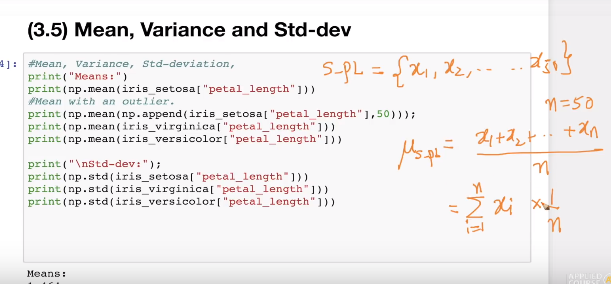
**Mean:**

Mean tells us the average or central tendency of the given data, below figure shows how is mean calculated.

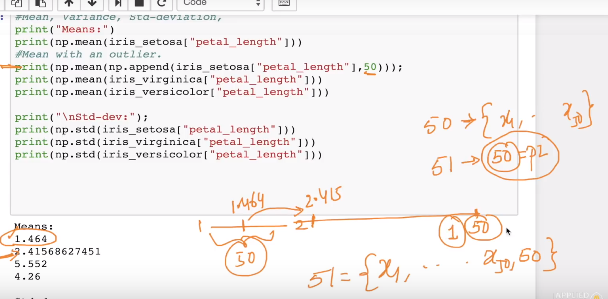


But there is a **limitation of Mean**  as, suppose in below fig ex: for setosa iris we have 50 points/data for **petal\_length** and for this we are getting average/mean as 1.464, and the value of petal\_length of setosa lies between 1 and 2 inclusive (as we’ve seen in PDF).

Now suppose someone append a new value to dataset for setossa and sets petal\_length value as 50 which is no where related to the actual petal\_length for setosa, so in such case this value is known as **outliers,**

Now just by adding a new petal\_length as 50, the mean increases from 1.464 to 2.415 which is far away from actual mean, which will mislead us in taking the actual decision.

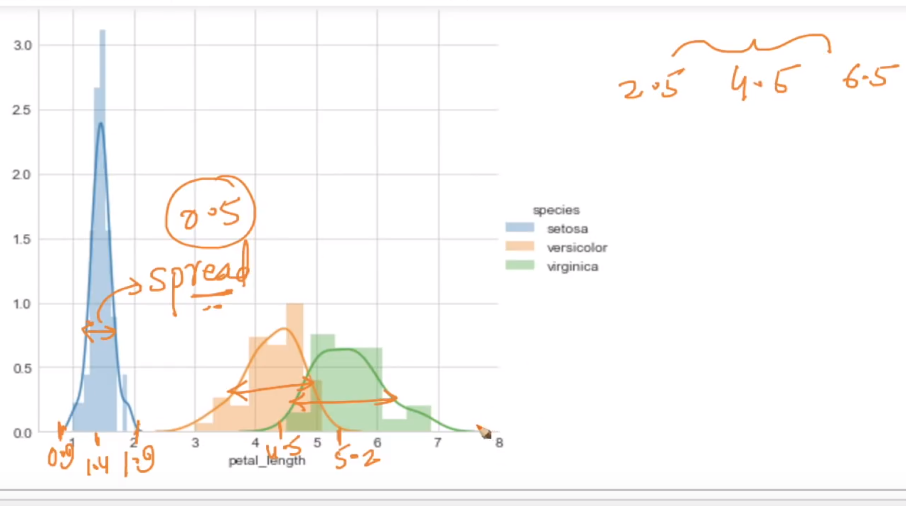
So this is the limitation of mean if dataset contains outliers then it will mislead us.

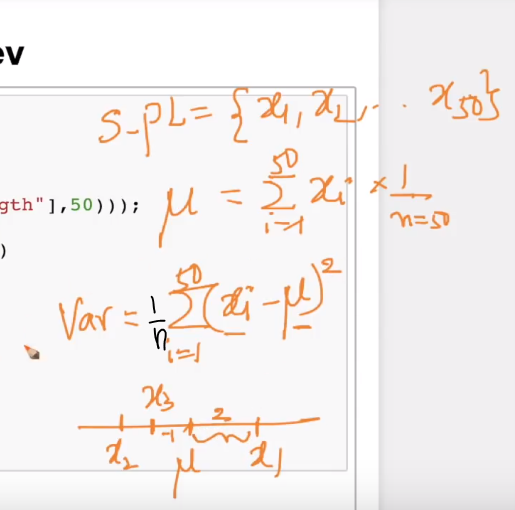


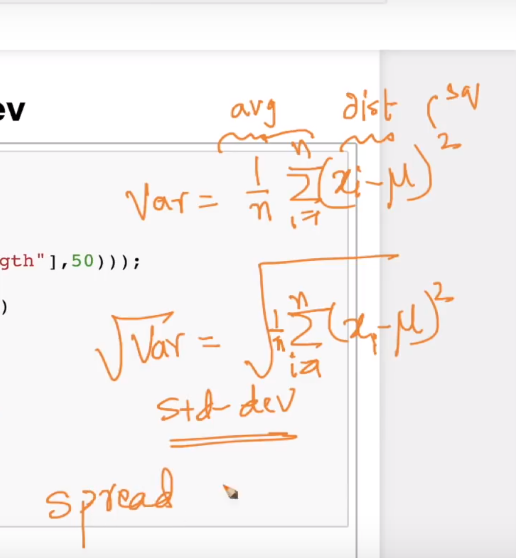
**Standard Deviation or Spread:**

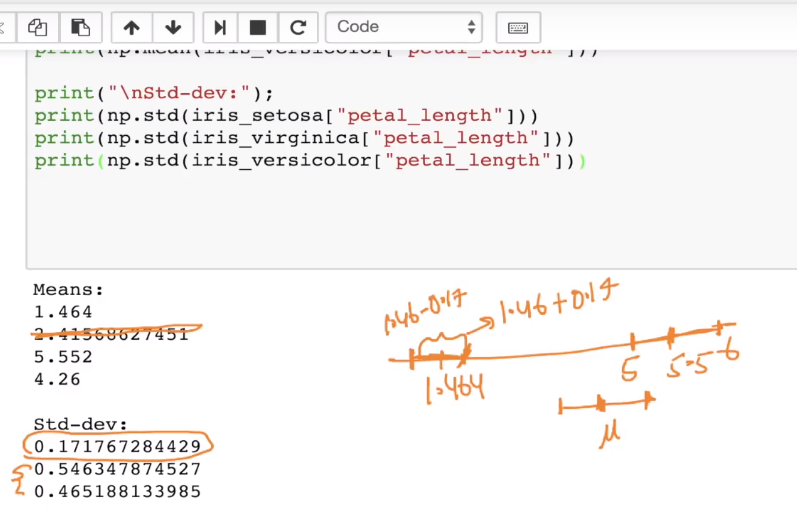
SD is used to find out the spread of data or how much the data is spreaded from mean or much till how much point the data can be from mean, basically SD tells the average of the distance of point from mean

Ex: in below fig for setosa the mean is 1.4, but the all flowers don’t have PL 1.4 some have less than it upto 0.9 and some have more upto 1.9, so what SD do it will give a number which on adding or subtracting, gives the range within which most of the of setossa flowers lies.









So, from above figure we can see mean for setosa is 1.464 and SD is 0.1717… so by seeing this data we can say that most of setosa lies between 1.464 – 0.1717 and 1.464 + 0.1717.

**SD too have same limitation of outliers as of mean, outliers can corrupt the value of SD, because eventually while calculating variance we are finding mean of distance of each point from mean.**