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

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

1. outcomes will be HHT, THH, HTH

so total outcomes are 3

As 3 coins are tossed 23=8

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
4. As two dice are rolled the minimum possibility is (1,1) then the sum will be 2. As mentioned it must be equal to zero. So the probability will be 0.
5. As two dice are rolled the sum should be less than or equal to 4.

Then the outcomes will be (1,1),(1,2),(2,1),(1,3),(3,1),(2,2).

The total possible outcomes are 6.

The total outcomes for two dice are 36

6/36 = 1/6.

1. Two dice are rolled the probability of sum must be divisible by 2 and 3

The possible outcomes are (1,5),(2,4)(4,2),(5,1),(6,6).

Possible outcomes are 6

=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?

1. E1.2 red ball

E2.2 green ball

E3.1 red and 1 green

Number of selecting 2 balls from two red balls n(E1) = 2C2 = 1

Number of ways of selecting 2 balls from 3 green balls n(E2) = 3C2 = 3 Number of ways of selecting 1 red ball and 1 green ball n(E3) = 2C1\*3C1 = 2\*3 = 6

Sample space S = Number of ways of selecting 2 balls from 7(2+3+2) balls n(S) = 7C2 = 21

n(E) = n(E1) + n(E2) + n(E3) = 1+3+6 = 10

Probability 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 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

A) Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

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\*0.120 = 3.09

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

* For Points, Score, Weight.

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

**Use Q7.csv file**

A)

| **Unnamed: 0** | **Points** | **Score** | **Weigh** |
| --- | --- | --- | --- |
|  |  |  |  |
| **0** | Mazda RX4 | 3.90 | 2.620 | 16.46 |
| **1** | Mazda RX4 Wag | 3.90 | 2.875 | 17.02 |
| **2** | Datsun 710 | 3.85 | 2.320 | 18.61 |
| **3** | Hornet 4 Drive | 3.08 | 3.215 | 19.44 |
| **4** | Hornet Sportabout | 3.15 | 3.440 | 17.02 |
| **5** | Valiant | 2.76 | 3.460 | 20.22 |
| **6** | Duster 360 | 3.21 | 3.570 | 15.84 |
| **7** | Merc 240D | 3.69 | 3.190 | 20.00 |
| **8** | Merc 230 | 3.92 | 3.150 | 22.90 |
| **9** | Merc 280 | 3.92 | 3.440 | 18.30 |
| **10** | Merc 280C | 3.92 | 3.440 | 18.90 |
| **11** | Merc 450SE | 3.07 | 4.070 | 17.40 |
| **12** | Merc 450SL | 3.07 | 3.730 | 17.60 |
| **13** | Merc 450SLC | 3.07 | 3.780 | 18.00 |
| **14** | Cadillac Fleetwood | 2.93 | 5.250 | 17.98 |
| **15** | Lincoln Continental | 3.00 | 5.424 | 17.82 |
| **16** | Chrysler Imperial | 3.23 | 5.345 | 17.42 |
| **17** | Fiat 128 | 4.08 | 2.200 | 19.47 |
| **18** | Honda Civic | 4.93 | 1.615 | 18.52 |
| **19** | Toyota Corolla | 4.22 | 1.835 | 19.90 |
| **20** | Toyota Corona | 3.70 | 2.465 | 20.01 |
| **21** | Dodge Challenger | 2.76 | 3.520 | 16.87 |
| **22** | AMC Javelin | 3.15 | 3.435 | 17.30 |
| **23** | Camaro Z28 | 3.73 | 3.840 | 15.41 |
| **24** | Pontiac Firebird | 3.08 | 3.845 | 17.05 |
| **25** | Fiat X1-9 | 4.08 | 1.935 | 18.90 |
| **26** | Porsche 914-2 | 4.43 | 2.140 | 16.70 |
| **27** | Lotus Europa | 3.77 | 1.513 | 16.90 |
| **28** | Ford Pantera L | 4.22 | 3.170 | 14.50 |
| **29** | Ferrari Dino | 3.62 | 2.770 | 15.50 |
| **30** | Maserati Bora | 3.54 | 3.570 | 14.60 |
| **31** | Volvo 142E | 4.11 | 2.780 | 18.60 |

