**Q1) Identify the Data type for the Following:**

|  |  |
| --- | --- |
| Activity | Data Types |
| Number of beatings from Wife | Integer |
| Results of rolling a dice | Discrete Integer |
| Weight of a person | Continuous Numerical(Float or decimal) |
| Weight of Gold | Continuous Numerical(Float or decimal) |
| Distance between two places | Continuous Numerical(Float or decimal) |
| Length of a leaf | Continuous Numerical(Float or decimal) |
| Dog's weight | Continuous Numerical(Float or decimal) |
| Blue Color | Categorical(Nominal) |
| Number of kids | Integer(assuming discrete count data) |
| Number of tickets in Indian railways | Integer(assuming discrete count data) |
| Number of times married | Integer(assuming discrete count data) |
| Gender (Male or Female) | Categorical(Nominal) |

**Q2) Identify the Data types, which were among the following Nominal, Ordinal, Interval, Ratio.**

|  |  |
| --- | --- |
| Data | Data Types |
| 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 | Circular (Not mentioned Data Type) |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

Answer:-

When three coins are tossed, there are 2^3 = 8 possible outcomes, as each coin can either land heads(H) or tails(T).

The outcomes with two heads and one tail are:

1. HHT
2. HTH
3. THH

So, there are three favorable outcomes.

The probability of getting two heads and one tail is the number of favorable outcomes divided by the total number of outcomes.

P(Two head and one tail) = Number of Favorable outcomes / Total number of cutcomes

P(Two head and one tail) = 3/8

P (Two heads and one tail) = N (Event (Two heads and one tail)) / N (Event (Three

coins tossed))

= 3/8 = 0.375 = 37.5%

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

**⦁** **Equal to 1**

**⦁** **Less than or equal to 4**

**⦁** **Sum is divisible by 2 and 3**

Answer:-

**Equal to 1:**

The minimum sum when rolling two dice is 2 (if both dice show 1).

There is no possible way to get a sum of 1.

Probability of sum equal to 1 = 0

**Less than or equal to 4:**

Possible sums less than or equal to 4 are 2, 3, and 4.

Sum = 2: (1, 1)

Sum = 3: (1, 2), (2, 1)

Sum = 4: (1, 3), (2, 2), (3, 1)

Total favorable outcomes = 6

Total possible outcomes = 6 \* 6

Probability = Number of favorable outcomes / Total possible outcomes= 6/36 = 16.66%

Probability of sum less than or equal to 4 = 1/6

**Sum is divisible by 2 and 3:**

The sums that are divisible by both 2 and 3 are 6 and 12.

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

Sum = 12: (6, 6)

Total favorable outcomes = 6

Total possible outcomes = 6 \* 6

Probability = Number of favorable outcomes / Total possible outcomes = 6/36=16.66%

**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?**

Answer: -

Total number of balls =7 balls

N (Event (2 balls are drawn randomly from bag) = 7! / 2! \* 5!

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

(2\*1) \* (5\*4\*3\*2\*1)

N (Event (2 balls are drawn randomly from bag) = (7\*6)/ (2\*1) = 21

If none of them drawn 2 balls are blue = 7 – 2 = 5

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

**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**

Answer:-

Expected Value=∑i (Xi \* Pi)

where Xi is the candy count and Pi is the probability for each candy count.

Let's calculate it:

Expected Value = (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120)

Expected Value = 0.015+0.80+1.95+0.025+0.06+0.24

Expected Value = 3.135

Therefore, the expected number of candies for a randomly selected child is approximately 3.145 candies.

