Assignment 10

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

1. Which test is used to analyse 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 proportion test
   2-sample (x > 0)
   No, the answer is incorrect.
   1.
   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 (a) and (c)
   No, the answer is incorrect.
   4.
   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 variance
   - 2-sample proportion test
   2-sample (0.000)
   No, the answer is incorrect.
   5.
   Accepted Answers:
   Chi-Square goodness of fit test

4. The null hypothesis for chi-square goodness of fit test is - 1 point
   - Population follows the specified distribution
   - Population does not follow the specified distribution
   - Population follows the normal distribution
   - Population does not follow the normal distribution
   No, the answer is incorrect.
   6.
   Accepted Answers:
   Population follows the specified distribution

5. Degrees of freedom can be calculated in chi-square Test of Independence using formula - 1 point
   Where, r denotes the number of rows, c denotes the number of columns in the contingency table.
   $$ \chi^2 = \sum \frac{(O-E)^2}{E} $$
   $$ \chi^2 = \frac{(O-E)^2}{E} $$
   No, the answer is incorrect.
   7.
   Accepted Answers:
   Degrees of freedom = (r-1)(c-1)

6. Degrees of freedom can be calculated in chi-square goodness of fit test using formula - 1 point
   Where, n denotes the number of categories, p denotes the number of proportions in sample data, r denotes the number of rows, and c denotes the number of columns in the contingency table.
   $$ \chi^2 = \sum \frac{(O-E)^2}{E} $$
   $$ \chi^2 = \frac{(O-E)^2}{E} $$
   No, the answer is incorrect.
   8.
   Accepted Answers:
   Degrees of freedom = (n-1)

7. Standardization of data is useful in clustering for - 1 point
   - To avoid the dependence on the choice of measurement units
   - Finding null variables
   - Standardization of data is never useful during clustering analysis
   - Both (a) and (b)
   No, the answer is incorrect.
   9.
   Accepted Answers:
   Finding null variables

8. Which are the properties strictly by the Euclidean metric and the Manhattan metric of a distance function, for all objects \( x, y \): 1 point
   - \( d(x, y) \geq 0 \)
   - \( d(x, y) = 0 \iff x = y \)
   - \( d(x, y) = d(y, x) \)
   - \( d(x, y) + d(y, z) \geq d(x, z) \)
   No, the answer is incorrect.
   10.
   Accepted Answers:
   Both (a) and (b)

9. 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 (c) and (d)
   No, the answer is incorrect.
   11.
   Accepted Answers:
   Both (c) and (d)