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

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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans: 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 2and 3

Ans: a) P(Sum=1) = 0

b)P(Sum<=4) = 0.166

c)P(Sum divisible by 2&3) =0.166

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: 0.476

Q6) The Expected No. of Candies for a random Child = 3.09

Q7)

1) Points

1. Mean = 3.596
2. Median = 3.695
3. Mode = 3.07, 3.92
4. Variance = 0.286
5. Standard Variation= 0.535
6. Range : Max= 4.93, Min= 2.76

2)Score

1. a) Mean = 3.21
2. Median = 3.32
3. Mode = 3.44
4. Variance = 0.957
5. Standard Variation= 0.978
6. Range : Max= 5.424, Min= 1.513

2)Weigh

1. a) Mean = 17.85
2. Median = 17.71
3. Mode = 17.02, 18.9
4. Variance = 3.19
5. Standard Variation= 1.78
6. Range : Max= 22.9, Min= 14.5

Inference

For all data sets, Mean=Median. So all them will be Normally Distributed.

Q8)

Q9)

1)Speed and Distance

a) Skewness of Speed = -0.117

b) Kurtosis of Speed = -0.508

c) Skewness of Distance= 0.806

d) Kurtosis of Distance= 0.405

2) SP and Weight

a) Skewness of SP= 1.611

b) Kurtosis of SP= 2.977

c) Skewness of Weight= -0.614

d) Kurtosis of Weight= 0.950

Skewness of Speed is Negative, ie it is Left Skewed and all the datas are concentrated on the right side.

Kurtosis of Speed is Negative, ie the curve has a flat peak.

Skewness and Kurtosis for Distance are positive. So it is Right Skewed and has a sharp peak.

Skewness and Kurtosis for SP are positive. So it is Right Skewed and has a sharp peak.

Skewness for Weight is Negative. So it is Left Skewed.

Kurtosis for Weight is positive. So it is has a sharp peak curve.

Q10) By looking at the Histogram we can infer that the is dataset is Right Skewed or the skewness is positive and that the datas are concentrated towards the right.

From the Boxplot we can conclude that more datas are towards Upper Quartile or Q3 and there are more outliers on the Q3 side than on Q1 side.

Q10)

From the histogram and box plot, we can infer that the data is positively skewed with more no of data points concentrated on the left side. Also w can infer that there are a lot of outliers towards the right side.

Q11)

Sample size, n=2000

Population size, N=3000000

Sample Average, x= 200

Sample Standard Deviation, s= 30

Since population std deviation is not given, we have to use t-test.

a) At 94% CI, t94,1999 = 1.75

Average Weight = [201.17, 198.82]

b) At 98% CI, t98,1999 = 2.32

Average Weight = [201.55, 198.44]

c) At 96% CI, t96,1999 = 2.05

Average Weight = [201.38, 198.62]

Q12) Mean = 41

Median = 40.5

Variance= 25.52

Standard Variation= 5.05

Q13) When mean=median, skewness=0 or a perfectly symmetric distribution.

Q14) When mean>median, skewness=positive or a positive skewed distribution

Q15) When median>mean, skewness=neghative or a negative skewed distribution

Q16) Positive kurtosis indicates a heavy-tailed distribution, ie. most of the datas are concentrated on the tails.

Q17) Negative kurtosis indicates a thin-tailed distribution, ie. the datas are almost equally distributed.

Q18)

a) From the box plot, we can infer that more data is spread out towards the Lower Quartile (Q1) region.

b) Negative or Left Skewed

c) IQR = 8

Q19) In Boxplot 1, the datas are more concentrated towards the median than Boxplot 2, while data in Boxplot are spread more. Also the minimum, maximum, IQR for Box plot 1 is smaller than Box plot 2.

Q20)

1. P(MPG>38) = 1-0.65=0.35=35%
2. P(MPG<40)=0.73=73%
3. P(20<MPG<50)=0.89=89%

Q21)

a) From Shapiro’s test,

p-value of MPG=0.176

since p-value>0.05, w can conclude that MPG **follows** a Normal Distribution.

b) From Shapiro’s test,

p-value of Waist=.0017, since p<0.05, Waist **doesn’t follow** Normal Distribution

p-value 0f AT=0.0063<.05, Adipose Tissue doesn’t follow Normal Distribution

Q22)

1. For 90% CI, z-value=1.645
2. For 94% CI, z-value=1.88
3. For 60% CI, z-value= 0.84

Q23)

1. For 95% CI and n=25, t-value=2.06
2. For 96% CI and n=25, t-value=2.17
3. For 99% CI and n=25, t-value=2.79

Q24)

t=x-µ/(s/√n)= -0.471

P(X<260)=.349=35%

Probabiity = 35%