Assignment  
Data Science Masters  
Q1. What is Statistics?

Ans- Statistics is branch of mathematics that deals with collection of data so it can help one understand the trend, downfall or upset of any event. It is use to make decision and predications on the base of data .

Basically, there are two types of Statistics-:

1. Descriptive statistics- It deals with data to summarise and analyse by using measures like mean, median and Standard deviation.
2. Inferential statistics- It is use to get population predications on the basis of sample.

Statistics is use in the field of science, engineering, medicine, business, social sciences,

and many others.

Q2. Define the different types of statistics and give an example of when each type might be used.

ANS-There are two are main types of statistics

1. Descriptive statistics- It deals with data to summarise and analyse by using measures like mean, median and Standard deviation.

Example- analyst collect data of footballer pulse rate to find average pulse rate of team and also classify best rate in the team.

2.Inferential statistics- It is use to get population predications on the basis of sample.

Example- suppose in a football team, doctor give certain vitamin to treatment group and not to a control group just to check that vitamin help to reduce of team or not then only the available knowledge can be use for team.

Q3. What are the different types of data and how do they differ from each other? Provide an example of  
each type of data.

1. Nominal data- category data but not use for ranking

Example- gender (male or female)

B ordinal data- category with ranking concept of data.

Example-TYPES OF SHOP (fruit, vegetables, diary) with more turnover (1st,2nd,3rd)

C. interval data- data whose interval of difference is always known with having zero point

Example -temperature or time in a clock.

D. ratio data – data with interval as well as real zero point.

Example- height

Basically difference is ordinal and nominal is base non parameters where interval and ratio is base on parameters.

4. Categorise the following datasets with respect to quantitative and qualitative data types:   
(i) Grading in exam: A+, A, B+, B, C+, C, D, E   
(ii) Colour of mangoes: yellow, green, orange, red   
(iii) Height data of a class: [178.9, 179, 179.5, 176, 177.2, 178.3, 175.8,...]   
(iv) Number of mangoes exported by a farm: [500, 600, 478, 672, ...

1. Qualitative
2. Qualitative
3. Quantitative
4. Quantitative

5.Explain the concept of levels of measurement and give an example of a variable for each level.

Concept of measurement are

* Ordinal level of measurement- use to measure data on the basis of data by categorising it and not ranking.

Example- gender(male or female.

* Nominal level of measurement- use to measure data on the basis of data by categorising it and ranking too.

Example- vegetable fruit seller in a market with their turnover ananlysis.

* Interval level of measurement:- use to measure data with pre-defined interval but with no real zero point.

Example- dates of any event.

* Ratio level of measurement- use to measure data with pre-defined interval but with real zero point.

Example- height .

6. Why is it important to understand the level of measurement when analysing data? Provide an  
example to illustrate your answer

Ans- it is important to understand different types of data measurement due to

* Data has it own types(category or non category)
* Diversity of data
* Availability of data is not from one source

Therefore there is need to measure different data as per the required situation so different types of measurement is needed

Example of situation- suppose one need data of percentage of student passing in std 10 in a state bihar and Jharkhand. Due to variety of data and different types of analysis like-:

1. Which state perform well in comparison
2. Drawback or any predictions
3. Nominal data like gender given exam.
4. Marks analysis by ratio measurement
5. Use of ordinal measurement to seek gender with best performance.
6. Interval measurement of having higher percentile in exam.

q.7.How nominal data type is different from ordinal data type.

|  |  |
| --- | --- |
| nominal data | ordinal data |
| 1. Categorised but not applicable for ranking | Categorised andt applicable for ranking |
| 2.non parametric test is use | 2. parametric and non parametric both test is use |
| Example- type of vendors in market | Example- types of vendor and who earn more. |

q.8 Which type of plot can be used to display data in terms of range?

Ans “BOX Plot” is use to display data in the rate’

It use IQR inter quartile range concept of having

Q1 ,Q2,Q3

BOX data is use to trace;-

1. Sets of data spread
2. Outliers in data set
3. Use by data analyts

Q,9 Describe the difference between descriptive and inferential statistics. Give an example of each  
type of statistics and explain how they are used.

. Descriptive statistics- It deals with data to summarise and analyse by using measures like mean, median and Standard deviation.

Example- analyst collect data of footballer pulse rate to find average pulse rate of team and also classify best rate in the team.

2.Inferential statistics- It is use to get population predications on the basis of sample.

Example- suppose in a football team, doctor give certain vitamin to treatment group and not to a control group just to check that vitamin help to reduce of team or not then only the available knowledge can be use for team.

Q.10 What are some common measures of central tendency and variability used in statistics? Explain  
how each measure can be used to describe a dataset

Ans- some measures of central tendency and variability are -:mean, median, and mode and range, standard deviation, and variance respectively.

Mean is just average of total dataset available just to give central value of the data

Median is just middle value in the dataset when put in the order.

Mode is the value that occur more in the data set.

Range is just difference between larger and smaller vale in the dataset

The standard deviation is a measurement of how much a dataset's values depart from its mean.

The average of the squared deviations of each result from the mean makes up the variance.

By giving a measurement of how widely distributed the data are around the mean, it is used to represent the variability of the data.

Outliers and extreme values have an impact on the variance and can make the data more variable.