# Shweta Jadhav Assignment No.1

**Branch-Andheri (Basic statistics level 1)**

Q1) Identify the Data type for the Following.

|  |  |
| --- | --- |
| 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 |

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 | Ratio |
| Weight | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |

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

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

**Ans:** when the three coins are tossed then the no of possible combination are = 8 P(S) = {HHH, HHT, HTH, THH, TTT, TTH, THT, HTT}

When the three coins are tossed then the probability that two heads and one tail are = 3

P (A) = {HHT, HTH, THH}

Probability = P (A) / P (S) = 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

**Ans:** Where the two Dice are rolled then the probability combination P(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)} = 36

* 1. Equal to 1 :

P (1) = 0/36 = 0

* 1. Less than or equal to 4 :

P(4) = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)} = 6 P (4) = 6/36 = 1/6

* 1. Sum is divisible by 2 and 3 :

P (2&3) = {(1,1),(1,2),(1,3)(1,5), (2,1),(2,2),(2,4)(2,6),

(3,1),(3,3),(3,5),(3,6), (4,2),(4,4),(4,5)(4,6),

(5,1),(5,3),(5,4),(5,5),(6,2),(6,3),(6,4),(6,6)} = 24

P (2&3) = 24/36 = 2/3

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:** we have,

Total no of balls = (2+3+2) = 7

n (s) = number of ways drawing 2 balls out of 7 7C2 = 7\*6 / 2\*1 = 21

n (e) = number of ways drawing 2 balls out of (2+3) 5C2 = 5\*4/2\*1 = 10

P (E) = n (e) / n (s) = 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 one candy = 0.015. Child B – probability of having 4 candies = 0.20 **Ans:** We have,

Child A – probability of having one candy = 0.015. Child B – probability of having 4 candies = 0.20

The Expected number of candies for a randomly selected child will be

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

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

= 3.09

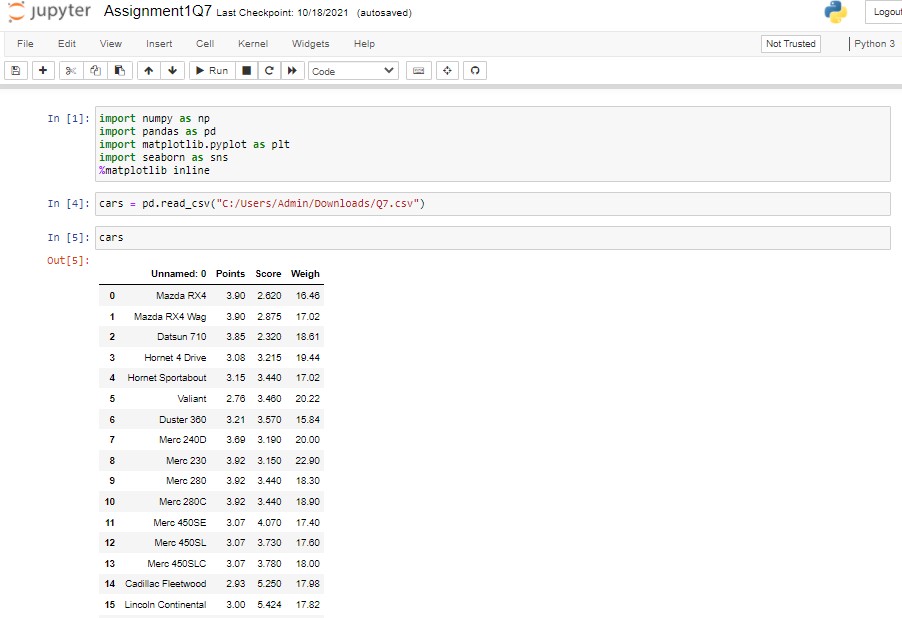
The Expected number of candies for a randomly selected child will be = 3.09

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

- For Points, Score, Weight

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

# Use Q7.csv file

**Ans:**

eronge = *‹fiox -min*

Point s\_range = c ars . Points .max( ) - cars . Point s .min( ) Point s\_range

S + 8. †| M t a c Run II C w ade In [9):



In [ 6] :

*».median*

*c ars* .nedia n{ )

*emeom*

cars.mean()

0' [fl: PoiutS 3.695

Score 3.325

weigh iI.7i0

In [ 8 ] !

e.mode

c ars . Point s . mode )

l 3.92

In [ s] :

ñ.mode /or sE sne

zn [ s8 ] :

ñ.mode /or i 'e Sqm

*c ars .Ve'zgh* .node( )

ou\*[ s8 ] : o 1y . gz

zn [ 31 ] : *e for yance*

cars .v ar ( )

” j u p yter Assignmentl Q7 Last c expi I: 10/18 2024 {autosaved} 

*e or dance c are .» a r (* )



In [ 31 ] :

Out[31] Points 8.285881

ScOKP 8. 957379

YJe:tgh Z. 193166

In [ 7 ] !

