COMP4433 Quiz 3

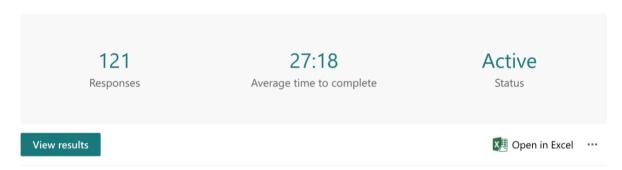
Submission deadline: 3:30pm 31 Oct 2023

Hi, Fu Lai Korris. When you submit this form, the owner will see your name and email address.	
1. Referring to slide 8 of "Clustering I" lecture notes, a dataset with n records and p attributes will form a nxp data matrix. how many dissimilarity values (i.e. d(i,j)) are needed to compute for the dissimilarity matrix? □ ⊘	
○ n*p	
0 (n * n - n)/2 Dissin d(i,i) v Dissin	nilarity matrix has nxn elements. The self dissimilarity values do not need to compute and there are n values. nilarity values are symmetric, i.e. $d(i,j)=d(j,i)$. So, the r is $(n * n - n)/2$
 2. This question is about distance measure for binary variables/attributes (i.e. slides 8-9 of "Clustering I" lecture notes. Which of the followings is true? Pick the best answer. □ ◊ ○ For a dataset with p asymmetric binary attributes, the range of Jaccard coefficient d(i,j) is from 0 to 1. 	
For a dataset with p symmetric binary attributes, the range of simple matching coefficient d(i,j) is from 0 to	
○ 1.	
For a dataset with p (asymmetric or symmetric) binary attributes, the sum of a, b, c, and d of the contingency table is equal to p, i.e., a+b+c+d=p.	
All of the above (i.e., the first, second and third statements above are correct).	
Only the second and the third statements are correct.	
	v obvious for the first two statements, i.e., $(b+c)/(a+b+c+d)$ or $)/(a+b+c)$ and the third statement $a+b+c+d=p$ which is the definition.
3. For a dataset with p categorical attributes, how many binary attributes will be created when one-hot encoding is used? Pick the best answer. \square	
It depends on the number of distinct attributes in each categorical attribute.	
It depends on whether the attribute values are floating point or integer type.	
It is equal to p.	The final answer should be $\sum_i n_i$ where n_i denotes the number of distinct attributes in categorical attribute i . So, only the first
None of the above	answer is valid.

COMP4433 Quiz#3

Some statistics for your reference:

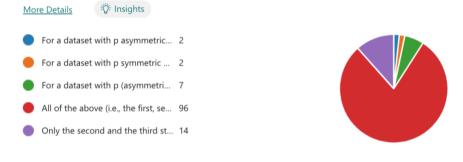
COMP4433 Quiz 3



1. Referring to slide 8 of "Clustering I" lecture notes, a dataset with n records and p attributes will form a nxp data matrix. how many dissimilarity values (i.e. d(i,j)) are needed to compute for the dissimilarity matrix?



2. This question is about distance measure for binary variables/attributes (i.e. slides 8-9 of "Clustering I" lecture notes. Which of the followings is true? Pick the best answer.



3. For a dataset with p categorical attributes, how many binary attributes will be created when one-hot encoding is used? Pick the best answer.



COMP4433 Quiz#3 2