|  |  |
| --- | --- |
| 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 | Nominal |
| Religious Preference | Ordinal |
| 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?

Soln: The probability of getting 2 heads and one tail

P(HHT)+P(THH)+P(HTH)

=1/8 +1/8+1/8

=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

Soln a) Total possible outcomes when 2 dice are rolled=36

There are no outcomes which corresponds sum is equal to 1.

Probability=0/36=0

b)Probability that sum is less than or equal to 4 =(1,3),(2,2),(3,1)

=P(1,3)+P(2,2)+P(3,1)

=1/36+1/36+1/36

=3/36

=1/12

c)Probability that sum is divisible by 2 and 3=(1,5),(3,3)(4,2),(5,1),(6,6)

=P(1,5)+P(3,3)+P(4,2)+P(5,1)+P(6,6)

=1/36+1/36+1/36+1/36+1/36=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?

Soln: Total number of balls=2+3+2=7

2 balls drawn at random= 7C2=7\*6/2\*1=21

None of the balls are blue=5C2=5\*4/2\*1=10

The probability that none of the balls drawn is blue=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

Soln: The 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\*0.120

=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.

Soln: Attached the script file(Q7Solution.ipynb)

**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?

Soln: Probability of choosing each patient=1/9

Expectedvalue=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

=145.33

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

**Cars speed and distance**

**Soln: Attached the script file(Q9\_aSolution.ipynb)**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Soln: Attached the script file(Q9\_bSolution.ipynb)**

**Use Q9\_b.csv**

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



Soln: In the above graph, most of the data is concentrated in the left hand side. The least observation is for the weight 400.The skewness is positive, hence it is right skewed histogram.



Soln: In the above graph, the whisker(range) between upper extreme and upper quartile is more compared to lower extreme and lower quartile which has less range. The outliers are present above the upper extreme. Inter quartile range(IQR) is not equally distributed. Most of the data is distributed over the right side if we visualize horizontally, hence it is left 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?

Soln: Attached the script file(Q11Solution.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?

Soln: Attached the script file(Q12Solution.ipynb)

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

Soln: The nature of the skewness is perfectly symmetrical, skewness is 0.

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

Soln: The nature of the skewness is right skewed, positive.

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

Soln: The nature of the skewness is left skewed, negative.

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

Soln: The positive kurtosis indicates that it has high peak and more number of observations are concentrated on peak.

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

Soln: The negative kurtosis indicates that the curve is flat and it has less number of observations concentrated.

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



What can we say about the distribution of the data?

Soln: The IQR is not equally distributed.

Lower quartile:10,Upper quartile:18

The whisker range between lower limit and lower quartile is more compared to whisker range between upper limit and upper quartile.

What is nature of skewness of the data?

Soln: Most of the data is concentrated on the left side, hence it is right skewed.

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

Soln: IQR = 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.

Soln: By observing both the plots, the whisker range for Boxplot 1 is low compared to Boxplot 2.The IQR is equally distributed and the distribution is symmetrical.

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)
  3. P (20<MPG<50)

Soln: a) 35% chance that P(MPG>38)

b)73% chance that P(MPG<40)

c)90% chance that P (20<MPG<50)

Attached the script file(Q20Solution.ipynb)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Soln: MPG of the cars follows normal distribution as mean and median

values are approximately same.

Attached the script file(Q21\_aSolution.ipynb)

1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist)

from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Soln: The Adipose Tissue(AT) and Waist Circumference(Waist) follows normal distribution as mean and median values for both are approximately same.

Attached the script file(Q21\_bSolution.ipynb)

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

interval, 60% confidence interval

Soln: Z score for 90% confidence interval- 1.64

Z score for 94% confidence interval- 1.88

Z score for 60% confidence interval- 0.84

Attached the script file(Q22Solution.ipynb)

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

Soln: degrees of freedom(df)= sample size-1

=25-1=24

t score for 95% confidence interval- 2.06

t score for 96% confidence interval- 2.17

t score for 99% confidence interval- 2.79

Attached the script file(Q23Solution.ipynb)

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

Soln: Attached the script file(Q24Solution.ipynb)