|  |  |
| --- | --- |
| 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) | Discrete |

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 | Discrete - Nomial |
| High School Class Ranking | Discrete - Nomial |
| Celsius Temperature | Continuous - Interval |
| Weight | Continuous - Ratio |
| Hair Color | Discrete - Ratio |
| Socioeconomic Status | Continuous - Interval |
| Fahrenheit Temperature | Continuous - Ratio |
| Height | Continuous - Ratio |
| Type of living accommodation | Discrete - Ordinal |
| Level of Agreement | Discrete - Interval |
| IQ(Intelligence Scale) | Discrete - Interval |
| Sales Figures | Discrete - Interval |
| Blood Group | Discrete - Ratio |
| Time Of Day | Continuous - Interval |
| Time on a Clock with Hands | Continuous - Interval |
| Number of Children | Discrete - Interval |
| Religious Preference | Discrete - Ratio |
| Barometer Pressure | Continuous - Ratio |
| SAT Scores | Continuous - Ratio |
| Years of Education | Discrete - Nominal |

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

Answer: The total possible outcome 23=8

HHH,HHT,HTT,THT,TTH,HTH,THH,TTT

Number of favorable outcomes =3

P(two head and one tail)=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

Answer:

1. Total number of outcomes=36

number of outcomes sum equals to 1 = 0

p(equal to 1) = 0/36 = 0

1. number of Outcomes lessthan or equal to 4 =6

P(<= 4) = 6/36 = 1/6  
c) number of outcomes Sum is divisible by 2 and 3 = 6  
 p(sum /2&3) = 6/36 = 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?

Answer:

There are 7 balls originally with 2 of them blue so the probability of the first ball not being blue is 5/7. This leaves 6 balls with 2 blue. The probability of the second ball not being blue assuming that the first wasn’t is 4/6. The probability that neither ball drawn was blue is (5/7)\*(4/6)=20/42=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:

Child A –probability of having 1 candy = 0.015.

Child B –probability of having 4 candies = 0.20

Expected number of candies for randomly selected child = (1\*0.015)+ (4\*0.20)+ (3\*0.65)+ (5\*0.005)+ (6\*0.01) +(2\*0.120) = 3.085

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

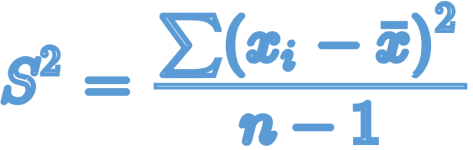
* For Points,Score,Weigh>

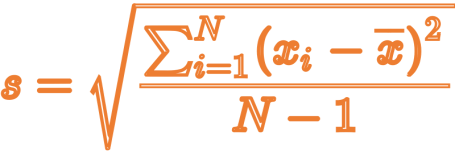
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

Answer:

Formulas:

1. Mean: Sum of terms/number of terms
2. median : the middle obersation of the data.
3. variance: 

4. Standard Devation : 

Points: Mean =3.596563, Median= 3.695, Mode= “numeric”,Variance= 0.2858814, Standard deviation= 0.5346787.

Score: Mean= 3.21725, Median= 3.325, Mode= “numeric”,Variance= 0.957379, Standard deviation= 0.9784574

**Note: Mean value are closer for both ‘Point’ and ‘Score’.**

Weight: Mean= 17.84875, Median= 17.71, Mode= “numeric”,Variance= 3.193166, Standard deviation= 1.786943

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:

Expected Value  =  ∑ ( probability  \* Value )

 ∑ P(x).E(x)

there are 9 patients

Probability of selecting each patient = 1/9

E(x)  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

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**

Skewness for speed= -0.117510, skewness value is negative so it is left skewed. Since magnitude is slightly greater than 0 it is slightly left skewed.

And for distance= 0.806895, right skewed (Positive) slight magnitude to right.

File name: Assignment-1-Q9\_a&b.ipynb

**SP and Weight(WT)**

**Use Q9\_b.csv**

Skewness for SP= 1.611450, skewness value is postive so it is right skewed. Since magnitude is slightly greater than 0 it is slightly right skewed.

And for WT= -0.614753, Left skewed (negitavie) slight magnitude to left.