*ifisto.ndar0* de'vt o\*iow

Score 0.978457

weigh I.786s43





Z7B0000 1.5¥30O0 14.5-EOOCO 25% Z.0800€O Z.ñ81Z?O 16.B02bO0 '50% S88g0£O T3250CO 17.7T00CO 75% Z.020£KO Z.610fKO 18.€K@£K@

In [ 11 ] !

” ” j u pyter Assignment1 Q7 cast c erxn int: 10/1 B 2021 (a tosaved}

Vé dgets Help

zn [ zz] :

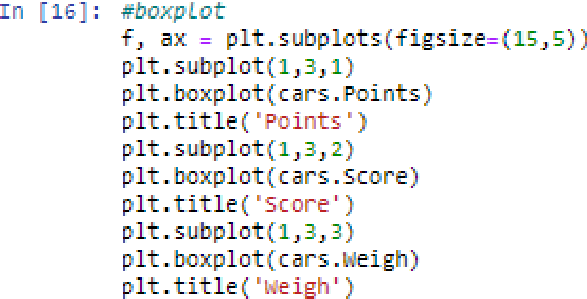
s core\_range=c ans . score . max( ) - cars . s cone. rrln ( }

s core range

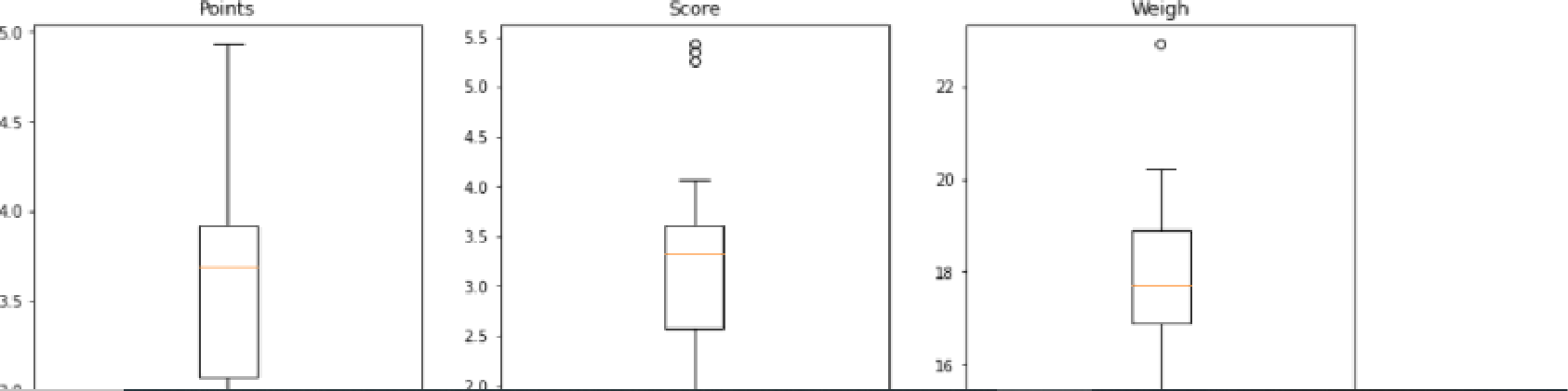
’ LDgOIJt



ou•E i• ] : 3 . slzeeeeeeeeeeees



0u^ § 1 E ] : Text ( B . 5, 1. 0, ' !geigh ' }

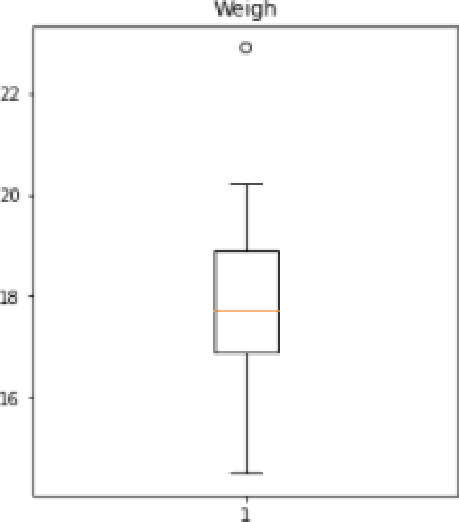
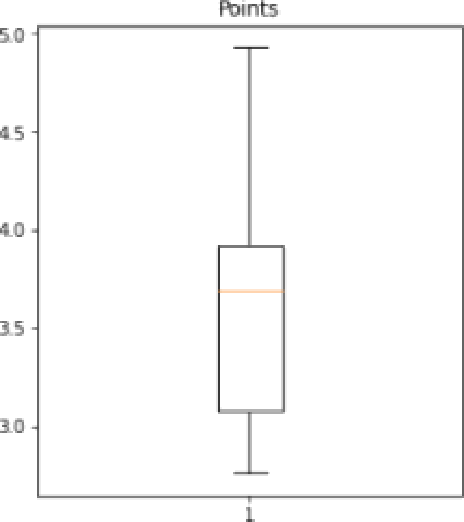


” ” j u p yter As signmen t1 Q 7 r»‹ c eokp«int: for IB nazi fautosaved› 

 Nat Trusted Pyth a n 3



ou\*[ as l ! text (e, s, i. e, ' ueigh ' )

ss



S0





1. we see I h at I h e data is co n cent rated around be dia n
2. There are no outliers
3. The distribution is Right skewed
4. For scar e dat a s et :
   1. The dat a i s c oncent rated around reed1 an
   2. There a re 3 outl iars
   3. The di s *trt* but:ton is Le-ft skewed

*c ) E or ue i gh* d at aset :

i )The dat a is concentrated around md ian z)Ther e is i out liar

Q8) Calculate Expected Value for the problem below

a) 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:** As we see, one of the patient is chosen at random and there are patients.

