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

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) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

**Ans.** Total outcomes = 8

Sample space = {HHH,HHT,HTH,THH,HTT,THT,TTH,TTT}

Probability (2 heads 1 tail) = 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.** Total outcomes of 2 dice = 36

Outcomes having sum is 1 = 0

Probability = 0

**B.** Outcomes having sum less than or equal to 4 = {(1,1), (1,2), (2,1), (2,2), (1,3), (3,1)}

Probability = 6/36 = 1/6

**C.** Outcomes having sum which is divisible by 2 and 3 = {(1,5), (5,1), (2,4), (4,2), (3,3)}

Probability = 5/36

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.** Total balls = 7, total non-blue balls = 5

No. of ways to draw 2 balls = 7!/2!(7 - 2)! = 21

No. of ways to draw 2 non-blue balls = 5!/2!(5 - 2)! = 10

Probability of drawing 2 non-blue balls = 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.** To find an expected no. of candies for randomly selected child,

E = 1 x 0.015 + 4 x 0.2 + 3 x 0.65 + 5 x 0.005 + 6 x 0.01 + 2 x 0.12

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

E = 3.09

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points,Score,Weight

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

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 of randomly chosen patient = Mean of the data

E = (108 + 110 + 123+ 134 + 135 + 145 + 167 + 187 + 199)/9 = 145.33 Pounds

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



**Ans.** The above histogram plot represents the distribution of the weights of chicks with respect to its frequency count. The following inferences can be drawn out by looking into the plot:

* The distribution is skew-symmetric, particularly right-skewed distribution.
* The bin size of the distribution is 50 (by weight).
* For such a right-skewed distribution, the longest bar gives the mode of data (ranging from 50 – 100 wgt) followed by the 2nd highest bar to its right which gives the median (100 – 150 wgt) and the next bar which gives the mean of the data (150 – 200 wgt).
* Having 1 peak, the data is unimodal.



**Ans.** The above box plot is represented for an arbitrary data. The following inferences can be drawn for the plot:

* The median (bold line inside the box) of the data lies near to the third quartile showing that the distribution has a negative skew or left-skewed distribution.
* The data also consists of several outliers outside the minimum value of the dataset.

**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?

**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?

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

**Ans.** When mean and median of a data are equal, then no skewness is present. Hence, it is a symmetric distribution.

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

**Ans.** When mean > median then the distribution is positively skewed.

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

**Ans.** When median > mean then the distribution is negatively skewed.

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

**Ans.** It indicates that the distribution is highly peaked and that the mode of the data is much higher.

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

**Ans.** It indicates that the peak of distribution is not too long so, the ends are evenly distributed. Hence, the distribution is much flatter than a normal distribution.

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



What can we say about the distribution of the data?

**Ans.** The distribution of the data is not normal because the median of the data is not at the center of the box.

What is nature of skewness of the data?

**Ans.** The nature of skewness is negatively skewed as the median lies near the 3rd quartile and the right whisker is shorter.

What will be the IQR of the data (approximately)?   
  
**Ans.** IQR (Area inside the box) = 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.

**Ans.** Some comments/inferences of the two plots are as follows:

* Both the plots have different size of distributions however, both are normal (symmetric) distributions and the median of both the data are the same. Hence, they both have similar central tendency.
* The interquartile range of box 2 is greater than that of box 1.
* The data in box 2 that located outside of the maximum and minimum of box 1 can be considered as outliers for box 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)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

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

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

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