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 | Nominal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ration |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| 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?

=HHT+HTH+THH

=1/8+1/8+1/8

=3/8

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

= Zero, as none of the outcomes will be equal to one.

1. Less than or equal to 4

= (1,3) (2,2) (3,1) =3/36

=1/12

1. Sum is divisible by 2 and 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?

=P(2r,3g,2b)

=P(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

=1\*.015+4\*.20+3\*.65+5\*.005+6\*.01+2\*.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.

Points:3.9, score, weight=16.46

Points:

Mean=3.596563, Median=3.695, Mode=” Numeric”, Variance=.2858814, Std deviation=.5346787

Score:

Mean=3.21725, Median=3.325, Mode=” Numeric”, Variance=0.9784574, Std deviation=.5346787

Weight:

Mean=17.84875, Median=17.71, Mode=” Numeric”, Variance=3.193166, Std deviation=1.78694

**Use Q7.csv file**

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?

Mean=1308 and Mue=163.5

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**

**Answer Code:**

import numpy as np

import pandas as pd

import statistics as st

Data=pd.read\_csv("Q9\_a.csv")

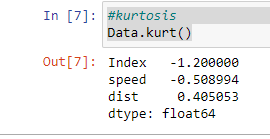
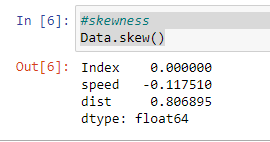
Data

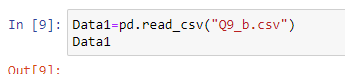
#skewness

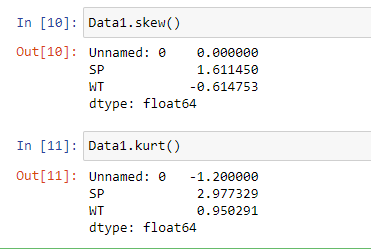
Data.skew()

#kurtosis

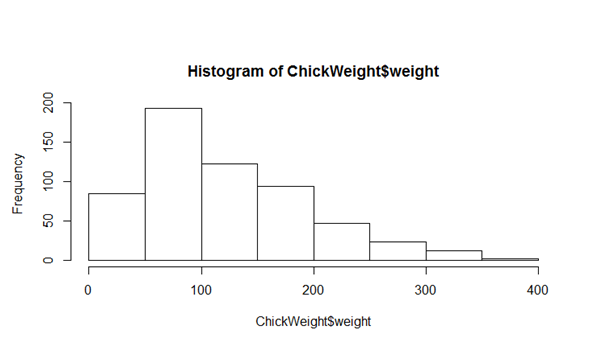
Data.kurt()

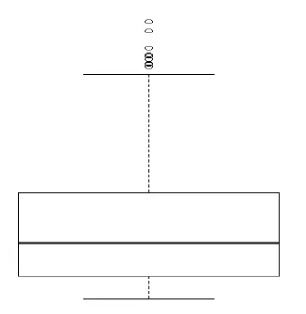






Q10) Draw inferences about the following boxplot & histogram





**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?

conf\_94 =stats.t.interval(alpha = 0.94, df=1999, loc=200, scale=30/np.sqrt(2000)) print(np.round(conf\_94,0)) print(conf\_94) For 94% confidence interval Range is [ 198.73 – 201.26] For 98% confidence interval

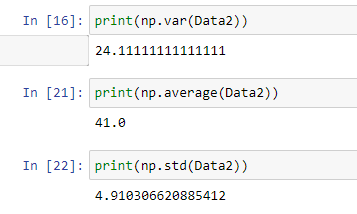
range is [198.43 – 201.56] For 96% confidence interval range is [198.62 – 201.37]

**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:



1. What can we say about the student marks?

Ans: we don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

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

Ans: No skewness is present we have perfect symmetrical distribution.

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

Ans: Skewness and tail is towards right

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

Ans: Skewness and tail is towards left

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

Ans: +ve kurtosis means the curve is more peaked and it is leptokurtic

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

Ans: -ve kurtosis means the curve will be flatter and boarder

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



What can we say about the distribution of the data?

Ans: The above Boxplot is not normally distributed the median is towards the higher value

What is nature of skewness of the data? What will be the IQR of the data (approximately)?

Ans: The data is a skewed towards left. The whisker range of minimum value is greater than maximum

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

Ans: The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 18 – 10 =8

Q19) Comment on the below Boxplot visualizations?



