|  |  |
| --- | --- |
| 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) | 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 | 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: Total number of outcomes possible are 8

P(2Head and 1 Tail)= 3/8 & 3/8 = 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:

1. When two dice are rolled the total number of possible outcomes are 36

When 2 dice are rolled the minimum sum possible is 2.

P(Sum Equal to 1=0)

1. P(sum<=4)= p(sum=3), p(sum=2), p(sum=4)

P(sum=2) There is only 1 way out of 36 to get the sum of 2: (1+1)

P(sum=3) There is only 2 ways out of 36 to get the sum of 3: (2+1) or (1+2)

P(sum=4) There is only 3 ways out of 36 to get the sum of 4: (2+2), (1+3) or (3+1)

1. P (sum is divisible by 2 and 3) = P(sum=6)+ P(sum=12)

P (sum=6) There are 5 ways out of 36 to get a sum of 6 =(1+5, 5+1, 2+4, 4+2, 3+3) = 5/36

P(sum=12) There is only one way out of 36 to get a sum of 12 = (6+6) =1/36

P(sum divisible by 2 and 3)= 5/36+ 1/36 = 6/35 = 0.16

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 number of the balls are 7.

5c2/7c2 = 10/21= 0.476

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, 0.24=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:**

**POINTS: Mean= 3.597**

**Median= 3.695**

**Mode= 3.92,3.07**

**Variance= 0.285**

**Range= 4.93 - 2.76=2.17**

**Std= 0.534**

**The difference between mean median and mode are very less so we can say that data is normally distributed.**

**Also the variance and standard deviation are less we can say that the data is symmetric around the mean.**

**SCORE:**

**Mean:3.217**

**Median:3.325**

**Mode:3.44**

**Range= 5.424-1.513= 3.911**

**Std=0.978**

**Var=0.957**

**The difference between the mean median and mode is very less so we can conclude that the data is normally distributed. The standard deviation and variance are less so we can conclude that most of the data lies symmetric around the mean.**

**WEIGH:**

**Mean= 17.848**

**Median= 17.71**

**Mode= 17.02**

**Std= 1.786**

**Variance= 3.193**

**Range= 22.9-14.5= 8.399**

**The difference between the mean median and mode are very less so we can conclude the data is normally distributed. The variance and standard deviation are away from the mean.**

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: P(x) =108+110+123+134+135+145+167+187+199/9 = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans: Skewness of speed = -0.11055**

**Kurtosis of speed = -0.673092**

**Skewness of distance = 0.7591268**

**Kurtosis of distance = 0.1193971**

1. **The value of skewness of speed is negative so we can conclude that the data is left skewed. The value of distance is positive so we can say the data is right skewed.**
2. **The value of speed in kurtosis is negative so we can say it is negative kurtosis with lighter tails and the value of distance in kurtosis is positive value so we can say it is positive kurtosis with heavy tails.**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans: Skewness of SP & WT = 1.581454, -0.6033099**

**Kurtosis of SP & WT = 5.723521, 3.819466**

1. **The skewness of SP is 1.58 that means it is right skewed and has outliers. The skewness of weight is -0.60 that means its moderately left skewed.**
2. **The kurtosis of SP is 5.72 which is high value so we can that the data tends to have outliers and heavy tails. The kurtosis of WT is 3.819 which also means it has outliers and heavy tails.**

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



Ans: Through the histogram we can understand the data is not evenly distributed most of the data is on the right side, So we can conclude this is right skewed data.

The chicken weight of 200 to 400 is less in frequency so we can conclude it as an outliers and also confirm with 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: n= 2000, population= 3000000, Average= 200, Sd=30

Since there is no standard deviation of the population we should go

With t value.

At 94% T value=1.88

CI= Sample mean+-t\*s/

200-1.88\*30/, 200+1.88\*30

=198.738, 201.261

At 98% t value is= 2.32

CI= 200-2.32\*30/ 200+2.32\*30/

=198.443, 201.556

At 96% t value is= 2.05

CI= 200-2.05\*30/ 200+2.05\*30

=198.610, 201.389

**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:

1. Mean=41

Median=40.5

Variance=25.52941

Standard deviation=5.052664

1. The average marks of the students are 41, the mean and median is very close we can say the data is almost symmetric. Only very few students got high marks.

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

Ans: The distribution will be symmetrical.

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

Ans: When the mean is greater than the median the distribution will be positively skewed.

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

Ans: When the median is greater than mean it will negatively skewed.

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

Ans: Positive kurtosis indicates peakedness of distribution with heavy tails.

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

Ans: Negative kurtosis indicates flat distribution with light tails.

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



What can we say about the distribution of the data?

Ans: From this boxplot we can understand the data is not symmetrical, the upper quartile and lower quartile are not equal. As most of the data is on lower quartile we can say this is a negatively skewed data.

What is nature of skewness of the data?

Ans: The data is left skewed because most the data lies in the lower quartile.

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

Ans: IQR = Upper quartile – Lower quartile = 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: In the boxplot one the upper quartile and lower quartile are not equal. The median for the both the boxplots are nearly equal. The boxplot 2 is normally distributed and since the boxplot one is not equal it will be a bit skewed.

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)

Answer:

1. P= 0.34748
2. P= 0.72945
3. P= 0.01316

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Mean= 34.42

Median= 35.12

Mode= 29.62

Skewness= - 0.1746343

The distribution seems to be normal.

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

Dataset: wc-at.csv

Ans: 1) ADIPOSE TISSUE

Mean:101.894

Median:96.54

Mode:121

Skewness: 0.576789

The mean, median and mode is somewhat close so we can say the distribution is symmetrical

WAIST CIRCUMFERENCE

Mean: 90.90183

Median: 90.8

Mode: 94.5

Skewness: 0.1322042

The mean, median and mode is almost the same so we conclude the distribution is symmetrical. The skewness is also

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

Ans: Z score of 90%= 1.64

Z score of 94%= 1.88

Z score of 60%= 0.84

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

Ans: T score of 95%= 2.064

T score 0f 96%= 2.172

T score of 99%= 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:

µ=270

x bar= 260

n= 18

s= 90

t= x-u/s

t= 260-270/90 =-0.47

Probability= 0.3221