|  |  |
| --- | --- |
| 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) | Nominal |

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 | Nominal |
| 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 | Ratio |
| 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: Number of outcomes= {HHH,HHT,HTH,THH,THT,TTH,TTT}

Outcome for getting two heads and one tail= {HTH,HHT,THH}

Probability of getting two heads and one tail = no. of outcomes/Total no. of outcomes = 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) **None**

b) Events =(1,1)(1,2)(1,3)(2,1)(2,2)(3,1)

P (Sum is less than or equal to 4) = no. of outcomes/Total no. of

outcomes= 6/36 = **0.16**

c) Events =(1,1)(1,2)(1,3)(2,1)(2,2)(3,1)

P (Sum is less than or equal to 4) = no. of outcomes/Total no. of

outcomes= 6/36 = **0.16**

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. 5) Two balls drawn randomly from bag

nCr = (n!) / (n-r)! = 7! / 2! \* 5! =

(7\*6\*5\*4\*3\*2\*1) /(2\*1) \* (5\*4\*3\*2\*1)=**21**

If the ball drawn randomly from bag is not blue = 7-2 = 5

None of the ball drawn is blue then = 5! / 2! \* 3! = (5\*4) / (2\*1) = **10**

P (None of the balls drawn is blue) = N (Event (None of the balls drawn is blue) /

N (Event (2 balls are drawn randomly from

bag)

= 10 / 21 = **0.47**

**P (None of the balls drawn is blue) =0.47**

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 6)

|  |  |  |
| --- | --- | --- |
| Candies count | Probability | Candies count\* Probability |
| 1 | 0.015 | 0.015 |
| 4 | 0.20 | 0.8 |
| 3 | 0.65 | 1.95 |
| 5 | 0.005 | 0.025 |
| 6 | 0.01 | 0.06 |
| 2 | 0.120 | 0.24 |

Expected number of candies for a randomly selected child

= 0.015 + 0.8+1.95+0.025+0.06+0.24

**= 3.09**

**Expected number of candies for a randomly selected child is 3.09**

Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

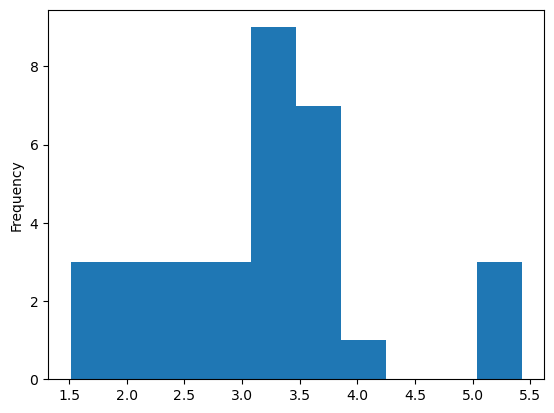
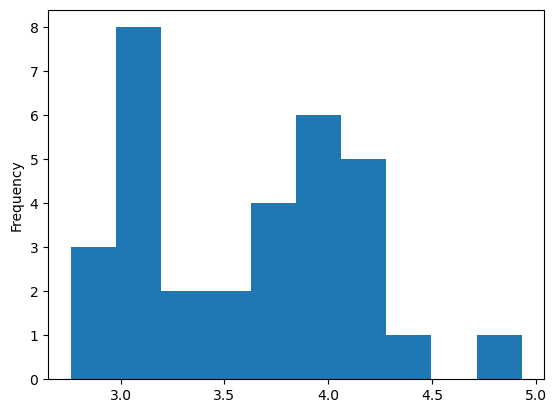
* For Points,Score,Weigh>

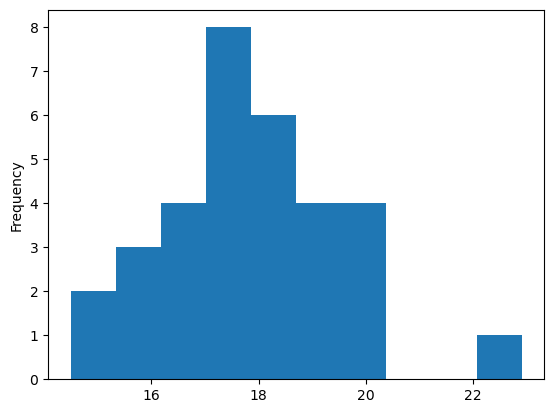
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

