statistic package

Kelvin Munsaka

October 31, 2023

1 section 1: Non-parametric Statistics (8.1, 8.2, 8.3)

1. which of the following tests is used to compare two independent samples distributed and the variance are not equal?

   Wilcoxon Rank-Sum Test

2. The wilcoxon Rank-Sum Test is a non-paramentric alternative to which parametric test?

   t-test

3. In a Chi-square Test for Independence, what is the null hypothesis?

   The variables are independent

4. You want to compare the median income of two different groups of people. Which non-parametric test should you use?

   Mann-Whitney U Test

2 Section 2: Advanced Data Analysis Techniques (Chapter 9)

5. In logistic regression, what type of variable is the dependent variable (the variable you are trying to predict)?

   Categorical variable

6. Which library is commonly used for time series analysis in Python?

   pandas

7. In time series analysis, what is a moving average used for?

   To smooth out noise and identify trends

8. Which of the following is a supervised learning algorithm used for binary classification in scikit-learn?

   K-Nearest Neighbors (KNN)

9. What is the purpose of training and testing datasets in machine learning?

   To evaluate the model's performance and generalization

10. Which of the following statements is true about overfitting in machine learning?

    Overfitting occurs when the model performs well on the training data but poorly on the testing data

11. Write Python code to perform the Kruskal-Wallis Test on the following dataset using the scipy library:

    $\text{data}_group1 = [23, 29, 32, 22, 28]$

$\text{data}_group2 = [19, 18, 24, 25, 21]$

$\text{data}_group3 = [16, 15, 14, 17, 20]$

```
from scripy.stats import kruskal

data_groupe = [23, 29, 32, 22, 28]
data_group = [19, 18, 24, 25, 21]
data_group = [16, 15, 14, 17, 20]

stat ,p = kruskal (data_group1, data_group2, data_group3)
print(p)
```