Probability of each patient = 1/9 We have formula,

Expected value = (probability \* value) P (x). E(x)

Now,

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 |

Therefore 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)]

= 145.33

# Q9) Calculate Skewness, Kurtosis & draw inferences on the following data Cars speed and distance

**Use Q9\_a.csv**

Inse n V/ dgets Help

I n [ 18 ] : Triport numpy as np

lripont pand as as pd

import seaborn as sns

%matp1otlib Online



I n [ 20] : df

Out[ 2s] !

|  |  |  |  |
| --- | --- | --- | --- |
| 4 | b | 8 | 10 |
| G | 7 | 10 | 18 |

|  |  |  |  |
| --- | --- | --- | --- |
| 10 |  |  | 28 |
| 12 | 1Z | 12 | 20 |





Zn [ 21] ! er/•e:i'.nerr

df . skew{ }

out[zz] i zndez e. eeeeee

dist o.se6ss5 dtype: floats

In [ ] : *ego e hence* •

1. is \‹e see S peed dist r ”abut hon is left skeyied (neg ative sk ewnes s
2. As yJe see D i st ance d z str i bu tion is right skewed ( pos itive skewne ss)



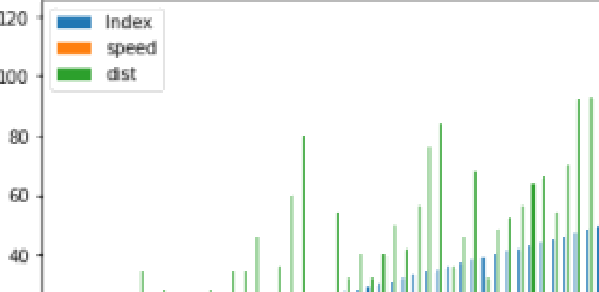
0u\*[ IE ] i znde z - 1. zeeeg•B

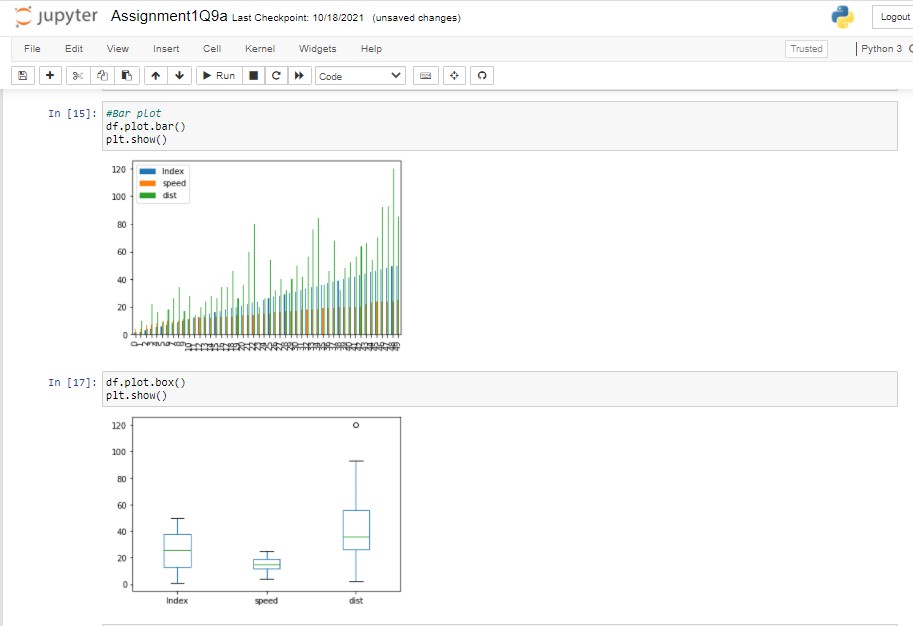
dtype: -fLoat64-



1. speed dist ri bu tion is -f 1 at ter I h an normal d z s tr1 but ion whlch is negative k u rtosis . z . Dlstanc e di st rlbut zon is peaked than *n area I* d1st r i bution yihich is pa s1ti 7e kurtosi s.

zn [ 1s ] : *eBar p'. a“*

d I , p lot . ba r )



