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

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

Ans. Three coin tossed so sample space S

S= {HHH,HHT,HTH,THH,TTH,THT,HTT,TTT}

n(S) = 8

let A be the set of two heads and one tail.

n(A) = 1

P(A) = n(A)/n(S)

=1/8

probability that two heads and one tail are obtained is ‘1/8’

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. Two dice are thrown simultaneously then number of observations can be 36.

Then sample space S={ *(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)}*

*n(S)=36*

1. *Let A = Sum is* equal to 1.

*n(A)=0*

*P(A) = n(A)/n(S)*

*= 0/36*

*P(A) = 0*

Probability that sum is equal to 1 is ‘0’.

1. Let B = Sum is Less than or equal to 4

n(B) = 6

P(B) = n(B)/n(S)

=6/36

P(B) =1/6

Probability that sum is less than or equal to 4 is ‘1/6’

1. Let C = Sum is divisible by 2 and 3

n(C)=6

P(C) = n(C)/n(S)

=6/36

P(C) =1/6

Probability that Sum is divisible by 2 and 3 is ‘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?

Ans. Total number of balls = 7.

Let S be the sample space.

Then, n(S) = Number of ways of drawing 2 balls out of 7

= 7 C 2

​ = (7!)/((7-2)! \* 2!) ​

=21

Let E = Event of drawing 2 balls, none of which is blue.

∴n(E)= Number of ways of drawing 2 balls out of 5 balls.

n(E)= 5 C 2

​ = (5!)/((5-2)! \* 2!) ​

​ =10

∴P(E)= n(E)/n(S)

=10/21

The probability that none of the balls drawn 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:**

Expected number of candies for a randomly selected child

=  1 \* 0.015  + 4\*0.20  + 3 \*0.65  + 5\*0.005  + 6 \*0.01  + 2 \* 0.12

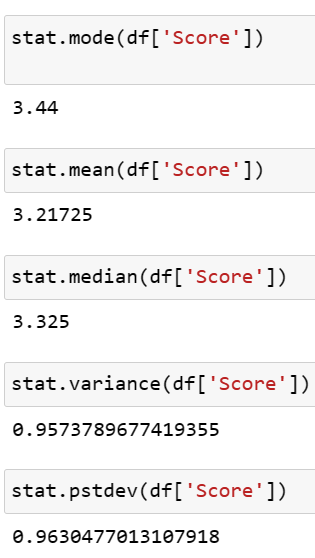
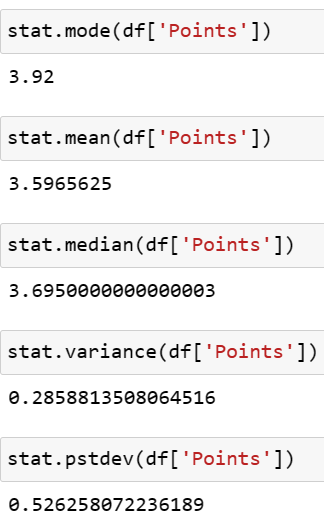
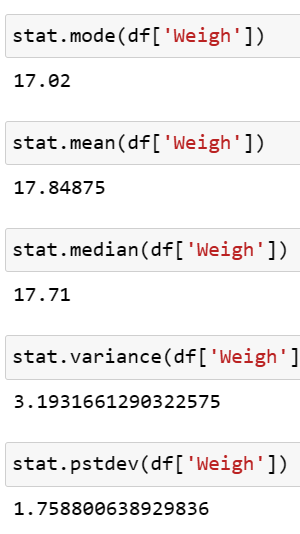
= 0.015 + 0.8  + 1.95 + 0.025 + 0.06 + 0.24

=  3.090

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.