Points

Mean = 3.59

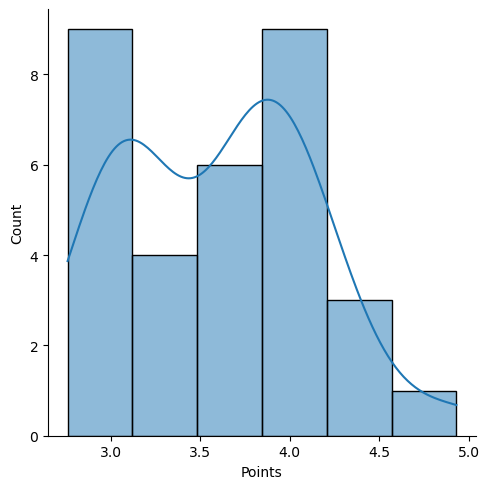
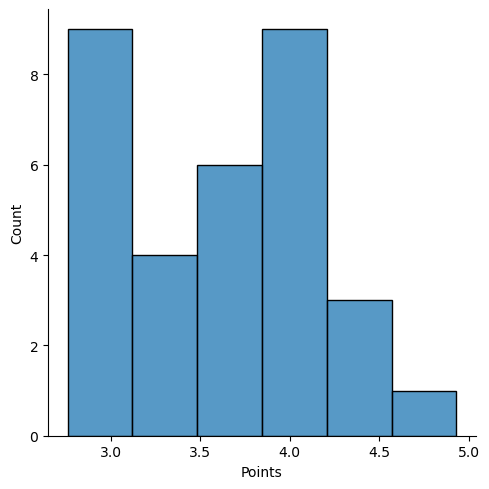
Median = 3.69

Mode = 3.92

Variance = 0.28

Standard deviation = 0.53

Range = 2.17



Score

Mean = 3.21

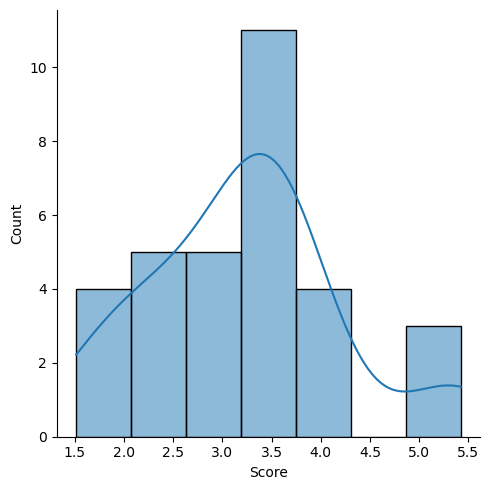
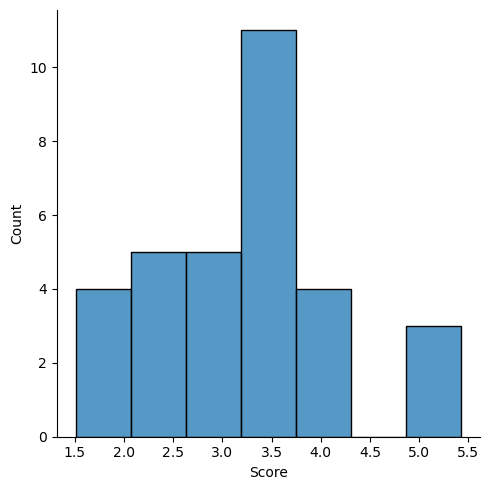
Median = 3.325

