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

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

ANS: Three coins are tossed at once.

The possible outcomes will be

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

We have to find the probability of two heads and one tail:

So, favorable outcomes =HHT,HTH,THH

The Number favourable Outcomes =3

Hence required probability is:

Probability to event to happen P(E)= Number favourable Outcomes/Total no of outcomes

Probability(2H & 1T)=3/8

The probability of getting two heads and one tails in the toss of three coins

Simultaneously is= 3/8 or 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: Sample Space are

X={HH,HT,TH,TT}

|  |  |
| --- | --- |
| Out Come | Probability |
| 0 | 1/4 |
| 1 | 1/2 |
| 2 | 1/4 |

Two dice thrown=36

1. Zero
2. Less than or equal to 4 is(1,3),(2,2),(3,1) and n(b)=3/36 or 1/12
3. Sum is divisible by 2 is 1

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

Then, n(S) = Number of ways of drawing 2 balls out of 7 =7C27C2 = 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 = 10

Therefore, 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: 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

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

=3.090

=3.09

Expected number of candies for a randomly selected child=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**

ANS:Inference:- The value of mean and median are nearly equal so we say that the cure is nearly **symmetric.**

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: there are 9 patients

Probability of selecting each patients=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**

**ANS:** Fromthe graph speed are linearly increasingbut in distance there are lots of variances.

**Use Q9\_a.csv**

**SP and Weight(WT)**

ANS: Fromthe graph SP are linearly increasingbut in WT there are lots of variances

**Use Q9\_b.csv**

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



ANS:Histogram:-

Chick weight data is right skewed or positively skewed.---- Yes

- More than 50% Chick Weight is between 50 to 150. ---- Yes

- Most of the chick weight is between 50 to 100. --- Yes

:- From the above histrogram we conclude that in data many or more than one **outlier**s are present.



ANS:The data is right skewed.

- There areoutliers at upper side

From the above Boxplot or Wisker plot we conclude that in data many **outliers** are present.

**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: In a distribution with**zero skew**, the mean and median are equal. Zero skew: mean = median

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

ANS: If the mean is greater than the median, the distribution is positively skewed. If the

mean is less than the median, the distribution is negatively skewed.

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

ANS**:  If the mean is greater than the median, the distribution is positively skewed.**

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

ANS: Positive values of kurtosis indicate that**distribution is peaked** and possesses

thick tails. An extreme positive kurtosis indicates a distribution where more of the

numbers are located in the tails of the distribution instead of around the mean.

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

ANS: Negative kurtosis. A distribution with a negative kurtosis value indicates that the distribution has**lighter tails than the normal distribution**

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



1).What can we say about the distribution of the data?

ANS: Boxplot Distribution The box plot distribution will explain**how tightly the data is grouped**, how the data is skewed, and also about the symmetry of data. Positively Skewed: If the distance from the median to the maximum is greater than the distance from the median to the minimum, then the box plot is positively skewed.

2).What is nature of skewness of the data?

**Positively Skewed:** If the distance from the median to the maximum is greater than the distance from the median to the minimum, then the box plot is positively skewed. Negatively Skewed: If the distance from the median to minimum is greater than the distance from the median to the maximum, then the box plot is negatively skewed.

3).What will be the IQR of the data (approximately)?   
  
ANS:Q3 (Upper Quartile) = 18

Q1 (Lower Quartile) = 10

Interquartile Range (IQR) = 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: First there are no outliers. 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. In above graph, two Boxplot are drawn of differents data points. Median of both Boxplot is same 265(appro.) . In 1st Boxplot the data set or sample are less than 2nd data set but both are symmetric . so we said that mean and median are same value.

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