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

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

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

Ans:

Probability for two heads and one tail is 3/8 i.e. 0.375

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

1. Equal to 1

Ans = 0 (i.e. not possible that sum always exceeds to one)

1. Less than or equal to 4

Ans = 6/36 i.e. 1/6

1. Sum is divisible by 2 and 3

Ans =

Nos. divisible by 2 =>(1,1),(1,3),(1,5),(2,2),(2,4),(2,6), (3,1),(3,3),

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

(6,4),(6,6).

=>18 possibilities

Nos. divisible by 3 => (1,2),(1,5),(2,1),(2,4),(3,3),(3,6),(4,5),

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

=> 12 possibilities

Probability = favorable outcome/ Total outcome

= [(18+12)-6]/36

= 24/36

= 2/3

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
* = 7C2
* = (7x6)/(2x1)
* =21
* Let E = Event of drawing 2 balls, none of which is blue.
* n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls.
* =5C2
* =(5x4)/(2x1)
* =10
* i.e 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 candies for a randomly selected child
* =  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
* =  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**

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  =  sum of ( probability  \* Value )

* there are 9 patients
* Probability of selecting each patient = 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

Expected Value of the Weight of that patient = 145.33

**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:

* Histogram:-
* Chick weight data is right skewed or positively skewed.
* More than 50% Chick Weight is between 50 to 150.
* Most of the chick weight is present between 50 to 100.



* Boxplot:
* The data is right skewed therefore in this distribution mean is greater than the median.
* There are 7 outliers 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?

**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: Then it is a symmetric distribution, where the mean is equal to the median, and the distribution has zero skewness.

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

* Ans: Then it is called positive skewness or right skewed distribution.

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

* Ans: Then it is called negative skewness or left skewed distribution.

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

* Ans: It indicates that the distribution has thick tails. That is more no. of data points are located at tails instead of around the mean.

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

* Ans: It indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

* Ans: The above Boxplot is not normally distributed the median is towards the higher value between 15 to 16.

What is nature of skewness of the data?

* Ans: The data is a skewed towards left. The whisker range of minimum value is greater than maximum.

What will be the IQR of the data (approximately)?  
  
  Ans: The Inter Quantile Range =Upper quartile – Lower Quartile

* = 18 – 10 =8

Q19) Comment on the below Boxplot visualizations?



* Ans:
* First of all there are no outliers in both of the boxplots.

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

Ans:

* Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker 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)
  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