

Python for data science assignment questions

Adithya Bijoy
PES2UG22EC008

1.The probability of two different events occurring at the same time is known as : -

- Marginal probability
- Conditional probability
- Joint probability
- Marginal and Joint probability

Ans : - Joint Probability

2. The command to add "Notebook" as the first element inside the first level of the list "Stationery" is :-

- `Stationery[0].append('Notebook')`
- `Stationery[0].insert(0,'Notebook')`
- `Stationery[0][1] = "Notebook"`
- `Stationery[0].extend('Notebook')`

Ans : - `Stationery[0].insert(0,'Notebook')`

3. Write a Python code snippet to create a 1D NumPy array of numbers from 10 to 50 with a step of 5. Find the mean, median, and standard deviation of the above array using NumPy functions.

```
import numpy as np
arr = np.arange(10, 51, 5)
mean_val = np.mean(arr)
median_val = np.median(arr)
std_val = np.std(arr)
print("Array:", arr)
print("Mean:", mean_val)
print("Median:", median_val)
print("Standard Deviation:", std_val)
```

4. What does the term “correlation” mean? How is it different from “covariance”? Explain the importance of hypothesis testing in data science.

Correlation refers to the statistical measure that expresses the strength and direction of a linear relationship between two variables. Its value lies between -1 and $+1$, where $+1$ indicates a perfect positive

relationship, -1 indicates a perfect negative relationship, and 0 means no linear relation. For example, there is usually a positive correlation between study hours and exam scores.

Covariance, on the other hand, measures how two variables vary together. A positive covariance means the variables increase together, while a negative covariance means one increases as the other decreases. However, unlike correlation, covariance is not normalized and its values can range from $-\infty$ to $+\infty$, making it harder to interpret. Thus, correlation can be thought of as a normalized form of covariance that is unit-free and easier to compare across datasets.

Hypothesis testing is important in data science because it provides a statistical way to make decisions and validate assumptions based on sample data. It helps us determine whether observed patterns are real or due to random chance. For example, in A/B testing, hypothesis testing is used to decide if a new website design truly improves user engagement compared to the old one. In short, it is a vital tool to support evidence-based conclusions in analytics and machine learning.