**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**

Answer: -

Mean for Points = 3.59, Score = 3.21 and Weigh = 17.84

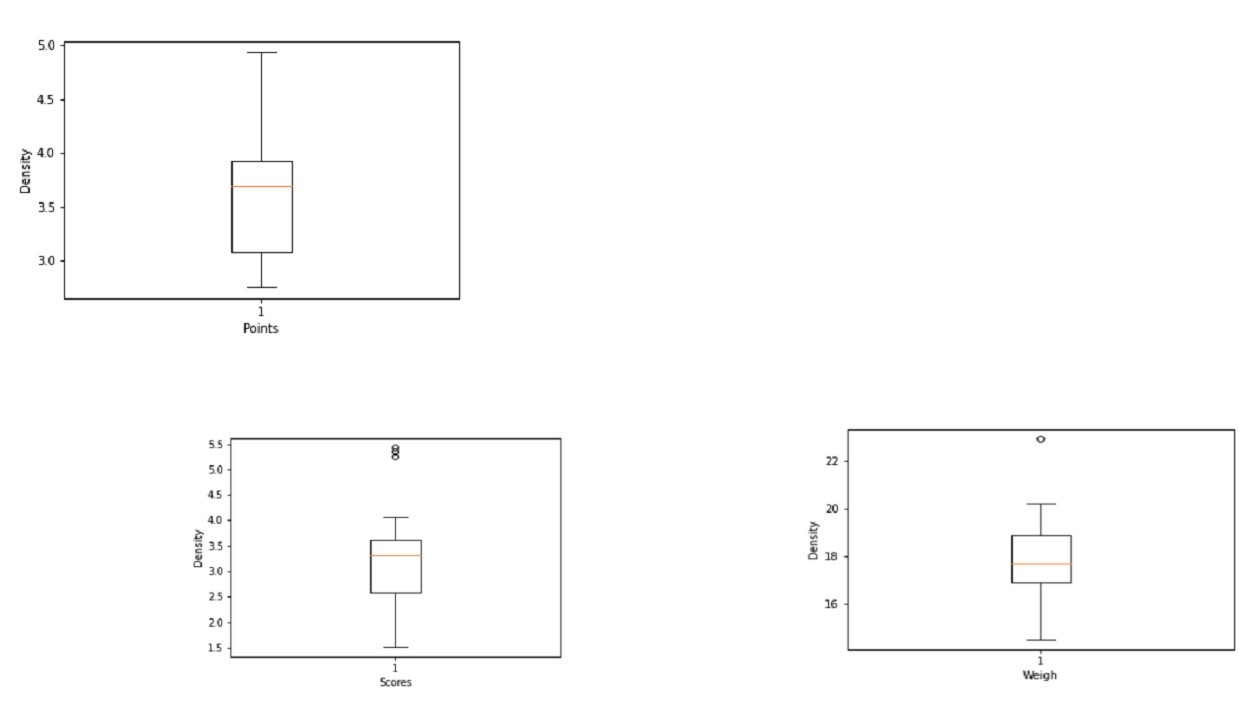
Median for Points = 3.69, Score = 3.32 and Weigh = 17.71

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**

**⦁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?**

Answer: -

The expected value (or mean) of a random variable is calculated by taking the sum of all possible values multiplied by their respective probabilities. In this case, since each patient has an equal chance of being chosen at random, the probability of selecting any specific weight is 1/9.

Let's calculate the expected value:

Expected Value=

1/9\*(108+110+123+134+135+145+167+187+199)

Expected Value=

1/9\*1308

Expected Value ≈ 145.33

So, the expected value of the weight of a randomly chosen patient is approximately 145.33 pounds. 221A

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

**Car's speed and distance**

**Use Q9\_a.csv**

Answer: -

q9a = pd.read\_csv("C:/Users/Moin Dalvi/Documents/EXcelR Study and Assignment Material/Data Science Assignments/Basic Statistics 1/Q9\_a.csv", index\_col = 'Index')

print('For Cars Speed', "Skewness value=", np.round(q9a.speed.skew(),2), 'and' , 'Kurtosis value=', np.round(q9a.dist.skew(),2))

For Cars Speed Skewness value= -0.12 and Kurtosis value= 0.81

print('Skewness value =', np.round(q9a.dist.skew(),2),'and', 'Kurtosis value =', np.round(q9a.dist.kurt(),2), 'for Cars Distance')

Skewness value = 0.81 and Kurtosis value = 0.41 for Cars Distance

**SP and Weight (WT)**

**Use Q9\_b.csv**

Answer: -

q9b =pd.read\_csv("C:/Users/Moin Dalvi/Documents/EXcelR Study and Assignment Material/Data Science Assignments/Basic Statistics 1/Q9\_b.csv")

print('For SP Skewness =', np.round(q9b.SP.skew(),2), 'kurtosis =',

np.round(q9b.WT.kurt(),2))

