|  |  |
| --- | --- |
| 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) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| 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?

Sample space = 8

P(HHT)= 3/8

NOTE: p(x) = no.of possible events/sample space

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 2and 3

sample space = 36

1. P(x =1) = 0/36 =0
2. P(x<=4) = 6/36
3. P(x=6,12) =6/36

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?

Sample space = 7c2 = 21

Possible event(none of the balls drawn blue) = 5c2 = 10.

P(x = none of the balls drawn blue) = 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)

P(no.of candies for randomly selected child) = (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120) =3.09

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**

mean for points,score,weight = 3.59,3.217,17.848

median for points,score, weigh = 3.695, 3.325, 17.71

mode for points = 3.07,3.92

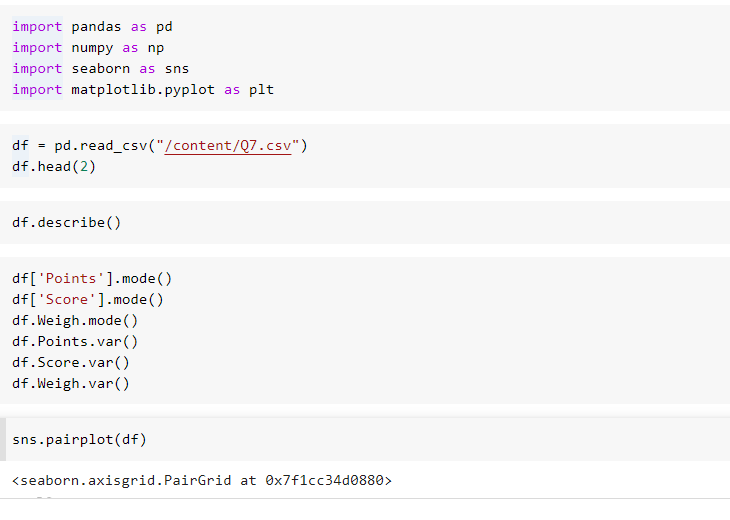
mode for score  = 3.44

mode for weigh  = 17.02,18.90

variance for points, score, weigh = 0.285, 0.957, 3.193

std for points, score, weigh = 0.5346, 0.978, 1.786

Range for points, score, weigh = 2.17, 3.911, 8.4



Inferences:

inferences based on pairplots :

if points are increasing scores are decreasing which means points and score are negatively correlated

inferences based on boxplots:

Points:

 1.there are no outliers in this variable

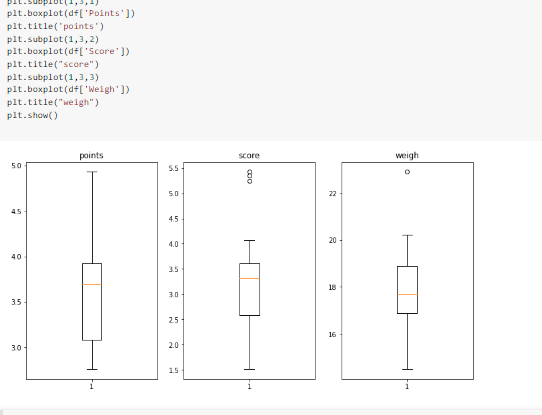
 2.right/postive skewed distribution

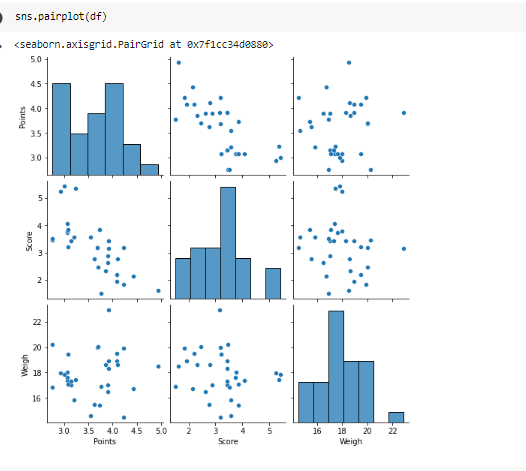
Score:

  1.found three outliers

Weigh:

  1.found 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?

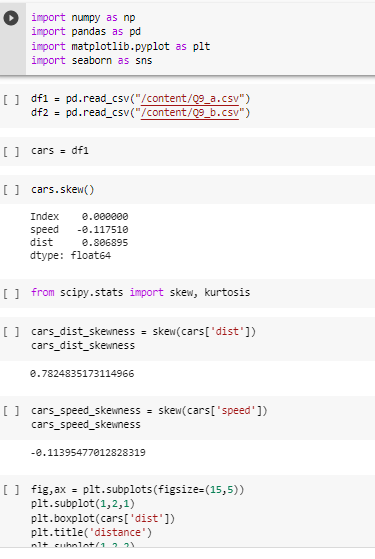
Ans)