Mode = 3.44

Variance = 0.957

Standard deviation = 0.97

Range = 3.911



Weight

Mean = 17.84

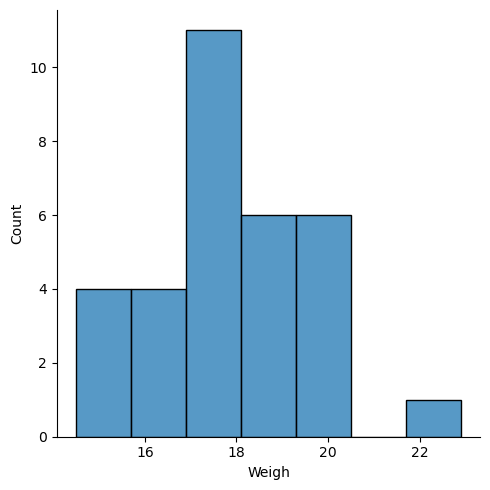
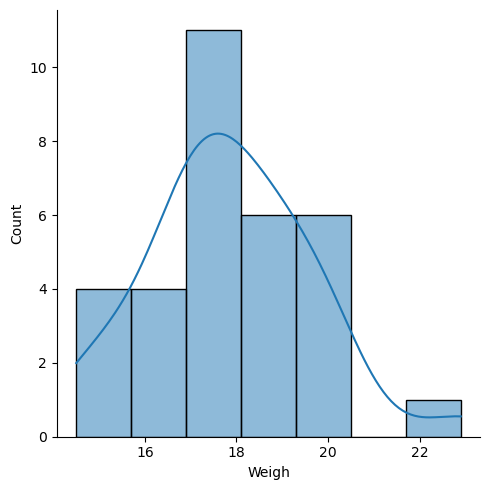
Median = 17.71

Mode = 17.02

Variance = 3.19

Standard deviation = 1.78

Range = 8.4

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?

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.

(108, 110, 123, 134, 135, 145, 167, 187, 199) / 9

= 145.33

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

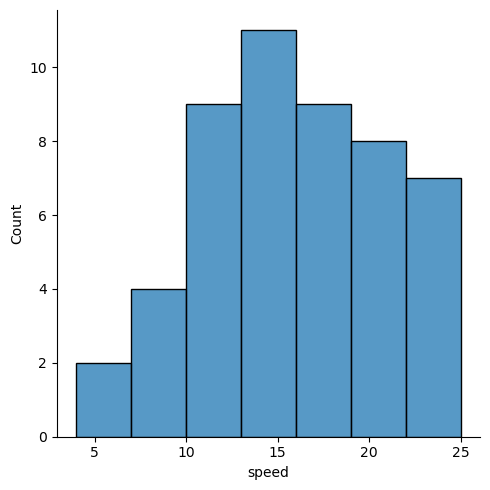
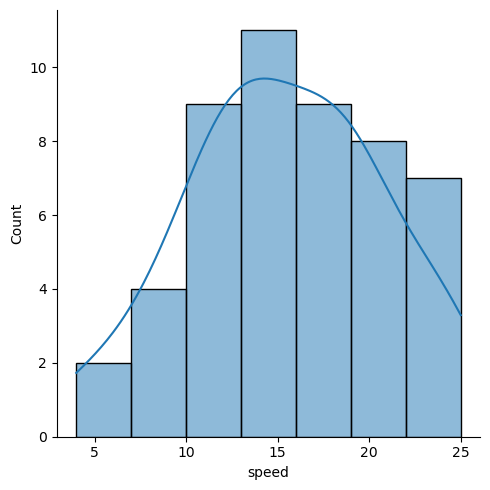
**Cars speed and distance**

