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 | Nominal |
| 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. 3/8

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) }

a) Let E be the event "sum equal to 1". There are no outcomes which correspond to a sum equal to 1, hence    
P(E) = n(E) / n(S) = 0 / 36 = 0

b) P(E)=6/36= 1/6

c) P(E)= (18 / 36) and (12/36)   
 = ½, 1/4

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 chances to draw ***2*** balls at random from ***7*** colored balls, **n(s)=7C2=21**  
let**E** be an event to draw ***2***balls other than **blue**.  
no of chances to draw two balls other than blue are, **n(E)=2C2+3C2+2C1.3C1=1+3+6=10**  
the probability that none of balls drawn is blue is ,p(E) = n(E)/n(S) = 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 value = X\*P(X)

X = candies count

P(X) = probability

Therefore expected value = (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.

Ans:

|  |  |  |  |
| --- | --- | --- | --- |
|  | points | Score | Weight |
| Mean | 3.59 | 3.21 | 17.84 |
| Median | 3.695 | 3.325 | 17.71 |
| Mode |  |  |  |
| Variance | 0.28 | 0.957 | 3.19 |
| Std deviation | 0.534 | 0.97 | 1.78 |

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. E(X) = sum(X)\*P(X)

= 1308\*(1/9)

= 143.88

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

**Cars speed and distance**

**Ans.: Speed and distance table**:

Skewness for speed is -0.844

When you see the skewness in histogram for speed and it is negatively skewed i.e, mean and median is not same for the speed

Skewness for distance is 1.217917

When you see the skewness in histogram for distance and it is positively skewed i.e., mean and median is not same for the distance

**Speed and weight:**

Skewness for speed is -0.407

When you see the skewness in histogram for speed and it is negatively skewed i.e, mean and median is not same for the speed

Skewness for weight is 1.287

When you see the skewness in histogram for speed and it is negatively skewed i.e, mean and median is not same for the speed

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



Ans. By seeing the boxplot we can tell that

* It is right skewed



Ans. By looking at boxplot we can say

* That there are outliers in the boxplot

**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. Mean (x) = 200

Std deviation = 30

Sample size (n) = 2000

Confidence interval @ 94% is (201.04, 198.95)

Confidence interval @ 98% is (201.38, 198.61)

Confidence interval @ 96% is (201.18, 198.81)

**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. i) mean = 41

Median =40.5

Mode = 41

Variance = 25.52

Standard deviation = 5.05

ii) The maximum no of students obtained marks in the range of 35 to 41.

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

Ans. When the values ofmean, median and mode are equal, there is no skewness.

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

Ans. When mean > median then the skewness will be positively skewed

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

Ans. When median > mean then the skewness will be negatively skewed

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

Ans. A distribution with a positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than the normal distribution.

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

Ans. A distribution with a negative kurtosis value indicates that the distribution has lighter tails and a flatter peak than the normal distribution.

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



(i)What can we say about the distribution of the data?

Ans. The data is not normally distributed, because the median is at the end of the data i.e., 15.5.

(ii) What is nature of skewness of the data?

Ans. The nature of skewness of the data is negatively skewed.

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

Ans. 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.

**Boxplot1**:

Upper quartile (Q3) = 275

Lower quartile (Q1) = 250

IQR = Q3-Q1 => 275-250=> 25

Median = 262.5

**Upper fence** = Q3+1.5(IQR)

= 275+1.5(25)

= 312.5

**Lower fence** = Q1-1.5(IQR)

= 250-1.5(25) => 215.5

**Boxplot2**:

Upper quartile (Q3) = 312.5

Lower quartile (Q1) = 225

IQR = Q3-Q1 => 312.5-225=> 87.5

Median = 262.5

**Upper fence** = Q3+1.5(IQR)

= 312.5+1.5(87.5)

= 443.75

**Lower fence** = Q1-1.5(IQR)

= 225-1.5(87.5)

= 93.75

**Inferences**: Here the averages or median of the two boxplots are same but the distribution of the data is different from each other.

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)

Ans. 0.65

* 1. P(MPG<40)

Ans.

c. P (20<MPG<50)

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 is not normally distributed but it is negatively skewed of -0.17, so the MPG of cars data is left skewed.

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

Dataset: wc-at.csv

Ans. AT and waist circumference are not following normal distribution, waist is positively skewed with skewness value as 0.13 and it is rightly skewed.

AT is also positively skewed with skewness value as 0.57 and is rightly skewed.

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

Ans. Z score at 90% confidence interval is 1.65

Z score at 94% confidence interval is 1.89

Z score at 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. . t score at 95% confidence interval is 2.06

t score at 96% confidence interval is 2.34

t score at 99% confidence interval is 2.79

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. Here we have given that,

Population man = mu= 270

n=18

x bar = 260

s = 90

t = (xbar-mu)/(s/sqrt(18))

t= -0.47

Required probability = 0.322

>pt(-0.47,df=17)

[1] 0.3221639