|  |  |
| --- | --- |
| 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 | Ordinal |
| 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:** No. of Favorable outcomes = 08

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

Total No. of possible outcomes = 03 (HHT, HTH, THH)

*# Probability of getting two heads and one tail:*

​ P = (Total No. of possible outcomes) / (No. of favorable outcomes) ​

P = 3/8​

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 = Dice 1, b = Dice 2

Total No. of possible outcomes = a x b = 6 x 6 = 36, see below

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SUM | 1,1 | 1,2 | 1,3 | 1,4 | 1,5 | 1,6 |
| 2 | 2,1 | 2,2 | 2,3 | 2,4 | 2,5 | 2,6 |
| 3 | 3,1 | 3,2 | 3,3 | 3,4 | 3,5 | 3,6 |
| 4 | 4,1 | 4,2 | 4,3 | 4,4 | 4,5 | 4,6 |
| 5 | 5,1 | 5,2 | 5,3 | 5,4 | 5,5 | 5,6 |
| 6 | 6,1 | 6,2 | 6,3 | 6,4 | 6,5 | 6,6 |
| 7 | 8 | 9 | 10 | 11 | 12 |  |

1. Total Favorable cases = 0 (for sum = 1)

Hence, probability is 0

1. Total Favorable cases = 6 (for sum ≤ 4)

Hence, probability is 6/36 = 1/6

1. Total Favorable cases = 6 (for sum is divisible by 2 and 3)

Hence, probability is 1/36

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:** As actual total balls = 07, amongst them 02 are blue. So, probability of first ball not being blue is 5/7. It lefts 06 ball with 02 blue. The probability of second ball not being blue is 4/6.

The probability that none of the balls drawn blue is (5/7 x (4/6) = 20/42 = 10/21.

P = 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:** 3.5

Expected Value = ∑ (probability \* Value)

∑ P(x). E(x)

Total = 06 Child

Probability of selecting each CHILD is= 1/6

Ex: 1, 4, 3, 5, 6, 2

P(x): 1/6, 1/6, 1/6, 1/6, 1/6, 1/6

Expected Value = (1/6) (1) + (1/6) (4) + (1/6) (3) + (1/6) (5) + (1/6) (6) + (1/6) (2)

= (1/6) (1+4+3+5+6+2) = (1/6) (21) = 3.5

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:** For Answer, Plz. Refer attached Jupyter Notebook named as Assignment-1.

**Inferences:** Out of 32 counts for different variables, it is observed that Point & score for vehicle getting high when weigh increases, and getting less when weigh decreases.

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 = ∑ (probability \* Value)

∑ P(x). E(x)

Total = 09 patients

Probability of selecting each patient is= 1/9

Ex: 108, 110, 123, 134, 135, 145, 167, 187, 199

P(x): 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9

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

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data Cars speed and distance? Use Q9\_a.csv**

**Ans:** *For more details refer attached Jupyter Notebook.*

Speed Distance

Skew: -0.11750986144663393 0.8068949601674215

Kurto: -0.5089944204057617 0.4050525816795765

**SP and Weight (WT) Use Q9\_b.csv**

SP WT

Skew: 1.6114501961773586 -0.6147533255357768

Kurto: - 2.9773289437871835 0.9502914910300326

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



**Ans:**

ChickWeight data is right skewed. It is observed that more than 50 percent ChickWeight is between 50 to 150. Outliers are found at upper 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?

**Ans:**

Sample meanof x = 200

Sample standard deviation of s = 30

Sample sizeof n = 2000



**Interval range:**

**i) 94% confidence interval:**

200-1 = 199 df

t = 1.896

200-1.896 x 30/2000 = 198.73

200+1.896 x 30/2000 = 201.27

**ii) 98% confidence interval:**

200-1 = 199 df

t = 2.0673

200-2.0673 x 30/2000 = 198.61

200+2.0673 x 30/2000 = 201.69

**iii) 96% confidence interval:**

200-1 = 199 df

t = 2.3452

200-2.3452 x 30/2000 = 198.43

200+2.3452x 30/2000 = 201.57

**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:** Mean = 41.0

Median = 40.5

Variance = 25.52

Standard deviation = 4.9103

**As per histogram, most of the student’s marks are between 38-40.** *For more details refer attached Jupyter Notebook.*

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

**Ans:** It means there is normal distribution, fairly symmetrical around mean. No lack of symmetry. No skewness in data.

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

**Ans:** Negative skewness. Distribution of data is on right side.

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

**Ans:** Positive skewness. Distribution of data is on left side.

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

**Ans:** Indicates thinner peaks and wider tails means heavy-tailed distribution.

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

**Ans:** Indicates wider peaks and thinner tails means light-tailed distribution.

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



-What can we say about the distribution of the data?

**Ans:** No normal distribution found.

-What is nature of skewness of the data?

**Ans:** Negative

-What will be the IQR of the data (approximately)?   
**Ans:** 18-10 = 08

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

**Ans:** For1) Data set

**Q-1:** 252, **Q3:** 278, **Min:** 244, **Median:** 262.5, **Max:** 287.5

**Interquartile range (Q3-Q1) =** 26

For2) Data set

**Q-1:** 224, **Q3:** 311, **Min:** 170, **Median:** 262.5, **Max:** 350

**Interquartile range (Q3-Q1) =** 87

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)

**Ans:** *For more details refer attached Jupyter Notebook.*

1. P(MPG>38):0.3475939251582705
2. P(MPG<40): 0.7293498762151616

c. P (20<MPG<50): 1.2430968797327613e-05

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:** Data is fairly symmetrical around the mean, i.e. fairly normally distributed.No outliers present in data. So that we can say that it is normally distributed. *For more details refer attached Jupyter Notebook.*

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

Dataset: wc-at.csv

**Ans:** *For more details refer attached Jupyter Notebook.*

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

**Ans:** *For more details refer attached Jupyter Notebook.*

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

**Ans:** *For more details refer attached Jupyter Notebook.*

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:** *For more details refer attached Jupyter Notebook.*