**Use Q9\_a.csv**

**A)** speed (left Skewed- negatively Skewed)

Skewness = -0.1139

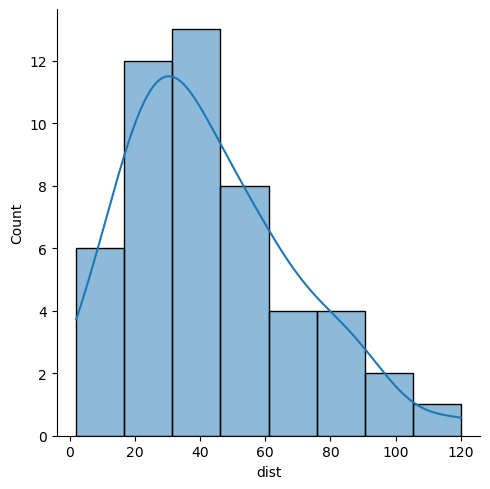
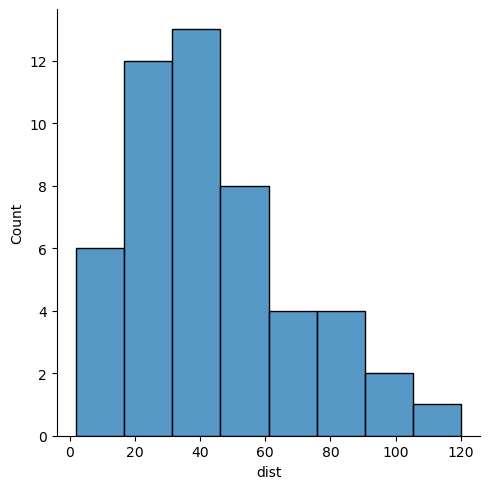
Kurtosis = -0.577

Distance (Right Skewed- positively Skewed)

Skewness = 0.7824

Kurtosis = 0.2480



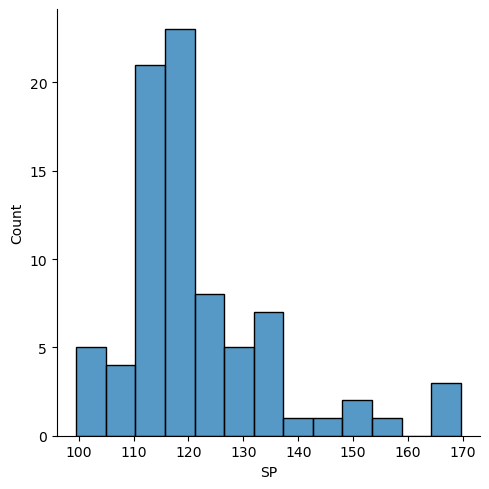
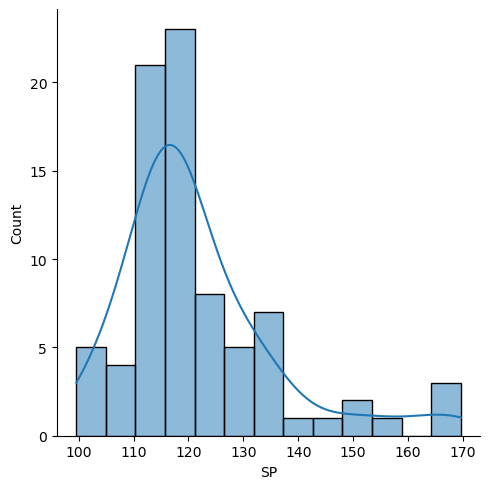
**SP and Weight(WT)**

