**Measures of Dispersion**

* What is dispersion
* Why is it needed
* The various measures of dispersion
* Variance
* Standard Deviation

**Let’s recap what central tendency is?**

A *measure of central tendency* gives us a rough idea where data points are centred.

**Does it give an idea how the data is spread? Does Mean, Median or Mode address that?**

But, in order to make better interpretation from the data, we should also have an idea how the data are scattered or how much they are bunched around a measure of central tendency.

**Illustration**

Let’s take a look at a data set and compute the Mean and Median

**Batsman A:** 30 91 0 64 42 80 30 5 117 71

**Batsman B:** 43 46 48 50 53 53 58 60 57 52

|  |  |
| --- | --- |
| **Batsman A** | **Batsman B** |
| 30 | 53 |
| 91 | 46 |
| 0 | 48 |
| 64 | 50 |
| 42 | 53 |
| 80 | 53 |
| 30 | 58 |
| 5 | 60 |
| 117 | 57 |
| 71 | 52 |

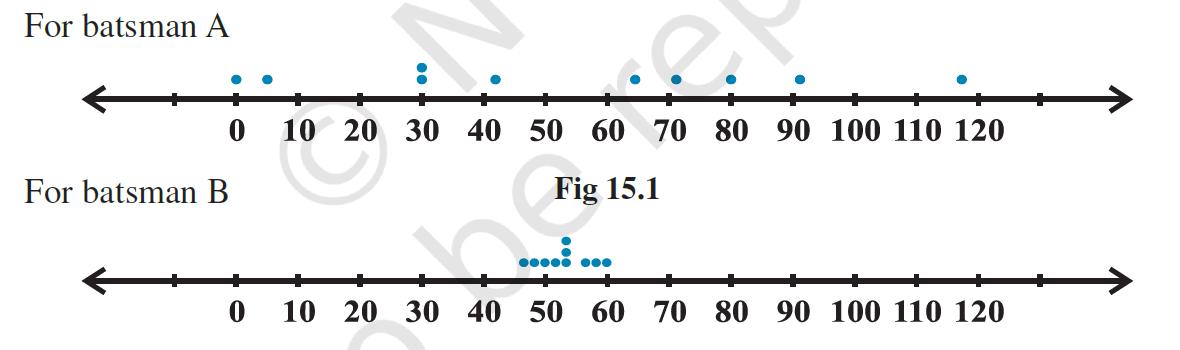
**Find out the averages**

|  |  |  |
| --- | --- | --- |
|  | **Batsman A** | **Batsman B** |
| **Mean** | **53** | **53** |
| **Median** | **53** | **53** |

**Dispersion Graphical Representation**

*How do we represent the dispersion (consistency in cricket)?*

**Let’s plot the scores on number line:**



**Theory**

We can see that the dots corresponding to batsman B are close to each other and are clustering around the measure of central tendency (mean and median), while those corresponding to batsman A are scattered or more spread out.

Thus, the measures of central tendency are not sufficient to give complete information about a given data.

Variability is another factor which is required to be studied under statistics.

Like ‘measures of central tendency’ we want to have a single number to describe variability. This single number is called a ‘measure of dispersion’.

We shall learn some of the important measures of dispersion and their methods of calculation for ungrouped and grouped data.

**The various measures of dispersion**

1. **Range**
2. **Interquartile Range**
3. **Standard Deviation**

**Range**

Range of a series = Maximum value – Minimum value

What is Range of Batsman A versus Range of Batsman B?