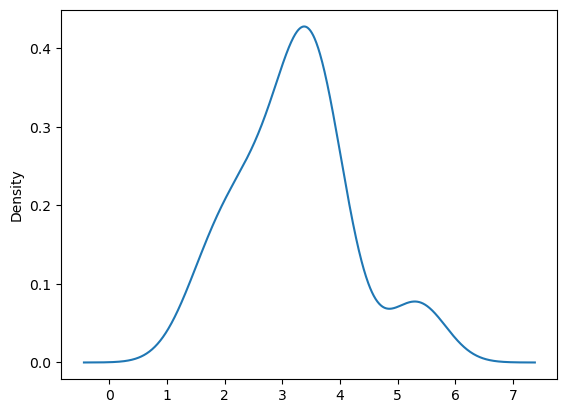
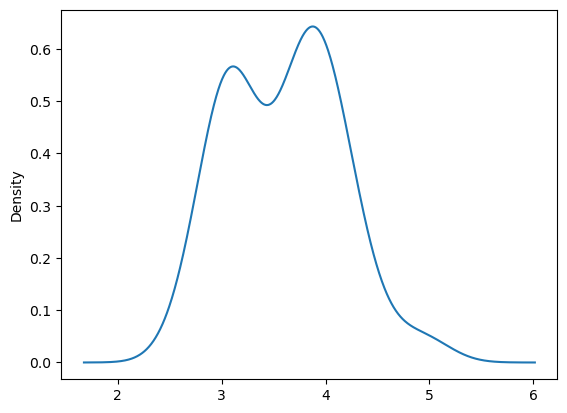
**Use Q7.csv file**

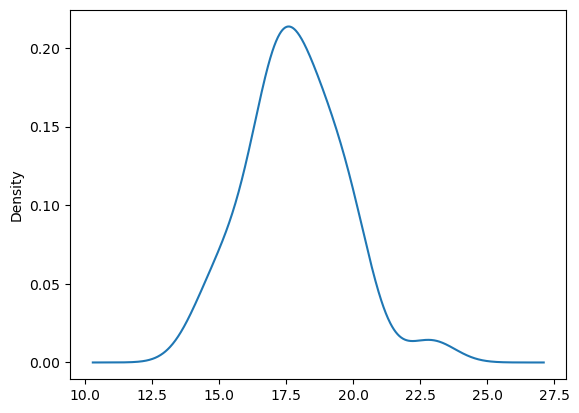
Ans 7)

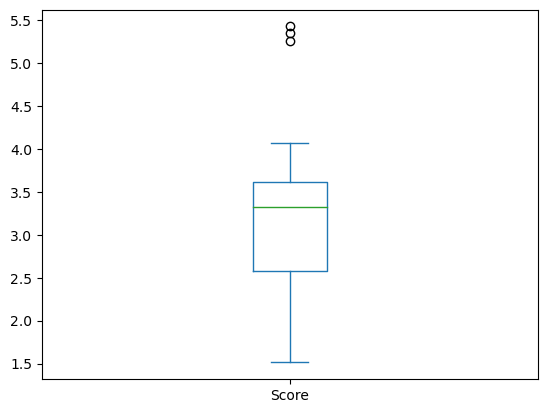
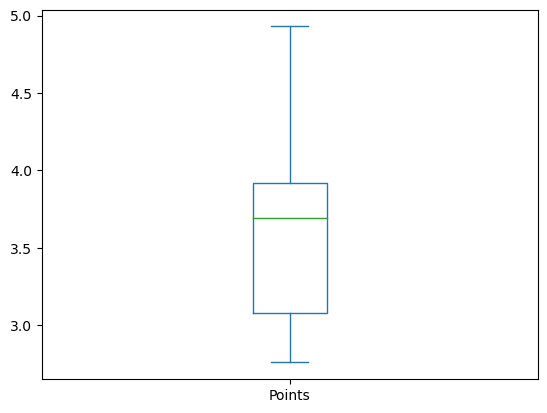
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | Standard Deviation | Range |
| Points | 3.59 | 3.69 | 3.92 | 0.29 | 0.53 | 2.17 |
| Score | 3.21 | 3.32 | 3.44 | 0.96 | 0.98 | 3.911 |
| Weigh | 17.85 | 17.71 | 17.02 | 3.19 | 1.79 | 8.4 |