# SP and Weight (WT) Use Q9\_b.csv

\_” j u pyte r As signmen II Q9b Last Checkpoint: 1 D/'â8/202â 



In [P]! import numpy as np

import pandas as pd

impori matp1otlib.pyplot as pit import seaborn as sns

%matp1otlib znline



In [ 6] : d f



se

2 105.401 2 I4 TO.400833





TB

81 1 ZO.84081 7 34.648615

81 rows \* 3 columns In [ 7 ] : *esfi'e:».n ess*

d f . skew( )

sP

dtype: I loat64

\_” j upyte r Assignment1 Q9b Usr Checkpoint: 1 D/'18^202T 



in [ 7 ] ! *ésker.n ess*

d I- . s kew ( )

o'u• [ 7 ] ! unnamed : e e .eeeeee

SP 5 . 61145g

dtype: -f Loat64 In [ 8 ] : *Tfi:ur tos is*

o›J#[a) : Unnamed: 0 -i.ze0e00 SP 2.977°:29

NT 0.95B29I

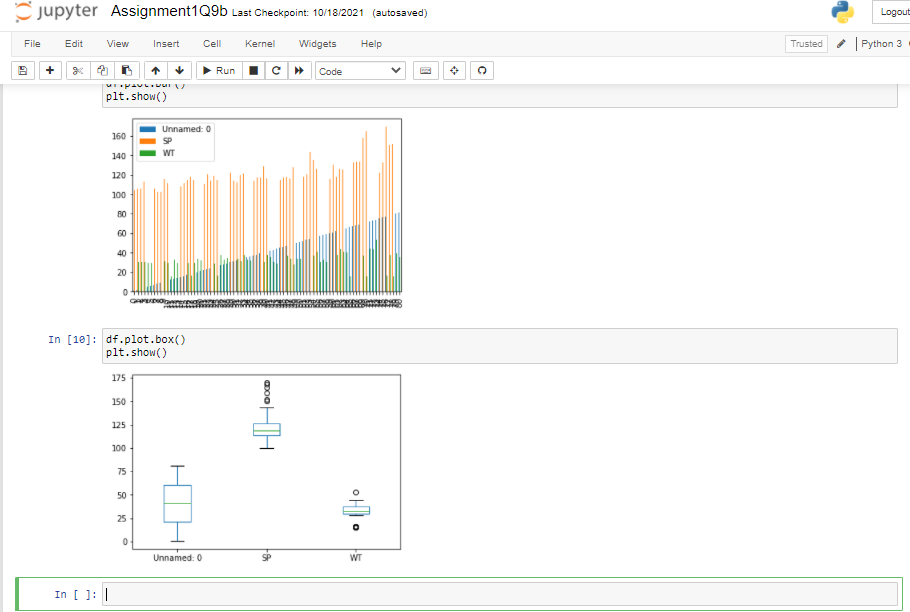
dtype: floats\*

In [ s ] i *a'Bar* p tod

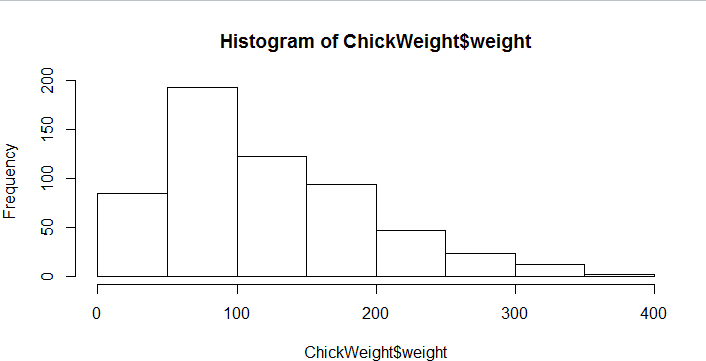


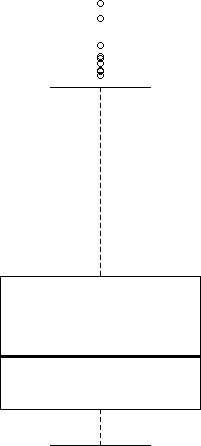
In [ie): df.plot.box()

plt. show()



