|  |  |
| --- | --- |
| 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) | Nominal |

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

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

A={HHT,HTH,TTH}

=3/8=0.375

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

1. Equal to 1

Not possible =0/36

1. Less than or equal to 4

A=(1,3)(2,2)(3,1) 3/36=0.833

1. Sum is divisible by 2 and 3

6/36=0.17

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?

S=7c2=21

D=5c2=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

Sum of (candidates\*prob)

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

=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 | Scoer | Weight |
| Mean | 3.59 | 3.21 | 17.84 |
| Median | 3.69 | 3.32 | 16.46 |
| Mode | 3.92 | 3.44 | 17.02 |
| Variance | 0.28 | 0.55 | 3.14 |
| Std.Deviation | 0.53 | 0.97 | 1.78 |
| Range |  |  |  |

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?

Each=1/9

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

=(1/9)(1308)

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Skew Kurt

speed = -0.117510 speed =-0.5089

distance =0.806895 distance =0.4050

dtpype : flot64 dtpype : flot64

**SP and Weight(WT)**

**Use Q9\_b.csv**

Skew Kurt

SP =1.6114 SP =2.977

WT =-0.614 WT =0.950

dtpype : flot64 dtpype : flot64

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



Right skewed

More than 50% chick weight is between 50 to 150

Most of chick weight is between 50 to 100



Right skewed

Outliers at upper side

**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?

X= mean

Std. deviation = 0.6708

94 =1.882

98 =2.326

96 =2.053

a)200(+/-)1.882\*0.6780=200(+/-)1.262

b)200(+/-)2.326\*0.6708=200(+/-)1.560

c)200(+/-)2.053\*0.6708=200(+/-)1.377

**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.

Mean=41 , median=40, variance=24.11, std deviation=4.9

1. What can we say about the student marks?

Chart, histogram

Description automatically generated

Avg. percentage 65

Only few students above 90%

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

Ans. Has zero skewness

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

Ans. Positive skewed

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

Ans. Negative skewed

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

Ans. Positive excess values of kurtosis indicate that a distribution

Is packed and possess thick tails. A leptokurtic distribution has a higher peak and taller tails than a normal distribution

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

Ans. Negative excess values of kurtosis indicate that a distribution is flat and has thin tails. A platykurtic distribution is flatter when compared with the normal distribution with fewer values in its shorter.

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



What can we say about the distribution of the data?

Ans. Data is distributed not normally distributed in left hand side between 10 to 18

What is nature of skewness of the data?

Ans. Negatively skewed

What will be the IQR of the data (approximately)?   
 Ans. IQR= Q3-Q1=18-10=8

Q19) Comment on the below Boxplot visualizations?



mean for small boxplot 1 and

mean of large boxplot 2 is at

approximately at 250

Both 1 and 2 are normally distributed

No outlier

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

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. 2P(MPG>38)

= mean(MPG)=34,42208

= sd(MPG) 9.131445

= 1- pnorm(38,mean (MPG), sd(MPG))

= 0.330

= 33%

Σ

P(MPG<40)

=pnorm(40,mean(MPG),sd (MPG))

=0.7293499

=72.3%

P (20<MPG<50)

=pnorm(50,mean(MPG), sd (MPG)) - pnorm(20,mean(MPG), sd(MPG))

= 0.955-0.057 =0.8988689

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

 Ans: MPG of cars 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: Adipose Tissue (AT) and Waist does not follow Normal Distribution

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

Z scores

=90%

= 95+2.5

=97.5

=qnorm(0.975) = 1.96

94%

Σ

= 94+4

=97

=qnorm(0.97) = 1.88

60%

= 60 + 20

= 80

= qnorm(0.80) = 0.841

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

TSCORE CALCULATION

T((1,alpha),(n-1))

Here n = 25

n-1=24

Hence t score values will be:

95%

= qt(0.975,24)

= 2.063899

96%

= qt(0.98,24)

= 2.171545

99%

= qt(0.995,24)

= 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

Sample size =18=n

Sample mean =260=x

Sample std. deviation=s=90days

= 260-270/90/SQRT(18)

=-10/9.487

=-1.054