Stats-1

Activity 1.1

A study has been conducted to examine whether the students who had applied for MAEEE entrance exam in 2020 are willing to write the exam in the coming week or not. A group of 10,000 students were randomly selected from the urban areas of India for this study. It is concluded that roughly 80 percent of the students want to write the exam.

Based on the above information, answer the following three questions.

1) Identify the sample and population from the statement given above.   **1 point**

○ The sample consists of students from all across India, and population consists of students from the urban areas of India.

○ The sample consists of students from the urban areas of India and the population consists of students from all across India.

○ The sample consists of students from the urban areas of India and the population consists of selected 10,000 students from the urban areas of India.

● The sample consists of the selected 10,000 students from the urban areas of India and the population consists of students from all across India.

2) Identify the correct statements from the following:

○ Aforementioned conclusion made on the basis of the sample is correct.

● Aforementioned conclusion is not correct because the sample is not a good representation of the population.

○ The sample is a good representation of the population.

○ None of the above.

3) What can you say about the nature of the study ? (more than one option can be correct)

☐ The study is descriptive.

☑ The study is inferential.

☐ The study is descriptive but not inferential.

☑ The study is inferential but not descriptive.

4) According to a news report, Facebook is partnering with external researchers from the fields of political science and social media to **1 point** examine the impact of social networking sites on the society during the 2020 U.S. Presidential elections. Between 200,000 to 400,000 users are expected to opt into the project to see how they behave on Facebook and Instagram.

Which branch of statistics does the above analysis come under?

- ● Inferential statistics
- ○ Descriptive statistics

5) Choose the correct statements about descriptive statistics from the following: (more than one option can be correct)

- ☐ Descriptive statistics is concerned with drawing conclusions from the data
- ☑ Descriptive statistics is concerned with describing and summarizing the data.
- ☑ Descriptive statistics can be performed either on a sample or a population.
- ☐ Descriptive statistics can only be performed on a sample.
- ☐ We must always take into account the possibility of chance while studying descriptive statistics.

6) Central Pollution Control Board wants to check the Air Quality Index (AQI) of all the major cities in India after Diwali to decide if air **1 point** pollution is reaching unhealthy levels. To do so, they have come up with a list of all the major cities and their respective AQI levels and want to determine which cities among the list have an AQI level greater than 100. What kind of statistics does the above analysis come under?

- ● Descriptive statistics
- ○ Inferential statistics

7) Suppose we are interested in studying the height of an average Indian adult above the age of 18. Due to practical constraints, it is not **1 poir** possible to collect the data for the whole country. So we decide to choose a sample of 10,000 people. Which of the following choices for this sample do you think is the most representative of the population?

- ○ A sample consisting of 10,000 people from Karnataka.
- ○ A sample consisting of 10,000 women from South India
- ● A sample consisting of 2000 adults from each part of India (i.e., from northern, southern, eastern, western, and middle parts of India)
- ○ A sample consisting of 10,000 people from all over the country who are above 70 kg but below 75 kg.

8) Which one of the following statements best describes the difference between population and sample? **1 point**

○ Sample is the entire group, while population is a subset of the sample which is used to describe the parameters of the sample.

○ Sample and population are synonyms of each other

◉ Population is the entire group, while sample is a subset of the population which is used to describe the parameters of the population

○ Population is all members of an organization while sample is the members of the organization who volunteer to participate in the study.

9) According to the International Cricket Council (ICC), a test cricket ball, when new, should not weigh more than 163 grams. The **1 point** manufacturers of Duke cricket ball suspected that the average weight of their product is more than 163 grams. To confirm this, 100 Duke cricket balls were selected. Based on this, choose the correct option from below.

○ The population is all cricket balls; sample is 100 Duke cricket balls

○ The population is all Duke cricket balls; sample is 100 cricket balls

◉ The population is all Duke cricket balls; sample is 100 Duke cricket balls

○ The population is all Duke cricket balls; sample is all Duke cricket balls

10) A director of an engineering institute wants to evaluate the performance of students of all years and all departments (Computer **1 point** Science, Mechanical Engineering, Electrical and Electronics Engineering, Electronics and Communication Engineering) in the institute. He randomly selects performance reports of 50 students from the first year Computer Science batch. Based on this, choose the correct option from below.

