Statistics for Data Science-1

Week 1 Graded Assignment Solution

The education minister wants to know the status of campus placements of B.Tech students in different engineering institutes of India. An analyst did a survey on the randomly selected four IITs of India and analysed the status of campus placements. Based on the information given, answer the questions (1), (2) and (3).

1. Identify the sample and population.

   (a) The sample consists of all the engineering institutes of India and the population consists of randomly selected four IITs of India.

   (b) The sample consists of all the IITs of India and the population consists of all the engineering institutes of India.

   (c) The sample consists of all IITs of India and the population consists of randomly selected four IITs of India.

   (d) The sample consists of four randomly selected IITs of India and the population consists of all the engineering institutes of India.

   Answer: d

   Solution:

   By definition, population is the entire collection of elements we are interested in. Here, the purpose of the survey is to know the status of campus placements of B.Tech students in different engineering institutes of India. Hence, the population will be all the engineering institutes of India.

   Also, sample is a subset of the population which is being studied. Since, the four IITs of India is selected to know the status of campus placement. Therefore, sample is four randomly selected IITs of India.

   Thus, the sample consists of four randomly selected IITs of India and the population consists of all the engineering institutes of India.

   Hence, option (d) is correct.

2. The report given by an analyst to the education minister about the status of campus placements states that "The campus placement of B.Tech students is 95% in the different engineering institutes of India". The given statement of analyst is based on which kind of statistical analysis?

   (a) Descriptive Statistics

   (b) Inferential Statistics

   Answer: b

   Solution:

Making conclusions from the sample data comes under inferential statistics. Here, analyst makes the conclusion in the report that "The campus placement of B.Tech students is 95% in the different engineering institutes of India" based on the information of four randomly selected IITs of India. Therefore, the study is inferential statistics.
Hence, option (b) is correct.

3. Is the conclusion of this study made by analyst on the basis of chosen sample reliable?

   (a) Yes

   (b) No

   Answer: b
   The objective of the survey is to know the status of campus placements of B.Tech students in different engineering institutes of India, but the institutes are selected only from IITs and not from different engineering institutes of India. Therefore, this sample is not a good representative of the population, as the status of the campus placement of B.Tech students could vary in various engineering institutes of India.
   Hence, option (b) is correct.

The data of five different types of fertilizers used by farmers of a village is tabulated in Table 1.1.G. Based on the information given, answer the questions (4), (5), (6), (7) and (8).

Fertilizers	Types of Fertilizers	Area of fields (In acres)	Types of Crops	Amount of fertilizers (In Kg)
Nitrogen	Inorganic	1	Rice	200
Phosphorus	Inorganic	2	Wheat	400
Manure	Organic	1.5	Potato	300
Compost	Organic	1.3	Rice	260
Potassium	Inorganic	1.6	Pulse	320

Table 1.1.G

4. Which of the following statements is/are true?

   (a) Inorganic is a case and Types of Fertilizers is a variable.

   (b) Rice is a case.

   (c) Manure is a case.

   (d) Amount of fertilizers is a variable.

(e) Nitrogen is a variable.

Answer: c, d

Solution:

Here, the specification data of the five different types of fertilizers used by farmers of a village are collected. So each specification (columns of the table) i.e. Fertilizers, Types of Fertilizers, Area of fields(In acres), Types of Crops and Amount of fertilizers(In Kg) is a variable.

Observation is an individual data point for which the entire data is being collected. So, here each value corresponding to which each of the specification noted is a case.

Thus, it is clear that Manure is a case.

Hence, options (c) and (d) are correct.

5. What is the scale of measurement of "Types of Crops"?

   (a) Ordinal Scale

   (b) Nominal Scale

   (c) Interval Scale

   (d) Ratio Scale

   Answer: b

   Solution:

   Types of Crops is a categorical variable and it has four different categories of crops, i.e., Rice, Wheat, Potato and Pulse which are just labels.

   Since, there is no particular order among the types of crops. Therefore, it has nominal scale of measurement.

   Hence, option (b) is correct.

6. What kind of variable is "Area of fields"?(More than one option can be correct)

   (a) Categorical

   (b) Numerical

   (c) Discrete

   (d) Continuous

   Answer: b, d

   Solution:

   Since Area of fields has numeric properties and can have arithmetic operations performed on it, it follows that Area of fields is a numerical variable. Moreover, it can take any value greater than 0 and have to measure in acres. Therefore, Area of fields is a continuous numerical variable.

   Hence, options (b) and (d) are correct.

7. What is the scale of measurement of "Amount of Fertilizers"?

   (a) Ordinal Scale

   (b) Nominal Scale

   (c) Interval Scale

   (d) Ratio Scale

   Answer: d
   Solution:
   Amount of Fertilizers can have a meaningful interval. It also has an absolute zero. Hence, it comes under the ratio scale of measurement.
   Hence, option (d) is correct.

8. Is the data given in Table 1.1.G structured or unstructured?

   (a) The data is structured

   (b) The data is unstructured

   Answer: a
   Solution:
   Since the data of the five types of fertilizers used by farmers of a village can be organized in a well-defined tabular form. Therefore, it comes under the structured data.
   Hence, option (a) is correct.

9. The data of Netflix subscribers at the end of year 2020 across different Asian countries is recorded. Based on this, choose the correct option:

   (a) It is time series data

   (b) It is cross-sectional data

   Answer: b
   Solution:
   Since the data of Netflix subscribers are recorded across different Asian countries at a same time (i.e., at the end of year 2020), not at different time-intervals. Therefore, the data collected is cross-sectional data.
   Hence, option (b) is correct.

10. Choose the incorrect statement(s):

    (a) Stock price of Titan is numeric and continuous variable.

    (b) Number of assignments submitted by a student has an interval scale of measurement.

    (c) Soccer positions (i.e. Defender, Midfielder, Forward) has an ordinal scale of measurement.

(d) The education level of a person has a nominal scale of measurement.

Answer: b, c, d

Solution:

Since stock price of Titan has numeric properties and can have arithmetic operations performed on it, it follows that Stock Price of Titan is a numerical variable. Moreover, it can take any value (any non-negative value). Therefore, Stock price of Titan is a numeric and continuous variable. Therefore, option(a) is correct.

Number of assignments submitted by a student can have a meaningful interval. It also has an absolute zero. Hence, it comes under the ratio scale of measurement. Therefore, option(b) is incorrect.

Soccer positions (i.e. Defender, Midfielder, Forward) are just labels. There is no particular order among positions. Thus, it has nominal scale of measurement. Therefore, option(c) is incorrect.

There is a particular order in the education level of a person like 12^{th}, Undergraduate, graduate etc. Thus, it has an ordinal scale of measurement. Therefore, option(d) is incorrect.

Hence, options(b), (c) and (d) are correct.