

Statistics for Data Science-1

Week 1 Graded Assignment Solutions

250 freshmen are randomly selected from the Electrical Engineering Department at IIT Madras for a survey that intends to determine the average score obtained by freshmen of IIT Madras in high-school. Based on this given information, answer questions (1), (2), (3) and (4):

1. What is the population?
 - a. The 250 selected students.
 - b. All students at IIT Madras
 - c. All freshmen at IIT Madras
 - d. All freshmen at IIT Madras, Electrical engineering department.

Answer: c

Solution:

By definition, population is the entire collection of elements we are interested in. Here, the survey is to determine the average score obtained by freshmen of IIT Madras in high-school. Hence, population will be all freshmen at IIT Madras.

Hence, option c is correct.

2. What is the sample?
 - a. The 250 selected students.
 - b. All students at IIT Madras
 - c. All freshmen at IIT Madras
 - d. All freshmen at IIT Madras, Electrical engineering department.

Answer: a

Solution:

Sample is a subset of the population which is being studied. Since the 250 freshmen of IIT Madras, Electrical engineering department is studied, the sample is the set of 250 selected students.

Hence, option a is correct.

3. What can you say about the nature of study?
 - a. The study is descriptive.
 - b. The study is inferential.

Answer: b

Solution:

Making conclusions from the data comes under inferential statistics. Here, we make the prediction of average score obtained by freshmen of IIT Madras in high-school based on the information of 250 randomly selected freshmen at IIT Madras, Electrical engineering department. Therefore, given study belongs to inferential statistics.

4. Is the conclusion of this study made on the basis of chosen sample reliable?
- a. Yes
 - b. No

Answer: b

Solution:

The survey is to predict the average score obtained by freshmen of IIT Madras in high school, but the sample is collected only from the freshmen of IIT Madras, Electrical engineering department, and not from all the freshmen at IIT Madras. Therefore, this sample is not a good representative of the population, as marks obtained in high school by the freshmen of IIT Madras could vary from department to department.

The data of second-hand cars is collected and shown in Table 1.1. The car seller used this data to predict the price of his car in the market.

Brand	Model Name	Fuel Type	Number of Cylinders	Mileage	Price (in USD)
Audi	A6 45 TFSI	Petrol	4	17.2	17000
Audi	A7 3.0 TFSI	Petrol	3	13.8	23210
Honda	Amaze VX CVT	Gas	4	19	16550
Hyundai	i10 Sportz	Gas	2	27.1	18000
Hyundai	i20 Magna	Diesel	4	15	18550
Honda	Jazz 1.5 i-MMD	Diesel	4	22	12000

Table 1.1

5. Which of the following statements is/are true:
- a. Jazz 1.5 i-MMD is a case and Brand is a variable
 - b. Mileage is a case
 - c. Gas is a variable
 - d. Honda is a case
 - e. Number of doors is a variable
 - f. Price is a variable

Answer: a,e,f

Solution:

Here, the specification data of second-hand cars is collected. So each specification (columns of the table) i.e. Brand, Model, Fuel Type, Number of Cylinders, Mileage and Price (in USD) is a variable. Observation is 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.

Hence, option a,e and f are correct.

6. What is the scale of measurement of Fuel type?

- a. Ordinal scale
- b. Nominal scale
- c. Ratio scale
- d. Interval scale

Answer: b

Solution:

Fuel type is a categorical variable. Also, there is no particular order among the types of fuel. So, it has a nominal scale of measurement.

Hence, option b is correct.

7. What is the scale of measurement of Number of Cylinders?

- a. Ordinal scale
- b. Nominal scale
- c. Ratio scale
- d. Interval scale

Answer: c

Solution:

Number of Cylinders can have a meaningful interval. It also has an absolute zero. Hence, it comes under the ratio scale of measurement.

Hence, option c is correct.

8. What kind of variable is Mileage?

- a. Numerical
- b. Continuous
- c. Discrete
- d. Categorical

Answer: a,b

Solution:

Since Mileage has numeric properties and can have arithmetic operations performed on it, it follows that Mileage is a numerical variable. Moreover, it can take any value greater than 0. Therefore, Mileage is continuous numerical variable.

Hence, option a and b are correct.

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

- a. The data is structured
- b. The data is unstructured

Answer: a

Solution:

Since the data of second hand car can be organised in a well defined tabular form. Therefore, it comes under the structured data.

Hence, option a is correct.

10. The price value of stocks are measured at a regular time interval of 5 minutes for 7 days. Based on this choose the correct option:

- a. It is a time-series data
- b. It is a cross-sectional data
- c. None of the above

Answer: a

Solution:

Since the price value of stocks are recorded over a period of time at regular intervals, the data collected is time series data.

Hence, option a is correct.

11. The CBSE has declared result for class 12th examinations last week. What is the scale of measurement of marks scored by a student?

- a. Ordinal
- b. Nominal
- c. Ratio
- d. Interval

Answer: c

Solution:

Marks scored can have a meaningful interval. It also has an absolute zero. Hence, it comes under the ratio scale of measurement.

Hence, option c is correct.

12. In an exam, there are 10 Multiple choice questions. If the answer marked is correct, he/she will be awarded 1 point and 0.25 marks will be deducted from the total score for every wrong answer. What is the scale of measurement of final score obtained by a student?
- a. Ordinal
 - b. Nominal
 - c. Ratio
 - d. Interval

Answer: d

Solution:

Final score obtained can have a meaningful interval. But, it do not have an absolute zero as final score obtained can be negative as well. Hence, it comes under the interval scale of measurement.

Hence, option d is correct.