○ The sample chosen is correct as it is collected randomly.

○ The sample chosen is wrong as it is not collected randomly

○ The sample chosen is correct as it is a subset of the population.

◉ The sample chosen is wrong as it is not a representative of the entire population

11) A construction company wants to start a house construction business in a city. Since there is a large number of houses, the company **1 point** collects a sample of houses from the city and uses the properties of the sample to study house prices in the city. Which branch of statistics does the above analysis come under?

○ Descriptive statistics

◉ Inferential statistics

12) An online opinion poll was conducted by a film reviewing website to rank Indian films from the 2000s. Among the 1000 films in the list, **1 point** Bhaag Milkha Bhaag received the maximum number of votes followed by 3 Idiots and Swades respectively. Which branch of statistics does the above analysis come under?

◉ Descriptive statistics

◯ Inferential statistics

Activity 1.2

1) Which of the following statement(s) is(are) true about structured data in a spreadsheet?

☑ Each observation must have its own row.

☑ Each variable must have its own column.

☑ For each variable the same type of value for each case is recorded.

☐ For each case the same type of value for each variable is recorded.
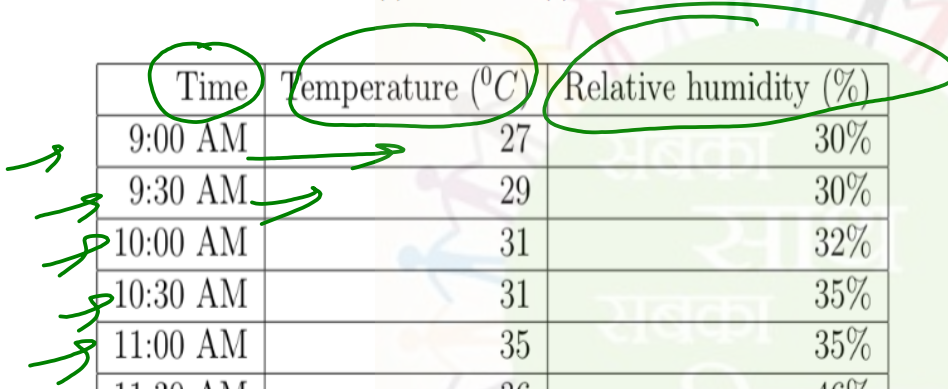
2) What kind of data is lyrics of a song?

◉ Unstructured data

◯ Structured data

3) Data is collected from people of every family living in a locality. The collected information is first name, second name, income, height, weight, blood group, house address, locality. Which among the following collected data is(are) not a variable(s)?

☐ First name

☐ Second name

☐ Income

☐ Height

☐ Weight

☐ Blood group

☐ House address

☑ Locality

4) The values of temperature and relative humidity of a room are measured for 24 hours, starting at 9 AM, at time interval of 30 minutes [Table 1.2.1]. What are the case(s) and variable(s) for this data?

Time	Temperature (^0C)	Relative humidity (%)
9:00 AM	27	30%
9:30 AM	29	30%
10:00 AM	31	32%
10:30 AM	31	35%
11:00 AM	35	35%
11:30 AM	36	46%
12:00 PM	38	47%
12:30 PM	39	50%
1:00 PM	38	51%
1:30 PM	38	53%
2:00 PM	35	54%
2:30 PM	31	55%
3:00 PM	31	57%
3:30 PM	29	58%
4:00 PM	28	60%

Table 1.2.1 Temperature and relative humidity variation in a day

○ Temperature is the case; relative humidity and time are the variables

○ Temperature is the case; relative humidity is the variable

○ Time is the case; temperature and relative humidity are the variables

○ Relative humidity is the case; temperature is the variable

● 9 AM, 9:30 AM, 10 AM are cases; time, temperature, and relative humidity are variables.

