|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Multiple |
| 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 | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Binary |

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 | Ratio |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Interval |
| Fahrenheit Temperature | Ratio |
| 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 | Ordinal |
| Number of Children | Interval |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Ans: 3/8

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1 ans: p(Equal to 1) = 0
2. Less than or equal to 4 ans: p(Less than or equal to 4) = 1/6
3. Sum is divisible by 2 and 3 ans: P(Sum is divisible by 2 and 3) = 6/39 = 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: 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: E(x) = 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.

**Use Q7.csv file**

ans: Mean(points) = 3.59

Mean (score) = 3.21

Mean (Weigh) = 17.84

Median (Points) = 3.695

Median ( Scores) = 3.325

Median ( Weigh) = 17.710

Mode ( Points) = 3.07,3.92

Mode ( Score) = 3.44

Mode (Weigh) = 17.02, 18.90

Variance (Points) = 0.285

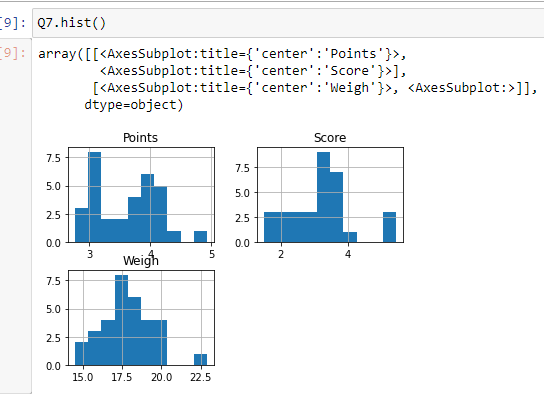
Variance (Scores) = 0.957

Variance (Weigh) = 3.193

Std (Points) = 0.534

Std (Scores) = 0.97

Std ( Weigh) = 1.786



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: (108+110+123+134+135+145+167+187+199) / 9 = 145.3

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

**Cars speed and distance**

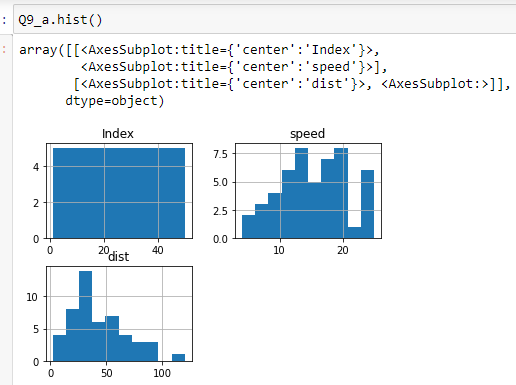
**Use Q9\_a.csv**

**Ans: Skewness of Speed data : - 0.1175**

**Skewness of Distance data : 0.8068**

**Kurtosis of Speed data : - 0.5089**

**Kurtosis of Distance data: 0.4050**



**SP and Weight(WT)**

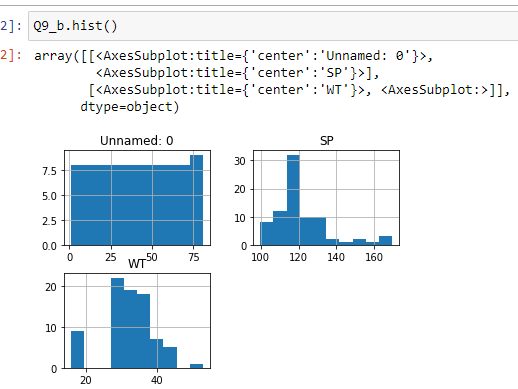
**Use Q9\_b.csv**

**Ans: Skewness for SP data = 1.6114**

**Skewness for WT data = - 0.6147**

**Kurtosis for SP data = 2.9773**

**Kurtosis for WT data = 0.9503**



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



Ans: From the above Histogram, we can conclude that, Distribution of the data is Right Skewed in nature and it is not Normally Distributed

From the given Boxplot representation we can say that the data has many Outliers and which is at the Upper Extreme part of Boxplot

**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: Confidence Interval (94%) = [ 134.895 , 265.107]

Confidence Interval (98%) = [122.725 , 277.274]

Confidence Interval (96%) = [130.209 , 269.790]

**Q12)** Below are the scores obtained by a student in tests

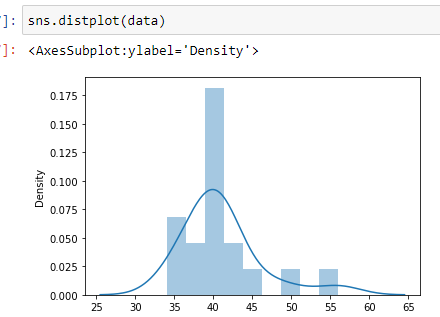
**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

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

Ans: Mean: 41, Median: 40.5, Variance: 25.529 , Std: 5.052

1. What can we say about the student marks?

Ans: Student’s marks are Normally distributed.



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

Ans: Skewness = 0

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

Ans: Positive Skewness

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

Ans: Negative Skewness

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

Ans: Distribution has heavier tails and sharper peak than Normal Distribution.

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

Ans: Distribution has lighter tails than Normal Distribution.

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



What can we say about the distribution of the data?

Ans: Distribution of the data is Left Skewed which means it not Normally Distributed.

What is nature of skewness of the data?

Ans: The data is left skewed in nature.

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

Ans: IQR = UQ – LQ

= 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: In both the boxplot, we can see that the median of the data is around 260. In the boxplot 1, the data are collected near the center or median of the distribution. And in boxplot 2, the data implies more variable 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)

Ans: (a). P(MPG>38) = 0.347

(b). P (MPG<40) = 0.729

(C). p(20<MPG<50) = 0.898

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Distribution of MPG data in dataset Cars.csv has 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

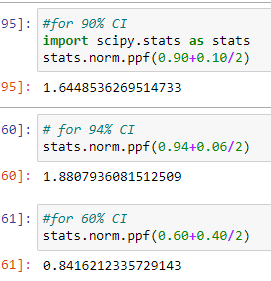
Ans: Waist data follows nearly Normal Distribution from the QQ-plot. AT data is not Normally Distributed and it is Skewed Right

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

Ans: Z scores of 90% confidence interval = 1.6448

Z scores of 94% confidence interval = 1.8807

Z scores of 60% confidence interval = 0.8416



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

Ans: t scores of 95% confidence interval = 2.063

t scores of 96% confidence interval = 2.017

t scores of 99% confidence interval = 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: t = (260-270)/(90/np.sqrt(18))

t = - 0.471

stats.t.cdf (-0.471 , df = 17)

= 0.3216