
Date		Session No	4
------	--	------------	---

Topic : Statistics in BI and Data Analytics - Distributions

Practice Questions

Question:

Consider the following dataset, where x represents a set of values and $P(x)$ represents their corresponding probabilities: Using Python, calculate the following statistical parameters based on the given data:

$x=[1,2,3,4,5]$

$P(x)=[0.1,0.2,0.3,0.25,0.15]$

1. **Mean** of the distribution.
2. **Variance** of the distribution.

Steps to solve:

- Use the given values for x and $P(x)$ to calculate the mean and variance.
- Write a Python script to compute these values and print the results.

Solution :

$x = [1, 2, 3, 4, 5]$

$P_x = [0.1, 0.2, 0.3, 0.25, 0.15]$

```
# Mean (Expected value)

mean = sum(x[i] * P_x[i] for i in range(len(x)))

# Variance

variance = sum(((x[i] - mean) ** 2) * P_x[i] for i in range(len(x)))

print(f"Mean: {mean}")

print(f"Variance: {variance}")
```

Question 1: Discrete Probability Distribution

Consider the following values for x and their corresponding probabilities $P(x)$:

- $x=[2,4,6,8]$
- $P(x)=[0.1,0.3,0.4,0.2]$

Using Python, calculate the **mean** and **variance** of the distribution.

Solution

```
x = [2, 4, 6, 8]

P_x = [0.1, 0.3, 0.4, 0.2]

# Mean (Expected value)

mean = sum(x[i] * P_x[i] for i in range(len(x)))

# Variance

variance = sum(((x[i] - mean) ** 2) * P_x[i] for i in range(len(x)))

print(f"Mean: {mean}")

print(f"Variance: {variance}")
```

Question 2: Weighted Dice Distribution

You are rolling a weighted die, where the probabilities of rolling each number are given by:

- $x=[1,2,3,4,5,6]$
- $P(x)=[0.05,0.15,0.3,0.2,0.2,0.1]$

Using Python, calculate the **mean** and **variance** of this weighted die distribution.

Solution

```
x = [1, 2, 3, 4, 5, 6]

P_x = [0.05, 0.15, 0.3, 0.2, 0.2, 0.1]

# Mean (Expected value)

mean = sum(x[i] * P_x[i] for i in range(len(x)))

# Variance

variance = sum(((x[i] - mean) ** 2) * P_x[i] for i in range(len(x)))

print(f"Mean: {mean}")

print(f"Variance: {variance}")
```

Question 3: Distribution of Students' Test Scores

The probability distribution of students' test scores in a small class is given by:

- $x=[50,60,70,80,90]$
- $P(x)=[0.1,0.2,0.4,0.2,0.1]$

Using Python, calculate the **mean** and **variance** of the students' test scores.

Solution

```
x = [50, 60, 70, 80, 90]

P_x = [0.1, 0.2, 0.4, 0.2, 0.1]

# Mean (Expected value)

mean = sum(x[i] * P_x[i] for i in range(len(x)))

# Variance

variance = sum(((x[i] - mean) ** 2) * P_x[i] for i in range(len(x)))

print(f"Mean: {mean}")

print(f"Variance: {variance}")
```

Homework Assignment

Question 1:

A company's employee salary distribution (in thousands) is given by:

- $x=[30,40,50,60,70]$
- $P(x)=[0.1,0.2,0.4,0.2,0.1]$

Using Python, calculate the **mean** and **variance** of the salary distribution.

Question 2:

The number of daily visitors to a website follows this distribution:

- $x=[100,200,300,400,500]$
- $P(x)=[0.05,0.15,0.5,0.2,0.1]$

Using Python, find the **mean** and **variance** of the daily visitor count.

Question 3:

A bag contains different colored marbles. The probability distribution of selecting a marble of each color is given by:

- $x=[1,2,3,4]$ (where 1 = red, 2 = blue, 3 = green, 4 = yellow)
- $P(x)=[0.2,0.3,0.4,0.1]$

Calculate the **mean** and **variance** of the distribution using Python.

END