5) Which of the following is(are) example(s) of unstructured data?

- ☑ Web pages
- ☑ Video files
- ☑ Image files
- ☐ Phone number
- ☐ Email address

6) Which of the following statement(s) is(are) true?

- ☐ All the structured data have units.
- ☐ Comments in a social media platform are structured data.
- ☑ Unstructured data cannot be used directly to perform statistical analysis.
- ☐ Mathematical operations can be performed on unstructured data.

In a survey, art students are asked to mention their favourite artist and the amount of time they spend in painting outside the classroom. The data collected are given in Table 1.2.2.

S.no.	Name	Gender	Favourite artist	Amount of time(hr)
1.	David	Male	Amrita Shergil	4
2.	Lana	Female	M.F. Hussain	6
3.	Riya	Female	Frieda Kahlo	1
4.	Aakash	Male	. . .	. . .
5.	Kabir	Male	Leonardo Da Vinci	2
6.	Shakshi	Female	Pablo Picasso	0
7.	Mohit	Male	Angolie Ela Menon	3
8.	Nisha	Female	Judith Leyster	3

Table 1.2.2: Art students survey

Based on the data in Table 1.2.2, answer the following three questions:

7) Is the data given in Table 1.2.2 structured or unstructured?

- ◉ The data is structured.
- ○ The data is not structured.

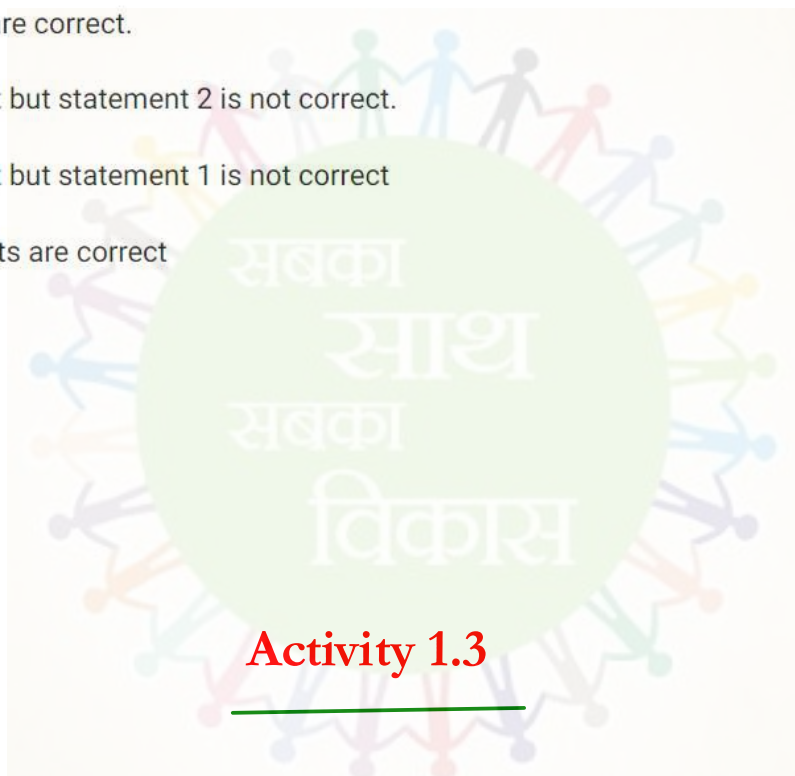8) Identify the variable(s) and observation(s) from the dataset given in Table 1.2.2.

☐ Name is an observation

☑ Name, Gender, Amount of time, and Favourite artist are variables

☐ Student and Name are variables

☑ Each student is an observation

9) Choose the correct statement(s) from the following:

Statement 1: Amount of time Aakash spends on painting is 0 hr.
Statement 2: We cannot make any conclusion about Aakash because data is not available to us.

○ Both the statements are correct.

○ Statement 1 is correct but statement 2 is not correct.

