|  |  |
| --- | --- |
| 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 |  |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation |  |
| Level of Agreement |  |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores |  |
| Years of Education | Ordinal |

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

**Ans** :- If we tossed 3 coin total result is 2^3 i.e.

TTT,HHH,THH,HTT,THT,HTH,TTH,HHT.

So, total probability of getting two heads and one tail are 3 i.e.

THH,HTH,HHT.

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** :- a)Probability of getting outcome equal to1 is 0.

b) Probability of getting outcome less than or equal to 4 is 6 ([(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)]).

c)Probability of getting total outcome whose sum is divisible by 2 and 3 are 24.

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** :- From the Question :-

Desired Outcome/Total Outcome

= 5/7.

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

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 :-**Finding Mean for Q7.csv

Mean

Points 3.596563

Score 3.217250

Weigh 17.848750

Median

Points 3.695

Score 3.325

Weigh 17.710

**Mode**

| **Points** | **Score** | **Weigh** |
| --- | --- | --- |
| **3.07** | 3.44 | 17.02 | 3.44 |  |

Std

Points 0.534679

Score 0.978457

Weigh 1.786943

**Variance**

**Points:** 0.2858813508064516

**Score:** 0.9573789677419355

**Weigh:** 3.1931661290322575

**Range**

**Points = Max-Min = 4.930000-2.760000 = 2.17**

**Score = Max- Min = 5.424000-1.513000 =** 3.9110000000000005

**Weigh = Max-Min = 22.900000-14.500000 =** 8.399999999999999

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:- Expected value of that person is 145.

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans :-**

**Skewness of speed :** -0.11395477012828319

**Skewness of Dist :** 0.7824835173114966

**Kurtosis of Speed :** -0.5771474239437371

**Kurtosis of dist :** 0.24801865717051808

**Inferences :- In data Q9\_csv the skewness of speed is < 0 i.e. it is left skewed**

**and also skewness of dist is > 0 i.e. it is right skewed and kurtosis is < 3 i.e. it has a low amount of data.**

**Here we can say that data is approximately normally distributed.**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans : -**

**Skewness of SP :** 1.5814536794423764

**Skewness of Weight :** -0.6033099322115126

**Kurtosis of SP :** 2.7235214865269244

**Kurtosis of Weight :** 0.8194658792266849

**Inferences :**

**It is right skewed and here data is not normally distributed we can see here the kurtosis is less than 3 i.e. here we expect low kurtosis and skewness is high .so, it impacts the stability and accuracy of our model and also sample data is not collected properly.**

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



We see that the chick weight from 50 to 100’s frequency is 200 and the lower weigt of chick’s frequency is to 10 . In the above histogram we can see that the data is skewed towerds right i.e. more data in the left side so this data shows that it is right skewed which is not good for the accuracy of our model.



According to the above boxplot we see that the data is left skewed i.e. most of the data points lies towerds right direction of the boxplot.And we also see some outliers present in the data set.

**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?

C.I . for 94% is [-2599 to 2999]

C.I . for 96% is [-2659 to 3059]

C.I . for 98% is [-2418 to 3118]

**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.

Mean : 41

Median : 40.0

**Variance** : 25.529411764705884

Std : 5.05266382858645

1. What can we say about the student marks?

Ans :- We can say that the some students scored below 41 marks and some students scored above 41 marks where the variance is 25 i.e. very high.

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

The data we collected is normally distributed.

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

It is right skewed.

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

It is left skewed

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

It shows that the data has a greater tail than the normal distribution.

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

It shows that the data has the lower tail than the normal distribution.

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



What can we say about the distribution of the data?

We can say that the data is skewed towards the right and we cannot see any outliers in the dataset. And the Q1 and Q3 of the data is between 10 to 18 and the median is approximately 17.

What is nature of skewness of the data?

It is right skewed.

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

IQR =(Q3 -Q1)  
IQR = 8  
  
Q19) Comment on the below Boxplot visualizations?



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

Ans :- Here we can see that the spread of boxplot 1 is from approx. 237.5 to 287.5 where the spread of datapoints for boxplot 2 is from 200 to 350.Both of the plots are normally distributed so we can say that the collected data is very good. We cannot see any outlier for the both data set. Due to the normal distribution the kurtosis and skewness is equal to 0 i.e. it cannot impact our model’s accuracy and stability.

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 :- Probability of MPG < 38 is 67%

* 1. P(MPG<40)

Ans :- Probability of MPG <40 is 74 %

* 1. P (20<MPG<50)
  2. Ans :- Probability of MPG <20 and > is 90%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans :- The standard deviation for MPG is 9.131445 which is high so we can say that the MPG of cars do not follow Normal Distribution.

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

Dataset: wc-at.csv

Ans :- The std of Waist is **13.559115982678826** and the std of AT is **57.29476272231215** which is very high. So, we can say that the waist is normally distributed with respect to AT.

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

Ans :- Z score for 90% C.I is 1.645

Z score for 94% C.I is 2.05

Z score for 60% C.I is 0.08

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

Ans :- t score for 95% C.I is 1.710

t score for 96% C.I is 1.828

t score for 99% C.I is 2.492

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 -: The probability of selected bulbs having average life no more than 260 days is 4%.