Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Data |
| Results of rolling a dice | Discrete Data |
| Weight of a person | Continuous Data |
| Weight of Gold | Continuous Data |
| Distance between two places | Continuous Data |
| Length of a leaf | Continuous Data |
| Dog's weight | Continuous Data |
| Blue Color | Discrete Data |
| Number of kids | Discrete Data |
| Number of tickets in Indian railways | Continuous Data |
| Number of times married | Discrete Data |
| Gender (Male or Female) | Discrete Data |

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender | nominal |
| High School Class Ranking | ordinal |
| Celsius Temperature | ratio |
| Weight | ratio |
| Hair Color | ordinal |
| Socioeconomic Status | ordinal |
| Fahrenheit Temperature | ratio |
| Height | ratio |
| Type of living accommodation | ordinal |
| Level of Agreement | interval |
| IQ(Intelligence Scale) | interval |
| Sales Figures | interval |
| Blood Group | ordinal |
| Time Of Day | interval |
| Time on a Clock with Hands | interval |
| Number of Children | interval |
| Religious Preference | ordinal |
| Barometer Pressure | ratio |
| SAT Scores | interval |
| Years of Education | interval |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?

Solution:

Sample space n(s) = {TTT, TTH, THT, THH, HTT, HTH, HHT, HHH}=8

n(E)= {THH,HTH,HHT}=3

Probability = no of favorable outcomes/ no of total number of outcomes

= 3/8

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

Let A be the event of getting equal to 1

n(s)= {(1,1) (1,2) (1,3) (1,4) (1,5) (1,6)……….(6,6)}

n(s)= 36

n(A)= 0

P(A)= 0/36

1. Less than or equal to 4

Let B be the event of getting sum equal to 4

n(B)={(2,2) (3,1) (1,3)}=3

P(B) = n(B)/n(s)

= 3/36

= 1/12

1. Sum is divisible by 2 and 3

Let C be the event of getting sum divisible by 2 and 3

n(C)={(1,5) (3,3) (4,2) (5,1) (2,4) (6,6)}

P(B)=n(C)/n(s)

= 6/36

= 1/6

Q5) A bag contains 2 red, 3 green and 2 blue balls. Two balls are drawn at random. What is the probability that none of the balls drawn is blue?

Sol:

Total no of balls = 2+3+7

= 7

S = two balls are drawn at random

n(s) = 7C2 = 7\*6/2\*1

= 42/2

= 21

A= none of the ball drawn is blue

n (A)= 5C2=5\*4/2\*1

= 20/2 =10

P(A)=n(A)/n(s)

= 10/21

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of count of candies for children (ignoring the nature of the child-Generalized view)

|  |  |  |
| --- | --- | --- |
| CHILD | Candies count | Probability |
| A | 1 | 0.015 |
| B | 4 | 0.20 |
| C | 3 | 0.65 |
| D | 5 | 0.005 |
| E | 6 | 0.01 |
| F | 2 | 0.120 |

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Sol:

Child A 1candy= 0.015

Child B 4candy= 0.20

Child C 3candy= 0.65

Child D 5candy= 0.005

Child E 6 candy= 0.01

Child F 2 candy= 0.120

=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

= 3.090

= 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weigh>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weight |
| Mean | 3.5965625 | 3.21725 | 17.84875 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode | 3.92 | 3.44 | 17.02 |
| Standard Deviation | 0.534678736 | 0.978457443 | 1.786943236 |
| Maximum | 4.93 | 5.424 | 22.9 |
| Minimum | 2.76 | 1.513 | 14.5 |
| Range | 2.17 | 3.911 | 8.4 |

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Sol:

Expected value = ∑(probability\*value)

∑p(x)E(x)

There are 9 patient

Probability of selecting each patient = 1/9

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Ex | 108 | 110 | 123 | 134 | 135 | 145 | 167 | 187 | 199 |
| P(x) | 1/9 | 1/9 | 1/9 | 1/9 | 1/91 | 1/9 | 1/9 | 1/9 | 1/9 |

E.V=(1/9)(108)+(1/9)(110)+(1/9)(123)+(1/9)(134)+(1/9)(135)+(1/9)(145)+

(1/9)(167)+(1/9)(187)+(1/9)(199)

= (1/9) (108+110+123+134+135+145+167+187+199)

= (1/9) (1308)

= 145.33

Expected value of the weight of that patient = 145.33

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
| **Index** | **Speed** | **Distance** |
| **Skewness** | -0.1175099 | 0.806895 |
| **Kurtosis** | -0.5089944 | 0.4050526 |

-Skewness for speed = -0.1175099, skewness value is negative so it is left skewed. -Since magnitude is slightly greater than 0 it is slightly left skewed.

-And for distance = 0.806895, right skewed (positive) slightly magnitude to right.

-Kurtosis for speed is -0.5089944, is less than 3 so it is platykurtic.

-Kurtosis for distance is 0.4050526, is less than 3 so it is platykurtic

**Q10) Draw inferences about the following boxplot & histogram**



Sol:

Here, Median is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points between Q1 and the bottom point.

**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 94%,98%,96% confidence interval?

Sol:

Degrees of freedom = 2000-1=1999

Confidence interval for 94% = 1.882

Confidence interval for 98% = 2.33

Confidence interval for 96% = 2.05

**Q12)** Below are the scores obtained by a student in tests

**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation.
2. What can we say about the student marks?

Sol:

Mean = 41

Median = 40

Variance = 27.05

Standard deviation = 5.20181

Q13) What is the nature of skewness when mean, median of data are equal?

Symetrical

Q14) What is the nature of skewness when mean > median ?

Right skewed.

Q15) What is the nature of skewness when median > mean?

Left skewed.

Q16) What does positive kurtosis value indicates for a data ?

The data is normally distributed and kurtosis is 0.

Q17) What does negative kurtosis value indicates for a data?

Negative kurtosis value indicates that wider and thinner tails.

Q18) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

Answer:Not Normally distributed

What is nature of skewness of the data?

Answer: Left skewed, median is greater than mean. So Negatively skewed

What will be the IQR of the data (approximately)?   
  
Answer: Here, will be the IOR of the data maybe 10 - 18

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Answer: In this boxplot this level is high in boxplot 2, mean and the median are equal ,and the distribution is symmetrical.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG of Cars for the below cases.

Answer: MPG <- Cars$MPG

* 1. P(MPG>38)=0.3475908
  2. P(MPG<40)=0.7293527

c. P (20<MPG<50)=0.01311818

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer:

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | Normally distributed |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| |  | | --- | |  | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Answer:

Waist Circumference is normally distributed

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| |  | | --- | |  | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | |  |  |  |  |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  | |  |  |  |  |  |  |  | |  |  |  |  |  |  |  | |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
| Adipose Tissue is normally distributed |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |  |

Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval

Answer:

|  |  |
| --- | --- |
| Confidence interval | Z scores |
| 60% | 0.8416212 |
| 90% | 1.644854 |
| 94% | 1.880794 |

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Answer:

|  |  |
| --- | --- |
| Confidence interval | T scores |
| 95% | 2.063899 |
| 96% | 2.171545 |
| 99% | 2.79694 |

Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Solution:

t – statistics for the data is given as follows:

x = mean of the sample of bulbs = 260

µ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

The probability that t<-0.471 with 17 degrees of freedom assuming the population mean is true, the t-value obtained with 17 degrees of freedom and t score of -0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of bulb is 300 days.