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

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 | ordinal |
| Level of Agreement | ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ordinal |

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

The tossing of three coins at a time, the number of possible outcomes {HHH, TTT, HHT,HTH, THT, TTH, THH, HTT}. So out of these outcomes we have the probability of getting twoheads and one tail is 3/8 = 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 : a

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?

1.red 2

2.Green 3

3.Blue 2

2+3+2 =7= 7C2

= 7\*6/ 2\*1= 42/2=2

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 \* 0.12 = 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**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Point** | **Score** | **Weight** |
| **Mean** | 3.5965625 | 3.21725 | 17.8487 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 0 3.07  1 3.92 | 0 3.44 | 0 17.02  1 18.90 |
| Variance | 0.285881 | 0.957379 | 3.193166 |
| Standard Deviation | 0.534679 | 0.978457 | 0.978457 |
| Range | 2.17 | 3.911 | 21.387 |
|  |  |  |  |

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?

There are 9 patientsProbability of selecting each patient = 1/9108, 110, 123, 134, 135, 145, 167, 187, 199 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

v

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans :Refer Q9 python file**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans :Refer Q9 python file**

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



**Ans: The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side**.



Ans: The boxplot 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?

Ans : Refer Q11 python file

**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?

Ans : Refer python file Q12

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

Ans :If Mean , median and mode are equal the skewness is zero.

And distribution is symmetric

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

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

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

Ans:When the median is greater than the mean, the distribution is left /negativeskewed

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

**Ans**

**1. Heavy Tails:** Positive kurtosis indicates that the tails of the distribution are fatter or heavier than those of a normal distribution. This means there is an increased probability of extreme values (outliers) in the dataset.

**2. Leptokurtic Distribution:** A distribution with positive kurtosis is often referred to as leptokurtic. It has a sharper peak and heavier tails compared to a normal distribution.

**3. Outliers:** The presence of positive kurtosis suggests that there may be more outliers or extreme values in the dataset. Extreme values can contribute to increased volatility and variability in the data.

**4. Tailedness:** Positive kurtosis is associated with a distribution that has more data points in the tails than a normal distribution. This can affect statistical analyses that assume a normal distribution.

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

ANS

Key points about negative kurtosis:

1. Lighter Tails: Negative kurtosis indicates that the tails of the distribution are less heavy or fatter than those of a normal distribution. This suggests a lower probability of extreme values (outliers) in the dataset.

2. Platykurtic Distribution: A distribution with negative kurtosis is often referred to as platykurtic. It has a flatter peak and lighter tails compared to a normal distribution.

3. Less Outliers: The presence of negative kurtosis suggests that there may be fewer outliers or extreme values in the dataset. The distribution has less variability in the tails.

4. Distribution Spread: Negative kurtosis is associated with a distribution that has less data in the tails than a normal distribution. This can result in a more spread-out or dispersed distribution.

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



What can we say about the distribution of the data?

Ans : there is no standard distribution

What is nature of skewness of the data?

Ans : Positive skewness

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

Ans:

# Sample data

data = np.array([2,4,6,8,10,12,14,16,18])

# Calculate Q1, Q3, and IQR

q1 = np.percentile(data, 25)

q3 = np.percentile(data, 75)

iqr = q3 - q1  
q1=6

q3=14

iqr=8

Q19) Comment on the below Boxplot visualizations?



Ans:

* There are no outliers.
* Both the box plot shares the same median ,app.in a range between 275 to 250
* Normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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

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

a.P(MPG>38)= 0.347

b.P(MPG<40)= 0.7294

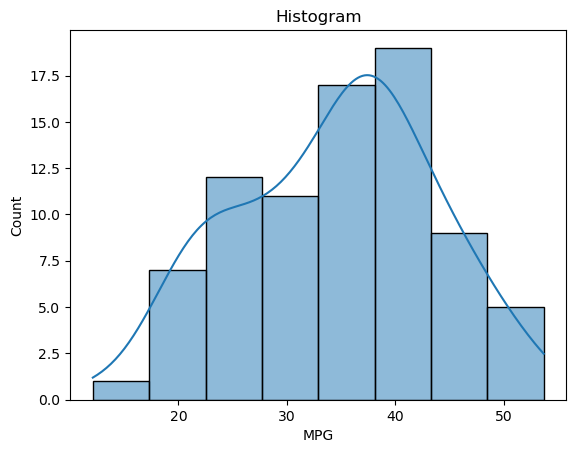
c.P (20<MPG<50) = 0.0131

Also refer Q20 python file

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

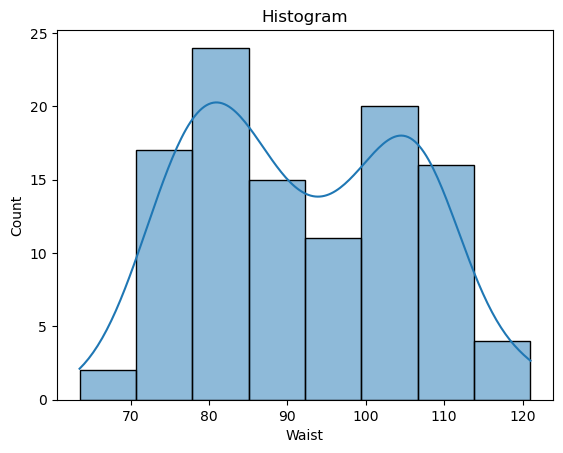
Dataset: Cars.csv

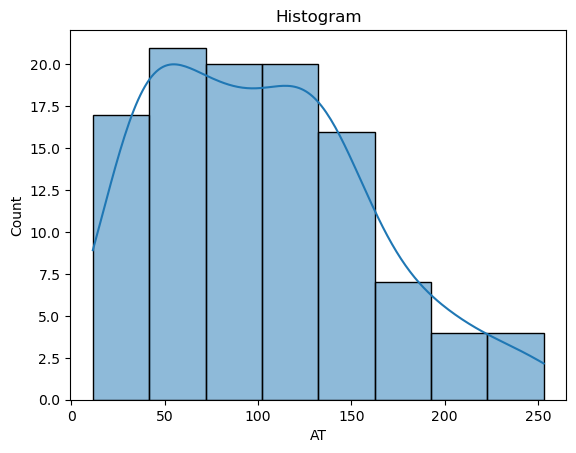
Ans : Approximtely normal distribution 

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

Dataset: wc-at.csv

Ans: wc-at does not follow normal distribution





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

ANS

90%= 1.6448

94%= 1.8807

60%= 0.841

Refer python file Q22

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

ANS 95%= 2.0638

96%= 2.171

99%= 2.796

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

Ans : 0.32 (refer python Q24 file)