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

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) | 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- P(x)=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- a) zero

b) 6/36 = 1/6 =0.167

c) 1/6 = 0.167

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- n(s) = 7c2 = 21

n(E) = 5c2 = 10

P(E) = 10/21 = 0.47

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- x\*P(x) = (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**

ANS- MEAN :-

POINTS = 115.09/32 = 3.59

SCORE = 102.952/32 = 3.22

WEIGH = 571.16/32 = 17.85

MEDIAN :-

POINTS = (3.69+3.7)/2=7.39/2=3.695

SCORE = (3.215+3.435)/2=6.65/2=3.325

WEIGH = (17.6+17.82)/2=35.42/2=17.71

MODE :-

POINTS = 3.92

SCORE = 3.44

WEIGH = 17.02

VARIANCE :-

POINTS = 8.862/32=0.2769

SCORE = 29.678748/32=0.9275

WEIGH = 98.98815/32=3.009338

STANDARD DEVIATION :-

POINTS = 0.526

SCORE = 0.9630

WEIGH = 1.7347

RANGE :-

POINTS = 4.93-2.76= 2.17

SCORE = 5.424-1.513= 3.911

WEIGH = 22.9-14.5= 8.4

COMMENT:- From the above scenario we can say that the MCT and MV of

Points and score are nearly same but Weigh is comparatively high .

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 – The expected value of random =sum of x / total no = (108+110+123+134+135+145+167+187+199)/9 = 1308/9 = 145.34

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

**Cars speed and distance**

**Use Q9\_a.csv**

ANS- Skewness= -0.40769

Kurtosis = 2.087

Left skewed and positive kurtosis

**SP and Weight(WT)**

**Use Q9\_b.csv**

ANS- Skewness = -1.28736

Kurtosis = 3.8189

Left skewed and right kurtosis

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



ANS- From the above histogram it is clear that distribution is asymmetric and right skewed or +ve skewed.

By seeing the Boxplot we can easily find the outliers ,upper quartile , lower quartile , median etc . The points which are above the upper extreme are the 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?

ANS- Confidence interval= sample mean(+/-) t- value\*standard error

Here sample mean= 200

T-value=% of CI \*Degree of freedom

Standard error=

**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- 1) MEAN = 41.1875

MEDIAN = 41

MODE = 41

VARIANCE = 26.77

STD = 5.17

1. From the above data we can say that the scores of students are not that

much good . The highest score is 56 and the lowest score is 34. The

average score is 41.19 .

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

ANS- When the mean and median of the data are equal that mean no

Skewness at all .

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

ANS- When mean > median , then the distribution is distributed on

right side that means data is positively skewed.

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

ANS- - When median > mean , then the distribution is distributed on

left side that means data is negatively skewed.

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

ANS- The positive kurtosis value indicates the data is heavily tailed distributed .

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

ANS- The negative kurtosis value indicates the data is slightly tailed distributed.

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 symmetrically distributed.

What is nature of skewness of the data?

ANS- The data is negatively skewed .

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

ANS- The IQR of the given data is approximately 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 plots are symmetrically distributed.

-> The range of Boxplot 2 is higher than Boxplot 1 .

-> Both Boxplots have same median with different distribution.

Median = 262.5 approx

-> IQR of Boxplot 2 > IQR of Boxplot 1

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)
  3. P (20<MPG<50)

ANS- MPG <- Cars$MPG

Here mean= 34.43 , std = 9.14

By applying below formula in python

From scipy import stats

stats.norm.cdf(x,loc=mean,scale=std)

we can get :-

P(MPG>38) = 34.25

P(MPG<40) = 36.22

P(20<MPG<50) = 30

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS- Its not normal distribution, slightly left skewed.

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

Dataset: wc-at.csv

ANS- wc-at$waist – follows normal distribution

wc-at$AT- right skewed

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

ANS- For 90% Area=( 1+0.90 )/2 = 0.95

From Z table Z score = 1.6+0.05 = 1.65

For 94% Area=( 1+0.94 )/2 = 0.97

From Z table Z score = 1.8+0.08 = 1.88

For 60% Area=( 1+0.60 )/2 = 0.8

From Z table Z score = 0.8+0.04 = 0.84

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

ANS- T Score where degree of freedom = sample size – 1 =25-1= 24

95% CI = 2.064

96% CI = 2.172

99% CI = 2.797

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- tvalue =(260-270)/(90\*sqrt(18)) = -0.4715

P value= stats.t.cdf(0.027,df=17) = 0.511