|  |  |
| --- | --- |
| 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 data |
| 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 | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Norminal |
| 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?

Ans:

To find the probability of getting two heads and one tail when three coins are tossed.

To get two heads and one tail, there are different ways this can happen:

1.HHT

2.HTH

3.THH

Now, let’s calculate the probability of each sequence.

1.HHT:

probability=(1/2)\*(1/2)\*(1/2) = (1/8)

2.HTH:

probability = (1/2) \*(1/2) \*(1/2) = (1/8)

3.THH:

probability = (1/2) \*(1/2) \*(1/2) = (1/8)

Total probability = (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 2and 3

**Ans:**

1. **Probability that the sum is equal to 1:**

The smallest possible sum of two dice is 2, so the probability of getting a sum of 1 is **0**.

1. **Probability that the sum is less than or equal to 4:**

There are 3 ways to get a sum of 2 (1+1, 2+1, 2+2), 2 ways to get a sum of 3 (1+2, 2+1), and 1 way to get a sum of 4 (2+2). So, there are a total of 3+2+1 = 6 ways to get a sum of 4 or less.

The total number of possible outcomes when rolling two dice is 6\*6 = 36.

Therefore, the probability of getting a sum of 4 or less is **6/36 = 1/6**.

**c)Probability that the sum is divisible by 2 and 3:**

To find the sum that is divisible by both 2 and 3, we need to find sums that are divisible by 6 (2 \* 3).

appropriate outcomes:

Sum is 6: (1, 5), (2, 4), (3, 3), (4, 2), (5, 1)

Sum is 12: (6, 6)

Thereare a total of 6 appropriate outcomes out of 36 possible outcomes.

Therefore, the Probability is= 6/36 = 1/6

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:

To find the probability that none of the balls drawn is blue, you can use the concept of combinations.

nCr = n! / (r! \* (n-r)!)

C(7, 2) represents the number of ways to choose 2 balls out of 7:

C(7, 2) = 7! / (2!(7-2)!) = (7 \* 6) / (2 \* 1) = 21

There are 5 non-blue balls (2 red and 3 green). You want to choose 2 of them.

C(5, 2) represents the number of ways to choose 2 balls from the 5 non-blue balls:

C(5, 2) = 5! / (2!(5-2)!) = (5 \* 4) / (2 \* 1) = 10

So, there are 10 ways to choose 2 balls without any of them being blue.

Now, to find the probability, you divide the number of favorable outcomes (choosing 2 non-blue balls) by the total number of possible outcomes (choosing any 2 balls):

Probability = (Number of Favorable Outcomes) / (Total Number of Possible Outcomes)

Probability = 10 / 21

So, the probability that none of the balls drawn is blue is 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: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.

Ans:

Points score weigh

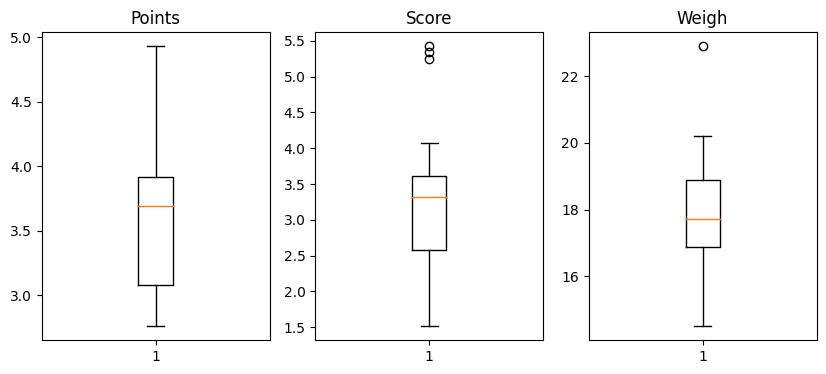
Mean 0.534679 0.978457 1.789643

Median 3.695 3.325 17.710

Mode 3.07 3.44 17.02

Variance 0.2763 0.9274 3.0933

Std 0.5346 0.9784 1.7869

Range 2.17 3.911 

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: 145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Cars speed skewness value=- 0.12 , kurtosis=0.81**

**Cars distance skewness value=0.81 , kurtosis=0.41**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Sp skewnesss value=1.61, kurtosis=0.95**

**Cars distance skewness value=0.81 , kurtosis=0.41**

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



**ANS**: The histogram is rightly skewed and tail is on right.Mean is greater than median and it has outliers



**ANS**:It has outliers on the maximum 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?

1.For 94% confidence interval Range is [ 198.73 – 201.26]

2.For 98% confidence interval range is [198.43 – 201.56]

3.For 96% confidence interval range is [198.62 – 201.37]

**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.

**ANS**: Mean: 41.0, Median: 41.0, Variance: 25.67, Standard deviation: 5.07

1. What can we say about the student marks?

**ANS**: we don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

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

ANS: When the mean, median, and mode of a dataset are equal, it indicates that the data is symmetrical. In this case, the skewness of the data is zero

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

ANS:When the mean is greater than median, the data is positively skewed,meaning it hasa tail o n the right side with some larger values pulling the average higher

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

ANS:The median is greater than the mean, the distribution is skewed to the lrft

Q16) What does positive kurtosis value indicates for adata ?