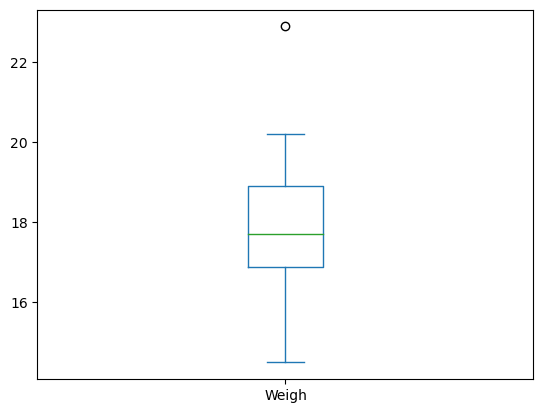












Score and Weigh has outliers as seen in the graph

In score there are three outliers and in weigh there is only one outlier

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 8) Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

Probability of selecting each patient = 1/9

Ex  108, 110, 123, 134, 135, 145, 167, 187, 199

P(x) 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**

**The Expected Value of the Weight of that patient = 145.33**

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

**Cars speed and distance**

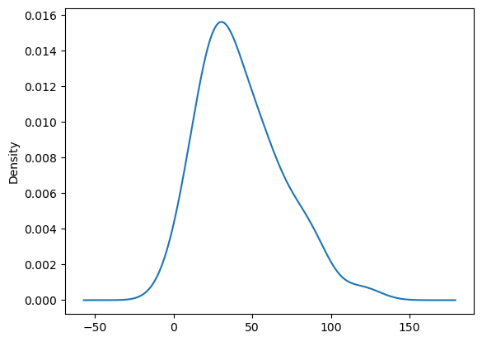
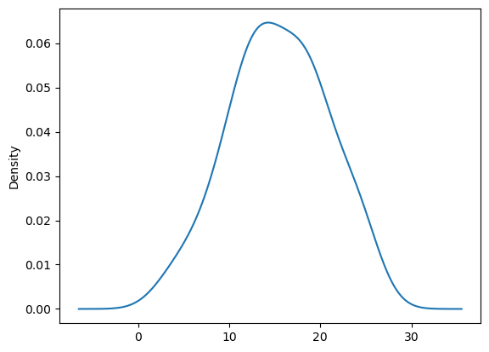
**Use Q9\_a.csv**

Ans 9 (a) Skewness for Speed = -0.11

Kurtosis for Speed = - 0.50

Skewness for Distance = 0.80

Kurtosis for Distance = 0.40



Speed Distance

* Skewness for speed is -0.11 which is less than 0 which means that data is negatively skewed data.
* Kurtosis for Speed is -0.50 which implies wider peak and thinner tails
* Skewness for Distance is 0.80 which is greater than 0 which means that data is positively skewed data.
* Kurtosis for Distance is 0.40 which implies heavier tails and a more peaked distribution

**SP and Weight(WT)**

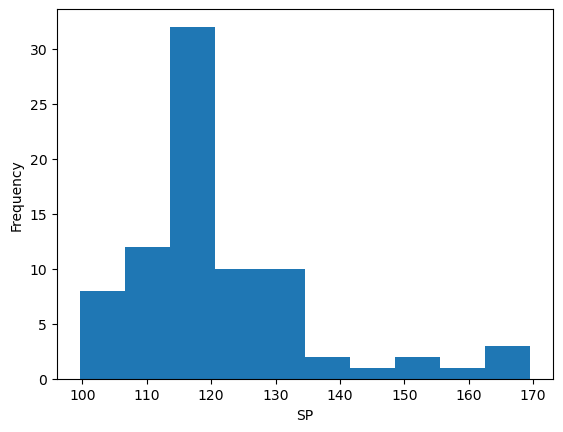
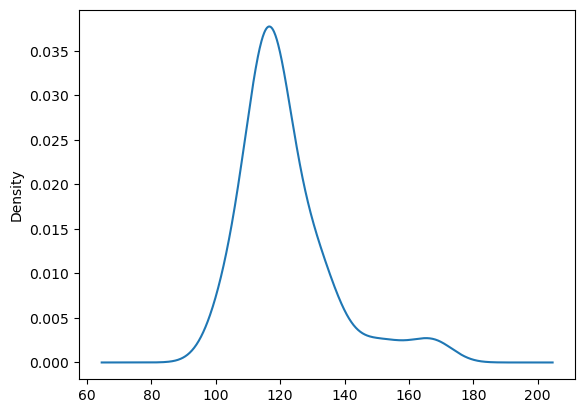
**Use Q9\_b.csv**

