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

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

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

Ans: The number of combinations which have two heads and one tail are:

HHT, HTH, TTH which makes them 3 in number:

P(Two heads and One Tail) = Number of desired outcomes

= 3 = 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: There are 36 possible outcomes when two dice are rolled, as each die has 6 possible outcomes. N(Event( Two dice rolled )) =6^2=36

a) The sum of the two dice cannot be equal to 1 since the lowest possible sum is 2. Therefore, the probability of getting a sum of 1 is 0. P(sum is Euqal to 1) =’0’

b) P(sum is less than or equal to 4) = (Number of Event outcomes / Total number of outcomes)

P = 6/36 = 1/6 = 0.16

So, the probabilaty that the sum of two dice is less then or equal to 4 is 0.06.

c) P(sum is divisible by 2 and 3) = (Number of Event outcomes / Total number of outcomes)

P = 6/36 = 1/6 = 0.16

So, the probabilaty that the sum of two dice is divisible by 2 and 3 is 0.06.

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: The total number of balls in the bag is 2 + 3 +2 = 7

* Number of non-blue balls=Number of red + number of green = 2 + 3 = 5.
* Probability of drawing a non-blue ball on the A1= Number of non-blue ball / Total number of ball = 5/7
* A1 = 5/7
* Number of non-blue balls now = Number of non-blue ball at the beginning -1 = 5 - 1 = 4
* Total number of balls now = Total number of balls at the beginning -1 = 7 - 1 = 6
* Probability of drawing a non-blue ball on the A2= Number of non-blue ball / Total number of ball = 4/6 = 2/3
* A2 = 2/3
* P = A1 \* A2 = 5/7\*2/3 = 10/21
* P = 10/21
* So he probability that none of the drown 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: 0.015+0.8+1.95+0.025+0.06+0.24 = 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 for Points = 3.59, Score = 3.21 and Weigh = 17.84

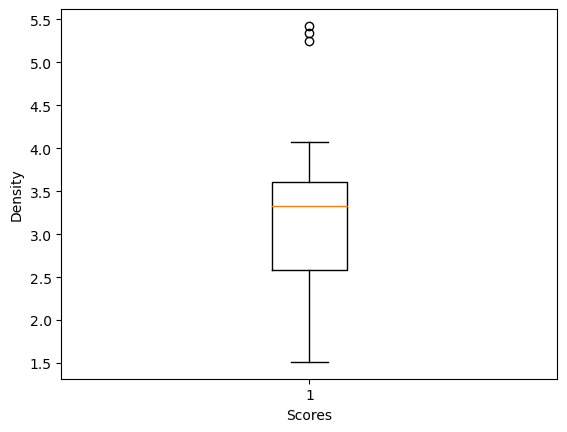
Median for Points = 3.69, Score = 3.32 and Weigh = 17.71

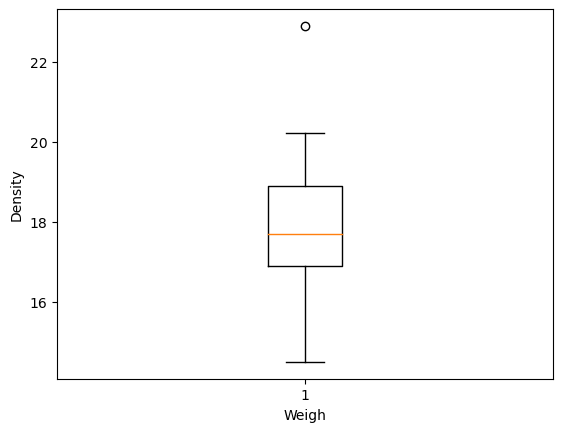
Mode for Points = 3.07, Score = 3.44 and Weigh = 17.02

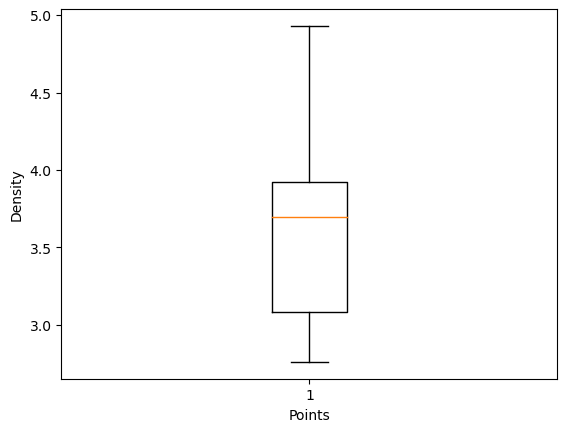
Variance for Points = 0.28, Score = 0.95, Weigh = 3.19

Standard Deviation for Points = 0.53, Score = 0.97, Weigh = 1.78

Range [Min-Max] for Points [3.59 – 4.93], Score [3.21 – 5.42] and Weigh [17.84 – 22.9]

****

****

****

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: Expected value = Sum (X \* Probability of X)

= (1/9)(108)+ (1/9)(110)+ (1/9)(123)+ (1/9)(134)+ (1/9)(145)+ (1/9)(167)+ (1/9)(187)+ (1/9)(199)

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:**

**[input] print('For Cars Speed', "Skewness value=", np.round(df1.speed.skew(),2), 'and' , 'Kurtosis value=', np.round(df1.speed.kurt(),2))**

**[Output]: For Car Speed Skewness value = -0.12 and Kurtosis Value= -0.51**

**[Input]:- print('Skewness value =', np.round(df1.dist.skew(),2),'and', 'Kurtosis value =', np.round(df1.dist.kurt(),2), 'for Cars Distance')**