**Use Q9\_b.csv**

**A)** SP (Right Skewed- positively Skewed)

Skewness = 1.581

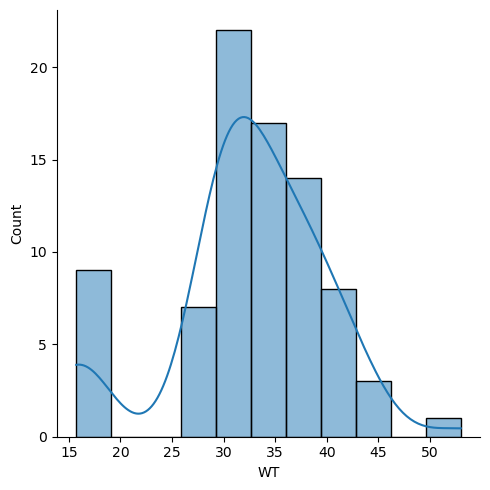
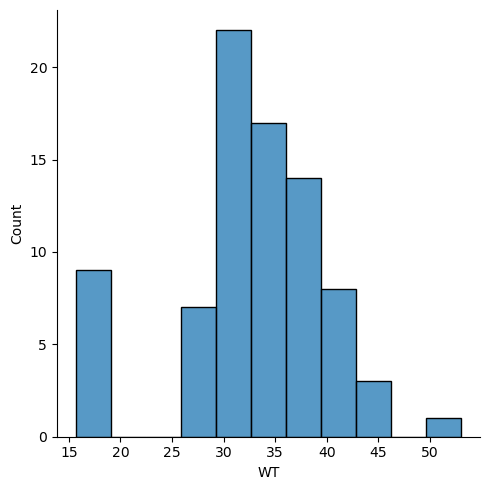
Kurtosis = 2.723

WT (Left Skewed- negatively Skewed)

Skewness = -0.603

Kurtosis = 0.819



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



1. The histograms peak has right skew and tail is on right.

Mean > Median. We have outliers on the higher side.



1. The box plot has outliers on the maximum side

**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?

A) random sample of 2,000 men from a population of 3,000,000 men

Mean = 200

Standard deviation = 30

Sample size = 2000

Scores of Confidence level for 94% = 1.89

Scores of Confidence level for 98% = 2.326

Scores of Confidence level for 96% = 2.053

Confidence level for 94% = (198.738325292158, 201.261674707842)

Confidence level for 98% = (198.43943840429978, 201.56056159570022)

Confidence level for 96% = (198.62230334813333, 201.37769665186667)

**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?

A)1) Mean = 41.0

Median = 40.5

Mode = 41

Variance = 24.111

Standard deviation = 4.9103

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

A) Mean = Median (Symmetric).

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

A) Mean > Median (Right Skewed).

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

A) Median > Mean (Left Skewed).

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

A) Positive kurtosis means the curve is more peaked and it is Leptokurtic.

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

A) Negative Kurtosis means the curve will be flatter and broader.

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

A)



What can we say about the distribution of the data?

A) Median is towards the higher value.

What is nature of skewness of the data?

A) Left skewed.

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

A) Q3 – Q1 = 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.

A)  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)
  2. P(MPG<40)
  3. P (20<MPG<50)

1. a)0.347

b)0.729

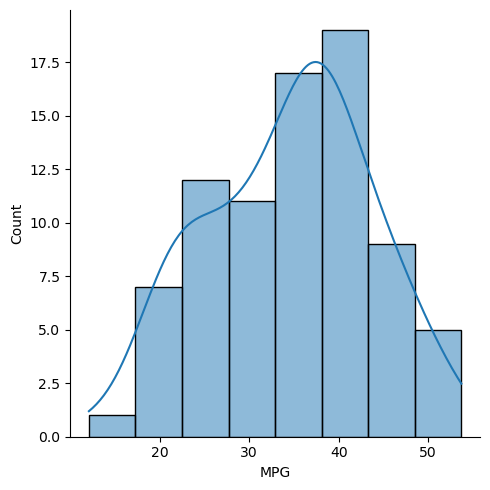
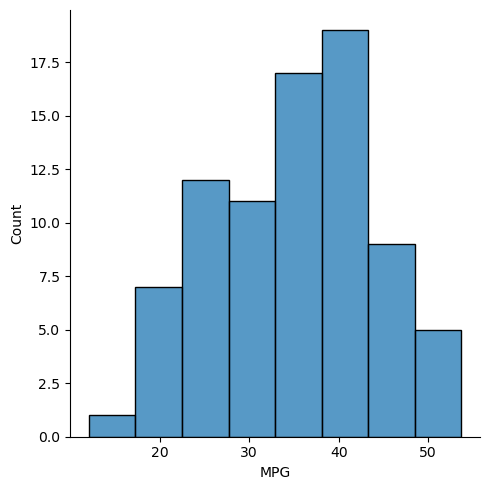
c)0.898

