**Session 4**

**Q1: Why measures of dispersion are important?**

Measures of dispersion tells us how far data points are from the mean value or average value. It is used to check the outliers, perform statistical tests and determine whether average is a good representative value or not.

**Q2: You have two samples from same population.**

**Sample 1: 1, 2, 3, 4, 5**

**Sample 2: 1, 2, 3, 4, 91**

**Please perform measures of central tendency. Which measures of central tendency is more important for sample 1 and which is more important for sample 2, and why?**



For sample 1, mean and median both are good. But for sample 2, median is preferred because mean is getting affected by the outlier value.

**Q3: You selected 2 samples from the same population. Standard deviation (sd) for sample 1 is 2.5 and sd for sample 2 is 2.1. Which sample is better representing your population and why?**

Sample 2 is better representative of population because the less the standard deviation, the less is deviation from the mean. Less variance give better results for statistical tests.

**Q4: Please calculate all measures of dispersion for sample 1 and sample 2 given in question 2. Which measures of dispersion is more appropriate and for which sample?**

Variance for sample 2 would be far greater than that for sample 1.



We generally use standard deviation because it has same units as that of observations. As we can see, sample 1 is clearly better representative of the sample.

**Q5: Please read given data (Actor.xlsx) and calculate measures of central tendency and measure of dispersion for all appropriate columns.**

**In jupyter notebook**