Statistics and Probability

Assignment

Q1. Given a sample, find the standard deviation and variance of the same.

Sample = [42 61 39 70 8 30 52 28 95 86 35 77 80 86 1 84 90 57 45 6 1 56 60 83 24 63 93 59 92 53]

    a. 28.44, 800.49
    b. 24.89, 811.59
    c. 28.94, 819.15
    d. 28.49, 811.95

Q2. For a given unsorted array, find the range of the array, and the Interquartile range(IQR) of the array.
array = numpy.array([17, 27, 80, 39, 72, 95, 52, 85, 8, 64, 36, 15, 62, 76, 66, 67, 59, 92])

    a. 87, 32.85
    b. 95, 26.25
    c. 87, 38.25
    d. 95, 35.28

Q3. In a skewed distribution, If the mean and median of a sample is 56, and 63 and the mode of the same is 65. The distribution graph of the sample will be ___?

    a. Positively Skewed
    b. Negatively Skewed
    c. Unskewed
    d. None of the Above

Q4. If the IQR range of the sample is 4.5, and IQR 1 and IQR 3 values are 4, 8.25 respectively. Which of the following are suspected outliers in the sample?

    a. 12, 3
    b. 9 , 2
    c. -4, 15
    d. -1, 13

Q5. In a normal distribution curve ___?
    a. The mean and median are less than the mode
    b. The mean and median are greater than the IQR range
    c. The AUC value is 1
    d. The graph is asymmetrical with negatively skewed or positively skewed distribution
Q6. Calculate the 95% confidence interval of the given sample using the t-distribution.
Sample = [78 65 20 69 36 81 85 71 44 71 47 37 89 25 73 55 80 52 46 77].

a. (50.400 , 69.699)
 b. (40.500 , 63.659)
c. (45.450 , 67.599)
d. (54.500 , 68.649)

Q7. If the mean value is 100 and standard deviation is 15 for a normally distributed sample.
Calculate the p-value from the given sample after performing the one sample z-test on the following
data.
Sample = [88, 92, 84, 64, 96, 97, 67, 97, 96, 99,115, 119, 119, 129, 130, 142, 102, 123, 134, 105]

a. 1.03
 b. 0.30
c. 1.30
d. 0.32

Q8. To calculate the p-value in a chi squared test, what among the following is not required?
a. Degrees of freedom
 b. Alpha Level
c. IQR value
d. None of the above

Q9. If the chi squared value is low, it means that ___?
a. The correlation is high between the two sets of data.
 b. The correlation is low between the two sets of data.
c. Indicates a statistically significant difference
d. None of the above

Q10. A card is drawn at random from a standard deck of cards. Recall that there are 13 hearts, 13
diamonds, 13 spades and 13 clubs in a standard deck of cards. • Let H be the event that a heart is
drawn, • Let R be the event that a red card is drawn and • Let F be the event that a face card is
drawn, where the face cards are the kings, queens and jacks.

a. If I draw a card at random from the deck of 52, what is P(H)?
 b. If I draw a card at random from the deck of 52, what is P(F)?
c. If I draw a card at random, and without showing you the card, I tell you that the card is red,
    then what are the chances that it is a face card?

a.   0.25, 0.23, 0.23
b.   0.50, 0.23, 0.23
c.   0.50 , 0.25, 0.25
d.   None of the above
Q11. A box contains three coins: two regular coins and one fake two-headed coin

(P(H)=1). a. You pick a coin at random and toss it. What is the probability that it lands heads up?

b. You pick a coin at random and toss it, and get heads. What is the probability that it is the two-headed coin?

a. 4/6 ,1/2
b. 2/3, 1/5
c. 2/5, 1/3
d. None of the above

Q12. In my town, it's rainy one third of the days. Given that it is rainy, there will be heavy traffic with probability 12, and given that it is not rainy, there will be heavy traffic with probability 14. If it's rainy and there is heavy traffic, I arrive late for work with probability 12. On the other hand, the probability of being late is reduced to 18 if it is not rainy and there is no heavy traffic. In other situations (rainy and no traffic, not rainy and traffic) the probability of being late is 0.25. You pick a random day.

a. What is the probability that it's not raining and there is heavy traffic and I am not late?

b. What is the probability that I am late?

c. Given that I arrived late at work, what is the probability that it rained that day?

a. 1/8 , 11/48, 6/11
b. 1/4, 11/22, 6/12
c. 1/2, 2/3, 4/12
d. 3/4, 2/4, 1/5

Q13. A nationalized bank has found that the daily balance available in its savings accounts follows a normal distribution with a mean of Rs. 500 and a standard deviation of Rs. 50. The percentage of savings account holders, who maintain an average daily balance of more than Rs. 500 is _____ ?

