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

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

Ans:- Number of possible outcomes = 23 = 8

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

We have to find the probability of two heads and one tail.

favourable outcomes = {THH, HTH, HHT}

Number of favourable outcomes = 3

Hence required probability is

Probability = Number of favourable outcomes/total number of possible outcomes.

Probability = 3/8

Hence when 3 coins are tossed, the probability of two heads and one tail is 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:- two dice are thrown here n  
n(s)=36  
a) equal to 1=probability is 0 because dice starts from 1  
b) less than or equal to 4=possible combination {((1,3),(2,2),(3,1))} therefore n(b)=3/36=1/12

c) Sum should be divisible by both 2 and 3

outcomes = (1, 5), (3, 3), (4, 2), (5,1), (6,6)

Therefore, Number of outcomes = 5

the probability that sum is divisible by 2 and 3 is 5/36

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 no of balls=2+3+2=7

Lets S be the sample space.

Then, n(S)=No of way drawing 2 balls out of 7=7C2=2

Lets E=Event of drawing 2 balls, none of which is blue.

n(E)=no of ways of drawing 2 balls out of (2+3)balls=5C2=10

therefore P(E)=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

**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 \* 0.12

= 0.015 + 0.8+ 1.95 + 0.025 + 0.06 + 0.24

= 3.090

= 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**

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:-** ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

Ex. 108, 110, 123, 134, 135, 145, 167, 187, 199

P(x) 1/9  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

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

= (1/9) (1308)

= 145.33

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**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



**Ans:-**The most of the data points are concerted in the range 50-100 with frequency 200.And least range of weight is 400 somewhere around 0-10.So the expected value the above distribution is 75.Skewness- we can notice a long tail towards right so it is heavily right skewed

**Ans:-**



Median is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points between Q1 and bottom point

**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:**

X+/-(Z1-α.σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

Confidence interval= 94%

(1-σ/2) = 1-0.03) =0.97

for confidence interval for 94% is 1.882

Confidence interval for 98%= 2.33

Confidence interval for 96% = 2.05

**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:-**

2) Average student marks is 41

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

**Ans:-** skewness is zero

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

**Ans:-** positively or right skewed

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

**Ans:**-Negatively or left skewed

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

**Ans:-** This shows that the distribution has heavier tails than the normal distribution.

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

**Ans:-**this shows that the distribution has lighter tails than the normal

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



What can we say about the distribution of the data?

**Ans:-** The distribution of data on the high end of the scale

What is nature of skewness of the data?

**Ans:-** The distribution is skewed left

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

**Ans:-** The IQR is 18 minus 10

Q19) Comment on the below Boxplot visualizations?



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

**Ans:** By observing both the plots whisker’s level is high in boxplot 2, mean and median are equal hence distribution is symmetrical.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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)

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