# Q10) Draw inferences about the following boxplot & histogram

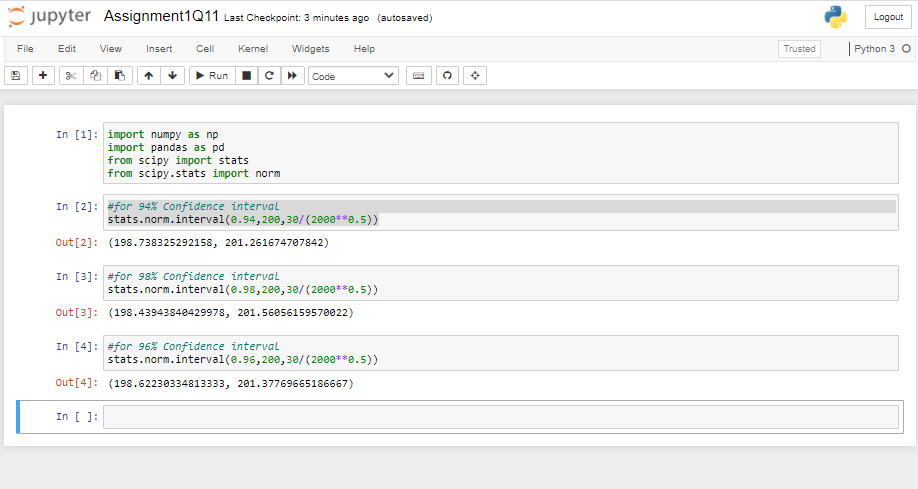




**Ans:** It is right side skewed or positively skewed.

**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:



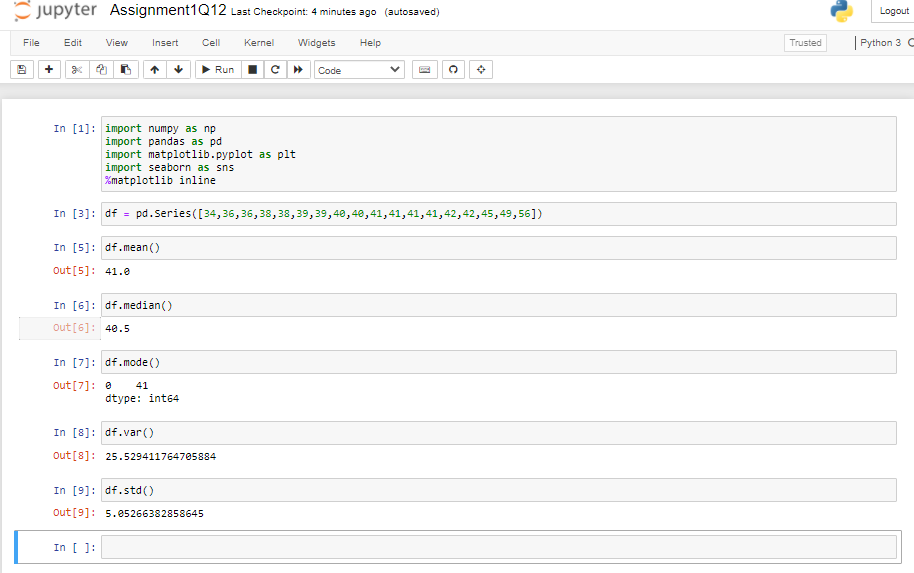
**Q12)** Below are the scores obtained by a student in tests

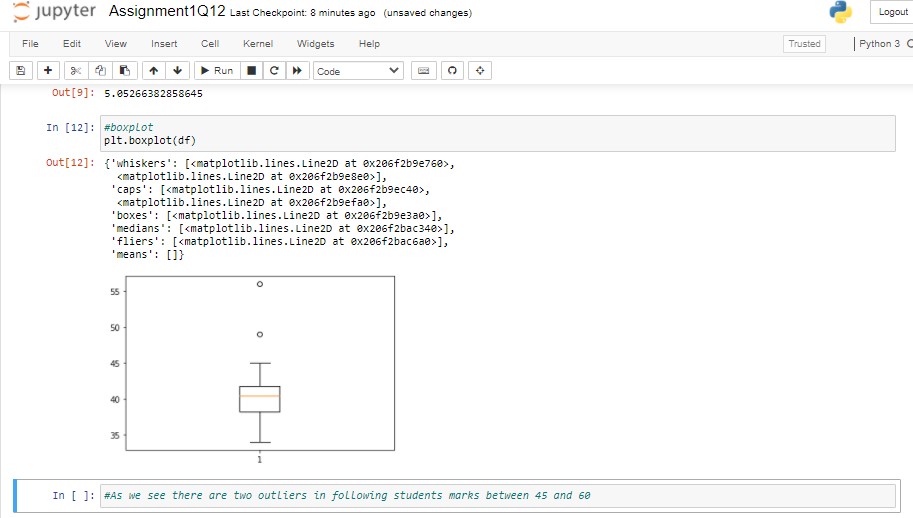
**34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56**

1. Find mean, median, variance, standard deviation. Ans: mean = 41

Median = 40.5

Variance = 25.52 Standard deviation = 5.05





1. What can we say about the student marks?

Ans: There are two outliers in student marks 49 and 56.

Q13) what is the nature of skewness when mean, median of data are equal? Ans: It shows the nature is Normalized Skewness

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

Ans: If mean>median then the nature of skewness is positively skewness.

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

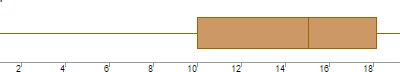
Ans: If median>mean then the nature of skewness is negatively skewness.

Q16) what does positive kurtosis value indicates for a data? Ans: Positive kurtosis value indicates that a Normal distribution.

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

Ans: Negative values of kurtosis indicate that the distribution is flat and has thin tails.

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



What can we say about the distribution of the data?

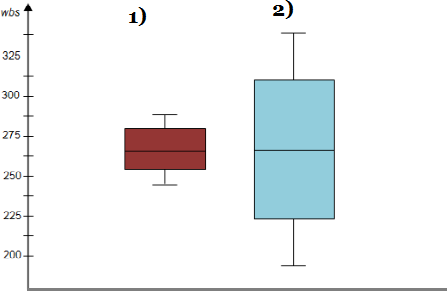
**Ans:** The data is distributed in skewed format. What is nature of skewness of the data?