File name: Assignment-1-Q9\_a&b.ipynb

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



Sol:

The most of the data points are concerated in the range 50-100 with frequency 200.And least range of weight is 400 somewere around 0-10.So the expected value the above distribution is 75.Skewness- we can notice a long tail towards right so it is heavily right skewed



 Sol: Median is less than mean right skewed and we have outlier on the upper side of box plot and there is less data points between Q1 and bottom point.

**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?

Sol: X+/-(Z1-α.σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

Confidence interval= 94%

(1-σ/2)= 1-0.03) =0.97

for confidene interval for 94% is 1.882

Confidence interval for 98%= 2.33

Confidence interval for 96% = 2.05

**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?

Sol: Mean= 41, Median= 40, variance= 24.111, Standard deviation= 4.910

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

Sol: Symetrical

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

Sol: Right Skewed

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

Sol: Left Skewed

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

Sol: The data is notmally distributed and kurtosis value is 0.

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

Sol: The distribution of the data has lighter tails and a flatter peaks than the normal distribution.

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



What can we say about the distribution of the data?

Sol: Let’s assume above box plot is about age’s of the students in a school. 50% of the people are above 10 yrs old and remainig are less.

  And students who’s age is above 15 are approx 40%.

What is nature of skewness of the data?

Sol: Left skewed, median is greater than mean.

What will be the IQR of the data (approximately)?   
Sol: Apporximately = -8  
  
Q19) Comment on the below Boxplot visualizations?



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

Sol: By observing both the plots whisker’s level is high in boxplot 2, mean and median are equal hence distribution is symetrical.

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)

c. P (20<MPG<50)

**Sol**: P(MPG>38) == 0.34759392515827137

P(MPG<40) == 0.7293498762151609

P (20<MPG<50) == 1.2430968797327491e-05

**Note: File Attached: Assignment-1-Q20 .ipynb**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol: File:- Assignment-1-Q21 a & b.ipynb

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

Dataset: wc-at.csv

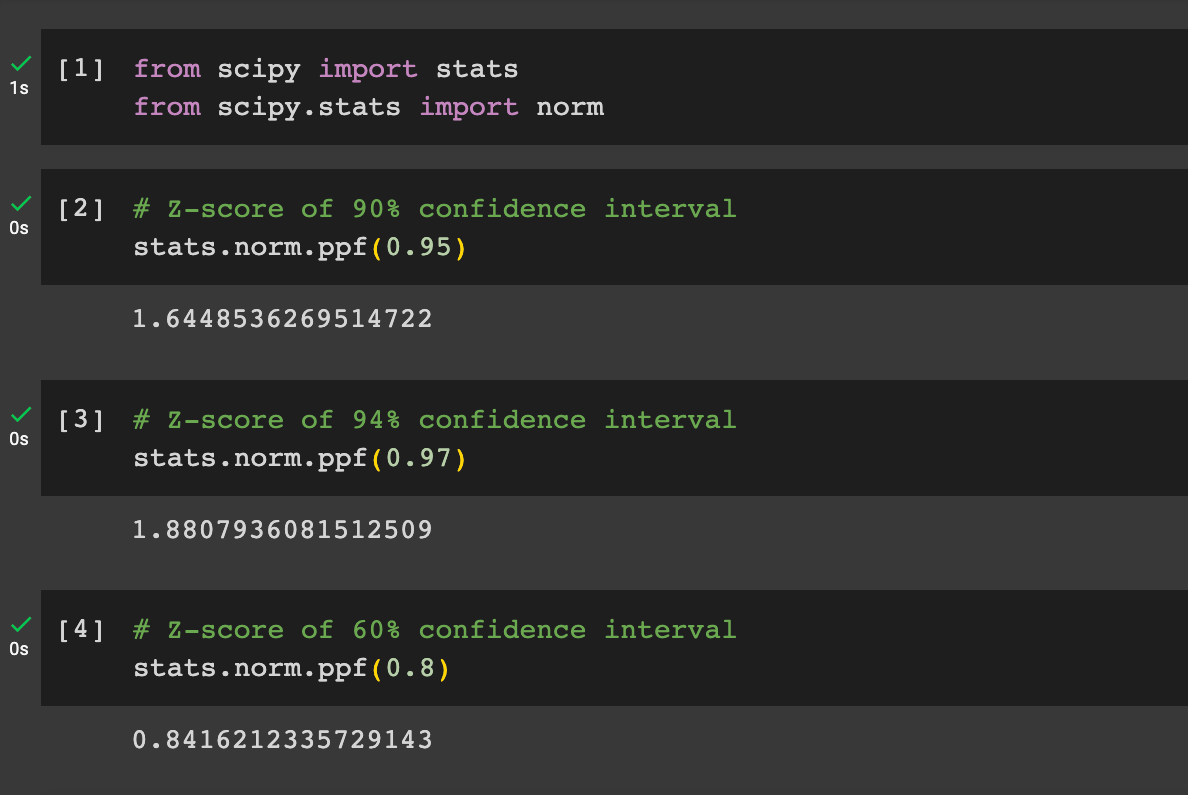
Sol: File:- Assignment-1-Q21 a & b.ipynb

