|  |  |
| --- | --- |
| 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 | 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 | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Answer-

Event- 3 Coins are tossed. Total Possible outcomes are as below:

(H, H, H), (H, H, T), (H, T, H), (T, H, H), (T, T, H), (T, H, T), (H, T, T), (T, T, T)

Total Number of outcomes = 8

Outcomes of two heads and one tail are obtained are as below:

(H, H, T), (H, T, H), (T, H, H)

Total Number of Outcomes of two heads and one tail are obtained = 3

probability = (No. of outcomes of two heads & one tail)/(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

Answer-

Total outcomes for the event= [ (1,1), (1,2), (1,3), (1,4), (1,5), (1,6), (2,1), (2,2), (2,3), (2,4), (2,5), (2,6), (3,1), (3,2), (3,3), (3,4), (3,5), (3,6), (4,1), (4,2), (4,3), (4,4),(4,5),(4,6), (5,1),(5,2),(5,3),(5,4),(5,5),(5,6), (6,1),(6,2),(6,3)(6,4),(6,5),(6,6)]

Total Number of outcomes = 36

Probability for :

1. Sum Equal to 1

No such Outcome found so Probability equal to **Zero**

1. Sum Less than or equal to 4

Outcomes= [ (1,1), (1,2), (1,3), (2,1), (2,2), (3,1),]

Total = 6

Probability = 6/36 = 1/6 = **0.167**

1. Sum is divisible by 2 and 3

Outcomes = [ (1,5), (2,4), (3,3), (4,2), (5,1), (6,6)]

Total = 6

Probability = 6/36 = 1/6 = **0.167**

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 no. of balls = 7

Total no. of events = 7c2 = (7\*6)/(2\*1) = 21

No. of events that ball is not Blue = 5c2 = (5\*4) / (2\*1) = 10

Probability = (10/21) **= 0.48**

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 no. of Candies = ∑ Xi \* Pi = (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

Answer-

Mean (Points) = 3.596563

Mean (Score) = 3.217250

Mean (Weigh) = 17.84875

Median (Points) =3.695

Median (Score) =3.325

Median (Weigh)= 17.710

Variance (Points)=0.285881

Variance (Score) =0.957379

Variance (Weigh) =3.193166

Std (Points)= 0.534679

Std (Score) = 0.978457

Std (Weigh) = 1.786943

Range (Points)= (2.76 - 4.93)

Range (Score) = (1.513 - 5.424)

Range (Weigh)= (14.5 - 22.9)

**Comment:** The mean & median values are close to each other. It means that values are close to the middle value of the range and there is no skewness in the graph and no possible outliers in the dataset.

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?

Answer : Probability of one of the patient chosen at random = 1/9

Expected Value = (108+110+123+134+135+145+167+187+199)\* 1/9 = **145.33**

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

**Cars speed and distance**

**Use Q9\_a.csv**

skewness(Q9\_a$speed) = -0.117510

skewness(Q9\_a$dist) = 0.806895

kurtosis(Q9\_a$speed) = 2.422853

kurtosis(Q9\_a$dist) = 3.248019

Comments-Negative skewness indicates that, slope is on left hand side,& positive skewness indicates that slope is on right hand side.

The values of skewness of speed & dist indicates that, slope of curve of dist is more as compared to the speed.

Also the values of kurtosis of both indicates that , both have outliers on the upper extreme

**SP and Weight(WT)**

**Use Q9\_b.csv**

**skewness(Q9\_b SP) = 1.611450**

**skewness(Q9\_b WT) = -0.614753**

**kurtosis(Q9\_b SP) = 2.977329**

**kurtosis(Q9\_b WT) = 0.95291**

Comments- Negative skewness indicates that, slope is on left hand side,& positive skewness indicates that slope is on right hand side.

The values of skewness of SP & Weight indicates that, slope of curve of SP is more as compared to the Weight.

Also the values of kurtosis of both indicates that , both have outliers on the upper extreme

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



Comments: The Histogram & Boxplot Shows that, most of the values of weight are in between 0-300.

Only some values are more than 300, it shows that the values which are greater than 300 are called as **Outliers**. Also the slope of curve on Histogram is on the right hand side which means that, the skewness of the data is **Positive.**

**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 : N=30,00,000, n=2000, X‾ = 200, s=30, 1-α= 0.97,0.99,0.98

Here, µ(std. devian of population) is unknown, so we use T test.

