|  |  |
| --- | --- |
| 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 | discrete |
| Length of a leaf | continuous |
| Dog's weight | continuous |
| Blue Color | discrete |
| Number of kids | discrete |
| Number of tickets in Indian railways | continuous |
| 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 | interval |
| Type of living accommodation | Ordinal |
| 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 | Nominal |

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

Possible outcomes (HHH,TTT,HHT,HTH,THH,HTT,THT,TTH)

Favorable outcomes (HHT,HTH,THH)

Probability=3/8

Q4) Two Dice are rolled, find the probability that sum is

* Equal to 1
* Less than or equal to 4
* Sum is divisible by 2 and 3
* Probability sum is equal to 1= 0
* Favorable outcome ((1,1), (1,2), (2,1), (1,3), (3,1), (2,2))

Probability of sum less than or equal to 4= 6/36=1/6

* Favorable outcome ((1,5), (5,1), (2,4), (4,2), (3,3), (6,6))

Probability of sum is divisible by 2 & 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?

probability that none of the balls drawn is blue=5c2/7c2=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

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\*0.120=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**

**Mean:**

Points 3.596563

Score 3.217250

Weigh 17.848750

**Median:**

Points 3.695

Score 3.325

Weigh 17.710

**Mode:**

Points 3.07

Score 3.44

Weigh 17.02

**Variance:**

Points 0.285881

Score 0.957379

Weigh 3.193166

**Standard deviation:**

Points 0.534679

Score 0.978457

Weigh 1.786943

**Range: min max**

Points 2.76 4.93

Score 1.513 5.424

Weigh 14.5 22.9

Weight has more variance as compare to points and score and points has least variance.

Q8) Calculate Expected Value for the problem below

* 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?

Expected value of weight of patients= (108+110+123+134+135+145+167+187+199)/9=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Skewness for speed= -0.11395477 skewness value is negative it means more weight in the right tail of the distribution.

Skewness for distance= 0.78248352 skewness value is positive it means more weight in the left tail of the distribution.

kurtosis for speed= -0.57714742 kurtosis<3 it is called platykurtic. As kurtosis is

negative it means it has wider peak.

kurtosis for distance= 0.24801866 kurtosis<3 it is called platykurtic. As kurtosis is

positive it means it has thinner peak.

**SP and Weight(WT)**

**Use Q9\_b.csv**

Skewness for SP= 1.58145368 skewness value is positive it means more weight in the left tail of the distribution.

Skewness for Weight= -0.60330993 skewness value is negative it means more weight in the right tail of the distribution.

kurtosis for SP= 2.72352149 kurtosis<3 it is called platykurtic. As kurtosis is

positive it means it has thinner peak.

kurtosis for Weight= 0.81946588 kurtosis<3 it is called platykurtic. As kurtosis is

positive it means it has thinner peak.

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





More data is in the range 50 to 100 with frequency 200 and least range 350 to 400 with frequency around 0 to 10.

As more data on right side so it is right skewed.

Median is less than mean so the distribution is right skewed. And outlier present in the upper side of the box plot and there is less data point 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?

dof=2000-1=1999

for 94% confidence interval =[200-(1.55\*30)/sqroot of 2000, 200+(1.55\*30)/sqroot of 2000]

=[198.961, 201.039]

for 98% confidence interval=[198.625, 201.375]

for 96% confidence interval=[198.827, 201.173]

**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**

* Find mean, median, variance, standard deviation.
* What can we say about the student marks?

Mean=41, median=40 variance=24.111 standard deviation=4.910

As standard deviation is large it means data is more spread. And outlier present in the data.

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

Answer: symmetrical distribution

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

Answer: Right skewed

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

Answer: left skewed

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

Answer: Positive values of kurtosis indicate that distribution is peaked and possesses thick tails.

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

Answer: A distribution with a negative kurtosis value indicates that the distribution has lighter tails and flatter peaks than the normal distribution.

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



What can we say about the distribution of the data?

The Median is not at the center of the Box and thewhiskers are not the same on both the ends then the data is not Normally Distributed.

What is nature of skewness of the data?

The Median lies closer to the Third Quartile and the whisker at the

upper end isshorter then it has a Negative Skew (Left Skew).

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

IQR=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.

Boxplot 2 has a longer length than Boxplot 1 which means that the dispersion of data is more in Boxplot 2 as compared to Boxplot 1.

The length of whiskers also gives an idea of the overall spread of data. The extreme values (minimum &maximum) gives the range of data distribution. Larger the range more scattered the data. Here Boxplot 2 has a larger range than Boxplot 1.

Also mean and median are same for both Boxplots hence the distribution is symmetric.

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

* P(MPG>38)

There are 65 values which is greater than 38

therefore P(MPG>38)= 65/81=0.80

* P(MPG<40)

There are 61 values which is less than 40

therefore P(MPG<40)= 61/81=0.753

* P (20<MPG<50)

There are 69 values which is greater than 50 and less than 20

therefore P(20<MPG<50)= 69/81=0.851

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

No MPG of car not follow normal distribution.

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

Dataset: wc-at.csv

No both waist and AT not follow normal distribution.

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

Z scores of 90% confidence interval=1.2815

Z scores of 94% confidence interval=1.5547

Z scores of 60% confidence interval=0.2533

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

t scores of 95% confidence interval=1.7108

t scores of 96% confidence interval=1.8280

t scores of 99% confidence interval=2.4921

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

s=18 , average=269, std=90

degree of freedom= 18-1=17

t=(260-270)/(90/sqroot18)

t=-0.46688

p(average life less than equal to 260)= 0.323