Q1) Identify the Data type for the Following:

|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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

Nominal, Ordinal, Interval, Ratio.

|  |  |
| --- | --- |
| Data | Data Type |
| Gender |  |
| High School Class Ranking | ONominalrdinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nomainal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation |  |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| 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

**Ans=** The total possibility outcomes 23=8

HHH,HHT,HTT,THT,TTH,HTH,TTT

Number of favorable outcomes are 3

P(two heads & one tail)=3/8

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

* Equal to 1

=zero, not any outcome will be equal to 1

* Less than or equal to 4
* =(1,3)(2,2)(3,1)

=3/36

=1/12

* Sum is divisible by 2 and 3

**Ans** =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?

**Ans** =jp(2Red, 3Green,2Blue)

=p(5/7,4/6)

=p(20/24)

=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** =Summationof all prababilities

=(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**

**Ans= A)Points:**

Mean=3.596563 , Median= 3.695, Mode=

Variance=0.2858814, Standed Deviation= 0.5346787

**B) Score:**

Mean=3.21725 , Median= 3.325, Mode=

Variance=0.957379, Standed Deviation= 0.9784574

**C)Weight:**

Mean=17.84875, Median= 17.71, Mode=

Variance=3.193166, Standed Deviation= 1.78694

Mean values of Points and Score are closer

Q8) Calculate Expected Value for the problem below

* 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

= ∑ ( probability \* Value )

∑ P(x).E(x)

there are 9 patients

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 1/9 1/9 1/9 1/9 1/9

Expected Value = (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**

**Ans=**

**A)Speed:**

1)Skewness =-0.1139548

2)Kurtosis =2.422853

Skewness for speed= -0.1139548, skewness value is negative so it is left skewed. Since magnitude is slightly greater than 0 it is slightly left skewed

**B)Distance:**

1)Skewness =0.7824835

2)Kurtosis =3.248019

Skewnessfor distance= 0.7824835, right skewed (Positive) slight magnitude to right.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**A)SP:**

1)Skewness =1.5814537

2)Kurtosis =5.723521

Skewnessfor SP = 1.5814537, right skewed (Positive) magnitude to right.

**b)Weight:**

1)Skewness =-0.6033099

2)Kurtosis =3.819466

Skewnessfor Weight= 0.6033099, right skewed (Positive) slight magnitude to right.

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



**Ans=** The most of the data points are concerated in the range 50-100 with frequency 200.

And range of weight is 0 to 400 .

So the expected value the above distribution is 75.

Skewness-we can notice a long tail towards right so it is heavily right skewed



**Ans=** Medican 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 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?

**Ans=** Degrees of freedom= 2000-1= 1999

Confidence interval= 94%

=(1-σ/2)= 1-0.03)

=0.97 for

Confidene interval for 94% =(198.5432,201.7295)

Confidence interval for 98%= (198.2704,201.7295)

Confidence interval for 96% = (198.4381,201.5618)

**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**

* Find mean, median, variance, standard deviation.

**Ans=** Mean:41

Median:40.50

Variance:25.52941

Standard Deviation:5.052664

* What can we say about the student marks?

**Ans=**

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

**Ans=**Symetrical

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

**Ans=** Positivley Skewed

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

**Ans=** Negatively Skewed

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

**Ans=** The data is normally distributed and kurtosis value is 0

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

**Ans=** Negative Kurtosis implies wider peak and thinner tail

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



* What can we say about the distribution of the data?

**Ans=** The 25% value is 10, 50% value is 15 and 75% value is 18.

* What is nature of skewness of the data?

**Ans=** Left Skewed, medain is greater than mean

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

**Ans=**Approximately 8

Q19) Comment on the below Boxplot visualizations?



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

**Ans=** From the above boxplot, Whishker's is higher of boxplot 2.

Mean and Median are likely to same , so the distribution is symetrical

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

* P(MPG>38)
* P(MPG<40)
* P (20<MPG<50)

**Ans=**  a)There are 33 observations in MPG which are greater than 38

b)There are 61 observations in MPG which are less than 40

c)There are 69 observations in MPG which are greater than 20 and less than 50

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans=** No, its not follow Normal distribution

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

Dataset: wc-at.csv

**Ans=** No,both the Adipose tissue and Waist Circumference does not follow normal distribution

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

**Ans=** Z score of 90% confidence interval is 1.65

Z score of 94% confidence interval is 1.55

Z score of 60% confidence interval is 0.85

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

**Ans=** For 95%= 2.06

For 96%= 2.17

For 99% = 2.8

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=**

t\_value=-0.4714

p\_value=0.3216

hypothesis test= Failed to reject the Null Hypothesis,So the Average life of a bulb is more than 260 days.