First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)

Prob\_MPG\_greater\_than\_38 = np.round(1 - stats.norm.cdf(38, loc= q20.MPG.mean(), scale= q20.MPG.std()),3) print('P(MPG>38)=',Prob\_MPG\_greater\_than\_38)

P(MPG>38)= 0.348

b.P(MPG<40)

prob\_MPG\_less\_than\_40 = np.round(stats.norm.cdf(40, loc = q20.MPG.mean(), scale = q20.MPG.std()),3) print('P(MPG<40)=',prob\_MPG\_less\_than\_40)

P(MPG<40)= 0.729

C.P (20<MPG<50)

prob\_MPG\_greater\_than\_20 = np.round(1-stats.norm.cdf(20, loc = q20.MPG.mean(), scale = q20.MPG.std()),3) print('p(MPG>20)=',(prob\_MPG\_greater\_than\_20)) p(MPG>20)= 0.943

prob\_MPG\_less\_than\_50 = np.round(stats.norm.cdf(50, loc = q20.MPG.mean(), scale = q20.MPG.std()),3) print('P(MPG<50)=',(prob\_MPG\_less\_than\_50)) P(MPG<50)= 0.956

prob\_MPG\_greaterthan20\_and\_lessthan50= (prob\_MPG\_less\_than\_50) - (prob\_MPG\_greater\_than\_20) print('P(20<MPG<50)=',(prob\_MPG\_greaterthan20\_and\_lessthan50)) P(20<MPG<50)= 0.013000000000000012

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Normally distributed.

Code(

import pandas as pd

from matplotlib import pyplot as plt

import seaborn as sns

cars=pd.read\_csv("Cars (2).csv")

cars

cars['MPG'].mean()

cars['MPG'].median()

cars['MPG'].mode()

cars['MPG'].hist()

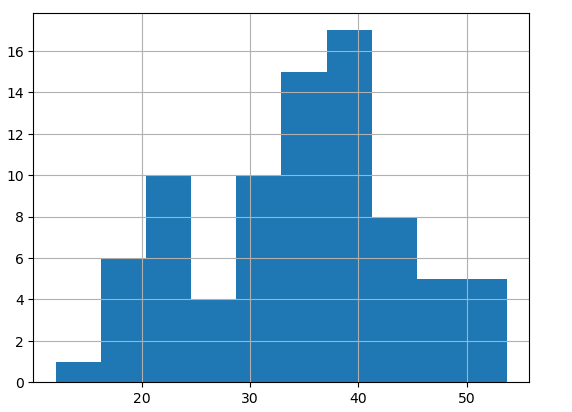
sns.distplot(cars['MPG'])

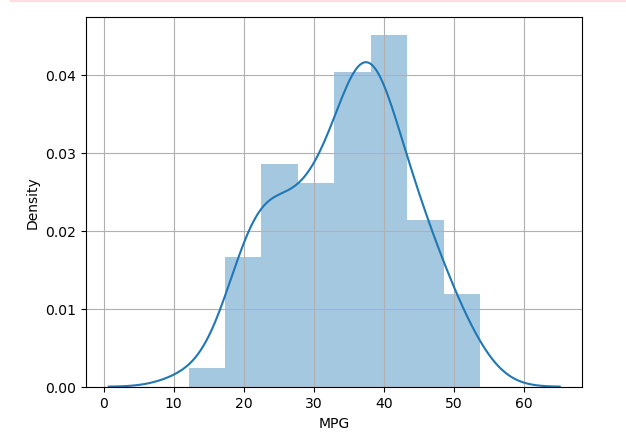
plt.grid(True)

plt.show()

cars['MPG'].skew()

cars['MPG'].kurt()





)

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

Dataset: wc-at.csv

Adipose Tissue (AT) and Waist does not follow Normal Distribution

Code(

Cars1=pd.read\_csv("wc-at.csv")

Cars1.head()

Cars1.mean()

Cars1.mode()

Cars1.median()

sns.distplot(Cars1['Waist'])

plt.show()

sns.distplot(Cars1['AT'])