* In ‘Weigh’ column Mean >median>mode the data is slightly right Skewed. From Std deviation we can say data is not spreaded
* In ‘Points’ column mode>median>mean so the data is left skewed. From std deviation we can say data is spreaded
* In ‘Score’ column mode>median>mean so the data is left skewed

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. Total number of patients= 9

Probability of choosing one patient =1/9

E(X) =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)

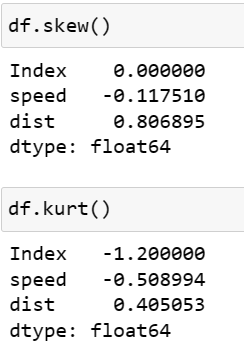
E(X) = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans.**

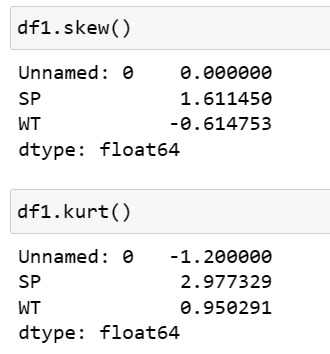
****

* Speed column data gives us almost symmetric distribution and negative kurtosis means data distribution is flatter than normal.
* Distance column data gives us moderated skewed distribution and positive kurtosis means distribution is more peaked than normal.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans.**

****

* SP column data gives us highly skewed distribution and positive kurtosis means distribution is more peaked than normal
* Weight column data gives us moderated skewed distribution and negative kurtosis means data distribution is flatter than normal.

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



Ans. The distribution of Chick weight is right or positive skewed. On the basis of skewedness we get an intuition of presence of outliers



Ans. The distribution of Given Box plot is right or positive skewed. On the basis of skewedness and we can clearly sees that some points are present above upper limit means presence of outlier.

**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**. n=2000, x-bar=200, SD=30

**94% Confidence Interval:**

Interval Estimate = x-bar ± Z\*Sd/sqrt(n)

=200 ± 1.88\*30/sqrt(2000)

**C.I.=[198.74 , 201.26]**

**98% Confidence Interval:**

x-bar = 200

Sd = 30

n = 2000

Interval Estimate = x-bar ± Z\*Sd/sqrt(n)

=200 ± 2.33\*30/sqrt (2000)

**C.I.=[198.44,201.56]**

**96% Confidence Interval:**

x-bar = 200

Sd = 30

n = 2000

Interval Estimate = x-bar ± Z\*Sd/sqrt(n)

=200 ± 2.05\*30/sqrt (2000)

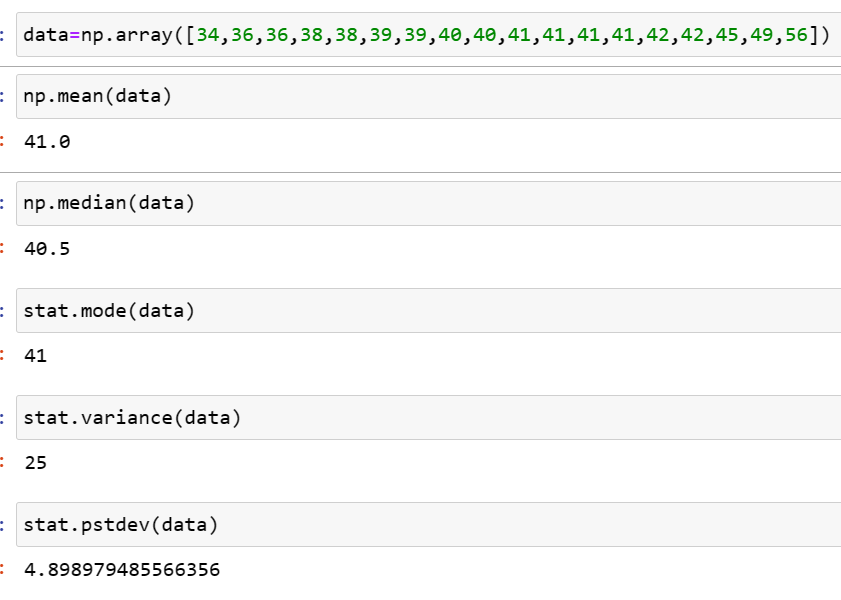
**C.I=[198.62,201.38]**

**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.

1. 

2.**Mean > Median, This implies that the distribution is slightly skewed towards right. No outliers are present.**

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

**Ans. When mean ,median and mode is equal then data is symmetric in nature**

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

**Ans. When mean>median then data distribution is positive or right skewed in nature**

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

**Ans. When mean<median then data distribution is negative or left skewed in nature**

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

**Ans. A positive value for the kurtosis indicates a distribution more picked than normal**

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

**Ans. A negative kurtosis indicates a shape flatter than normal**

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



What can we say about the distribution of the data?