Confidence Interval = X‾ +/- t(1-α) (n-1) \*s/sqrt(n)

For 1-α=0.97 & n-1 = 1999

So in python we can calculate t(1-α) (n-1) = 1.88

So C.I. = 200+/-1.88(30/sqrt(2000))=200+/-1.26

= 198.74 – 201.26

For 1-α=0.98 & n-1 = 1999

So in python we can calculate t(1-α) (n-1) = 2.06

So C.I. = 200+/-2.06(30/sqrt(2000))=200+/-1.38

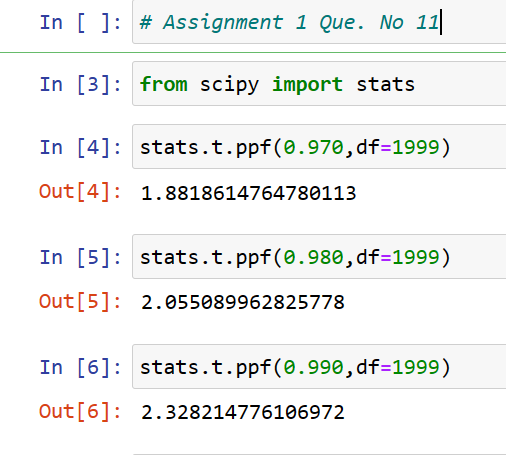
= 198.62 – 201.38

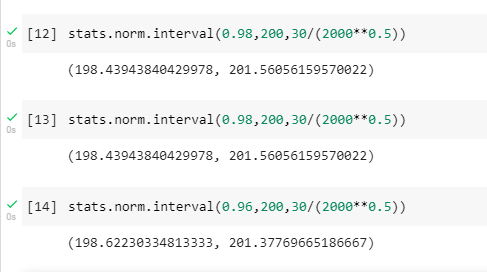
For 1-α=0.97 & n-1 = 1999

So in python we can calculate t(1-α) (n-1) = 2.33

So C.I. = 200+/-2.33(30/sqrt(2000))=200+/-1.56

= 198.44 – 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.
2. What can we say about the student marks?

s=c(34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56)

Mean(s) = 41

Median(s) = 40.5

Var(s) = 25.52

SD(s) = 5.05

Comments:

The average scores obtained by students is 41. Also most of the students obtained scores within range of 34-45.

It means that only 2 students got higher marks which can be termed as Outliers.

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

**Ans**- When Mean and Median of data is equal , the given data is normally distributed and there is **no skewness present** in the data and the curve of Skewness is **symmetric.**

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

**Ans**- When mean of data is greater than median of data , then the skewness of data is Positive. It implies that the slope of curve of Skewness is on the right hand side. Also there are Outliers presents on the **Upper Extreme**.

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

**Ans**- When mean of data is less than median of data , then the skewness of data is Negative. It implies that the slope of curve of Skewness is on the left hand side. Also there are Outliers presents on the Lower Extreme.

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

**Ans**- The Positive Kurtosis value indicates that, the values of data are above the curve of data. It means that it has Outliers above the Upper Extreme.

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

**Ans-** The Negative Kurtosis value indicates that, the values of data are below the curve of data. It means that it has Outliers above the Lower Extreme.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

**Ans**-

1. The Boxplot shows that most of the values of data are in between 10- 18.

The data has values which are less than 10, which is also called as Lower Extreme.

2. The Value of Skewness of data is Positive. The Curve has slope on the right hand side.

The IQR of Data : 10 - 18

Q19) Comment on the below Boxplot visualizations?



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

**Ans-** Both the boxplot have IQR. Boxplot 1 has 50 % values which are in between 250-275 & Boxplot 2 has 50% values which are in between 225-300.