**IQR – Interquartile Range**

**Quartile is nothing but dividing a data set into 4 equal parts**

**IQR is the difference between the beginning of the 4th part less the ending of the of the 1st part.**

*How do we achieve this?*

**Steps:**

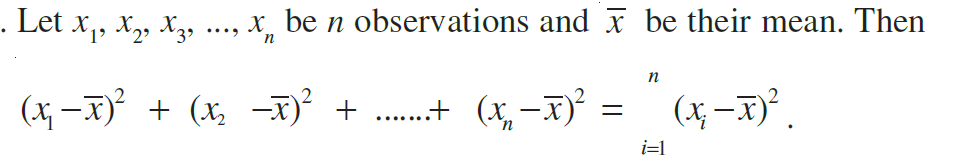
* Order the series
* First, find the median position.
* Divide the dataset in to two halves
* Next, find the median position of the first half and the second half.
* Next, The median position of the second half is called Q3
* Next, The median position of the first half is called Q1
* IQR = Q3 – Q1

Take the following data as an example

Data: 13 19 34 45 15 10 33 30 19 21 44 28 11 36 16 60 47 23 10 9

**Variance**

It’s the Mean sum of squared deviations



**Illustration**

Let us take the set A of six observations: 5, 15, 25, 35, 45, 55.

Find the mean

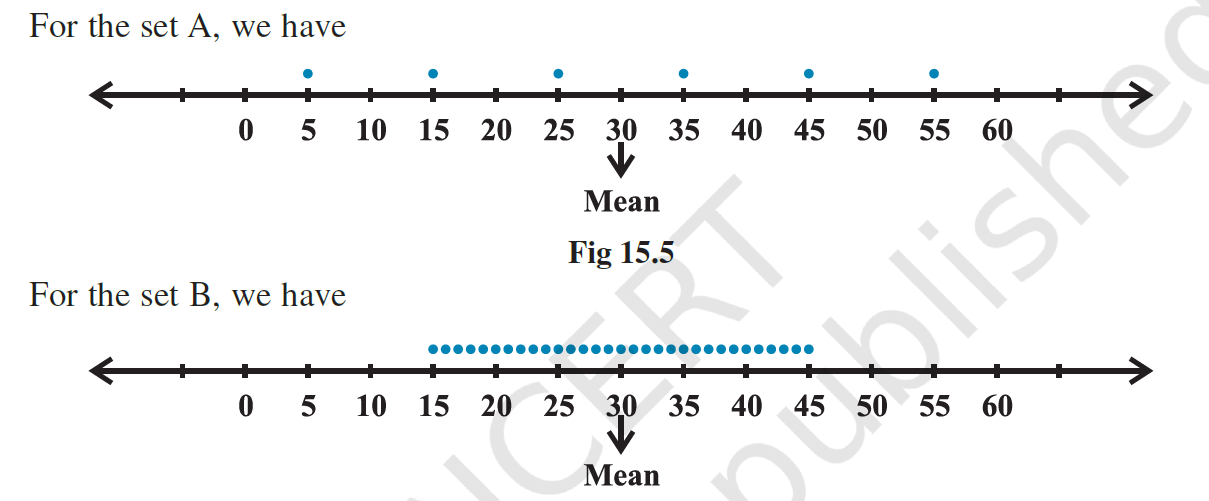
Mean is 30

SSD = 1750

Mean of SSD = 291.667 (1750/6)

Let us take set B has 7 observations: 27 28 29 30 31 32 33

Find the Variance:



**Standard Deviation**

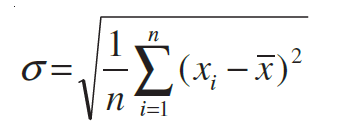
*So, what is the drawback of Variance?*

In the calculation of variance, we find that the units of individual observations *x*i and the unit of their mean *x* are different from that of variance, since variance involves the sum of squares of (*x*

i– *x* ).

For this reason, the proper measure of dispersion about the mean of a set of observations is expressed as positive square-root of the variance and is called *standard deviation*.

Therefore, the standard deviation, usually denoted by , is given by



**Illustration**

Standard Deviation of a Frequency Distribution Table

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| x | 4 | 8 | 11 | 17 | 20 | 24 | 32 |
| f | 3 | 5 | 9 | 5 | 4 | 3 | 1 |

S.D = 6.77