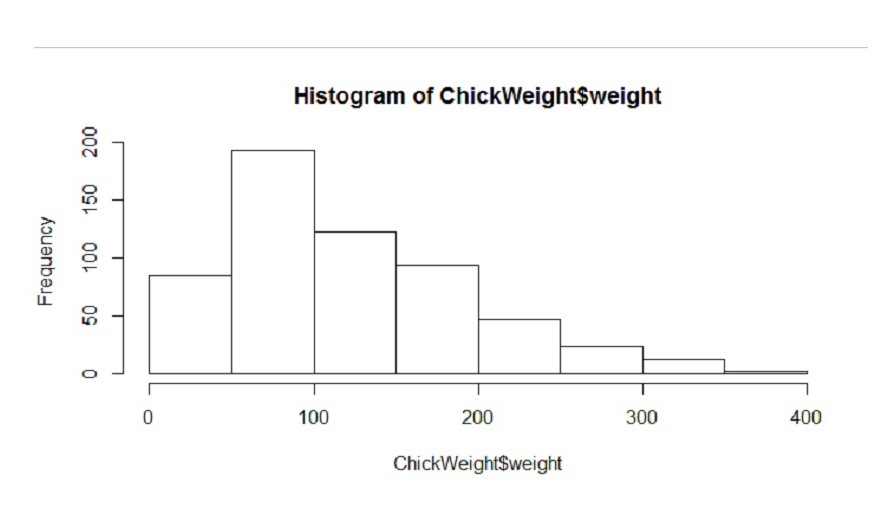
For SP Skewness = 1.61 kurtosis = 0.95

print('For WT Skewness =', np.round(q9b.SP.skew(),2), 'Kurtosis =',

np.round(q9b.WT.kurt(),2))

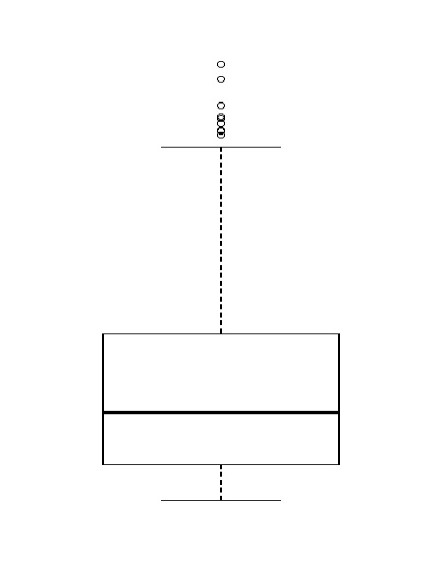
For WT Skewness = 1.61 Kurtosis = 0.95

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



Answer:-

The histogram's peak has a right skew and tail is on right. Mean > Median. We have outliers on the higher side.



Answer:-

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?**

Answer: -

Given:

Sample mean (x): 200 pounds

Standard deviation (s): 30 pounds

Sample size (n): 2000 men

We'll use the formula for the margin of error (

ME = Z \* s/√n

And then calculate the lower and upper bounds of the confidence intervals:

Lower Bound = x - ME

Upper Bound = X + ME

For a 94% confidence interval:

Z for 94% confidence level is approximately 1.88.

ME = 1.88 \* 30/√2000 = 1.261

Lower Bound = 200 - ME = 200 - 1.261= 198.74

Upper Bound = 200 + ME= 200 + 1.261= 201.26

For a 96% confidence interval:

Z for 96% confidence level is approximately 2.05.

ME = 2.05 \* 30/√2000 = 1.375

Lower Bound = 200 - ME = 200 - 1.375= 198.625

Upper Bound = 200 + ME= 200 + 1.375= 201.375

For a 98% confidence interval:

Z for 98% confidence level is approximately 2.33.

ME = 2.33 \* 30/√2000 = 1.56

Lower Bound = 200 - ME = 200 - 1.56= 198.44

Upper Bound = 200 + ME= 200 + 1.56= 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**

**⦁** **Find mean, median, variance, standard deviation.**

Answer: -

Mean (Average):

Mean = ∑Scores / Number of Scores

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

Mean = 738 / 18 = 41

Median:

Arrange the scores in ascending order: 34, 36, 36, 38, 38, 39, 39, 40, 40, 41, 41, 41, 41, 42, 42, 45, 49, 56.

9th = 40 , 10th= 41

Median = (9th = 10th) / 2 = (40+41) = 40.5

Variance:

Variance = ∑(Score−Mean) ^2 / Number of Scores

Variance = [(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)^2] / 18 = 24.111

Standard Deviation:

Standard Deviation = √Variance

Standard Deviation = √24.11 = 4.91

**⦁** **What can we say about the student marks?**

Answer: -

We don’t have outliers and the data is slightly skewed towards the right because mean is greater than median.

