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Class | Mid-value X | Frequency f | = *x-*5.2 | f\*│ | ∑ f \* x |
| 2-4 | 3 | 3 | 2.2 | 6.6 | 9 |
| 4-6 | 5 | 4 | 0.2 | 0.8 | 20 |
| 6-8 | 7 | 2 | 1.8 | 3.6 | 14 |
| 8-10 | 9 | 1 | 3.8 | 3.8 | 9 |
| Total |  | N=10 |  | 14.8 | 52 |

=∑ fi\*x/ ∑ fi is the mean. = 52/10 = 5.2

MD= 1/N ∑f│ X-A│= 14.8/10 = 1.48

**Merits and demerits of M.D**

**Merits**

1. It is clearly defined and easy to understand.
2. Based on all observations better than range and quartile deviation.
3. Averaging the absolute deviation from an average removes the irregularities in the distributions.
4. It is less affected by extreme values.
5. Since it is based on deviation about an average it provides better measure for comparison.

**Demerits**

1. We ignore the signs of deviations while calculating M.D which is illogical.
2. It is not used sociological studies.
3. It cannot be used in open ended classes.

**Standard Deviation**

It is denoted by the symbol and is defined as positive square root of average of squared distances of the observations from the mean. It is given by

=√ 2 =( (∑(x-µ)2 )/N) = sqrt (∑ (x2 / N) - µ2)



Where x= is the observation

µ = population mean

N =Total number of elements in population

= population standard deviation

**Variance**

Variance ( 2) is given by 2 = (∑ (x2 / N) - µ2).

**Uses of the Standard Deviation**

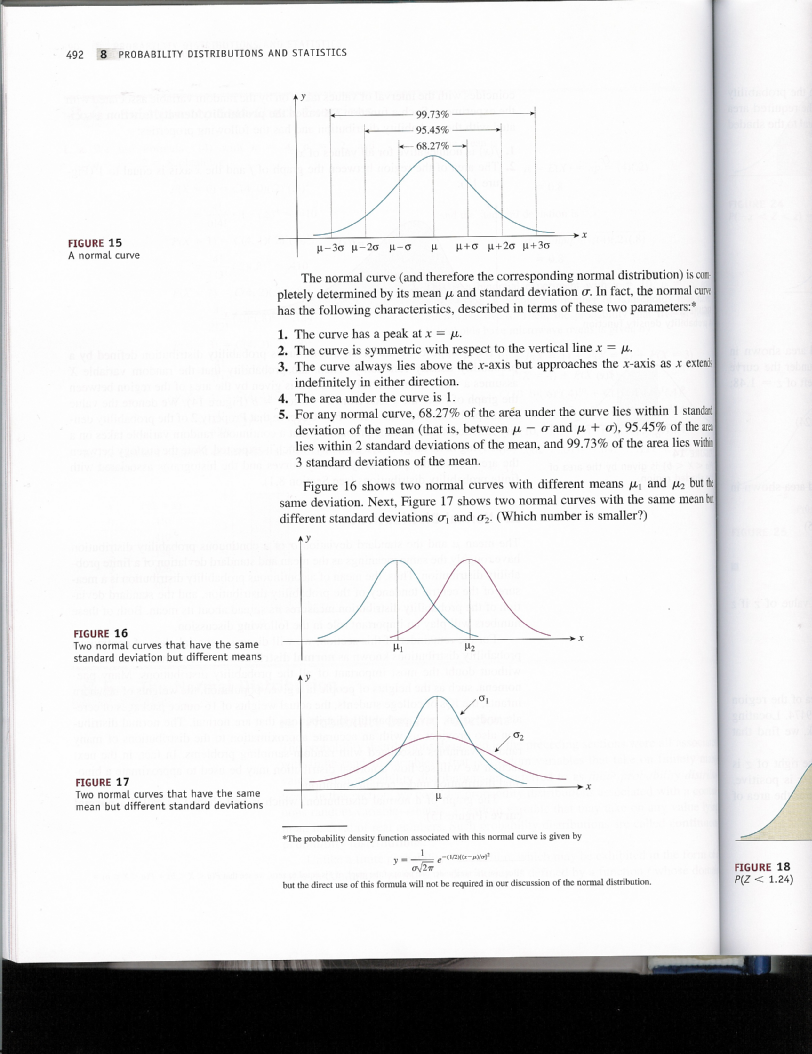
**Standard** deviation enables us to determine with accuracy where the values of a frequency distribution are in relation to the mean. According to the Russian mathematician P.L.Chebyshev

1. About 68% of the values in the population will fall within ±1 standard deviation from the mean.
2. About 95% of the values lies with in ± 2 standard deviations from the mean.
3. About 99% of the values will be in an interval ranging from 3 standard deviations below the mean.