**Ans:** The nature of the skewness of data is Left side skewed What will be the IQR of the data (approximately)?

**Ans:** Q3-Q1 = 18 – 10

= 8 is IQR

Q19) Comment on the below Boxplot visualizations?



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

**Ans:** The box plot 1 designed with range = 3 The second one range is = 1.5

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)

# Ans:

1. P(MPG>38):

1-pnorm (38, 34.422, 9.13144) = 0.3475908

1. P(MPG<40):

pnorm (40, 34.422, 9.13144) = 0.7293527

c. P (20<MPG<50):

pnorm (50, 34.422, 9.13144)-(1-pnorm (20, 34.422, 9.13144))

=0.01311818

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv

” j upyter



Iz [1]:

In [2]:

cars =

Assignment1Q21a) cast cuaxnoi»t: 18 mi «tes ago (autosaved) 

VG dgets H eip

impont numpy as np impol-t pandas as pd

import mat plotllb.pyp1ot as p It impont sea born as s ns

2mat plotI i b inline



|  |  |  |  |
| --- | --- | --- | --- |
| N | 5g 50.010401 | 02 105.401284 | @.488832 |
| 2 | 55 50.01£401 | 62 105.4612B4 | @.1035B7 |
| 4 | 53 50.WZ32 | 62 10d.4012B4 | 29.BB64 49 |
|  |  | TO 1B0. i08g ¥3 | 18.13204T |
| 7B | 203 34.C00C00 | l20 IU.508l2¥3 | 15.T09gZ5 |
| 7B | 29g 18.8g3733 | T4B 107.8444 Ci | GB.4Z30g i |
| BO | 2 g 12.TO-12B3 | 107 UB.840B¥7 | 34.648g15 |



IM [4]: *+ñtstogrom*

sns.h\*stplot(cars.NPG, 1abel='cars-LPG') plt.x1abel(’NPG'}

plt.v1abelf'oensity')

” ” j upyter



Assignmentl Q21a) cost cuaxp>i»t› 10 mi»tes ago {autosaved} 

VG dgets Help

G + ¥z ? i &



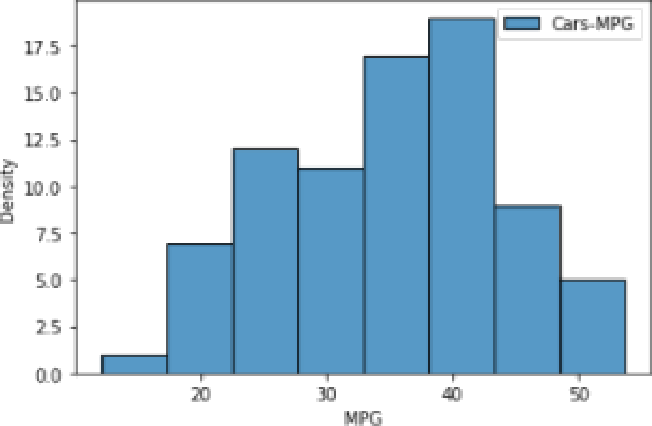
•$• • • 6 Run M W R Dode

IM [4]:

chis*togrom*

s ns . h:tstp1ot (cars .<Ps, la bet= 'ca *rs - R'fi ’* ) p1t .x1abe1{ ' NPG ’ }

p1t .y1abe1{ 'Density' ) p1t . legend{) ;



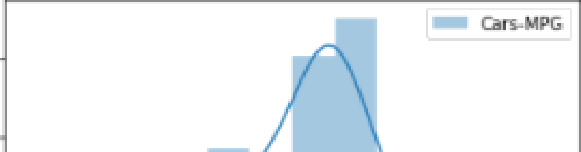
sns.distplot(cars.NPs, label='cars-LPG') plt.x1abel('NPG')

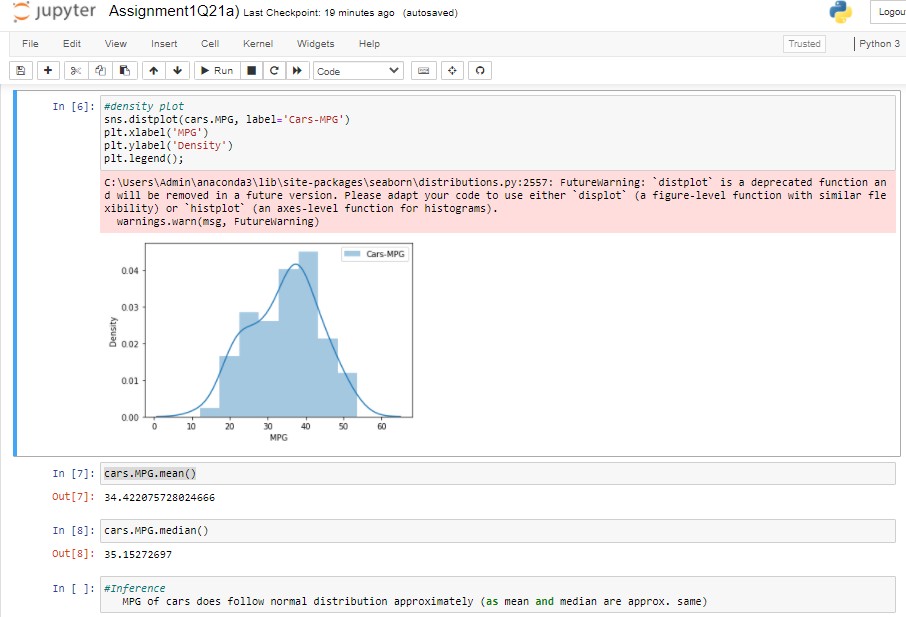
plt.y1abel('Density’)

plt.1egend();