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

Answer: -

No skewness is present, we have a perfect symmetrical distribution.

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

**:-**

Skewness and tail is towards Right

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

**:-**

Skewness and tail is towards left

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

**:-**

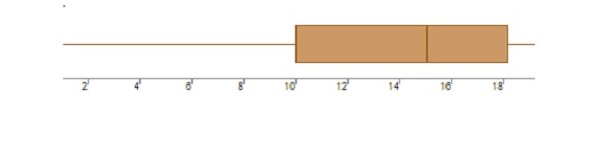
Positive kurtosis means the curve is more peaked and it is Leptokurtic

**Q17) What does negative kurtosis value indicate for a data?**

**:-**

Negative Kurtosis means the curve will be flatter and broader.

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



**What can we say about the distribution of the data?**

Answer: -

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

**What is the nature of skewness of the data?**

Answer: -

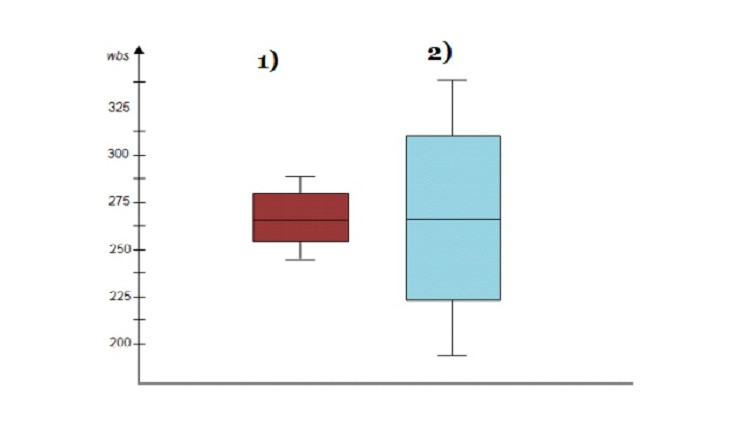
The data is a skewed towards left. The whisker range of minimum value is greater than maximum.

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

Answer: -

The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 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.**

Answer: -

: 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**

**⦁** **P(MPG>38)**

Answer: -

Prob\_MPG\_greater\_than\_38 = np.round(1 - stats.norm.cdf(38, loc= q20.MPG.mean(), scale= q20.MPG.std()),3)

print('P(MPG>38)=',Prob\_MPG\_greater\_than\_38)

P(MPG>38)= 0.348

⦁ **P(MPG<40)**

Answer: -

prob\_MPG\_less\_than\_40 = np.round(stats.norm.cdf(40, loc = q20.MPG.mean(), scale = q20.MPG.std()),3)

print('P(MPG<40)=',prob\_MPG\_less\_than\_40)

P(MPG<40)= 0.729

⦁ **P (20<MPG<50)**

Answer: -

prob\_MPG\_greater\_than\_20 = np.round(1-stats.norm.cdf(20, loc = q20.MPG.mean(), scale = q20.MPG.std()),3)

print('p(MPG>20)=',(prob\_MPG\_greater\_than\_20))

p(MPG>20)= 0.943

prob\_MPG\_less\_than\_50 = np.round(stats.norm.cdf(50, loc = q20.MPG.mean(), scale = q20.MPG.std()),3)

print('P(MPG<50)=',(prob\_MPG\_less\_than\_50))

P(MPG<50)= 0.956

prob\_MPG\_greaterthan20\_and\_lessthan50= (prob\_MPG\_less\_than\_50) - ( (prob\_MPG\_greater\_than\_20)

print('P(20<MPG<50)=',(prob\_MPG\_greaterthan20\_and\_lessthan50))