◉ Statement 2 is correct but statement 1 is not correct

○ None of the statements are correct

Activity 1.3

1) In order to forecast tomorrow's temperature in Mumbai, the India Meteorological Department collects the temperature, humidity and rainfall data of Mumbai for all days from the past three months. What kind of data is this?

◉ Time series data

○ Cross-sectional data

2) What kind of variables are comparative forms of adverbs (like more, less; good, better and best etc.)?

◉ Categorical variables

○ Numerical variables

3) What kind of variables are date and day of the week?

○ Categorical and numerical variables respectively

○ Both are numerical variables

◉ Both are categorical variables

○ Numerical and categorical variables respectively

4) What kind of variable is latitude?(More than one option can be correct)

☑ Numerical

☐ Categorical

☑ Continuous ✓

☐ Discrete

$1, 2, 3, 4, 5, 6$

$\frac{1}{2}, \frac{1}{4}$

5) During the Great Depression, several countries across the world experienced a rise in the rate of unemployment. The unemployment rate between the years 1929 and 1942 in the US was recorded on a monthly basis. Determine whether the data is time series or cross-sectional. **1 point**

◉ The data is time series ✓

○ The data is cross-sectional

6) Number of days a group of students take to finish a school project is a:

◉ Quantitative and Discrete variable

○ Quantitative and Continuous variable

○ Qualitative variable

○ Quantitative but not discrete or continuous variable

7) What kind of variable is the qualification of a candidate sitting for a job interview?

◉ Categorical/ Qualitative

○ Numerical/ Quantitative

○ Numerical and discrete

○ Numerical and continuous

8) Suppose one is interested in calculating the weight of a bag (in kg) before leaving for the airport. What kind of variable is the weight of the bag?

○ Quantitative and discrete

◉ Quantitative and continuous

○ Qualitative or categorical

Activity 1.4

1) A courier service is interested in knowing customer feedback. The customers are being given the following options to choose from.
Not satisfied
Somewhat unsatisfied
Neutral
Satisfied
Very Satisfied
What scale of measurement has been used here?

◉ Ordinal

○ Nominal

○ Interval

○ Ratio

2) For which of the following scale(s) of measurement is rank or order meaningful?

- ☑ Interval scale

- ☑ Ratio, interval and ordinal scale

- ☐ Nominal scale

- ☑ Ordinal scale

3) A survey has been conducted to determine the amount of time (in minutes) a student below the age of 12 spends on playing video games. The students are being given the following options to choose from.
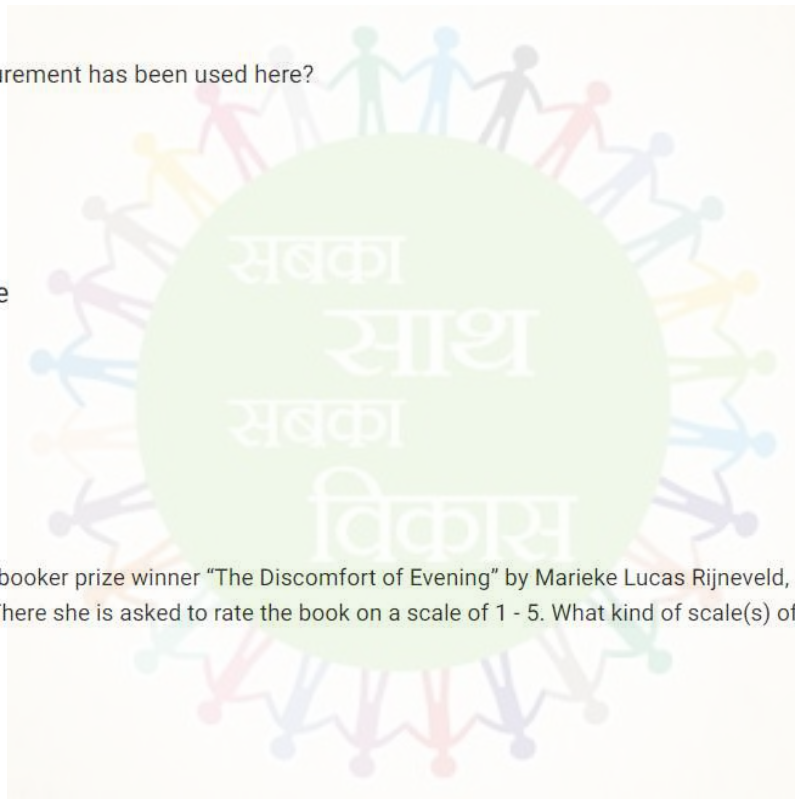Less than 30
30-45
45-60
More than 60
What kind of scale(s) of measurement has been used here?

