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

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

**Ans:** **3 coins are tossed**- {HHH, TTT, THT, HHT, TTH, HTH, HTT, THH}

S=8

**\* Probability that two heads and one tail are obtained**.

A {HHT, HTH, THH}

A=3

P(A)=A/S=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:** **Two dice are rolled**

{(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)}

S=36

1. **Equal to 1**

=0

**b) Less than or equal to 4**

(1,1) (1,2) (1,3) (2,1) (2,2) (3,1) =6

b=P(B)/P(S) =6/36 =0.16

**c)Sum is divisible by 2 and 3**

(1,1) (1,3) (1,5) (1,2) (1,2) (2,2) (2,4) (2,6) (3,1) (3,3) (3,5) (3,6) (4,2) (4,4) (4,5) (4,6) (5,1) (5,3) (5,4) (5,5) (6,2) (6,3) (64) (6,6)

C=24

P(C)/P(S) =24/36 =0.66

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:** Total number of balls  
= (2 + 3 + 2)  
= 7  
Let S be the sample space  
Then, n(S) = Number of ways of drawing 2 balls out of 7

n(S)=7C2

n(S)= (7×6) (2×1) n(S)=21

Let E = Event of 2 balls, none of which is blue  
n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(E)=5C2

n(E)= (5×4) (2×1)

n(E)=10

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

**Ans** – Expected number of randomly selected child

=1\*0.015+4\*0.20+3\*0.65+5\*0.05+6\*0.01+2\*0.120

=0.015+0.8+1.95+0.025+0.06+0.24

=3.090

=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**

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.42 | 3.14 | 16.86 |
| Median | 3.65 | 3.27 | 17.51 |
| Mode | 3.92 | 3.44 | 17.02 |
| Std. Dev | 0.91 | 1.10 | 4.11 |
| VAR | 0.83 | 1.23 | 16.91 |
| RANGE | 4.64 | 4.46 | 21.11 |

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:** 145.33

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

Cars speed and distance

**Use Q9\_a.csv**

**Ans-** csv file jupyter notebook(.ipynb)

|  |  |  |
| --- | --- | --- |
|  | Skew | Kurt |
| speed | -0.11 | -0.50 |
| dist | 0.80 | 0.40 |

**SP and Weight (WT)**

**Use Q9\_b.csv**

csv file jupyter notebook ( .ipynb )

|  |  |  |
| --- | --- | --- |
|  | Skew | Kurt |
| SP | 1.61 | 2.97 |
| WT | -0.61 | 0.95 |

Q10) Draw inferences about the following boxplot & histogram



**ANS :**

1. Majority of the Chicks has weight in range 50 – 100, followed by 100 -150 and 150 – 200
2. The data is positively Skewed
3. 3 Categories can be define (Under weight >50; Avg weight 51 – 150; Over weight <150)



**ANS:**

1. Data has outliers
2. Data is positively skewed

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**- csv file jupyter notebook (.ipynb )

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?

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 5.052664 |
| Standard deviation | 25.52941 |

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

**Ans:** When mean, median of data is equal is call zero skewness OR normal distribution

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:** Positive kurtosis value indicates that thinner peak and wider tails (Peakedness).

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

**Ans:** Negative kurtosis value indicates that wider peak and thinner tails (peakedness).

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



What can we say about the distribution of the data?

**Ans**: Not normally distributed. Data is more concentrated towards right side

What is nature of skewness of the data?

**Ans:** Negative skewness

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

**Ans:** IQR data is 8 (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. Data is Normally Distributed. No Outliers. Center around 262.5. Comparatively, first graph has less range
2. Data is Normally Distributed. No Outliers. Center around 262.5

Comparatively, second graph has more range

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) = 0.3475939251582705
  2. P(MPG<40) =0.7293498762151616
  3. P (20<MPG<50) =1.2430968797327613e-05

**Ans**- Jupyter notebook

Q 21) Check whether the data follows normal distribution a) Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv

**Ans:** Jupyter notebook

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

**Ans**-

|  |  |
| --- | --- |
| Confidence Interval | Z-score |
| 90% | 1.6448536 |
| 94% | 1.8807936 |
| 60% | 0.8416212 |

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

|  |  |
| --- | --- |
| Confidence Interval | T-score |
| 95% | 2.0638985 |
| 96% | 2.1715446 |
| 99% | 2.7969395 |

**Ans-**

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 = -0.4714,  
Degree of freedom = 17   
P (t) = 0.3216725