X

![swayam logo] (https://swayam.gov.in)   ![NPTEL logo] (https://swayam.gov.in/nc_details/NPTEL)

sai.doc45@gmail.com ⌄

NPTEL (https://swayam.gov.in/explorer?ncCode=NPTEL) » Business analytics and data mining Modeling using R (course)

☰

Course outline

How does an NPTEL online course work? ()

Week 0 ()

Week 1 ()

Week 2 ()

🔵 Lecture 6 PARTITIONING PROCESS (unit? unit=28&lesson=29)

🔵 Lecture 7 VISUALIZATION TECHNIQUES (unit? unit=28&lesson=30)

🔵 Lecture 8 VISUALIZATION TECHNIQUES PART-2 (unit? unit=28&lesson=31)

🔵 Lecture 9 VISUALIZATION

Week 2 : Assignment 2

The due date for submitting this assignment has passed.
Due on 2023-02-08, 23:59 IST.

Assignment submitted on 2023-02-08, 11:36 IST

1) Which of the following is true about data and data preprocessing?   **1 point**

○ Raw data is the data obtained after executing pre-processing steps.

○ Pre-processed data is the original form of data.

◉ Raw data is the main source of data.

○ None of the above.

Yes, the answer is correct.
Score: 1

Accepted Answers:
Raw data is the main source of data.

2) Which of the following statement is wrong about dataset partitioning?   **1 point**

○ Partitioning helps in addressing bias in the model.

◉ Partitioning creates multiple subsets of data which are then used for data visualisation.

○ Partitioning can be used to create a validation set which is then used to finetune and improve the model.

○ None of the above.

Yes, the answer is correct.
Score: 1

Accepted Answers:
Partitioning creates multiple subsets of data which are then used for data visualisation.

3) Which of the following statements indicate the true advantages of using panel data *1 point* over pure cross-sectional or pure time series data?
I: Panel data can model both the common and individual differences in the population.
II: Panel data packs more information and variability than pure cross-sectional or pure time series data
III: Panel data allows the analyst to estimate fewer statistical effects than pure cross-sectional or pure time series data.

○ I and II only.

○ II only.

○ None of the above.

○ I only.

Yes, the answer is correct.
Score: 1

Accepted Answers:
I and II only.

4) Which of the following data is the outcome of data cleaning and data preparation *1 point* steps?

○ Raw data.

◉ Pre-processed data.

○ Synchronised data.

○ None of the above.

Yes, the answer is correct.
Score: 1

Accepted Answers:
Pre-processed data.

5) When a dependent variable is restricted to a certain range of values while the *1 point* independent variables are allowed to have their corresponding values, this step is most accurately described as:

○ Censoring

◉ Truncation.

○ Discretization

○ None of the above.

No, the answer is incorrect.
Score: 0

Accepted Answers:
Censoring

6) Which of the following steps are executed in the data exploration phase? *1 point*

☑ Plot various data visualisations to shortlist the data mining techniques to be used.

☑ Free-form exploration of data to help the formal analyses.

☑ Identify bin sizes to convert continuous variables into categorical variables.

☐ None of the above.

Yes, the answer is correct.
Score: 1
Accepted Answers:
Plot various data visualisations to shortlist the data mining techniques to be used.
Free-form exploration of data to help the formal analyses.
Identify bin sizes to convert continuous variables into categorical variables.

7) Which of the followings is best suited to compare the distributions of more than one **1 point** group in the data?

- ◉ Box plots.
- ○ Histogram.
- ○ Line graph.
- ○ None of the above

Yes, the answer is correct.
Score: 1
Accepted Answers:
Box plots.

8) Which of the following statements are not the most accurate descriptions of data **1 point** visualisation?

- ☐ Box plots are useful to understand the distribution of the data and missing values.
- ☑ Bar graphs are suitable for time series data.
- ☑ Line graphs are not suitable for time series data.
- ☐ None of the above

Yes, the answer is correct.
Score: 1
Accepted Answers:
Bar graphs are suitable for time series data.
Line graphs are not suitable for time series data.

9) What does a large standard deviation suggest? **1 point**

- ◉ The values are widely distributed and the mean may not be a reliable measure of central tendency.
- ○ The values are widely distributed and the median would be an unreliable measure of the central tendency.
- ○ Values are not normally distributed.
- ○ None of the above.

Yes, the answer is correct.
Score: 1
Accepted Answers:
The values are widely distributed and the mean may not be a reliable measure of central tendency.

10) Histograms, line charts, and scatter plots are all types of: **1 point**

- ◉ 2D diagram.
- ○ 1D diagram

○ 3D diagram.

○ None of the above.

Yes, the answer is correct.
Score: 1

Accepted Answers:
2D diagram.