- ○ Interval scale

- ○ Ratio scale

- ○ Ratio and Interval scale

- ● Ordinal scale

4) Katheryn has recently read the booker prize winner "The Discomfort of Evening" by Marieke Lucas Rijneveld, and decided to rate this **1 point** book on the Goodreads website. There she is asked to rate the book on a scale of 1 - 5. What kind of scale(s) of measurement is(are) being used here for rating?

- ● Interval scale

- ○ Ordinal scale

- ○ Ratio scale

- ○ Both interval and ratio scale

- ○ Both interval and ordinal scale

5) What can you say about the scale(s) of measurement being used for describing hair color such as blonde, brunette, red, and black?
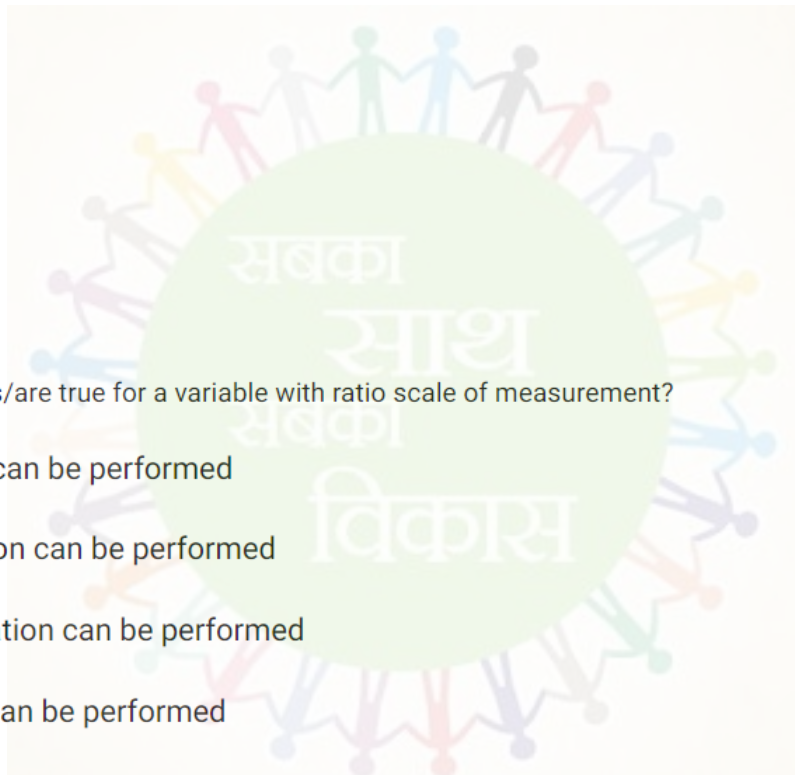
○ The scale of measurement is ordinal

○ The scale of measurement is both ordinal and nominal

◉ The scale of measurement is nominal

6) What is the scale of measurement for phone number?

○ Ordinal

◉ Nominal

○ Interval

○ Ratio

7) Which of the following is/are true for a variable with ratio scale of measurement?

☑ Addition operation can be performed

☑ Subtraction operation can be performed

☑ Multiplication operation can be performed

☑ Division operation can be performed

☑ The variable with ratio scale of measurement can have zero as one of its values

8) If addition, subtraction can be performed on a variable, then the scale(s) of measurement of the variable could be:

☑ Ratio

☐ Ordinal

☑ Interval

☐ Nominal