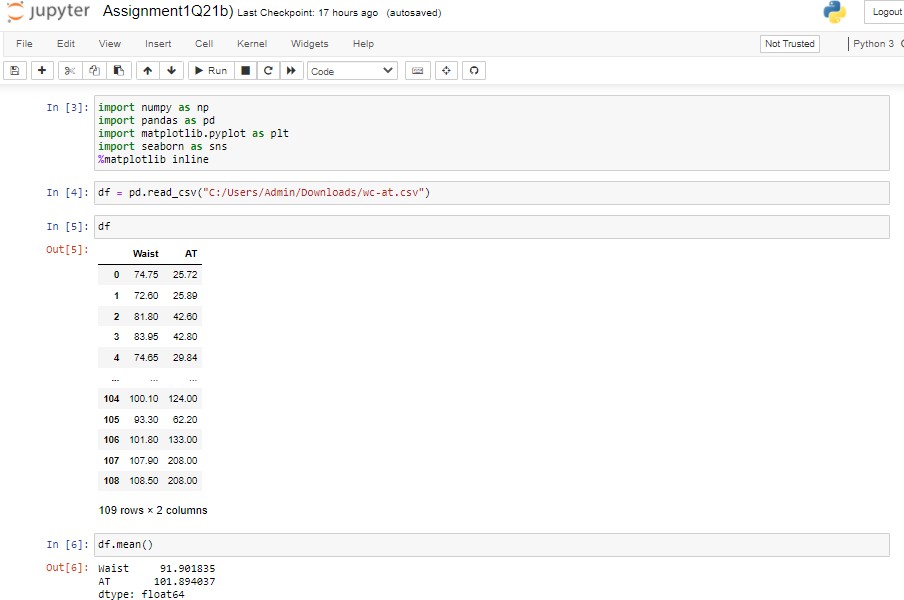
c: \Users\Adsin\anaconda3llib\site-packages\seabornldistributions.py:zss7: FutureWarnIng: ’distplot’ is a deprecateo function an d viii be removed 5n a future version. Please adapt your code to use e5ther ’displot’ (a figure-level function with similar fle sibility) or ’lustplot’ (an axes-level function for lustograns).

warnings.varn(nsg, Futurevarn5ng)





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

Dataset: wc-at.csv



c!’',Users...;Jmin..anaconda3’..lib’',site-packages’..seabci n’...distr iD'utizns .py::557: Fut'Jre..'arniM°: d-stplot’ 1s a Jeprecatec {uMction ar› c .-11 be iemo.°d in a futu-° 'e-sicn. Please adapt vcui code to use either d-splo" ia -igure-l°.'e1 OJnction ’..iU s-Sila- fle xib-lits) :+ ’h-stplot’ icn axcz-Le.'cl function f:r *histog-ars:.*

:.a+ rings...'c-n{rsg, Futurehai ning]



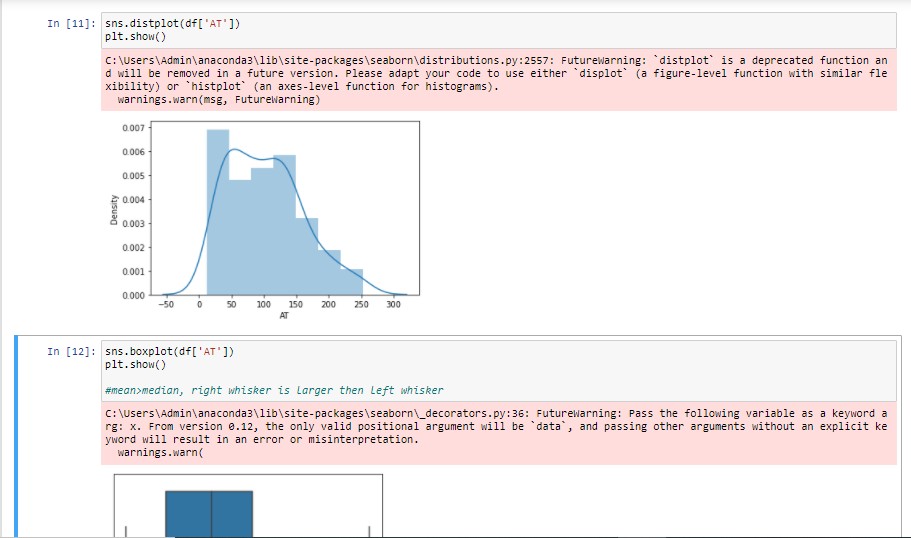
c : '',u s e r s '...ifl min '.. a n a c one a 3 '',1 ib '', s i te - pa c kcg.•s''.seabcr n’-.. d .• c nc\*c rs . pv : se : -'ut ui enar n i n g : Pas s \* t .• 'o I 2c ..'i rig '.' a i iz b 2 .• z s a ke 'i' o r d c rg! x. :’cm '...e’s-or› 0.i2, the ord’ .•alid pzsi#icnal ar.ument ..'ill be ’cata’, and pa>>-ng o'tei s’,omer›#> ..'ithout sn e›:elicit ke ',.’ucrJ ,iIL i’e>JLt in an sr-or -r misir›terpi’etat-on.