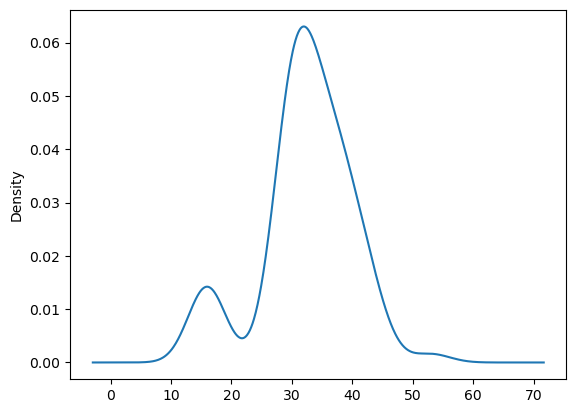
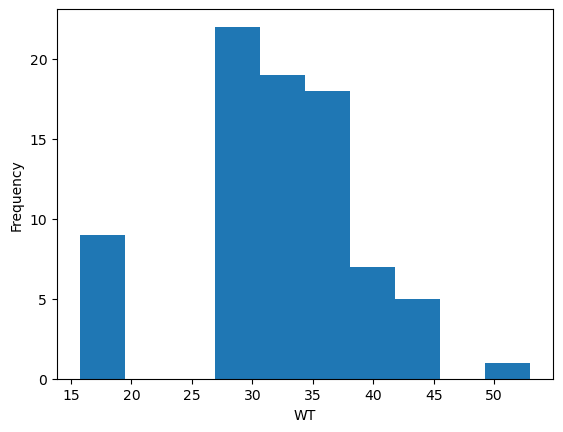
Ans 9 (b) Skewness for SP = 0.61

Kurtosis for SP= 2.97

Skewness for WT = -0.61

Kurtosis for WT = 0.95



* Skewness for SP = 0.61 which is greater than 0 which means that data is positively skewed data.
* Kurtosis for SP is 2.97 which implies heavier tails and a more peaked distribution
* Skewness for WT is -0.61 which is less than 0 which means that data is negatively skewed data.
* Kurtosis for Distance is 0.95 which implies heavier tails and a more peaked distribution

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



* The data is positively skewed data. In this case mean >= median>=mode
* Majority of the chick weight in range between 50-100



* Data has outliers
* Data is positively skewed

**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 11) Sample mean = 200

Standard Deviation = 30

Sample size = 2000

1. For CI 94%

**CI 94%** = **x̄ ± z \* ơ / √n**

= 200 **± 1.555 \* 30/√2000**

**= 200 ± 1.555 \* 0.670**

**= 200 ± 1.043**

**CI 94%=(201.043 , 198.956)**

**Upper limit = 201.043**

**Lower limit = 198.956**

1. **For** CI 98%

**CI 98%** = **x̄ ± z \* ơ / √n**

= 200 **± 2.054 \* 30/√2000**

**= 200 ± 2.054\* 0.670**

**= 200 ± 1.376**

**CI 98%=(201.376, 198.623)**

**Upper limit = 201.376**

**Lower limit = 198.623**

1. **For** CI 96%

**CI 96%** = **x̄ ± z \* ơ / √n**

= 200 **± 1.751 \* 30/√2000**

**= 200 ± 1.751\* 0.670**

**= 200 ± 1.173**

**CI 96%=(201.173, 198.827)**

**Upper limit = 201.173**

**Lower limit = 198.827**

**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 12 ) Mean = (Sum of all observations/ Total number of observations)

=34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56/18

= 738/18

**= 41**

Median = Middle most term of an ordered dataset

Arrange the sample dataset into ascending order

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

**= 40.5**

standard deviation= s2 = (xi - **x̄**)2 /n-1

= (xi – 41)/n-1

= (34-41)2 +(36-41)2 + (36-41)2+ (38-41)2+(38-41)2+(39-41)2 +(39-41)2+(40-41)2+(40-41)2+(41-41)2+(41-41)2+(41-41)2+(41-41)2 + (42-41)2 +(42-41)2 + (45-41)2 + (49-41)2 + (56-41)/ 17

= 49 + 25 + 25 + 9 + 9 + 4 + 4 + 1 + 1 + 0 + 0 + 0 + 0 +1 + 1 + 16 + 64 + 225 /17

standard deviation **= 25.529**

Variance = **5.052**

**Mean =41**

**Median = 40.5**

**Standard deviation =25.529**

**Variance = 5.052**

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

Ans 13): Skewness = 0. Perfectly symmetric bell-shaped curve

It will be normally distributed data

If the mean is equal to the median in a set of numbers, then this

distribution is a symmetric distribution

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

Ans 14 ) Distribution is positively skewed .

