Basic Statistics

1) If the variance of a variable/column is 0 then what does it mean? Can we use that variable for our analysis?

Answer: The variance helps us to understand about the data in variable that how much each record is far away from the mean. If variance value is more it represents that record is far away from mean. If variance value is low, then the record is close to the mean of the data.

So coming to the answer if variance of variable is zero means that every data present in the variable is constant or similar data i.e. I mean every record has same value. Even if you calculate the mean the value of mean generated is similar to data present in variable.

We can't use this variable as every record is constant so we can't get any insight from analysing that variable so better to ignore that variable.

2) Calculate mean, median, mode, variance and standard deviation for column A

Ans:

I done calculations in Excel,

Mean step procedure:

1)calculate total value of A columns by adding

2)Divide total value **72** with number of data **11,** we will get the answer

Median step procedure:

1)First arrange the data set in increasing order

2)Calculate the 50th percentile position by dividing noof data with 2 I got answer as **5.5**

3)I considered 6th position of data set which is **7.**

Mode step procedure:

1)Calculate count of every number that is repeated.

2)As we saw in figure 1 the 7 number has highest frequencies so mode is **7**

Fig1

A	A	
7	5	count of 5
6	5	3
7	5	count of 6
7	6	1
8	7	count of 7
5	7	5
8	7	count of 8
7	7	2
7	7	
5	8	
5	8	
72	5.5	
mean	median	mode
6.545455	7	7

The mean,median and mode are the measures of central tendency helped to summerise data and gives ability to understand the data. Now I will tell when to use mean,median and mode

The mean are sensitive of outliers if any outliers are present in data the mean will changes according to it, so if the data was not in skew format then it is appropriate to use mean.

The median doesn't has any interfearence with outliers as it caluclated by using 50th persentile rule. Even if data has skewness we can use median.

Steps for variance

1)caluclate mean of data set and subtract it with each record as done in fig2

2)square all of the xi-u colums and then add whole of it

3)After adding divide it with number of data then you will get variance.

Formula=sigma[(xi-u)^2]/n

A	x1-u	(x1-u)^2
7	0.454545	0.206612
6	-0.54545	0.297521
7	0.454545	0.206612
7	0.454545	0.206612
8	1.454545	2.115702
5	-1.54545	2.38843
8	1.454545	2.115702
7	0.454545	0.206612
7	0.454545	0.206612
5	-1.54545	2.38843
5	-1.54545	2.38843
6.545455	sum	12.72727
	variance	1.157025

Fig 2

Steps for standard deviation

1)do until 2 steps of above done for caluclating variance

2)divide it with n-1(10) i.e. 1.27

3)now square root of 1.27 we get answer **1.128**

Formula=sqrt([sigma(xi-u)^2]/(n-1))

Variance represents how far the data was spread far away from mean and standard deviation represents how much data deviates from the actual mean.

3) In a group of 12 scores, the largest score is increased by 36 points. What effect will this have on the mean of the scores?

Ans: The mean value will increase as it is influenced by value of data so it shifts to right

4) Explain the difference between Data (Singular) and Data (Plural) with examples?

Ans: The single value represented in data of any type like name, number, date is called singular data. For example, in fig2 the single value like 8,0.234 etc. are called as singular data.

The set of data that represents in variable is called as plural data. For example, in fig 2 A, xi-u called ad data plural.

5)How the inferential statistics helps to make decisions out of it?

Ans: Statistical inference is the act of generalizing from a sample to a population with calculated degree of certainty. We will make decisions by analysing sample data of a whole population. For example I took survey on “Consumer perception on Edwisor company” I can take sample of students from Bangalore and can apply statistics to get some insights on it. By this I will make descision on whole data.