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

1. MPG(Left skewed – Negatively Skewed)



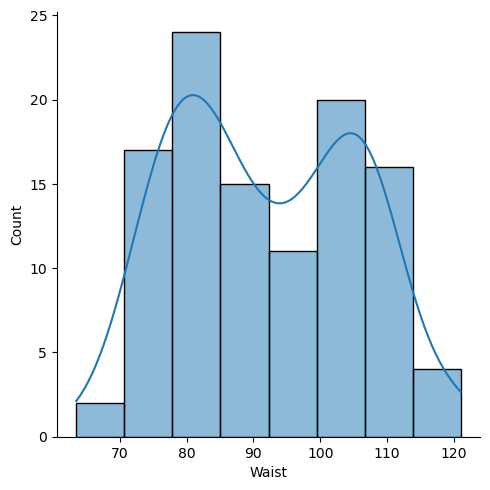
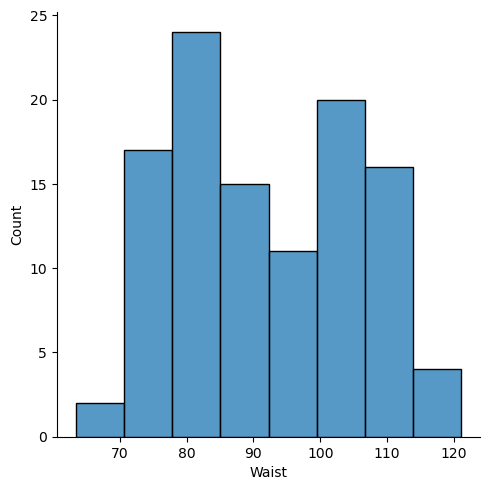
b)Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

1. Waist (Right Skewed- positively Skewed)

Skewness = 0.132

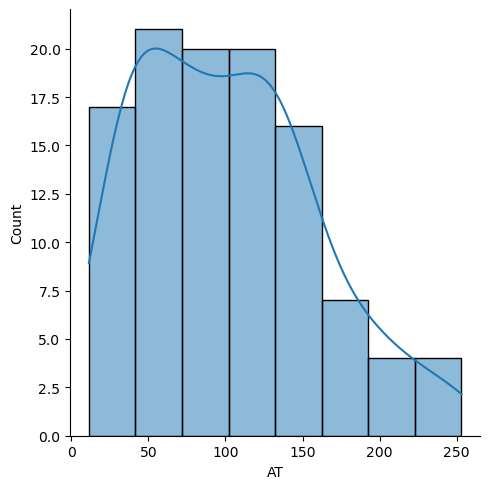
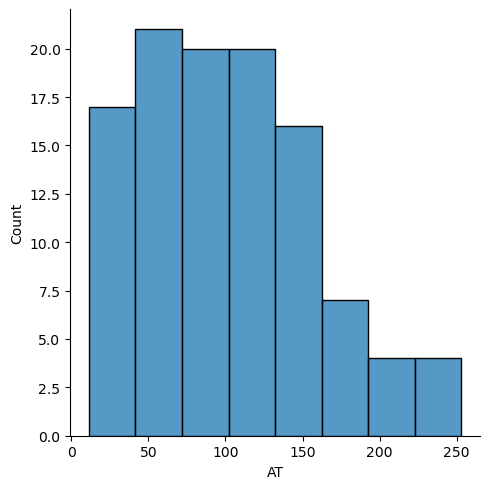
Kurtosis = -1.107



1. AT (Right Skewed- positively Skewed)

Skewness = 0.5767

Kurtosis = -0.327



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

A) Z scores of 90% confidence interval = 1.64

Z scores of 94% confidence interval = 1.88

Z scores of 60% confidence interval = 0.85

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

A) T scores of 95% confidence interval = 0.830

T scores of 96% confidence interval = 0.8315

T scores of 99% confidence interval = 0.8351

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

A) 32%