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

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 | nominal |
| Level of Agreement | ordinal |
| IQ(Intelligence Scale) | interval |
| Sales Figures | ratio |
| Blood Group | nominal |
| Time Of Day | ordinal |
| Time on a Clock with Hands | interval |
| Number of Children | ratio |
| Religious Preference | nominal |
| Barometer Pressure | interval |
| SAT Scores | interval |
| Years of Education | ratio |

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

{HTH, HHT,THH} =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

ANSWERS:

1. 0
2. 6/36 =1/6
3. 24/36=2/3

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: ncr = n!\r!(n-r)!

=7!/2!(7-2)!

=21

Probability of drawn 2 balls out of 5 balls(2 red,3 green)

5c2 = 5!/2!(5-2)!

= 10

Prob of none of ball drawn blue : No. of ways drawn 2 ball out of 5/No. of ways drawn balls out of 7

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

=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

=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?

Weight of patient 108,110,123,134,135,145,167,187,199

Probability of selecting each patient 1/9

Exp value : 1/9(108+110+123+134+135+145+167+187+199)

=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: Histogram has + ve skewness and above 350 weight there is outlier**

**As shown in box plot there is lot of outliers**



**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?

Ans: mean =41

Median = (40+41)/2=40.5

Variance = 24.11

Std dev = 4.91

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

Ans : if the distribution is symmetric, then mean =median and distribution has zero skewness. If the distribution is both symmetric and unimodal, then the

mean =median=mode

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

Ans : mean mode and median can be used to figure out if you have - ve or + ve skewed distribution. If the mean is less than median, then the distribution is – ve skewed.

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

Ans : -ve skewed

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

Ans : + ve kurtosis value indicates that a distribution is peaked and possess thick tails. An extreme + ve kurtosis indicates a distribution where more of the values are located in the tails of the distribution rather than around the mean.

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

Ans : - ve kurtosis indicates that the 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: Not symmetric and not normally distributed

What is nature of skewness of the data?

Ans : - ve skewness

What will be the IQR of the data (approximately)?   
Ans : 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 : There are not outliers in the box plots

It is normally distributed 0

The mean is approximately 262

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

Ans: 1-α = 95%

α = 5 %

α/2 = 2.5 %

Area to the left = 95+2.5=97.5

Therefore, z score for 0.975 is 1.96

For 94 % confidence interval

1-α = 94%

α = 6 %

α/2 = 3 %

Area to the left = 94+3=97 %

Therefore, z score for 0.97 is 1.88

For 60 % confidence interval

1-α = 60 %

α = 40 %

α/2 = 20 %

Area to the left = 60+20=80 %

Therefore, z score for 0.80 is 0.84

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

Ans: n =25

df = n-1=24

1-α = 95 %

α = 5 %

α/2 = 2.5 %

Area to the left = 95+2.5 =97.5

Therefore, t score for 24 is 2.064

n = 25

df = n-1=24

1-α = 96 %

α = 4 %

α/2 = 2 %

Area to the left = 96+2 =98

Therefore, t score for 24 is 2.492

n =25

df = n-1=24

1-α = 99 %

α = 1 %

α/2 = 0.5 %

Area to the left = 99+0.5=99.5

Therefore, t score for 24 is

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: 1

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom