Score: **84%**

No. of questions: **10**

Correct answer: **9**

Incorrect answer: **1**

Show incorrect attempt only ⬤

Question 1 ✅   1 Mark

Chi square test is used for which pair of variables?

A	categorical-categorical	✅

B	continuous-categorical	

C	continuous-continuous	

Correct Answer: A. categorical-categorical

Chi square test is used when we have a pair of categorical variables.

Question 2 ✅   2 Marks

Which of the following function can be used to print the frequency table for a variable in Python?

‹ **Prev**
Quiz
(machine–
learning/screen/6144)

Next ›
Video
(machine-
learning/screen/6252)

A   value_counts

C   sum

D   frequency_table

Correct Answer: A. value_counts

In Python, we can use value_counts function to print the frequency table for a variable.

Question 3 ⊗   3 marks

Considering probability distribution - If mode is greater than median then the distribution is classified as?

A   Variable model

B   Positively skewed   ⊗

C   Negatively skewed

D   Constant model

Correct Answer: C. Negatively skewed

If the mode is greater than the median, this means that the peak of the distribution (mode) will be to the right whereas the median is on the left side of the mode. Hence, the distribution is left skewed or negatively skewed.

Question 4 ✔

1 Mark

If you are asked to predict the probability of a person passing the exam, then this problem is of type?

‹ Prev
Quiz
(machine-
learning/screen/6144)

Next ›
Video
(machine-
learning/screen/6252)

A	Classification

B	Regression

C	Clustering

D	None of the above

Correct Answer: A. Classification

Since the target variable is passing the exam, it will only have 2 values, i.e. pass or fail and hence it is a classification problem.

Question 5 ✓

2 Marks

Why do we have to identify the variables?

A	Because techniques like supervised learning requires identification of dependent variable

B	Data processing techniques vary for different types of variables

C	Both A and B ✓

D	None of the above

Correct Answer: C. Both A and B

There are different techniques based on the type of variables. Also, the processing techniques for different types of variables are different. So, both of the above answers are correct.

Prev
Quiz
(machine-learning/screen/6144)

Next
Video
(machine-learning/screen/6252)

Question 6 ✅

3 marks

The missing values have no relation to the variable in which missing values exist and neither with the other variables in the dataset. Which type of missing value is it?

A	Missing Completely at Random (MCAR)	✅

B	Missing at Random (MAR)

C	Missing Not at Random (MNAR)

D	None of the above

Correct Answer: A. Missing Completely at Random (MCAR)

Since there is no relation between the missing values and the variable having missing values neither between the missing values and other variables, it is an MCAR type, i.e. Missing Completely at Random.

Question 7 ✅

1 Mark

At what stage do we list down all the factors that affect our objective in the problem description?

A	Data exploration

B	Problem definition

C	Hypothesis generation	✅

D	Model building

Correct Answer: C. Hypothesis generation

Once we have defined the problem statement, the next step is to list down all the factors that can affect our objective and it is called as hypothesis generation.

Prev
Quiz
(machine-learning/screen/6144)

Next
Video
(machine-learning/screen/6252)

Question 8 ✓

2 Marks

Which measure of central tendency is most affected by outliers?

A	Mean	✓

B	Median

C	Mode

D	All of the above

Correct Answer: A. Mean

Adding a single outlier say 1000 in the series (say 1,2,3,4,5,5,6,7,8) will change the mean drastically, whereas the mode and median will remain the same. Hence, mean is most affected by outliers.

Question 9 ✓

3 marks

What is the use of bivariate analysis?

A	To find the relation between the target and independent variables

B	To find the relation between two independent variables

C	To examine all the continuous variables

D	Both A and B	✓

❮ **Prev**
Quiz
(machine-
learning/screen/6144)

Next ❯
Video
(machine-
learning/screen/6252)

Correct Answer: D. Both A and B

Bivariate analysis is used to find the relation between target and independent variables as well as the relation between two independent variables. Examining all the continuous variables comes under univariate analysis.

Module Test

Question 10 ✓

1 Mark

Which of the following data type represents continuous variable(s)?

A	int64	✓
B	object	
C	Both A and B	

Correct Answer: A. int64

int64 represents continuous variable whereas object represents categorical variable.

Prev
Quiz
(machine-learning/screen/6144)

Next
Video
(machine-learning/screen/6252)