|  |  |
| --- | --- |
| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

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) | Ordinal |
| Sales Figures | Ordinal |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Ratio |

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

Ans:Total outcomes=8=(HHH,HHT,HTH,HTT,THH,THT,TTH,TTT)

Favorable outcomes=3(HHT,HTH,THH)

P(2H,1T)=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

Ans:totaloutcomes=36[(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(3,1) ,(3,2),(3,3),(3,4),(3,5),(3,6),(4,1)(4,2)(4,3)(4,4)(4,5),(4,6),(5,1)(5,2)(5,3)(5,4)(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)]

a) P(Sum=1) =0/36

b) P(Sum<=4) =6/36=1/6

c)P (Sum is divisible by 2 and 3) =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?

Ans: Total non-red balls=5, Total Red balls=2, Total Balls=7, Taking Ways =2

Total samples=7C2, Number of favorable outcomes that is no blue ball =5C2

P (No Balls drawn is blue) =5C2/7C2=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

Ans: Expected number of candies for a randomly selected child=**1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=3.09** So We can take it as 3 Candies.

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**

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: Mean=(108+110+ 123+ 134+ 135+ 145+ 167+ 187+ 199)/9=145.333

When we choosing in random -3 TO +3 Sigma are more probability of getting so

Mean we can take so aprox 145.33.But the exact we cant take.Each weight have equal probability of 1/9.

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



Ans:from histogram more numbers around 200 have chick chick weight $ weight 50-100.Its right skewed histogram.

From Boxplot its upper limit is very high and not symmetric.Outliers are more in upper limit.Median is less so that more datas are concentrated in down side or in less values.

**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,Wavg=200,STD sample=30,

n>2000 we can use z or t but std of population not given so we can use t

so t:

for 94% delta=+-1.31299206 so interval(198.687-201.313)

for 98% delta=+-1.65157 so interval(198.348-201.6515)

for 96% delta=+-1.44248 so interval(198.557-201.44248)

**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:mean=41,median=40.5,variance=24.1111,STD=4.910307

Average score of student is 41 ,STD is 4.9 which is less so the marks of students are not more distributed but it concentrated little near to the mean and we can conclude that students are performing good overall.

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

Ans:symmetric distribution

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

Ans:right skewed(positively skewed)

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

Ans:left skewed(negatively skewed)

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

Ans:The datas have high risky in nature.That is the both side of leptokurtosis have higher outliers.In business we frequently avoid liptokurt based on risk management.

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

Ans:Not high risk Data.Both side of platykurtic have flat tails.So the outliers are less.Its good for investment.

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



What can we say about the distribution of the data?

Ans:Right skewed data,And is platykurtic.The datas are not more concentrated about the mean.And the data is not more risky task.

What is nature of skewness of the data?

Ans:Right skewed data

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

Ans: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: boxplot2 have more inter quartile data range compared to data Boxplot 1.But in Boxplot 1 the data are more concentrated and inter quartile range is less. So, the Boxplot 1 is leptokurtic compared to Boxplot 1 and is risky data sets.

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)

Ans:a.34.75%

b.72.935%

c.89.886%

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans:Yes it follows

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

Dataset: wc-at.csv

Ans:YES BOTH FOLLOWS NORMAL DISTRIBUTION CURVE

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

Ans: 90% = (1+ci)/2 = (1+0.9)/2-🡪1.65

94% = (1+0.94)/2🡪1.89

60 % = (1+0.6)/2🡪0.85

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

Ans:CI=95%,DF=25-1=24

@95%=2.064

@96%=2.1715

@99%=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

Ans:n=18

Xavg=260

STD=90

H0🡪<=270

P(18 randomly selected bulbs would have an average life of no more than 260 days )=68.082%