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

Q1) Identify the Data type for the Following:

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 | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Ans:

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

Probability of two heads and one tail = = 0.375

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

1. Equal to 1
2. Less than or equal to 4
3. Sum is divisible by 2 and 3

Ans:

S = {(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),

(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),

(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),

(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),

(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),

(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)}

1. P(E1) = = 0
2. P(E2) = = 1.66
3. P(E3) = = 1.66

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?

Ans: Total number of balls (S) = 2+3+2=7

Number of ways drawing 2 balls n(S) = = 21

Number of ways drawing 2 balls which is not blue n(E) = = 10

P(E) = =

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

Ans: 0.015+0.8+1.95+0.025+0.06+0.24 = 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**

Ans:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.5965625 | 3.217250 | 17.848750 |
| Median | 3.69500 | 3.325 | 17.71 |
| Mode | 3.07,3.92 | 3.44 | 17.02,18.90 |
| Standard deviation | 0.285881 | 0.957378 | 1.78694 |
| Variance | 0.5346787 | 0.9784574 | 3.19316 |
| Range | 2.17 | 3.911 | 8.3999 |

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?

Ans:

Expected value = (×108) + (×110) + (×123)+ (×134)+(×135) (×145)+ (×167)+ (×187)+ (×199)

=

= 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**

Ans:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Speed | Dist | SP | WT |
| Skewness | -0.1175 | 0.8068 | 1.6114 | -0.6147 |
| Kurtosis | -0.5089 | 0.4050 | 2.9773 | 0.9502 |

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



Ans: The distribution is positively skewed. The values are more concentrated towards the right side. It has got a very long right tail. Here mode<median<mean. This distribution has got outliers.



Ans: There is the presence of outliers and they have values greater than the upper bound.

**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?

Ans: Sample mean(x) = 200

Sample size (n)=2000

Sample standard deviation (S) = 30

Confidence level = (x-×,x+×)

Confidence level at 94% = (198.68,201.31) [t=1.968]

Confidence level at 98% = (198.43,201.56) [t=2.330]

Confidence level at 96% = (198.61,201.38) [t=2.0673]

**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?

Ans:

Mean = = 41

Median = = 40.5

Variance = 24.11

Standard Deviation = 4.91

1. Most of the students have scored 41 marks. The data points are deviated from the mean at 4.91

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

Ans:The Distribution is perfect and we don’t have any skewness

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

Ans: The distribution is positively skewed and the tail is on the right side

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

Ans: The distribution is negatively skewed and the tail is on the left side

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

A kurtosis value greater than 3 will indicate positive kurtosis and the distribution will be heavy tailed distribution. The curve will be more peaked and it is called Leptokurtic.

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

A kurtosis value lesser than 3 will indicate negative kurtosis and the distribution will be short tailed distribution. The curve will be more flatter and it is called Platykurtic.

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



What can we say about the distribution of the data?

Ans: The median in the aforementioned Boxplot is not normally distributed and is moving upward.

What is nature of skewness of the data?

Ans: The data is negatively skewed.

What will be the IQR of the data (approximately)?

Ans: IQR = 18-10 = 8  
  
  
Q19) Comment on the below Boxplot visualizations?



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

Ans: Both the boxplots have same median approximately 260. The datas are normally distributed and there is no skewness present. Also there are no outliers present in both plots.

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.

MPG <- Cars$MPG

* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Ans: a)P(MPG>38) = 0.347

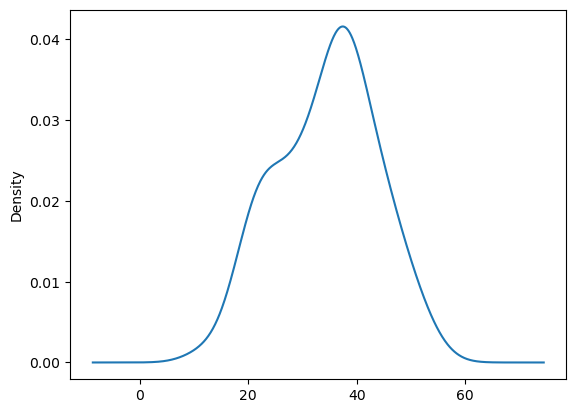
1. P(MPG<40) = 0.729
2. P(20<MPG<50) = 0.013

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:

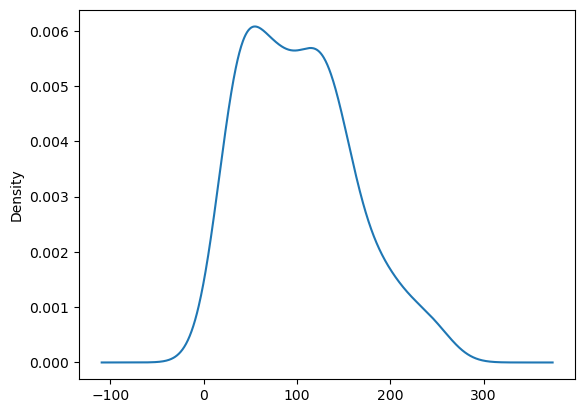
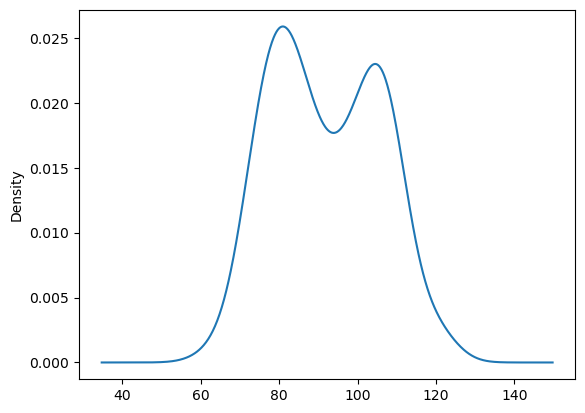


Yes it follows normal distribution.

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

Dataset: wc-at.csv

ANs:



No it does not follow normal distribution.

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

Ans:

Z scores of 90% confidence interval= 1.644

Z scores of 94% confidence interval=1.880

Z scores of 60% confidence interval=0.841

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

Df=24

t scores of 95% confidence interval= 2.264

t scores of 96% confidence interval=2.171

t scores of 99% confidence interval=2.797

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

Ans:

Sample mean(x) = 260

Population mean(X)=270

Sample standard deviation(S)= 90

n=18

t=(x-X)/()

= (260-270)/() = -0.471

Probability = 0.3218