|  |  |
| --- | --- |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Ans:

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

Favorable outcomes are: (HHT, HTH, THH)

Probability 3/8

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

(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).

1. Equal to 1 🡺 Ans: 0
2. Less than or equal to 4 🡺 Favorable outcomes are: 6

Probability is: 6/36 🡺 1/6

1. Sum is divisible by 2 and 3 🡺 Favorable outcomes are: 22

Probability is: 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:

sample space =7c2 = 21. Let event E be none of the balls is blue = all balls are either red or green or both. n(E) = 5c2 =10 p(E) =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 = 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

Ans = 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.

**Ans:**

Points: 'Mean': 3.6, 'Median': 3.7, 'Mode': 3.07, 'variance': 0.29, 'standard deviation': 0.53, 'range': [2.76, 4.93]

Score: 'Mean': 3.22, 'Median': 3.32, 'Mode': 3.44, 'variance': 0.96, 'standard deviation': 0.98, 'range': [1.513, 5.424]

Weigh: 'Mean': 17.85, 'Median': 17.71, 'Mode': 17.02, 'variance': 3.19, 'standard deviation': 1.79, 'range': [14.5, 22.9]

Inference: Score and Weigh have Outliers in Upper range, All the data set are fairly symmetrical because skewness of the data sets lies between -0.5 to 0.5. Outliers affects the Score and Weigh value so they are not reliable.

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:**

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 that patient = 145.33

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

**Cars speed and distance**

**Skewness:**

**Speed:** -0.11750986144663393

**Distance:** 0.8068949601674215

**Kurtosis:**

**Speed:** -0.5089944204057617

**Distance:** 0.4050525816795765

**Inferences:** Speed is fairly Symmetrical because skewness lies between -0.5 to 0.5 and Distance is positively skewed.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Skewness:**

**SP:** 1.6114501961773586

**WT:** -0.6147533255357768

**Kurtosis:**

**SP:** 2.9773289437871835

WT: 0.9502914910300326

**Inferences:** SP is Highly Skewed because skewness lies above 1.0 and Distance is Moderately skewed because skewness lies above 0.5.

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



Inference: Data is positively skewed as seen in histogram and outliers are present in upper range as seen in boxplot.

**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:

Standard error = standard deviation/squareroot(n)

Standard error = 30/ = 0.671

For 94%

stats.t.ppf(0.97,1999)

1.8818614764780113

s = 1.88\*30/np.sqrt(2000)

1.2611423393098813

200-1.26

198.74

200+1.26

201.26

For 98%

stats.t.ppf(0.99,1999)

2.328214776106972

s = 2.32\*30/np.sqrt(2000)

1.5563033123398535

200-1.55

198.45

200+1.55

201.55

For 96%

stats.t.ppf(0.98,1999)

2.055089962825778

s = 2.05\*30/np.sqrt(2000)

1.3751818061623704

200-1.37

198.63

200+1.37

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?

ANS:

a1 Mean = 41, Median = 40.5, Variance = 25.5 , Standard Deviation = 5.05

a2 The data has positive skewness and outliers are present in data

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

Ans: Zero Skewness

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

Ans: Positive Skewness

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

Ans: Negative Skewness

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

Ans: Distribution is peaked and has thick tails; more data values are located near the mean and less data values are located on the tail.

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

Ans: Distribution is Flatter and has no thick tails, more data values are located near the tail and less data values are located on the mean.

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



What can we say about the distribution of the data?

ANS: It is observed that mean of observed data is < median hence distribution is left skewed

What is nature of skewness of the data?

ANS: It is observed as the data is negatively skewed

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: Range of 2nd box plot is greater than that of 1st boxplot, median of both of the boxplot is equal and both of the boxplot are free from skewness

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)

Ans: 0.347

* 1. P(MPG<40)

Ans: 0.729

* 1. P (20<MPG<50)

Ans: 0.0131

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS: Distribution is not Normal

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

Dataset: wc-at.csv

ANS: Distribution is not Normal

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

Z score = 90%: 1.6448536269514722 , 94% : 1.8807936081512509,

60%: 0.8416212335729143

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

T score= 95%: 2.0638985616280205, 96%: 2.1715446760080677,

99%: 2.796939504772804

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

Probability value is 0.32167253567098353