**Ans. This distribution is asymmetric or not Normal**

What is nature of skewness of the data?

**Ans. Above data left or negative skewed**

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

**Ans. IQR=Q3-Q1**

**=18-10**

**IQR=8 approximately**

Q19) Comment on the below Boxplot visualizations?



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

**Ans. Above both box plots are seems to be normally distributed Mean and mode of both boxplot is same.**

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

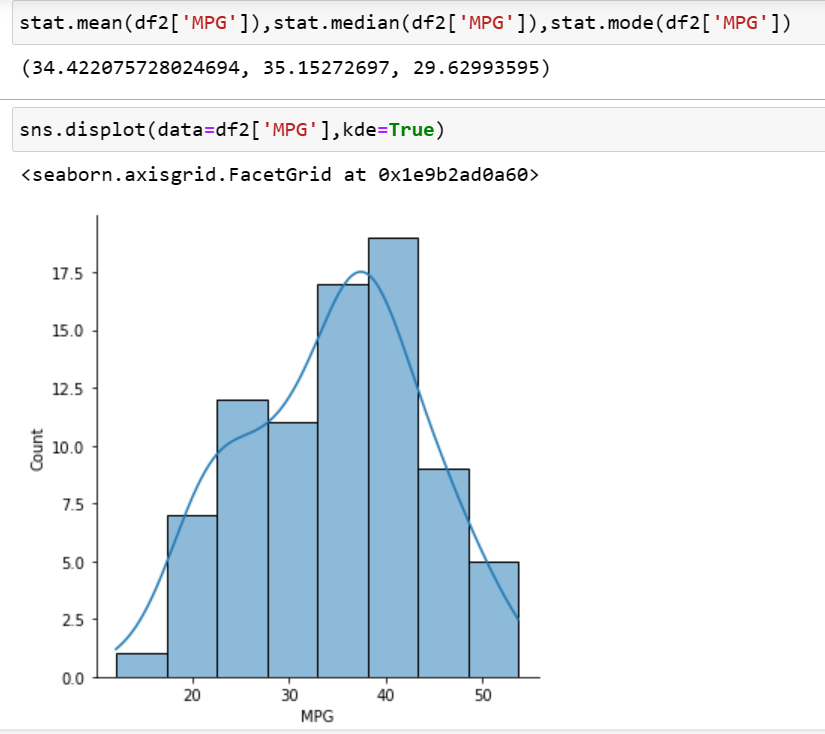
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

Ans.

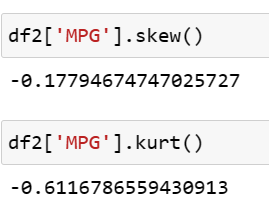
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

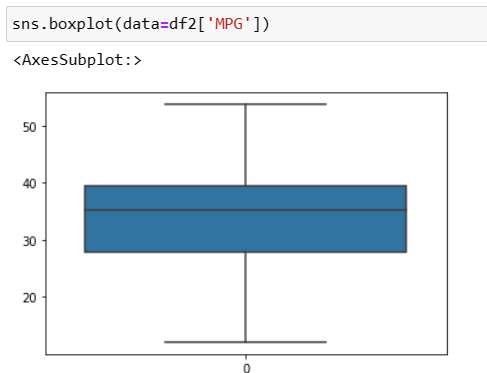


On the basis of mean, median and mode this distribution is near to normal.



Skewness of above distribution is between -0.5 to 0.5 so we can say it is near to normal distribution.

