|  |  |
| --- | --- |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

**Ans**- If three coins are tossed, the sample space is

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

n(S)=8

Let A be an event such that two heads and one tail are obtained

A={HHT,HTH,THH}

n(A)=3

probability that two heads and one tail are obtained = n(A)/n(S) =3/8

**Required probability=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-** If two dice are rolled, the sample space is

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),}

n(S)=36

**a)**Let A be an event such that sum is equal to 1

A={} , since there is no such sample point which have sum equal to 1

n(A)=0

probability that sum is equal to 1= n(A)/n(S) = 0/36

**Required probability=0**

**b)**Let B be an event such that sum is less than or equal to 4

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

n(B)=6

probability that sum is less than or equal to 4 = n(B)/n(S) = 6/36

**Required probability=1/6**

**c)**Let C be an event such that sum is divisible by 2 and 3

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

n(C)=6

probability that sum is is divisible by 2 and 3= n(C)/n(S) = 6/36

**Required probability=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?

Ans- total no. of balls = 2+3+2 = 7

Two balls are drawn from 7 balls at random

Possible combinations= == =21

If two balls drawn are not blue, then possible no. of combinations= = = 10

**Required probability=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

Ans- Expected number of candies = ∑

=0.015+0.80+1.95+0.025+0.06+0.240 =3.09

Expected number of candies is 3.

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- i) Mean of Points = 3.59656

Mean of Score =3.21725

Mean of Weigh =17.84875

ii)Median of Points = 3.695

Median of Score = 3.325

Median of Weigh=17.71

iii)Mode of Points =3.07 and 3.92

Mode of Score =3.44

Mode of Weigh= 17.02 and 18.90

iv) Variance of Points = 0.285881

Variance of Score =0.957378

Variance of Weigh =3.193166

v) Standard Deviation of Points = 0.53467

Standard Deviation of Score =0.978457

Standard Deviation of Weigh =1.786943

vi) Range of Points = 2.17

Range of Score =3.911

Range of Weigh =8.399

Hence, Points and Weigh are bimodal

Points has very less variance.

From the range, data points of Weigh are widely distributed.

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- The expected value is given by ∑/n

Hence expected value of weight of patient= 145.334 pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans-** skewness of speed= -0.1175

kurtosis of speed= -0.5089

Speed is negatively skewed and platykurtic.

skewness of Distance= 0.8069

kurtosis of Distance= 0.4051

Distance is positively skewed and leptokurtic.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans-** skewness of SP= 1.6115

kurtosis of SP= 2.9773

SP is positively skewed and leptokurtic.

skewness of WT= -0.61475

kurtosis of WT= 0.9503

SP is negatively skewed and leptokurtic.

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



Ans- From the given histogram ,data is negatively skewed . And outliers are present.

From the boxplot, outliers are present on higher observations.

**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.- We have,

N=3000000, n=2000,

Xbar=200, s.d.=30

94% Confidence Interval is (198.9570, 201.0429)

98% Confidence Interval is (198.6223, 201.3776)

96% Confidence Interval is (198.8256, 201.1743)

**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 is 41.0

Median is 40.5

Mode is 41

Variance is 24.1111

Standard deviation is 4.9103

As Mean >Median, data is positively skewed. There are no outliers present.

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

Ans. The data is said to be symmetric or have absence of skewness.

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

Ans. The data is said to be positively skewed.

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

Ans. The data is said to be negatively skewed.

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

Ans. The distribution curve is more peaked than normal bell-shaped curve.

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

Ans. The distribution curve is flattered than normal bell-shaped curve.

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



What can we say about the distribution of the data?

Ans- The data is not symmetrically distributed. More data points having higher values of observations.

What is nature of skewness of the data?

Ans- The data is negatively skewed.

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

Ans.- Inter Quartile Range is

IQR =Q3-Q1 =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.- The data points in Boxplot 1 are mostly having central values of observation, In Boxplot 2 , data points are distributed in wide range.

IQR of Boxplot 2 is much higher than Boxplot 1.

Both Boxplot have same median and symmetrically distributed data points.

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.

1. P(MPG>38) =0.3476
2. P(MPG<40) =0.7293
3. P (20<MPG<50) =0.8989

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans- From histogram, boxplot and normal probability plot, MPG of Cars does not follow 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.- From histogram, boxplot and normal probability plot, Waist Circumference and Adipose Tissue do not follow Normal Distribution.

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

Ans- z score for 90% confidence interval = -1.6449

z score for 94% confidence interval = -1.8808

z score for 60% confidence interval = -0.8416

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

Ans- t score for 95% confidence interval = -2.0638

t score for 96% confidence interval = -2.1715

t score for 99% confidence interval = -2.7969

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.

Ans-Required Probability is 0.321672.