A measure called standard score gives us the number of standard deviations, an observation lies below or above the mean. If X symbolises the observation

the standard score is given by

(**x- µ) / σ**



**Merits and demerits of Standard Deviation**

**Merits**

1. Standard deviation is by far the most important and widely used measure of dispersion.
2. It is most suitable for mathematical treatment.
3. It’s satisfies all most all properties laid down for an ideal measure of dispersion.

Sample Standard deviations = Sqrt (∑ ()2 / (n-1))

Sample variance s2 = ∑ ()2 / (n-1))

**Skewness:** literal meaning of skewness is lack of symmetry. We study skewness to have an idea about shape of the curve while we can draw with help of frequency distribution. It helps to determine the nature and extent of the concentration of the observation towards the higher or lower values of the variable.

Various Measures of skewness

Sk = mean-median

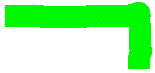
Sk = mean-mode

Sk = Q3+Q1 - 2mode.

Karl Pearson’s coefficient of skewness



**Sk =**



A quiet often mode is ill defined and is thus quite difficult to locate in such situation we use

Mo =3Md-2M

Therefore sk =

Theoretically Karl Pearson’s coefficient of skewness lies between **±3** but these limits are rarely attained in practice.



The skewness is zero if M=Mo=Md in other words for symmetrical distribution mean mode and median coincide.

sk > 0 if M > Mo and sk>0 if M > Md.

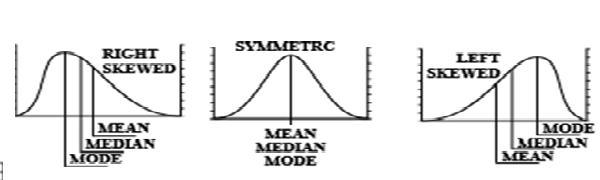
Sk > 0 if Md > Mo i.e sk > 0 if M > Md > Mo or Mo < Md < M

i.e.

the distribution is positively skewed.

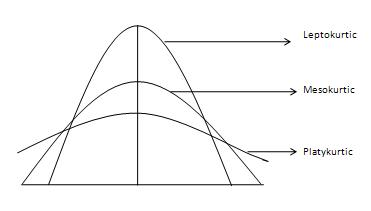
If the distribution is negatively skewed the inequality is reversed i.e.

**sk < 0 if M < Md < Mo**





**Kurtosis**



So far, we have studied a three different methods central tendency, dispersion and skewness to describe characteristics of frequency distribution.

Despite knowing all these three measures characterising a distribution cannot be done completely. For example, the above picture shows three distributions which are symmetric about mean and have the same variation, but they have different Peakedness. Karl Pearson introduced a concept called kurtosis or convexity of the curve. Kurtosis helps us to have an idea about the shape and nature of heapedness of a frequency distribution i.e. kurtosis helps us in understanding flatness or Peakness of the curve.

Kurtosis is the measure of the thickness or heaviness of the tails of a distribution. The kurtosis of a distribution is in one of three categories of classification:

* Mesokurtic
* Leptokurtic
* Platykurtic

We will consider each of these classifications in turn

## Mesokurtic

Kurtosis is typically measured with respect to the normal distribution. A distribution that has tails shaped in roughly the same way as any normal distribution, not just the standard normal distribution is said to be mesokurtic. The kurtosis of a mesokurtic distribution is neither high nor low, rather it is a baseline for the two other classifications.

Besides the normal distribution, binomial distributions for which p is close to 1/2 are considered to be mesokurtic.

## Leptokurtic

A leptokurtic distribution is one that has kurtosis greater than a mesokurtic distribution. Leptokurtic distributions are sometimes identified by peaks that are thin and tall. The tails of these distributions, to both the right and the left, are thick and heavy. Leptokurtic distributions are named by the prefix "lepto" meaning "skinny."

There are many examples of leptokurtic distributions. One of the most well-known leptokurtic distributions is [students](https://www.thoughtco.com/student-t-distribution-table-3126265) t- distribution

## Platykurtic

The third classification for kurtosis is platykurtic. Platykurtic distributions are those that have slender tails. Many times, they possess a peak lower than a mesokurtic distribution. The name of these types of distributions come from the meaning of the prefix "platy" meaning "broad."

All uniform distributions are platykurtic. In addition to this, the discrete probability distribution from a single flip of a coin is platykurtic.