***Session\_1***

***Example***: A college dean is interested in learning about the average age of faculty. Identify the basic terms in this situation.

The ***population*** is the age of all faculty members at the college.

A ***sample*** is any subset of that population. For example, we might select 10 faculty members and determine their age.

The ***variable*** is the “age” of each faculty member.

**One *data*** would be the age of a specific faculty member.

**The *data*** would be the set of values in the sample.

The ***experiment*** would be the method used to select the ages forming the sample and determining the actual age of each faculty member in the sample.

The ***parameter*** of interest is the “average” age of all faculty at the college.

The ***statistic*** is the “average” age for all faculty in the sample.

***Central Tendency :***

Think about how you describe a single piece of numerical data. This is usually done in terms of its value. For example, in order to describe the number 2, you might put up two fingers or you might say 2 = 1 + 1. How would you describe a group of data? It would not be beneficial to use your fingers in this instance. Nor is it beneficial to simply add the data together. **However, you can describe a group of data in a single value by using measures of central tendency.**

**Variance and Standard Deviation**

There are many examples of variance and standard deviation. Variance, standard deviation, range, inter-quartile range are all measures of spread of data. That is the tell you how far away data is from the middle/center of the data.

Example 1: Stock market or other investment returns. The stock market has return on average 7% per year. This does not mean that every year you get a 7% return, some years are more and some years are less. This variability (called volatility in stock terms) is an example of variance and standard deviation.

Example 2: A manufacturing machine fills up water bottles with 8oz. Now the machine does not fill up each water bottle with exactly 8oz, there is a little variation due environmental factors (weather, placement of bottle, etc.). This variation is example of variance or standard deviation.

Example 3: You drive to work every and take the same route. There is both variation in the time is takes you to get to work (traffic, stop light timing etc.) and variation in the amount of gas you use (same factors of traffic, stop light time affect this). All of this variability can be measured with variance and standard deviation.

Basically, anywhere you see movement away from the center of something it can be measured with variance or standard deviation (for numerical things anyway).

***Standard Deviation and Coefficient of Variation:***

**Standard deviation basically means , how much deviation is there in data-set from it’s mean,**

**So , you have idea about standard deviation and Coefficient of variation**

* *But you want to know difference let’s take a example and you will know why coefficient of variation is necessary.*

let say you want to choose a player in your team , you have 2 contender , after comparing both , you conclude that both are exceptionally good, but you can choose only 1, so you think of applying statistics , i.e standard deviation.

Let say you take into account both the players goals stats , how much they have scored in last 30 matches ,

**Player A**

**Mean = 60**

**Standard deviation = 18**

**player B**

**Mean = 48**

**standard deviation = 12**

Now looking at mean , you think player A has scored more goal than B and difference of Standard deviation of A from B is also not much higher so you decide to Take A .

Based on the mean and Standard deviation you end up taking Player A , but wait a minute is the comparison viable just by looking at mean and Standard deviation.

How will you know the comparison was justified?

Then comes the “**coefficient of variation”**in picture, it simplify the data set by giving a value which is easily comparable i.e Coefficient of variation is used to compare two data set , basically spread of data, How much Data is spread from it’s mean , by taking ratio of S.D and mean we find out Coefficient of variation.

Coefficient of variation = (S.D / mean) \*100

For A it is 30 % and For B it is 25%

so For A dispersion is higher than B , Basically B is better than A .

Selecting B is better.

I hope you understood the difference , Both Standard deviation and Coefficient of variation are measure of spread of data from it’s mean , for practical purpose they have different uses , as you have seen in above example , how decision could have gone wrong just by looking at mean and Standard deviation :)

Statistics cannot be learned by cramming formulas, one should study by applying formula in real life and understanding the rationale behind formulas, why it is used , how it is derived.

**Standard Deviation :**

If for an exam students got an average of 12 but a large standard deviation that means that the marks are spread around 12: it is likely to observe very good grades and also bad ones.  On the other hand, if you have an average of 12 but a small standard deviation, that means that the marks are more narrowly spread around the mean: most people will get grades close to 12 and few of them will get grades far away from the mean.

<https://study.com/academy/lesson/covariance-correlation-equations-examples.html>

**chi square and Analysis of Variance (ANOVA)**

The chi square and Analysis of Variance (ANOVA) are both inferential statistical tests. Inferential statistics are used to determine if observed data we obtain from a sample (i.e., data we collect) are different from what one would expect by chance alone. A more simple answer is that we want to determine if the relationships among variables or differences between groups that we see in our sample data are occurring in the entire population.

That said, chi square is used when we have two categorical variables (e.g., gender and alive/dead) and want to determine if one variable is related to another. In ANOVA, we have two or more group means (averages) that we want to compare. In an ANOVA, one variable must be categorical and the other must be continuous. For example, we may want to examine if marijuana use (0 to 25 times) differs by grade level (9th grade, 10th grade, 11th grade).