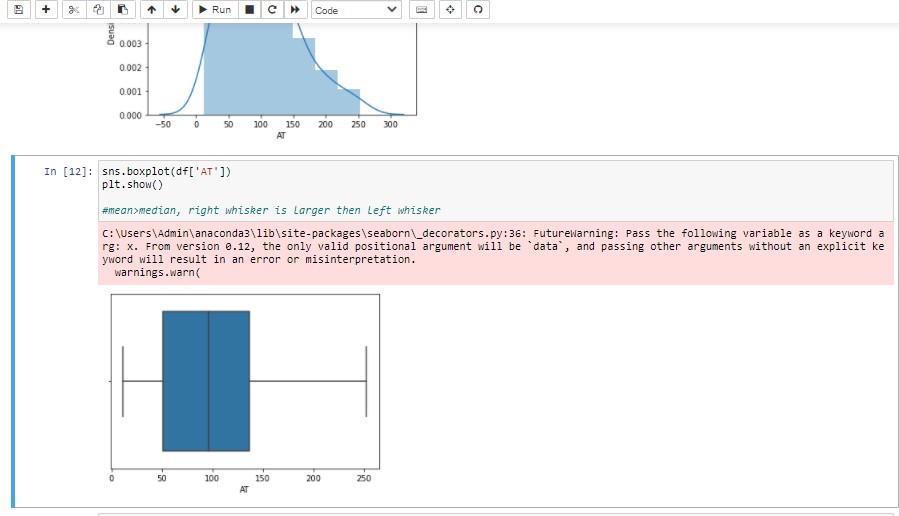


c! ',Users ..;Jmin .anaconda3 ..lib',site- packages..seabcrn’..distrib'ut-:ns.py:2ss7! Future...'arning: d-stplot is a Jeprecatec {uMctioM ar› c ”.'z be i” e no'.' .°d in a £-u t'u-.° '.' e -been . P?.°ase adapt .cui” cod.° to 'use eith.°i” ” d sy I at t a ' if° u i” e - L.°^.'e T'un ct ion '..”1 IN. such la - £-be x i b T to } i h st plot i\* n aae - L.°'.'e1 -funct ion F n *h st og - ar s\* .*

”.. a i ning s . ..'a -n t *r* s g, F ut ur-et.ai n ing}







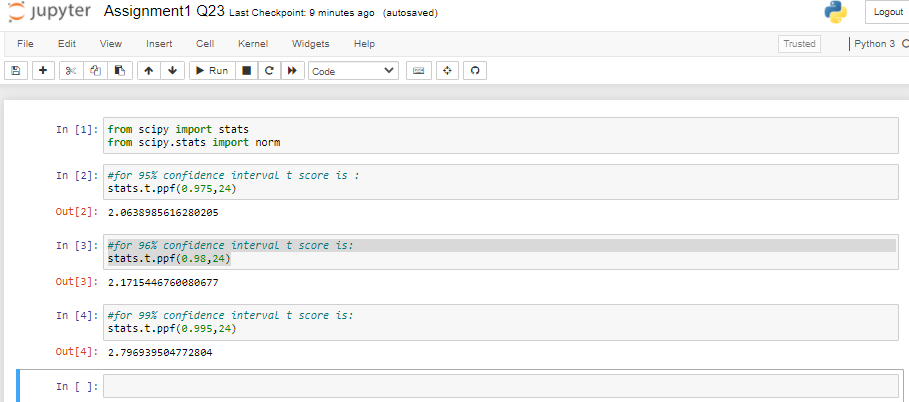
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval .

# C:\Users\Admin\Pictures\Screenshots\Screenshot (50).pngAns:

|  |  |
| --- | --- |
| Cofidence Interval | Z-score |
| 60% | 0.84161212 |
| 90% | 1.644854 |
| 94% | 1.880794 |

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

# Ans:



|  |  |
| --- | --- |
| Confidence Interval | T-score |
| 95% | 2.063899 |
| 96% | 2.171545 |
| 99% | 2.79694 |

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

**Ans:**

We have the following data:

x = mean of the sample of bulbs = 260 μ = population mean = 270

s = standard deviation of the sample = 90 n = number of items in the sample = 18

t = [260-270 | 90/sqrt of 18 ] We get, t = - 0.471

We have formula for, degrees of freedom is n - 1, so we will get 18-1 =17 We want t-distribution with 17 degrees of freedom.

So the probability of t < - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained With 17 degrees of freedom and a t score of - 0.471.

So the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.