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)

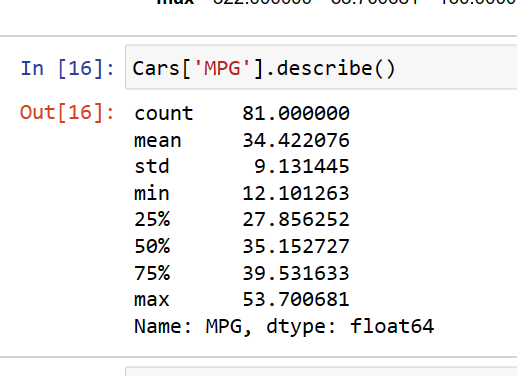
c. P (20<MPG<50)

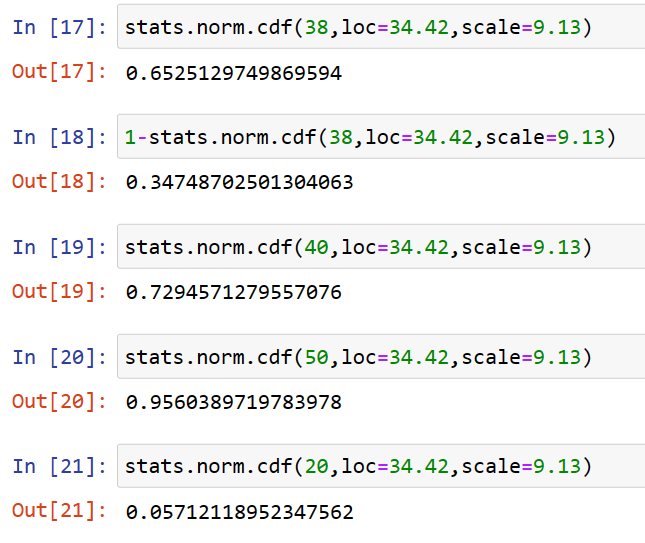
Answer : From python we can calculate mean= 34.42 , std deviation = 9.13

a. P(MPG>38) = 0.3474

b. P(MPG<40)= 0.7295

c. P(20<MPG<50) = P(MPG<50)- P(MPG<20) = 0.9560-0.0571 = 0.8989





Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: We consider sample of 10 values.

So by standardization we get µ = -6.7 E^16 & sigma= 1

So, it follows 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

Answer: We consider sample of 10 values of Adipose Tissue (AT).

So by standardization we get µ = -2.05 E^16 & sigma= 1

So, it follows Normal Distribution.

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

1. For 90% CI 1-α = 0.95

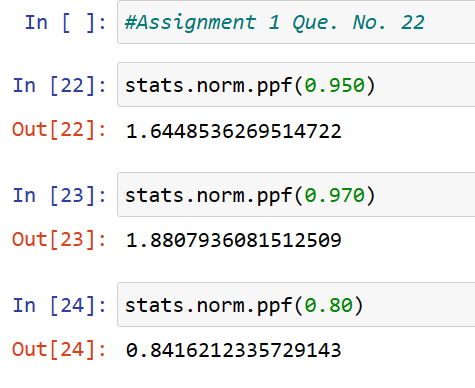
**Z Score = 1.64**

1. For 94% CI 1-α = 0.97

**Z Score = 1.88**

1. For 60% CI 1-α = 0.8

**Z Score = 0.84**



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

Answer:

1. For 95% CI

1-α = 0.975 & n-1=24

t Score = 2.06

1. For 96% CI

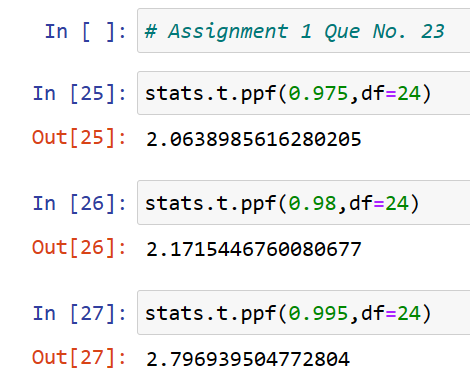
1-α = 0.98 & n-1 = 24

t Score = 2.17

1. For 99% CI

1-α = 0.995 & n-1 = 24

t Score = 2.79



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

Given : n= 18 , x‾ = 260 , S = 90 , µ = 270 , df = 17

Answer-

Here, Sigma not given. So we used T test.

T= (x‾ - µ)/(S/sqrt(n)) = (260-270)/(90/sqrt18) = -0.4714

So from python : Probability = 0.32167

