|  |  |
| --- | --- |
| 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 | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Sol – P(H H T) + P(H T H) + P(T H H)

=1/8 + 1/8 + 1/8

= 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

Sol-

1. There is no outcome that whose addition will be one so the probability

Is 0/36, zero probability

1. (1,3) (2,2) (3,1),(1,1),(2,1),(1,2) = 3 outcomes 6/36 = 1/6
2. (1,5) (2,4) (3,3) (4,2) (5,1) (6,6) = 6 outcomes 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?

Sol-

P(2R,3G,2B)

P(5/7)/P(5/3) =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

Sol- Child A – probability of having one candy = 0.015

Child B - probability of having one candy = 0.20

Sol – Expected number of candies for randomly selected child =

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**

Sol –



**Points:** Mean = 3.596563, Median = 3.695, Mode = 3.07, 3.92

Variance = 0.2858814, Standard Deviation = 0.5346787, Range = 2.76 4.93

**Score:** Mean = 3.21725, Median = 3.325, Mode = 3.44

Variance = 0.957379, Standard Deviation = 0.9784574, Range = 1.513 5.424

**Weigh:** Mean = 17.84875, Median = 17.71, , Mode =17.02 , 18.90

Variance = 3.193166, Standard Deviation = 1.786943, Range = 14.5 22.9

**comment :** 1) **“Points”** and “**Score”** these two columns have mean and median close to each other but for “**Weigh”** it’s slightly different.

2) **“Points”** and “**Weigh”** are Bimodal

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?

Sol-

Probability of selecting each patient = 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

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

**Cars speed and distance**

**Use Q9\_a.csv**

Sol- 

-**Skewness** for **“speed”** = -0.11, skewness value is negative, so it is left skewed.

-**Kurtosis for “speed”** = -0.508, kurtosis value is less than Normal Kurtosis that is 3.

- **Skewness** for **“Distance”** = 0.80, skewness value is positive, so it is right skewed.

-**Kurtosis for “Distance”** = 0.405, kurtosis value is nearly equal to Normal Kurtosis that is 3.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Sol- **

**-**-**Skewness** for **“SP”** = 1.611, skewness value is positive, so it is right skewed.

-**Kurtosis for “SP”** = 2.97, kurtosis value is close to Normal Kurtosis that is 3, so SP is normal .

- **Skewness** for **“Weight”** = -0.61, skewness value is negative, so it is left skewed.

-**Kurtosis for “Weight”** = 0.95, kurtosis value is less than Normal Kurtosis that is 3.

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



Sol-

1. The most of datapoints are in the range 50-100 with frequency 200
2. And least range of weight is 400 some were around 0-10
3. Skewness- we can notice a long tail towards right so it is heavily right skewed.



1. The data is right skewed
2. There are 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?

Sol-

The **information given**is: Sample meanof X 200

Sample standard deviation of S =30

Sample size n = 2000

X+/-(t1\*S/sqrt(n))

Degree of freedom = 2000-1 = 1999

Considering a **94%** confidence level

**T1 = 1.56**

**= 200+/-(1.56\*30/ sqrt of 2000)**

**= (197.06, 202.93)**

Considering a **96%** confidence level

T1 = 1.75

= **200+/-(1.75\*30/ sqrt of 2000)**

**= (196.40, 203.60)**

Considering a **98%** confidence level

T1 = 2.05

**= 200+/-(2.05\*30/ sqrt of 2000)**

**=(195.22, 205.77)**

**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.

Sol- Mean = 41, Median = 40.5, Variance = 25.52941

Standard Deviation = 5.052664

1. What can we say about the student marks?

Sol – we can say that mean and median are closer to each other, seems no outlier present.

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

Sol- symmetrical, equally oriented

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

Sol- Right Skewed, +ve Value

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

Sol- Left Skewed , -ve Value

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

Sol- The data is sharper and has long tails

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

Sol- The data is not Sharper and under the normal distribution with short tail.

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



What can we say about the distribution of the data?

Sol-

most of observations are having value above 10

Observations whose value is above 15 are 40%

What is nature of skewness of the data?

Sol- Left skewed, median is greater than mean.

What will be the IQR of the data (approximately)?   
Sol -   
approximately= 8

Q19) Comment on the below Boxplot visualizations?



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

Sol- By observing both the plots the whiskers level are same from the Q1 and Q3 to their respective upper and lower limit and again mean and median are also equal so we can say distribution is symmetrical

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.

Sol-



MPG <- Cars$MPG

* 1. P(MPG>38)

0.34

Sol- There are 33 observations in MPG which are greater than 38

* 1. P(MPG<40)

0.72

Sol- 61 observations in MPG which are lesser than 40

* 1. P (20<MPG<50)

-0.013

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

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

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

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