Q1) Identify the Data type for the Following:

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

Q2) Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

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

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

**Answer:-**

**🡺** Three Coins are tossed.

HHH, HHT, HTH, HTT, THT, THT, TTH, TTT

=>Two heads and one tail are obtained.

HHT, HTH, THT

* PROBABILITY =3\8

**The probability that two heads and one 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**

**Answer:-**

1. **Equal to 1:-**

S=36, A= (1,1), (1.2)

P(A)=(A\S) => 0/36

**P(A)**=> 0

**NOTHING IS EQUAL TO 1 BECAUSE (1,1), (1.2) =>1+1=2,1+2=3**

**b) Less than or equal to 4:-**

=> S=36, B= (1,1), (1,2), (1,3), (2,1), (2,2), (3,1)

B=6, S=36

P(B)=B/S = 6/36

🡺 P(B)= **1\6**

**c) Sum is divisible by 2 and 3:-**

**=>** S=36, C= (1,5), (3,3), (4,2), (5,1), (6,6)

C= 5, S=36

P(C)=C/S = 5/36

1. 🡺**P(C)= 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?

**Answer:-**

🡺Total number of balls

= 2 + 3 + 2 = **7**

🡺 Number of ways of drawing 2 balls out of 7

=7C2 = (7 × 6) / (2 × 1) = 42/2 = **21**

🡺Number of balls other than blue = **5**

 🡺Number of ways of drawing 2 balls out of 5

= 5C2 = (5 × 4) / (2 × 1) = 20/2 = **10**

**Probability = 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

**Answer:-**

🡺 (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

**Expected number of candies for a randomly selected child is 3.090**

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**

**Answer:-**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Median** | **Mode** | **Variance** | **Standard deviation** | **Range** |
| **Points** | 3.596563 | 3.695 | 3.891 | 0.534679 | 0.534679 | 2.17 |
| **Score** | 3.596563 | 3.325 | 17.43 | 0.957379 | 0.978457 | 3.91100 |
| **Weight** | 17.84875 | 17.710 | 17.43 | 3.193166 | 1.786943 | 8.39999 |
| **Dtype** | float64 | float64 | float64 | float64 | float64 | **-------** |

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?

**Answer:-**

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**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer:-** 1) Cars speed and distance

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurttosis** |
| **Car speed** | -0.1139548 | 2.422853 |
| **Distance** | 0.782485 | 3.248019 |

**Answer:-** 2) SP and Weight(WT)

|  |  |  |
| --- | --- | --- |
|  | **Skewness** | **Kurtosis** |
| **SP** | 1.581454 | 5.723521 |
| **WEIGHT(WT)** | -0.6033099 | 3.819466 |

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

**Answer:-**



**The skewness of this data is positively skewed**

**The most chick weight is between 50 to 100**



**There are outliers above upper extreme**

**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?

**Answer:-**

**At 94% C.I = [0.479,0521]**

**At 98% C.I = [0.474, 0.526]**

**At 96% C.I = [0.477, 0.523]**

**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?

**Answer:-**

1. **Find mean, median, variance, standard deviation**.

=> Mean = **41.0**

Median=**40.5**

Variance=**25.529**

Std. Deviation=**5.052**

1. **What can we say about the student marks?**

=> the average score of students Marks is = **41**

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

**Answer:- Skewness will not exist**

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

**Answer:- Negatively Skewed**

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

**Answer:- Positively Skewed**

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

**Answer:- Heavy tailed distribution and the data is less spread.**

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

**Answer:- Low tailed distribution and the data is densely spread**

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



1. What can we say about the distribution of the data?
2. What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
  
  
**Answer:-**

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

**Ans**:-

* The min value of data is 2
* Lower limit is -2 (lower limit = Q1 – 1.5(IQR)
* Upper limit is 30(Upper Limit = Q3 + 1.5(IQR)
* 50% of data lies between 10 to 18.3
* Most of the data is present between 10 to 15.2(Which is from first quartile to second quartile )

1. What is nature of skewness of the data?

**Ans:-** its negatively skewed

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

**Ans:-** 8

Q19) Comment on the below Boxplot visualizations?



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

**Answer:-**

* + - Data 1 has lower IQR then data 2
    - Mean is common in both data which is 262.5
    - There are no outliers in both data
    - Both the data is normally distributed
    - Second data is big compare to First (which means the upper limit and lower limit is bigger in Second data)

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)

**Answer:-**

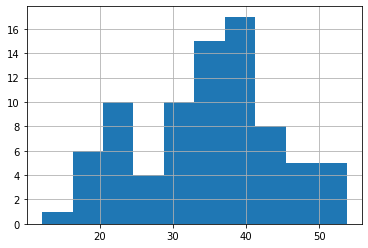
* + P(MPG>38)
    - **Ans:-** 0.347598
  + P(MPG<40)
    - **Ans:-** 0.729353
  + P (20<MPG<50)
    - **Ans:-** 0.013119

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Answer:-**

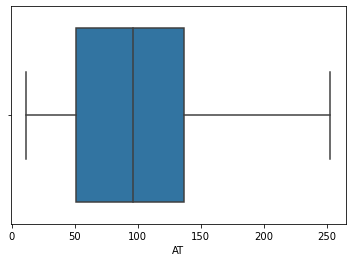


**from above we can say the data is fairly distribution**

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

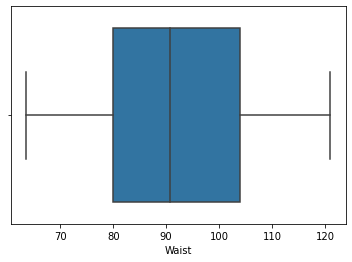
Dataset: wc-at.csv

**Answer:-** (i)



**AT is positively skewed**

(ii)



**Waist fairly normal distributed**

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

**Answer:-**

|  |  |
| --- | --- |
| Confidence interval (C.I.) | Z scores |
| 90% | 1.645 |
| 94% | 1.88079 |
| 60% | 0.841621 |

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

**Answer:-**

|  |  |
| --- | --- |
| Confidence interval (C.I.) | Z scores |
| 95% | 2.06 |
| 96% | 2.171 |
| 99% | 2.7969 |

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

**Answer:-**

the probability that 18 randomly selected bulbs would have an average life of no more than 260 days will be =  **0.32167**