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

Sol: 90%== 1.6448536269514722

94%== 1.8807936081512509

60%== 0.8416212335729143

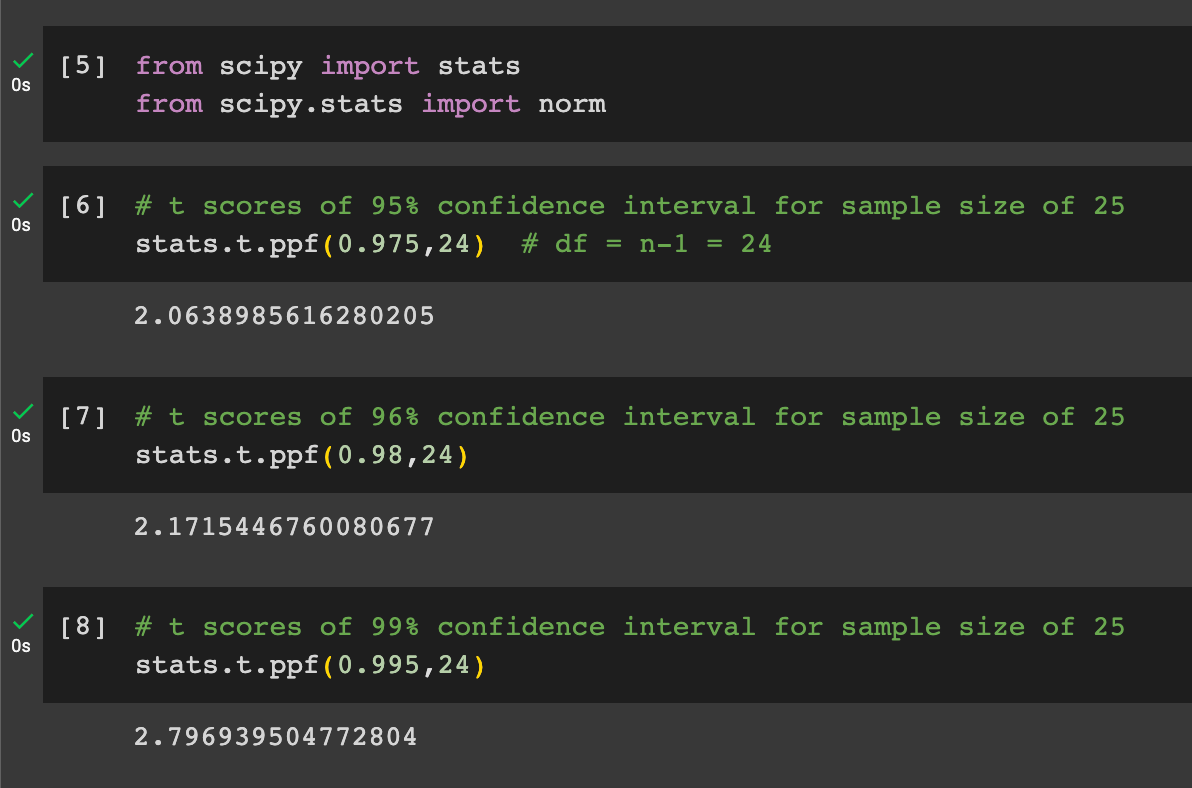


Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Sol: 95% == 2.0638985616280205

96%== 2.1715446760080677

99%== 2.796939504772804



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

Sol:

