|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Data |
| Results of rolling a dice | Discrete Data |
| Weight of a person | Continuous Data |
| Weight of Gold | Continuous Data |
| Distance between two places | Continuous Data |
| Length of a leaf | Continuous Data |
| Dog's weight | Continuous Data |
| Blue Color | Discrete Data |
| Number of kids | Discrete Data |
| Number of tickets in Indian railways | Discrete Data |
| Number of times married | Discrete Data |
| Gender (Male or Female) | Discrete Data/Qualitative |

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

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

=>Total Outcome

{HHH,HHT,HTH,THH,THT,HTT,TTH,TTT}=8

* Probability =3/8 = 0.375

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

1. Equal to 1 => 0/36=0
2. Less than or equal to 4 => favorable outcomes=6

{(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}

Probability = 6/36 = 1/6

1. Sum is divisible by 2 and 3

=>{(1,1),(1,2),(1,3),(1,5)

(2,1),(2,2),(2,4),(2,6)

(3,1),(3,3),(3,5),(3,6)

(4,2),(4,4),(4,5),(4,6)

(5,1),(5,3),(5,4),(5,5)

(6,2),(6,3),(6,4),(6,6)

} 16/36 =4/9

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?

**= Total balls = 7**

**Total No .of drawing balls (7\*6)/(2\*1)=21**

**Drawing of 2 balls except none of these balls are blue**

**=(5\*4)/(2\*1) = 10**

**Probability = 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

**= 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**

**= 0.015+0.8+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**

**= In above dataset Points has no outliers ,but Weigh and Score has outliers.**

**There is no case has the variable 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?

**= There are 9 patients in clinic**

**Then probability =1/9**

**Expected value = summation of(probability \* value)**

**Expected value = (1/9)(108+110+123+134+135+145+167+187+199)**

**= 145.33**

**Expected value of weight of the patient is 145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**= Distance is positively skewed where as speed Is negatively skewed, thus distance has distribution of datta concentrated on the left whereas WT has distribution on the right.**

**Both distance and speed has positive kurtosis.**

**Use Q9\_b.csv**

**= SP is positively skewed where as WT is negatively skewed .Thus SP has distribution data on the left where as WT has distribution on the right.**

**Both SP and WT has positive kurtosis.**

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



= **The histogram peak has in right skew and tail is on right.we have outliners on the higher side.**

**The boxplot has outliners on the upper scale**

**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?

= The Answer is in Jupyter notebook.

**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=> Maximum number of students scored between 35 – 45 marks**

**The data is not normally distributed**

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

**= zero skewness ,If the distribution is symmetric, then the mean is equal to the median , and the distribution has zero skewness**

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

**= The skewness is positive & the mean will be greater than the median.**

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

**= The skewness is positive, Data distributed to right**

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

**= It indicates that distribution is peaked and possesses thick tails.**

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

**= It indicates the distribution has lighter tails than the lighter distribution**.

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



What can we say about the distribution of the data?

**= The boxplot is not symmetric and towards right side.**

What is nature of skewness of the data?

= negative skewness

What will be the IQR of the data (approximately)?   
= IQR data is (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.

= The data is normally distributed. There are no outliers. The center is around

260 above,First plot is less range .

= The data is normally distributed. There are no outliers. The center is around

260 above ,Second plot ha more range

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)

c. P (20<MPG<50)

= The Answer is in Jupyter notebook.

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

= MPG is in 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

= Both AT and Waist doesn’t follow Normal Distribution.

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

= t=-0.471

P(t)=0.3216

t<-0.471 with 17 degree of freedom

Then the probability of the bulbs lasting less than 260 days on average of 0.3216 assuming the mean life of the bulb is 300 days.

= The Answer is in Jupyter notebook.