**Assignment Name - Concepts of Statistics**

**Answers**

(Please note that I have not copied the problem statements. The answers to the all the five problems statements are only furnished here)

**Answer to problem statement 1**

If the variance of a variable/column is 0, then it means that amount of variability present in that dataset is also zero which actually means that the numbers in the dataset used to find the mean and its variance are all same numbers.

When a variance value is zero, as such, since it tells that there is no variability between the numbers in the set of the data, it does not give us any useful information to understand any pattern or trend about the data under study. Therefore, a dataset or variable where the variance is zero is not useful for our analysis.

**Answer to problem statement 2**

(The answer is based on the column A of the problem statement 2)

**Measures of Central Tendency**

**Mean**

Mean = the sum of all the values ÷ Number of values

= 72÷11

=6.55

**Median**

Median is the 50th percentile of a data arranged either in ascending or in descending order.

The data when arranged in ascending order looks like this:

5,5,5,6,7,7,7,7,7,8,8

Therefore the number at the 50th percentile or in the middle is 7. So, the median is 7.

**Mode**

Mode is the score that occurs most number of times in a set of data.

So, the number that appears in our dataset is 7, and, therefore the mode is 7

**Measures of dispersion**

**Variance**

Variance =

Mean (µ) is 6.55 as calculated above under ‘Mean’.

So the working is:

|  |  |  |
| --- | --- | --- |
| **A** | **x-µ** | **(x-µ)²** |
| 7 | 0.45 | 0.2025 |
| 6 | -0.55 | 0.3025 |
| 7 | 0.45 | 0.2025 |
| 7 | 0.45 | 0.2025 |
| 8 | 1.45 | 2.1025 |
| 5 | -1.55 | 2.4025 |
| 8 | 1.45 | 2.1025 |
| 7 | 0.45 | 0.2025 |
| 7 | 0.45 | 0.2025 |
| 5 | -1.55 | 2.4025 |
| 5 | -1.55 | 2.4025 |
| **72** | **-0.05** | **12.7275** |

Variance =

= 1.157045

If the given data is assumed to be a sample, then the denominator would be N-1 instead of N as above in the formula for population. In that case the variance will be:

Variance (if sample data) = = 1.27275

**Standard Deviation**

(Assuming that the data is a sample)

S =

The values for all the variables of this formula are already calculated above in the working for ‘Variance’:

S =

=

=

= 1.128162

**Answer to problem statement 3**

In a group of 12 scores, if the largest score is increased by 36 points, it is obvious that the total number of values (i.e., N) does not change, which remains at 12.

So, if the total sum of the scores before increasing the largest score by 36 is, for example:

Mean = = 5

If the largest score is increased by 36, then the sum also will be increased by 36 as shown here:

Mean = = = 8

Therefore the change in the mean is an increase in the mean by 3 which now is 8.

**Answer to problem statement 4**

Data singular and Data plural are two of many statistical jargons.

**Data Singular**

Data Singular is the value of the variable associated with one element of a population or sample. This value could be a number, a word, or a symbol. For example, if we have set of data of a public school showing the columns (variables) as follows:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **ABC Public school data of higher secondary exam results** | | | | |
| **Academic year** | **Total number of students** | **Students scored above 90%** | **Students scored between 81-90%** | **Students scored less than 80%** |
| 2017-18 | 1000 | 205 | 380 | 415 |
| 2016-17 | 998 | 190 | 375 | 433 |
| 2015-16 | 1005 | 199 | 383 | 423 |
| 2014-15 | 990 | 185 | 370 | 435 |
| 2013-14 | 985 | 180 | 371 | 434 |
| 2012-13 | 956 | 182 | 356 | 418 |

The data singular in this case can be a value falling under the variable (column) against one academic year. For example, the value 1000 under the ‘Total number of Students’ which appear against the academic year 2017-18 can be called a data singular.

**Data Plural**

A data plural is the set of values collected for the variable from each of the elements belonging to the sample. Here, in the above example, the whole set of values from 1000 to 956 shown under the variable ‘Total number of Students’ which appear against the academic year 2017-18 is an example of data plural.

**Answer to problem statement 5**

If a study of large amount of data to be carried out, for example, a survey, it will be extremely difficult, time-consuming and very expensive to cover the entire population. Therefore, a sample which represents the entire population is selected and this selected sample is studied using descriptive statistical methods and then uses that result to draw conclusions about the entire population. This method of making inferences or drawing conclusions from the sample data is called inferential statistics. Thus inferential statistics helps us to study a sample and apply those results to the entire population and helps us to make decisions regarding the entire population.