Course outline

course work?

Week 1

Week 2

Week 3

Week 4

Week 5

Week 6

Week 7

Week 8

Week 9

Week 10

Test

Week 11

Week 12

Feedback

Download Videos

Text Transcripts

Chi - Square Test of Independence - I

Ohi-Square Test of Independence - II

Chi-Square Goodness of Fit

Cluster analysis: Introduction-

Clustering analysis: part II

Important data files

O Quiz : Assignment 10

O Solution : Assignment 10

How does an NPTEL online





Announcements

About the Course

Ask a Question

Progress Mentor

Unit 11 - Week 10

NPTEL » Data Analytics with Python

	due date for submitting this assignment has passed. er our records you have not submitted this assignment.
1) V	hich test is used to analyze the frequencies of two variables with multiple categories to determine whether the two
	Chi Square Test of Independence
	2 – way ANOVA
	2 – sample z proportion test
	Both (a) and (c)
No, t	he answer is incorrect.
Acce	pted Answers: Square Test of Independence
2) T	he null hypothesis for test of independence is –
0	Two proportions are the same
	Two proportions are not the same
0	Two proportions are different
	Both (b) and (c)
No, t	he answer is incorrect.
	pted Answers: proportions are the same
3) V	hich test compares expected (theoretical) frequencies of categories from population distribution to the observed
	tion to determine whether there is a difference between what was expected and what was observed -
0	Chi Square goodness-of-fit test
	Chi Square Test of Independence
	2 – sample z proportion test
0	2 – sample z proportion test ANOVA
O No, t	ANOVA he answer is incorrect.
No, t	ANOVA he answer is incorrect. e: 0
No, t Scor	ANOVA he answer is incorrect.
No, t Score Acce Chi S	ANOVA he answer is incorrect. e: 0 pted Answers:
No, t Score Acce Chi S	ANOVA he answer is incorrect. e: 0 pted Answers: equare goodness-of-fit test
No, t Scor Acce Chi S	ANOVA he answer is incorrect. e: 0 pted Answers: equare goodness-of-fit test he null hypothesis for chi square goodness-of-fit test is –
No, t Score Acce Chi S	ANOVA he answer is incorrect. e: 0 pted Answers: equare goodness-of-fit test he null hypothesis for chi square goodness-of-fit test is – Population follow the specified distribution
No, t Score Acce Chi S	he answer is incorrect. e: 0 pted Answers: Equare goodness-of-fit test he null hypothesis for chi square goodness-of-fit test is – Population follow the specified distribution The population does not follow the specified distribution
No, t Score Acce Chi S	ANOVA the answer is incorrect. the null hypothesis for chi square goodness-of-fit test is — 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 Population does not follow the normal distribution he answer is incorrect.

As per our records you have not submitted this assignment.	Due on 2020-04-08, 23:59	
 Which test is used to analyze the frequencies of two variables with multiple categories to determine whether the ty 	vo 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		
O 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	1 point
distribution to determine whether there is a difference between what was expected and what was observed -		
Chi Square goodness-of-fit test		
Chi Square Test of Independence 2 – sample z proportion test		
O 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
O Population follow the specified distribution		
The population does not follow the specified distribution		
O Population follow the normal distribution		
O 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	1 point
Where k denotes the number of categories, p denotes the number of parameters in sample data, r denotes the number of columns.	of rows and c denotes the number of	f
○ k-1-p ○ (r-1)(c-1)		
r+c-1		
onone 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 columns.	of rows and c denotes the number of	f
○ 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
		1 point
7) Clustering analysis is- Supervised learning technique Unsupervised learning technique		1 point
Supervised learning technique		1 point
Supervised learning technique Unsupervised learning technique		1 point
Supervised learning technique Unsupervised learning technique Statistical tool		1 point
Supervised learning technique Unsupervised learning technique Statistical tool Both (a) and (b) No, the answer is incorrect. Score: 0 Accepted Answers:		1 point
Supervised learning technique Unsupervised learning technique Statistical tool Both (a) and (b) No, the answer is incorrect. Score: 0		1 poin
Supervised learning technique Unsupervised learning technique Statistical tool Both (a) and (b) No, the answer is incorrect. Score: 0 Accepted Answers:		
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 –		
Supervised learning technique Unsupervised learning technique Statistical tool Both (a) and (b) No, the answer is incorrect. Score: 0 Accepted Answers: Unsupervised learning technique		
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 – To avoid the dependence on the choice of measurement units Finding unit-less variables Standardization of data is never useful during clustering analysis		
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 — 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)		
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 – To avoid the dependence on the choice of measurement units Finding unit-less variables Standardization of data is never useful during clustering analysis		
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 – 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:		
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 – 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)		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 – 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:		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 – 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)		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 − 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 or d(i, j) ≤ 0		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 – 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 or		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 — 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 of the d(i, j) ≤ 0 d(i, j) ≤ 0 d(i, j) ≥ d(i, h) + d(h, j)		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 — 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 of the distance function function functions for the distance function functio		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 — 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 of the d(i, j) ≤ 0 d(i, j) ≤ 0 d(i, j) ≥ d(i, h) + d(h, j)		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 − 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 o d(i, j) ≤ 0 d(i, j) ≥ d(i, h) + d(h, j) d(i, i) ≠ 0 No, the answer is incorrect. Score: 0 Accepted Answers:		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 – 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 o d(i, j) ≥ d(j, i) d(i, j) ≥ d(j, h) + d(h, j) d(i, j) ≠ 0 No, the answer is incorrect. Score: 0		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 — 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 of d(i, j) ≤ 0 d(i, j) ≥ d(i, h) + d(h, j) d(i, j) ≠ 0 No, the answer is incorrect. Score: 0 Accepted Answers: d(i, j) = d(j, i)	bjects i, j, and h	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 — 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 o d(i, j) ≤ 0 d(i, j) ≥ d(i, h) + d(h, j) d(i, i) ≠ 0 No, the answer is incorrect. Score: 0 Accepted Answers: d(i, j) = d(j, i)	bjects i, j, and h	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 − 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 o d(i, j) ≤ 0 d(i, j) ≥ d(j, h) + d(h, j) d(i, i) ≠ 0 No, the answer is incorrect. Score: 0 Accepted Answers: Both (a) and (b) 10) Pick the right statement about the selection of variables − Variables which do not contain any relevant information are worse than useless	bjects i, j, and h	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 − 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 of d(i, j) ≤ 0 d(i, j) ≤ d(i, h) + d(h, j) d(i, i) ≠ 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 − Variables which do not contain any relevant information are worse than useless The selection of "good" variables is a nontrivial task	bjects i, j, and h	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 − 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 o d(i, j) ≤ 0 d(i, j) ≥ d(j, h) + d(h, j) d(i, i) ≠ 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 − Variables which do not contain any relevant information are worse than useless	bjects i, j, and h	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 – 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 o d(i, j) ≤ d(i, h) + d(h, j) d(i, j) ≤ d(i, h) + d(h, j) d(i, j) = d(j, i) 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 – 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.	bjects i, j, and h	1 point
Supervised learning technique Unsupervised learning technique Statistical tool Both (a) and (b) No, the answer is incorrect. Score: 0 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 of d(i, j) ≤ 0 d(i, j) ≤ d(i, h) + d(h, j) d(i, j) ≠ 0 No, the answer is incorrect. Score: 0 Accepted Answers: Both (a) and (b) 10) Pick the right statement about the selection of variables − 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)	bjects i, j, and h	1 point