P(20<MPG<50)= 0.013000000000000012

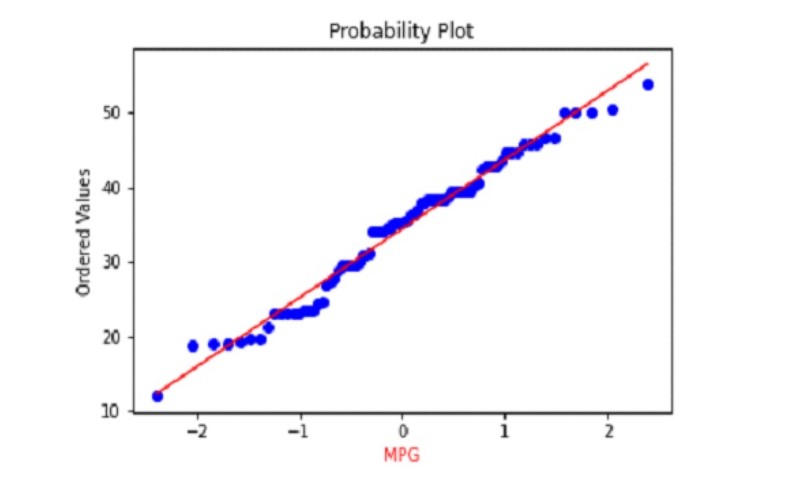
**Q 21) Check whether the data follows normal distribution**

**⦁Check whether the MPG of Cars follows Normal Distribution**

**Dataset: Cars.csv**

Answer: -

⦁ MPG of cars follows normal distribution

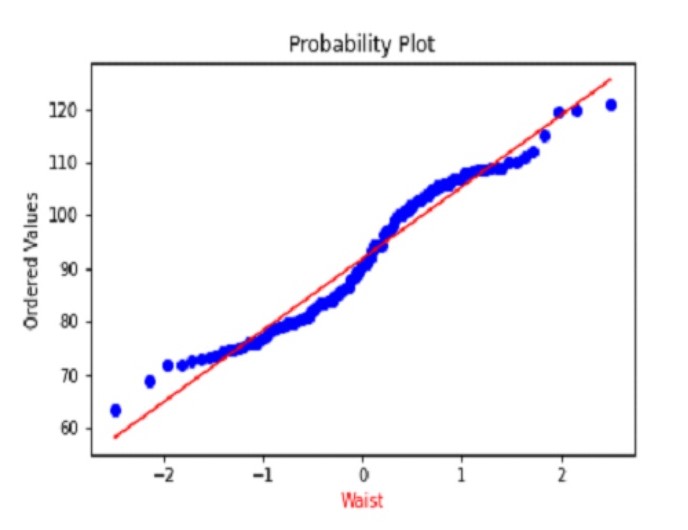


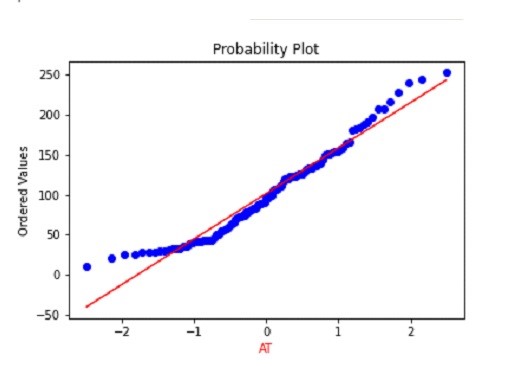
⦁**Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution**

**Dataset: wc-at.csv**

Answer: -

Adipose Tissue (AT) and Waist does not follow Normal Distribution





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

Ans:

# z value for 90% confidence interval

print('Z score for 60% Conifidence Intervla =',np.round(stats.norm.ppf(.05),4))

Z score for 60% Conifidence Intervla = -1.6449

# z value for 94% confidence interval

print('Z score for 60% Conifidence Intervla =',np.round(stats.norm.ppf(.03),4))

Z score for 60% Conifidence Intervla = -1.8808

# z value for 60% confidence interval

print('Z score for 60% Conifidence Intervla =',np.round(stats.norm.ppf(.2),4))

Z score for 60% Conifidence Intervla = -0.8416

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

Answer: -

# t score for 95% confidence interval

print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.025,df=24),4))

T score for 95% Confidence Interval = -2.0639

# t value for 94% confidence interval

print('T score for 94% Confidence Inteval =',np.round(stats.t.ppf(0.03,df=24),4))

T score for 94% Confidence Inteval = -1.974

# t value for 99% Confidence Interval

print('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.005,df=24),4))

T score for 95% Confidence Interval = -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**

Answer: -

import numpy as np

Import scipy as stats

t\_score = (x - pop mean) / (sample standard daviation / square root of sample size)

(260-270)/90/np.sqrt(18))

t\_score = -0.471

stats.t.cdf(t\_score, df = 17)

0.32 = 32%