|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuos |
| Weight of Gold | Continuos |
| Distance between two places | Both |
| Length of a leaf | Both |
| Dog's weight | Continuos |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ordinal |

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

Probability is : 3/8

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

1. Equal to 1 : 0 (bcoz the least sum is 2)
2. Less than or equal to 4 : 1/6
3. Sum is divisible by 2 and 3 : 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?

N(s)=7c2=42/2=21

N(e) = 5c2 =20/2=10

P=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

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.

|  | **Points** | **Score** | **Weigh** |
| --- | --- | --- | --- |
| **mean** | 3.596563 | 3.217250 | 17.848750 |
| **std** | 0.534679 | 0.978457 | 1.786943 |
| **mode** | Points | Score | Weigh |
| **0** | 3.07 | 3.44 | 17.02 |
| **1** | 3.92 | NaN | 18.90 |

Range:

Points 2.170

Score 3.911

Weigh 8.400

**MEDIAN:**

Points 3.695

Score 3.325

Weigh 17.710

**VARIANCE:**

Points 0.285881

Score 0.957379

Weigh 3.193166

1)we calculated the mean median mode standard deviation and variance for

the 3 columns and mean and median are closer to each other.

2)Mode of score is uni-modal,for points and weigh it is Bi-modal.

3)Range is calculated by subtracting max and min values and displayed the

Results.

4)There is high standard deviation and variance for weigh column whe compa

Red to points and score columns

**Use Q7.csv file**

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?

145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SKEWNESS:**

speed -0.117510 1)For speed column the skewness is negative or slight

dist 0.806895 left skewed.

2)For dist column the skewness is positive or slight

Right skewed.

**KURTOSIS:**speed -0.508994 1)For speed column kurtosis is negative which is not

dist 0.405053 good.

2)For dist column Kurtosis is positive which is good.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**SKEWNESS:**

SP 1.611450

WT -0.614753

1)For sp column the skewness is positive or slight right skewed.

2)For WT column the skewness is negative or slight left skewed.

**KURTOSIS:**SP 2.977329

WT 0.950291

1)For sp column kurtosis is positive which is good.

2)For WT column Kurtosis is positive which is not as good as WT.

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



1)The median lies in the 75 for the chickweight.

2)The data is right skewed.



1)The box plot has outliers above the Max value.

2)The Middle line in the IQR denotes median of the data.

**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?

**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?

1) Mean 41.000000

Std 5.052664

Median 40.5

Var 25.529412

2)The most no of student scoring lies in 40-41 by mean and median.

Max marks:56.00 Min marks:34.00

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

Normally distributed or symmetric.

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

Positive skewed.

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

Negative skewd

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

Sharp curve for positive Kurtosis and indicates very good results

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

Flat curve for negative Kurtosis and indicates bad results.

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



What can we say about the distribution of the data?

Median lies in 15

Lower quartile is 1

Upper quartile is 20

What is nature of skewness of the data?

Negative or left skewed

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

Q19) Comment on the below Boxplot visualizations?



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

1)The median of boxplot 1 is 267.5 where as for 2nd plot it is 260 approx.

2)Boxplot 2 is having higher iqr than 1st boxplot

3)Boxplot 1: lower quartile:240,upper quartile:280

4)Boxplot 2: lower quartile:200,upper quartile:350

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)

0.34759393

* 1. P(MPG<40)

0.72934988

* 1. P (20<MPG<50)

0.89886892

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

No it is not normal distributed having a positive skewness of 9.131

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

Dataset: wc-at.csv

No it is not normally distributed

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

90% CI = 1.644

94% CI = 1.880

60% CI = 0.8416

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

95% t score CI= 2.0638

96% t score CI= 2.1715

99% t score CI= 2.7969

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

**0.3218**