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

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

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

P(H=2&T=1) =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
4. P=0

b){(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)}=6/36=1/6

c){(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}=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?

Total balls=7(n), 2balls randomly drawn=7!/(2!\*5!) = 7\*6\*5!/(2\*1)\*(5!) =42/2= 21

None of blue = 7-2 = 5(n), 2balls randomly drawn = 5!/(2!\*3!) = 5\*4\*3!/2\*1\*3! = 10

P(None balls drawn blue) = 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

=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**

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?

Total patients =9

Expected value = 1/9\*[108+110+123+134+135+145+167+187+199]

=1/9\*(1308) = 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**



1. It is positively skewed or right skewed, it’s outlier has high value.
2. Mean>Median



1.The boxplot shows outliers on the higher or 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?

1.Import libraries=import numpy as np, from scipy import stats

2.conf94=stats.t.interval(alpha=0.94,df=1999,loc=200,scale=30/np.sqrt(2000))

conf94

[198.73 to 201.26]

3.conf98= stats.t.interval(alpha=0.98, df=1999, loc=200, scale=30/np.sqrt(2000))

Conf98

[198.43 to 201.56]

4.conf96=stats.t.interval(alpha=0.96, df=1999, loc=200, scale=30/np.sqrt(2000))

Conf96

[198.62 to 201.37].

**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?

1. Mean = 41, Median = 40.5, Variance = 25.52, Standard Deviation=5.05.

2. It shows Mean>Median, so the data is positively skewed towards right side, there are no outliers.

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

Ans. When there is Mean = Median, then it shows symmetrical distribution and skewness is absent.

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

Ans. If Mean > Median, then it shows Right or positive skewness.

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

Ans. If Median > Mean, then it shows Left or negative skewness.

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

Ans. Positive kurtosis shows curve will be more peak.

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

Ans. Negative kurtosis shows curve will be flat and no peakness.

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



What can we say about the distribution of the data?

Ans. The boxplot shows the Median has high value, data is not distributed normally.

What is nature of skewness of the data?

Ans. The boxplot show the Left skewness or negatively skewed, whisker range of minimum is more than maximum.

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

Ans. The IQR = Upper Quartile – Lower Quartile = 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. 1. There is no any outliers in both boxplot.

2.There is no skewness in minimum and maximum whisker range.

3.They are normally distributed and have median range between 250 to 275.

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) Z score for 90% C.L. = (alpha=5%=0.05),

Z at alpha = 0.05 from z table = -1.644

2)Z score for 94% C.L. = (alpha=3%=0.03),

Z at alpha = 0.03 from z table = -1.881

3) Z score for 60%C.L. = (alpha=20%=0.2),

Z at alpha =0.2 from z table = -0.841

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

Ans. 1) T-Scores for 95% = -2.063

2) T-Scores for 94% =-1.974

3) T-Scores for 99% =-2.796

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. T-Score = (Mean – pop mean)/[SD/sqrt(Sample size)]

T-Score = -0.471

Prob = 0.32= 32%