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

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

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

Ans- Three coins are tossed, the no of possible combinations are 2^3=8

Combinations are **:** HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.

No of combinations have two heads and one tail are:

HHT, HTH, TTH

There the 3 numbers

So, the probability of 2 heads and 1 tail is 3/8 = 0.375

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-When two dies are rolled then the combinations are

(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)- Equal to 1 = 0

b)- Less than or equal to 4 = 6/36 =1/6

c)- 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- No of balls = (2+3+2) = 7

No of drawing 2 balls out of 7 = 7c2

= (7\*6)/(2\*1)

= 21

No of drawing 2 balls out of by 5 = 5c2

= (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

Ans – Expected no 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

Expected number of candies for a randomly selected child = 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- There are 9 patients.

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

Expected value = ∑ ( Probability \* Value)

=∑ P(x).E(x)………………….(1)

Probability of the patient = 1/9

By using eq 1,

Expected Value = (1/9)(108)+(1/9)(110)+(1/9)(123)+(1/9)(134)+

(1/9)(135)+(1/9)(145)+(1/9)(167)+(1/9)(187)+(1/9)(199)

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

=(1/9)\*(1308)

=145.33

Expected Value of the 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)**

**Use Q9\_b.csv**

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



Ans – In Histogram peak has right skew and tail is on right. We have outliers on the higher side.

In boxplot has outliers on the maximum side.

**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?

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

Ans – Skewness can be positive, negative .

When mean, median of data are equal there is no skewness.

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

Ans – If the mean is greater than the median, the distribution is positively skewed.

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

Ans – Skewness and tail is towards left

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

Ans - Positive values of kurtosis indicates that distribution is peaked and possesses thick tails.

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

Ans – A distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

Ans – The data is not normaly distributed. Median is towards higher value.

What is nature of skewness of the data?

Ans – The data is skewed towards left. The whisker range of minimum is greater than minimum.

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

Ans – The Inner Quantile Range = Q3 -Q1= 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 - Both the boxplots shares the same median i.e. approximately in a range between 275 to 250 and they are normaly distributed with zero to no skewness on either whisker range. Also there are no outliers.

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)

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