# Unit 11 - Week 10

Course outline

How does an NPTEL online course work?

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

Week 9

Week 10

- Chi - Square Test of Independence - I
- Chi-Square Test of Independence - II
- Chi-Square Goodness of Fit Test
- Cluster analysis: Introduction-I
- Clustering analysis: part II
- Important data files
- Quiz : Assignment 10
- Solution : Assignment 10

Week 11

Week 12

Download Videos

Feedback

Text Transcripts

## Assignment 10

**The due date for submitting this assignment has passed.**
**As per our records you have not submitted this assignment.**

**Due on 2020-04-08, 23:59 IST.**

1) Which test is used to analyze the frequencies of two variables with multiple categories to determine whether the two variables are independent? *1 point*

- Chi Square Test of Independence
- 2 – way ANOVA
- 2 – sample z proportion test
- Both (a) and (c)

No, the answer is incorrect.
Score: 0
Accepted Answers:
Chi Square Test of Independence

2) The null hypothesis for test of independence is – *1 point*

- Two proportions are the same
- Two proportions are not the same
- Two proportions are different
- Both (b) and (c)

No, the answer is incorrect.
Score: 0
Accepted Answers:
Two proportions are the same

3) Which test compares expected (theoretical) frequencies of categories from population distribution to the observed (actual) frequencies from distribution to determine whether there is a difference between what was expected and what was observed - *1 point*

- Chi Square goodness-of-fit test
- Chi Square Test of Independence
- 2 – sample z proportion test
- ANOVA

No, the answer is incorrect.
Score: 0
Accepted Answers:
Chi Square goodness-of-fit test

4) The null hypothesis for chi square goodness-of-fit test is – *1 point*

- Population follow the specified distribution
- The population does not follow the specified distribution
- Population follow the normal distribution
- Population does not follow the normal distribution

No, the answer is incorrect.
Score: 0
Accepted Answers:
Population follow the specified distribution

5) Degrees of freedom can be calculated in chi square Test of Independence using formula – *1 point*
Where k denotes the number of categories, p denotes the number of parameters in sample data, r denotes the number of rows and c denotes the number of columns.

- k-1-p
- (r-1)(c-1)
- r+c-1
- none of these

No, the answer is incorrect.
Score: 0
Accepted Answers:
(r-1)(c-1)

6) Degrees of freedom can be calculated in chi square goodness-of-fit test using formula – *1 point*
Where k denotes the number of categories, p denotes the number of parameters in sample data, r denotes the number of rows and c denotes the number of columns.

- k-2-p
- k-1-p
- (r-1)(c-1)
- r+c-1

No, the answer is incorrect.
Score: 0
Accepted Answers:
k-1-p

7) Clustering analysis is- *1 point*

- Supervised learning technique
- Unsupervised learning technique
- Statistical tool
- Both (a) and (b)

No, the answer is incorrect.
Score: 0
Accepted Answers:
Unsupervised learning technique

8) Standardization of data is useful in clustering for – *1 point*

- To avoid the dependence on the choice of measurement units
- Finding unit-less variables
- Standardization of data is never useful during clustering analysis
- Both (a) and (b)

No, the answer is incorrect.
Score: 0
Accepted Answers:
Both (a) and (b)

9) Which are the properties satisfy by the Euclidean metric and the Manhattan metric of a distance function, for all objects i, j, and h *1 point*

- $d(i, j) \leq 0$
- $d(i, j) = d(j, i)$
- $d(i, j) \geq d(i, h) + d(h, j)$
- $d(i, i) \neq 0$

No, the answer is incorrect.
Score: 0
Accepted Answers:
$d(i, j) = d(j, i)$

10) Pick the right statement about the selection of variables – *1 point*

- Variables which do not contain any relevant information are worse than useless
- The selection of "good" variables is a nontrivial task
- The selection of "good" variables is a trivial task
- Both (a) and (b)

No, the answer is incorrect.
Score: 0
Accepted Answers:
Both (a) and (b)