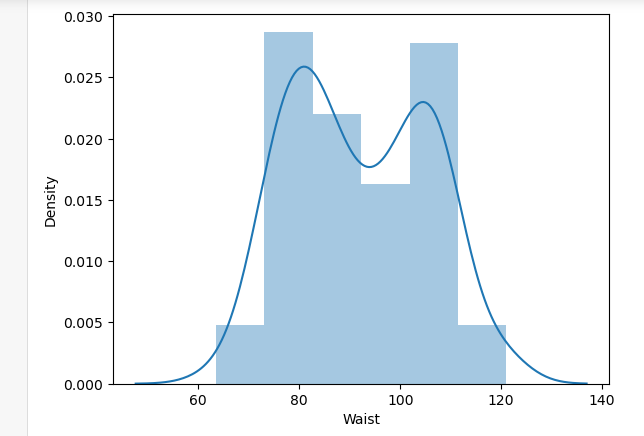
plt.show()

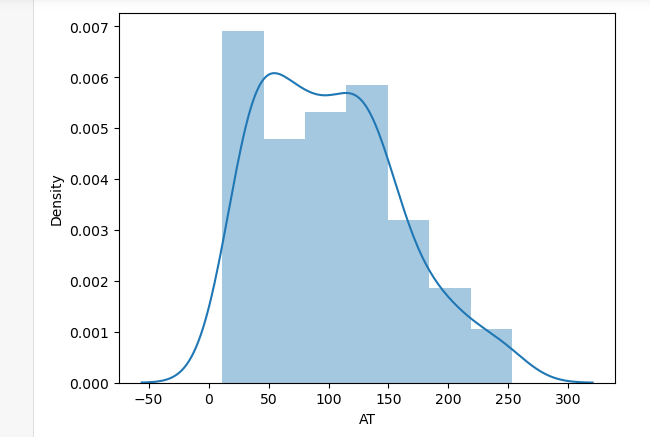
sns.boxplot(Cars1['AT'])

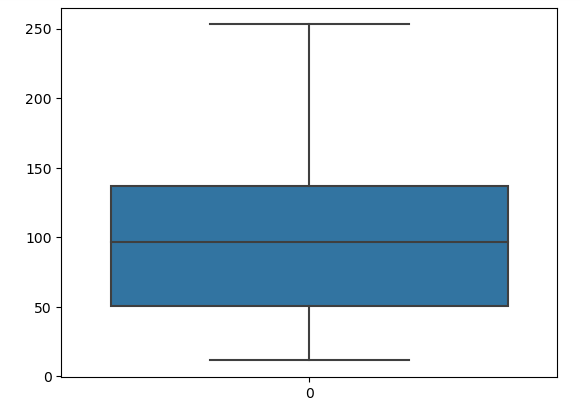
plt.show()

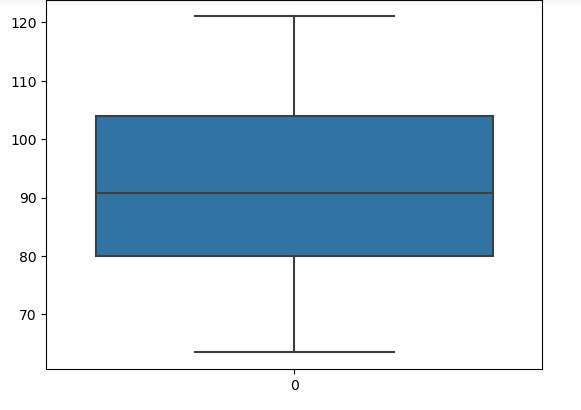
sns.boxplot(Cars1['Waist'])

plt.show()



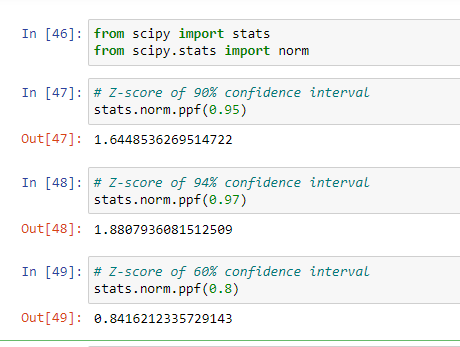




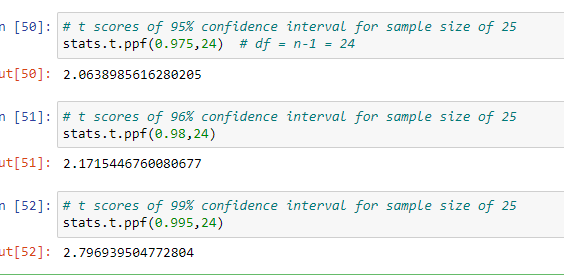


)

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