**[Output]:- Skewness Value = 0.81 and Kurtosis Value = 0.41 for Cars Distance**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:-**

**[Input]:- print('For SP Skewness =', np.round(df2.SP.skew(),2), 'kurtosis =', np.round(df2.SP.kurt(),2))**

**[Output]:- For SP Skewness = 1.61 Kurtosis =2.98**

**[Input]:- print('For WT Skewness =', np.round(df2.WT.skew(),2), 'Kurtosis =', np.round(df2.WT.kurt(),2))**

**[Output]:- For WT Skewness = -0.61 Kurtosis = 0.95**

**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:-

[Input]:- stats.norm.interval(0.94,200,30/(2000\*\*0.5))

[Output]:- 198.73 , 201.26

[Input]:- stats.norm.interval(0.96,200,30/(2000\*\*0.5))

[Output]:- 198.62 , 201.37

[Input]:- stats.norm.interval(0.98,200,30/(2000\*\*0.5))

[Output]:- 198.43 , 201.56

**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 =25.52 and Standard Deviation =5.05

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: No skewness is present we have a perfect symmetrical distribution

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

Ans: Skewness and tail is towards Right

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

Ans: Skewness and tail is towards left

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

Ans: Positive kurtosis means the curve is more peaked and it is Leptokurtic

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

Ans: Negative Kurtosis means the curve will be flatter and broader

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



What can we say about the distribution of the data?

Ans: The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data?

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

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

Ans: IQR is the inter quartile range. Here Q1=10, Q2= 14.7, Q3=18

IQP = Q3- Q1 = 8

8

Q19) Comment on the below Boxplot visualizations?



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

Ans: First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)

1. P(MPG>38)

Ans: [Input] -1-stats.norm.cdf(38,df3.MPG.mean(),df3.MPG.std())

[Output] 0.34759392515827137

1. P(MPG<40)

Ans:- [Input] stats.norm.cdf(40,df3.MPG.mean(),df3.MPG.std())

[Output] 0.7293498762151616

1. P (20<MPG<50)

Ans:- [Input] stats.norm.cdf(0.50,df3.MPG.mean(),df3.MPG.std())-stats.norm.cdf(0.20,df3.MPG.mean(),df3.MPG.std())

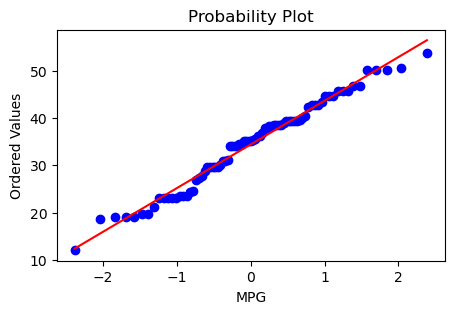
[Output] 1.243096879732713

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

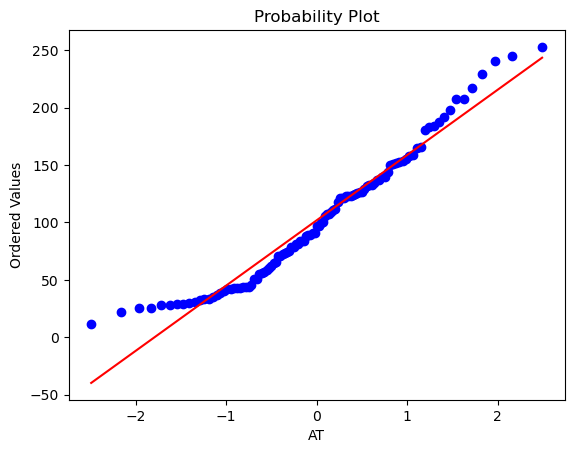
Ans:- MPG



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

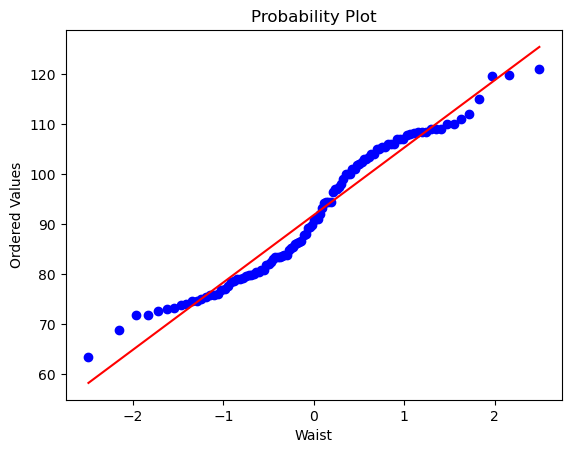
Dataset: wc-at.csv

Ans:- AT



Ans:-

Waist



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

Ans:- 90% confidence interval

[Input] stats.norm.ppf(0.95)

[Output] 1.6448

Ans:- 94% confidence interval

[Input] stats.norm.ppf(0.95)

[Output] 1.8807

Ans:- 60% confidence interval

[Input] stats.norm.ppf(0.95)

[Output] 0.8416

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

Ans:- 95% confidence interval

[Input] stats.t.ppf(0.975,24)

[Output] 2.0638

Ans:- 96% confidence interval

[Input] stats.t.ppf(0.98,24)

[Output] 2.1715

Ans:- 99% confidence interval

[Input] stats.t.ppf(0.995,24)

[Output] 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

Ans:- from scipy.stats import t

mu = 270

x\_bar = 260

s = 90

n = 18

t\_score = (x\_bar - mu) / (s / (n\*\*0.5))

df = n - 1

probability = t.cdf(t\_score, df)

print("t-score:", t\_score)

print("Probability:", probability)

[Output]:- t-score : -0.4714

Probability: 0.3216