a. 0.5
b. 0.4
c. 0.3
d. 0.1

Q14. A day's production of 850 manufactured parts contains 50 parts that do not meet customer requirements. Suppose two parts are selected from the batch, but the first part is replaced before the second part is selected. What is the probability that the second part is defective? (Denoted as B) given that the first part is defective (denoted as A)?

a. 0.003
b. 0.03
c. 0.3

d. None of the above

Q15. Which of the following statements is/are correct with respect to the below shaded part in the image:
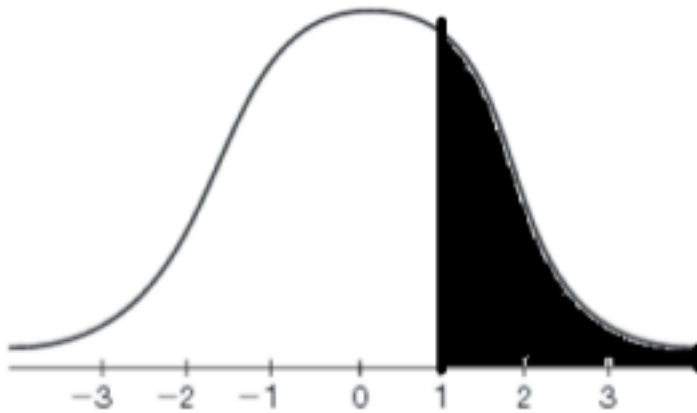
a. All data that is greater than 1.

b. All data that is either one or more standard deviations above the mean value. c. All data between 1 and 3.

d. All data that is greater than mean.

Q16. Which of the following options is/are area under the standard normal distribution curve?

a. 0.1

b. 25

c. 0.5

d. 1

Q17. Which of the following statements is/are true with respect to the below image:
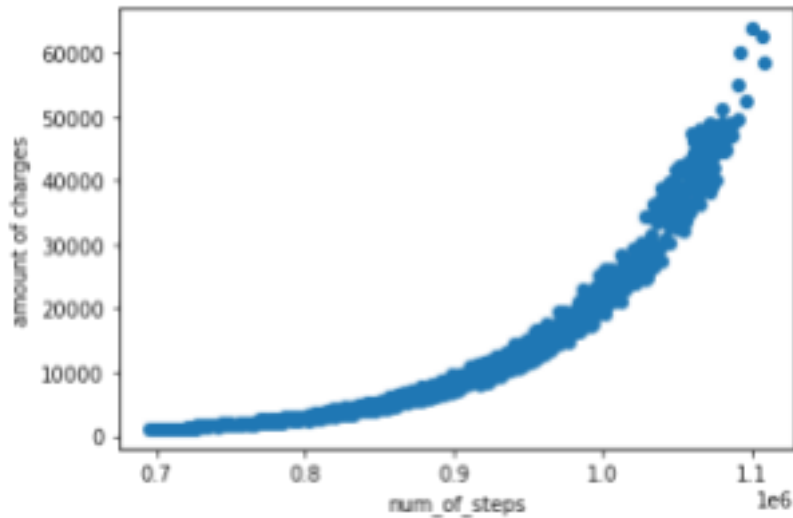
a. The amount of charges is not having a linear relationship with the number of steps. b. The amount of charges is exponentially increasing with respect to the number of steps. c. Both of the above.

d. None of the above.

Q18. Which of the following statements is/are true with respect to the below image
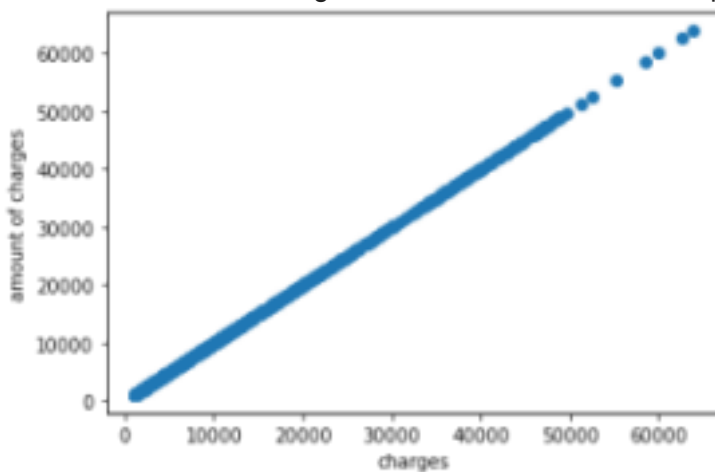
a. The amount of charges is having a linear relationship with number of steps.

b. The amount of charges is increasing with respect to the number of steps.

c. Both of the above.

d. None of the above.