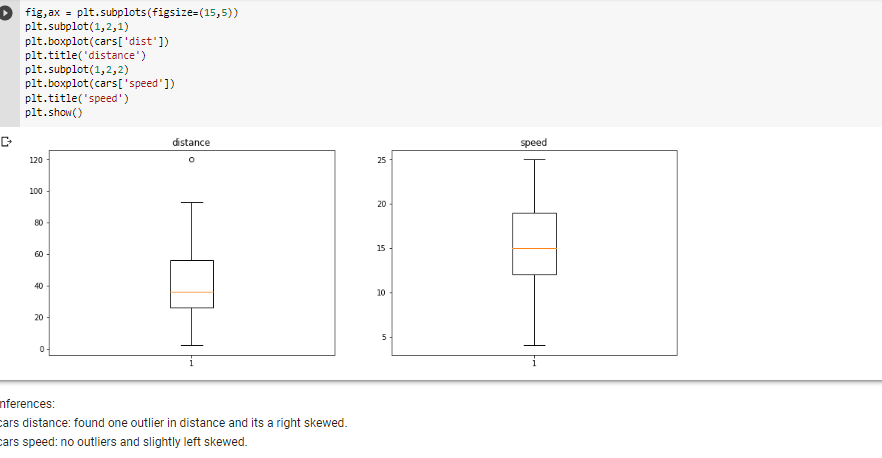
P(x=weight of the randomly chosen patient) = 1/9(108+110+123+134+135+145+167+187+100) = 145.33

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

**Cars speed and distance**

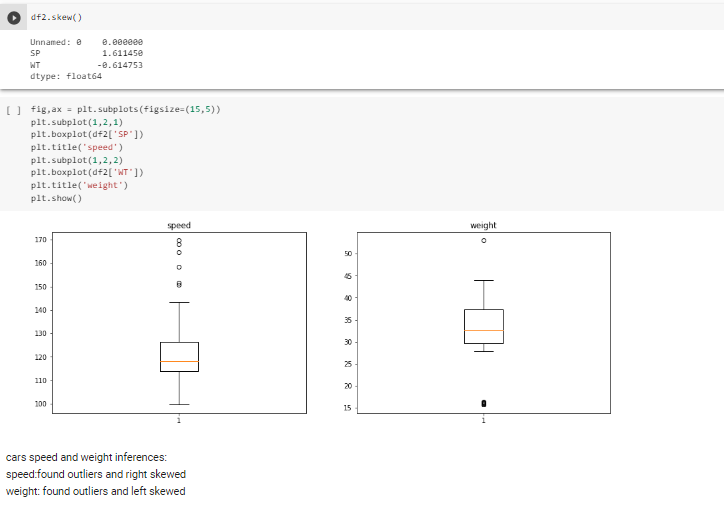
**Use Q9\_a.csv**

****

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

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



Ans) Histogram inferences: right skewed and most of the points lies between 50 to 150.



Ans)

Boxplot inferences: outliers observered on maximum side and right 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)sample size(n) = 2000 population size(N) = 3000000**

Sample avg(x) = 200pounds

Sample std(s) = 30

**Formula for confidence interval:**

**Stats.t.interval(1-alpha,df=n-1,loc=samplemean,scale=samplestd/sqrt(n))**

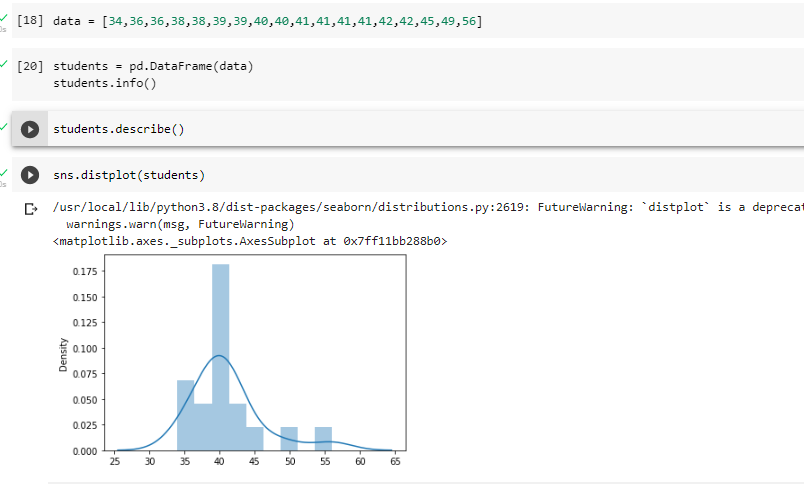
****

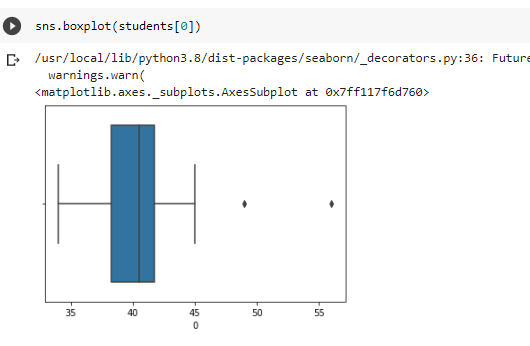
**Q12)**Below are the scores obtained by a student in tests

**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: Mean =41, Median =40.5, Variance =25.52 and Standard Deviation =5.05





we can we say about the student marks is right skewed and found two outliers.