ANS: A positive kurtosis value indicates that a data distribution has heavier tails and a more peaked center than a normal distribution.

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

ANS:A negative kurtosis value indicates that a data distribution has lighter tails and a flatter center 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 boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

ANS: the data is skewed towards the left. The whisker range of minimum value is greater than the maximum

What will be the IQR of the data (approximately)?   
ANS:IQR(inter quartile range)=upper quartile- lower quartile , 18-10=8  
  
Q19) Comment on the below Boxplot visualizations?



ANS: Here there is a representation of 2 box plots in which box plot 2) is highly distributed across the plane and 1) is slightly less distributed. (variance)

Whiskers in these diagrams also show this.100% of the data is spread across values from 350 in 2 whereas its spread in range 250-290 app x in 1

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

Ans:Here when we compare box plot 1 with box plot 2, we can say that the data in boxplot 1 is widely spread. Here the main inference is that since the data range varies high in box plot 2 it is hard to make a prediction in box plot 2. The median in the 2box plots is equal. And the data spread is both of them are symmetrical.

Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

* 1. P(MPG>38)

ANS:0.347593

* 1. P(MPG<40)

ANS:0.729349

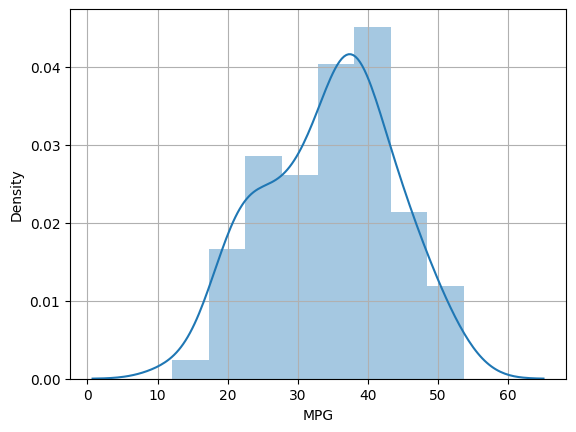
* 1. P (20<MPG<50)

ANS:0.89886

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

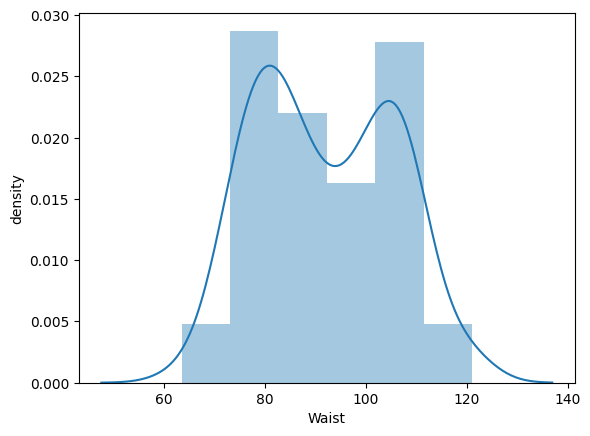
Dataset: Cars.csv

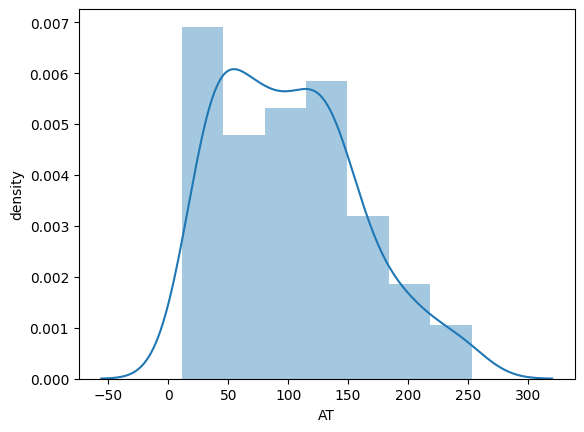
ANS: The MPG of cars appears to be approximately normally distributed.

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

Dataset: wc-at.csv

ANS: The AT and Waist values from the wc-at.csv dataset appear to be approximately normally distributed.





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

ANS:

1.Z scores of 90% confidence interval=1.644853

2.Z scores of 94% confidence interval=1.880793

3. .Z scores of 60% confidence interval=0.8416521

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

ANS:

1.t scores of 95% confidence interval=2.063898

2. t scores of 96% confidence interval=2.171544

3. . t scores of 99% confidence interval=2.796939

Q 24**)**A Government companyclaims 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

# ANS

# import math

# from scipy.stats import t

# # Define your values

# sample\_mean = 260

# population\_mean = 270

# sample\_std = 90

# sample\_size = 18

# # Calculate the t-score

# t\_score = (sample\_mean - population\_mean) / (sample\_std / math.sqrt(sample\_size))

# # Calculate the probability using the t-distribution

# df = sample\_size - 1  # Degrees of freedom

# p\_value = t.cdf(t\_score, df)

# print("The probability is:", p\_value)

# The probability is: 0.32167253567098364