Q19. Find t-statistic for the following samples given that the mean of the distribution is 8. Sample: [5 3 3 5 5 6 6 7 3 8]

(a) -5.3045
(b) 0.0005

(c) 3.1623
(d) -0.0026

Q20. There are two machines for spray coating at desired thickness. Their variability should be similar if they need to be used for the same applications. To test this, various samples were collected from both machines and the sample covariance of machine 1 and 2 are 1.84 and 6.36 respectively. Comment on whether these machines can be used for the same application if the machines were truly built to have the same variance. Perform the test at 95% significance given n1 = 24 n2 = 25

(a) Accept null hypothesis, variance of both the machines are same
(b) Accept null hypothesis, variance of both the machines are different
(c) Reject null hypothesis, variance of both the machines are same
(d) Reject null hypothesis, variance of both the machines are different

Q21. In the contact process for the production of sulphuric acid, temperature of not more than 450°C is desired. The reactor temperature was measured 24 times and it was found that the mean temperature was 450.3°C and the variance was 1.0. Find whether the reactor is operating at the desired condition? Perform the test at 95% significance.

(a) Reject null hypothesis, reactor operating at the desired condition
(b) Accept null hypothesis, reactor not operating at the desired condition
(c) Accept null hypothesis, reactor operating at the desired condition
(d) Reject null hypothesis, reactor not operating at the desired condition

Q22. Find f-statistic for the following sample sets generated from the distributions with variance 7 and 11 respectively.
Sample set 1: [-3 3 5 9 2 5 10 5 2 6], Sample set 2: [6 7 8 7 3 12 7 10 6 11].

(a) 9.731
(b) 3.0497
(c) 3.527
(d) 0.7178

Q23. Two biased coins are tossed 126 times. 36 trials are both heads and 44 trials are both tails out of these 126 trials. Among the rest, 16 trials have head in the first coin while tail in the second coin and 30 trials with tail in the first coin and head in the second coin. Find the probability that the coin 2 gives a tail given that the first coin gave a head.

(a) P(C2=T|C1=H) = 0.308
(b) None of these
(c) P(C2=T|C1=H) = 0.476
(d) P(C2=T|C1=H) = 0.625

Q24. A short circuit tester is being used in a battery assembly line to test the batteries for short circuit.

The battery will be rejected if the test is positive and the battery will be accepted if the test is negative. 91.74% of the non-defective batteries were accepted and 3.19 % of the defective batteries also were accepted. In addition, 93.73 % of the batteries are non-defective and 2.27% of the batteries are defective. If a battery is rejected, what is the probability that the battery is non-defective? (a) 0.023

    (b) 0.917
    (c) 0.555
    (d) 0.937

    Correct Answer: 0.555

Q25. A total of 69 percent of Indians eat Dosai,24 percent of Indians eat Dhokla, and 11 percent eat Dosai and Dhokla. Then, what percentage of Indians eat neither Dosai nor Dhokla? (a) 0.17

    (b) 0.18
    (c) 0.11
    (d) 0.82

Q26. Two balls are randomly drawn one after the other without replacement from a bowl containing 12 and 8 black balls. If the first ball drawn is white, what is the probability that second ball drawn is black?

    (a) 0.63
    (b) 0.4
    (c) 0.42
    (d) 0.6

Q27. Two biased coins are tossed 144 times. 50 trials are both heads and 39 trials are both tails out of these 144 trials. Among the rest, 25 trials have head in the first coin while tail in the second coin and 30 trials with tail in the first coin and head in the second coin. Calculate the marginal probabilities for both the coins.

    (a) C1H=0.521, C1T=0.479, C2H=0.556, C2T=0.444
    (b) None of these
    (c) C1H=0.521, C1T=0.556, C2H=0.479, C2T=0.444
    (d) C1H=0.556, C1T=0.479, C2H=0.521, C2T=0.444

Q28. The outcomes of an experiment are {e, f, g, h} with probabilities 0.39, 0.24, 0.77, and 0.88, respectively. Let E and F denote the events {e, f} and {f, h}, respectively. Compute (a) P(E∩F) and (b) P(E ∪ F)

    (a) 0.24,1.51
    (b) 1.51,0.24
    (c) 1.12,1.51
    (d) 0.24,0.63