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

Ans) No skewness

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

Ans) right skewed/postive skewness, tail towards right

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

Ans) left skewed/negative skewness, tail towards left

Q16) What does positive kurtosis value indicates for adata ?

Ans) postive kurtosis means high peaked value and variance will be low

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

Ans) negative kurtosis means low peaked value and vairance will be low and the curve will look like flat.

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



What can we say about the distribution of the data?

Ans) it is not symmetric and mean is not equal to median also so we can say distribution is not normally distributed.

What is nature of skewness of the data?

Ans) Left skewed distribution

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

Ans) IQR = upper quartile – 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.

Inferences:

1. The range of boxplot 1 is small with respect to box plot 2
2. 50% quartile(median) is almost same for both box plots
3. No skewness observed in both plots

Q 20) Calculate probability from the given dataset for the below cases

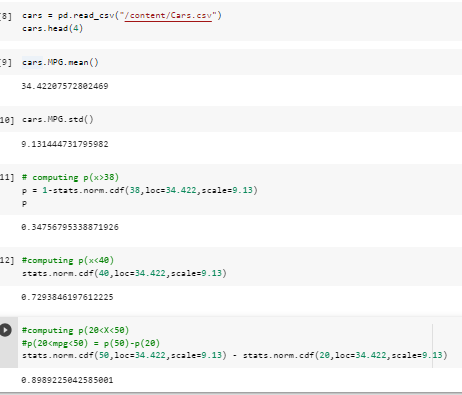
Data \_set: Cars.csv

Calculate the probability of MPG ofCars for the below cases.

MPG<- Cars$MPG

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

c. P (20<MPG<50)



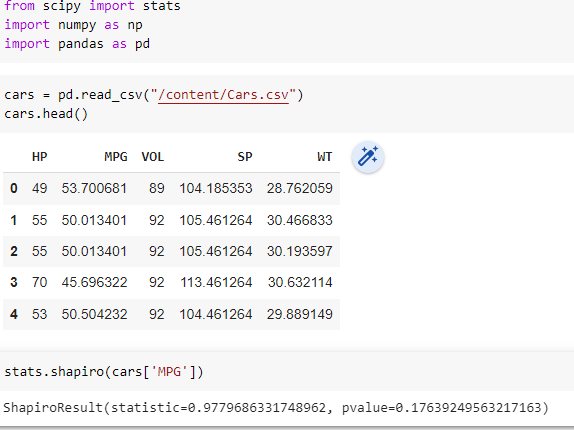
1. P(MPG>38) = 0.3475
2. P(MPG<40) = 0.729

c. P (20<MPG<50) = 0.898

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

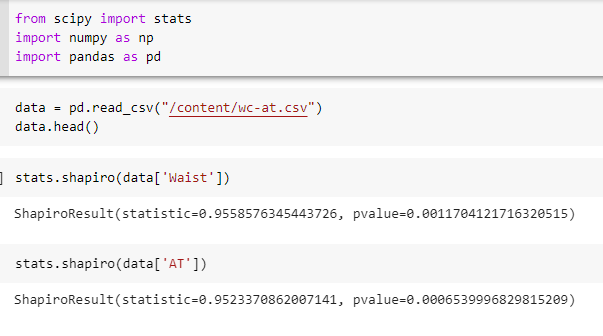


Pvalue > 0.05 which means reject null hypothesis.

Alternative hypothesis is cars[‘MPG’] 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



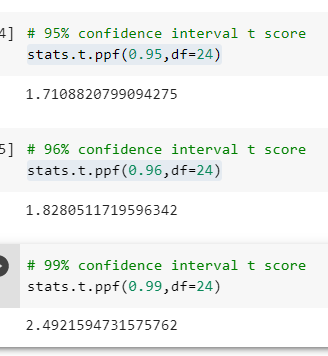
Both Waist and AT p values are less than 0.05

So fail to reject null hypothesis

Null hypothesis = Waist and AT are not follwing normal distribution.

Q 22) Calculate the Z scoresof 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 companyclaims 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) sample size (n) = 18

avg = 260

std = 90

computing critical value(t) why because we don’t know population std.

formula for computing t value:

t =(samplemean – population mean)/(sample std/sqrt(n))

computing p values for critical values, as below

from scipy import stats

p­\_value = stats.t.cdf(t\_value, df = n-1)

p\_value = 0.32 =32%

