

NPTEL - PYTHON FOR DATA SCIENCE

Assignment - 0 Solution

1. A coin is thrown 3 times. What is the probability of getting at least one head?

Solution: a)

Sample space= {HHH, TTT, HHT, HTH, THH, TTH, THT, HTT}

Probability of getting at least one head = $7/8$

2. Few computer rooms in the school are denoted by the letters A to C. Staff records the number of classes held in each room during the first term. What kind of graph would be appropriate to present the frequency distributions of these data?

Solution: c)

Bar plot is appropriate for plotting the frequency distribution of categorical data

3. Calculate the median for the given set of values

16, 1, 8, 7, 20, 22

Solution: d)

1, 7, 8, 16, 20, 22

$(8+16)/2 = 12$

4. Which of the following statement(s) is/are TRUE with respect to data?

Solution: c)

Data will contain both qualitative and quantitative variables

5. The steps involved in data analytics are given in random order. Order the steps correctly.

Solution: b)

Steps involved in data analytics are

I. Preparing questionnaire II. Collection of data III. Preprocessing the data IV. Analysis of the data

6. Two dice are rolled simultaneously, what is the probability that the product of

two numbers obtained is an odd number?

Solution: c) 0.25

$$E = \{(1, 1), (1, 3), (1, 5), (3, 1), (3, 3), (3, 5), (5, 1), (5, 3), (5, 5)\}$$

$$n(E) = 9 ; n(S) = 36$$

$$\text{prob} = 9/36$$

$$\text{Ans: } 0.25$$

7. If A and B are two mutually exclusive event, which of the following expression is correct?

Solution:

$$\text{b) } P(A \cap B) = 0$$

When two events are mutually exclusive the probability of the two events occurring at the same time is 0.

Using the sample data, Data=(1,5,8,9,5,2,3) answer the following questions 8 & 9

8. The range value for the Data is__

Solution:

$$\text{d) } 8$$

$$\text{Range} = (\text{maximum value} - \text{minimum value})$$

$$\text{In this case, max} = 9 ; \text{min} = 1$$

$$\text{Range} = (9-1) = 8$$

9. The average value of the Data is

Solution:

$$\text{a) } 4.71$$

$$(1+5+8+9+5+2+3)/7=4.71$$

10. Which of the following are a part of data science?

Solution:

$$\text{d) All of the above}$$

Data visualization, data exploration and data cleaning are a part of data science

11. Which of the following programming language(s) is/are popularly used for data science?

Solution:

a) R, c) Python

R and Python are popularly used programming languages for data science

12. Three unbiased coins are tossed. What is the probability of getting at most two heads?

Solution:

Here $S = \{TTT, TTH, THT, HTT, THH, HTH, HHT, HHH\}$

Let, E = event of getting at most two heads

Then $E = \{TTT, TTH, THT, HTT, THH, HTH, HHT\}$

$$P(E) = n(E)/n(S)$$

The probability of getting at most two heads = $7/8$

13. A dice is rolled twice. What is the probability of getting sum 9?

Solution:

Possible event = $6 \times 6 = 36$

Favourable outcomes = $\{(3, 6), \{4, 5\}, \{5, 4\}, (6, 3)\}$

Thus, probability of getting sum 9 = favourable/total

$$= 4/36$$

$$= 1/9$$

14. Which of the following is a parameter in statistics?

Solution:

b) a population characteristic

15. Which of the following divides the number of data points into four quarters?

Solution:

c) Quartiles

Quartiles divides the number of data points into four quarters.