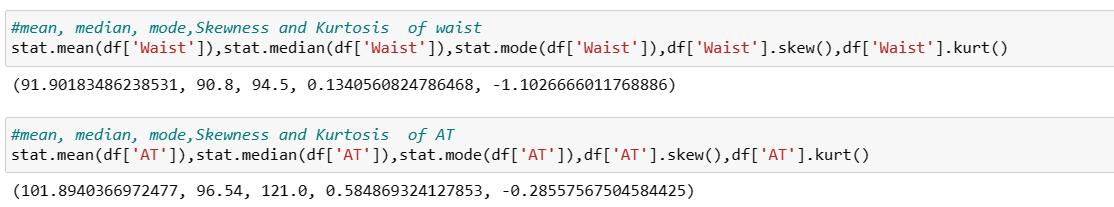
We can check by using Box plot

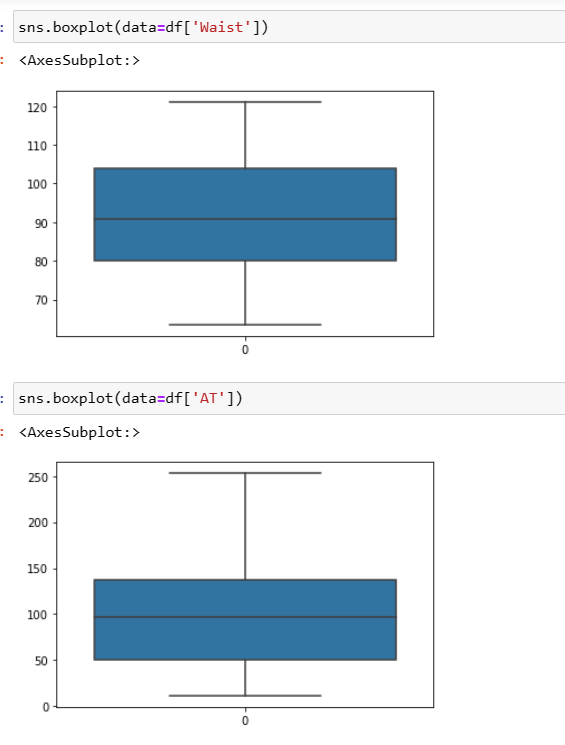
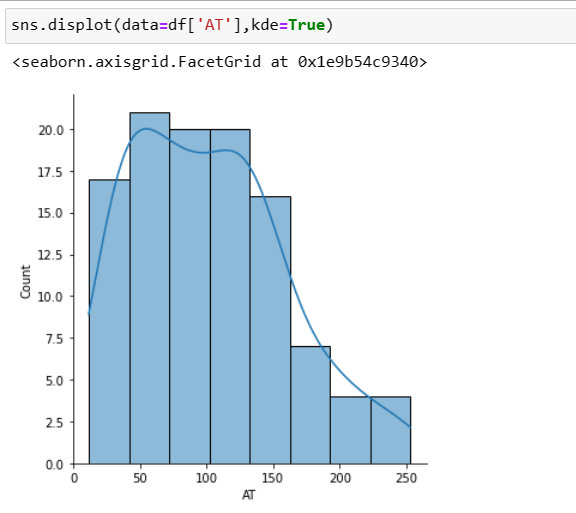
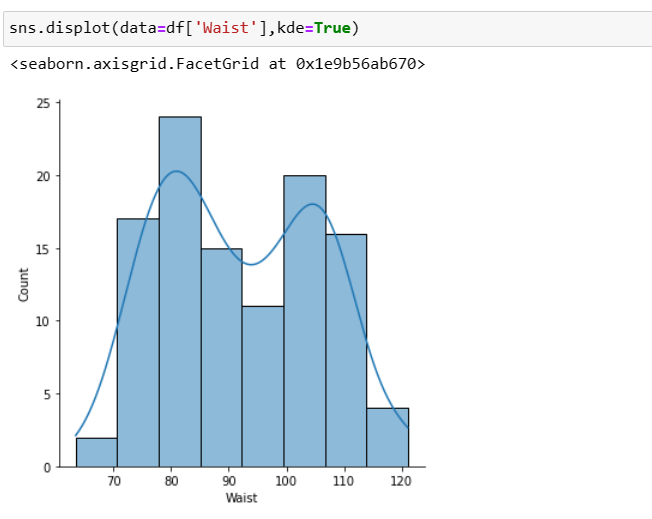


Median is slightly shifted but both whiskers are same length so we can say it is slightly normal

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

Dataset: wc-at.csv



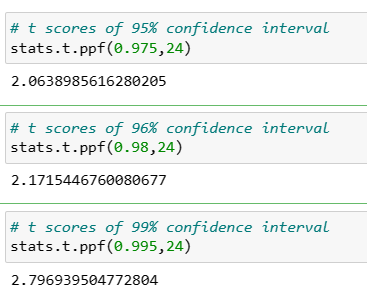


**On the basic of box plot we can say that ‘Waist’ column distribution is normal and ‘AT’ is slightly right skewed**

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



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



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