Data will be distributed more on the left.

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

Ans 15) Distribution is negatively skewed .

Data will be distributed more on the right.

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

Ans 16) **Heavier Tails:** Distribution has more values in the tails that are farther from the mean.

**Sharper Peak:** The central part of the distribution is more peaked than a normal distribution.

**More Outliers:** Indicates a higher likelihood of extreme values or outliers in the dataset.

**Kurtosis > 0:** The kurtosis value is greater than zero.

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

Ans 17) **Lighter Tails:** Distribution has fewer values in the tails that are farther from the mean.

**Less Peaked:** The central part of the distribution is less peaked than a normal distribution.

**Fewer Extreme Values:** Suggests a lower likelihood of extreme values or outliers in the dataset.

**Kurtosis < 0:** The kurtosis value is less than zero.

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



What can we say about the distribution of the data?

Ans: The data is not symmetric.

Data is distributed more on the right-side

What is nature of skewness of the data?

Ans: Skewness is Negative

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

ANS: IQR = Q3-Q1

IQR = 18-10

**IQR = 8**

Q19) Comment on the below Boxplot visualizations?



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

ANS:

1. Data is Normally Distributed

No outliers

Boxplot 1 has lesser range as compared to Boxplot2

1. Data is Normally Distributed

No outliers

Boxplot 2 has more range as compared to Boxplot1

Q 20) Calculate prosbability 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: 20 ) a) P(MPG>38) :

P(MPG>38) = 1-(P(MPG<38))

z = (x − μ)/ σ

z=38-34.42208/9.131445

z=0.6517

P(MPG>38) = 1-(P(MPG<38))

P(MPG>38) =1-0.6517

**P(MPG>38) =0.3483**

1. P(MPG<40)

z= (x − μ)/ σ

z=40-34.42208/9.131445

z=0.6108

**P(MPG<40)= 0.7291**

c)P(20<MPG<50) :

z = (x − μ)/ σ

z=20-34.42208/9.131445

z=-1.57939

P-value from Z-Table:

P(x<20) = 0.057124

P(x>20) = 1 - P(x<20) = 0.94288

P(MPG<50)

z = (x − μ)/ σ

z=50-34.42208/9.131445

z=1.705

**P(MPG<50) = 0.9554**

P(20<MPG<50)= Probability of MPG less than 50 – Probability pf MPG more than 20

P(20<MPG<50)= 0.9554 - 0.94288

**P(20<MPG<50)=0.012**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans a) MPG is 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

And b) Adipose Tissue (AT) and Waist Circumference(Waist) doesn’t follow normal distribution.

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

Ans 22) For 90%

Significant level at 5 % = 0.05

z at α =0.05 from z table will be

**z = 1.645**

For 94%

Significant level at 3 % = 0.03

z at α =0.03 from z table will be

**z = 1.880**

For 60%

Significant level at 20 % = 0.2

z at α =0.2 from z table will be

**z = 1.841**

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

Ans 23 ) For 95%

Significant level

α = 1-c =1-0.95=0.05

Sample size n = 25

t α/2 =  t 0.05/2 = t 0.025

**t α/2 = 2.045**

For 96%

Significant level

α = 1-c =1-0.96=0.04

Sample size n = 25

t α/2 =  t 0.04/2 = t 0.02

**t α/2 = 2.492**

For 99%

Significant level

α = 1-c =1-0.99=0.01

Sample size n = 25

t α/2 =  t 0.01/2 = t 0.005

**t α/2 = 2.797**

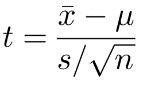
Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Ans 24 ) calculating t- statistics for data



|  |
| --- |
|  |

t= 260-270/ 90/**√**18  
 t=-1 \***√**2 /3

**t=-0.471**

**P(t)= 0.3216**