

| 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 | Categorical |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Categorical |

Q1) Identify the Data type for the Above :

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 | Interval |
| Blood Group | Nominal |
| Time Of Day | Interval |

| | |
|----------------------------|----------|
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

ANS:- $\frac{3}{8}$.

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

- a) Equal to 1
- b) Less than or equal to 4
- c) Sum is divisible by 2 and 3

ANS:- a) $\frac{0}{36}$

b) $\frac{6}{36} = \frac{1}{6}$

c) $\frac{6}{36} = \frac{1}{6}$

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:- $\frac{10}{21}$

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:- 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

ANS :-

| | Mean | Median | Mode | Variance | STD | Range |
|--------|---------|--------|--------------|----------|--------|-------|
| Points | 3.5965 | 3.6950 | 3.07, 3.92 | 0.2858 | 0.5346 | 2.17 |
| Score | 3.2172 | 3.325 | 3.44 | 0.9573 | 0.9784 | 3.911 |
| Weigh | 17.8487 | 17.71 | 17.02, 18.90 | 3.1931 | 1.7869 | 8.4 |

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:- Weight of that person = $(108+110+123+134+135+145+167+187+199) / 9$
= 145.33 (pounds)

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

Ans :-

Q9_a.csv

- a) Skewness of car-speed = - 0.118
- b) Kurtosis of car-speed = - 0.509
- c) Skewness of distance = 0.807
- d) Kurtosis of distance = 0.405

The speed of car has negative skewness and negative kurtosis.

The distance has positive skewness and positive kurtosis .

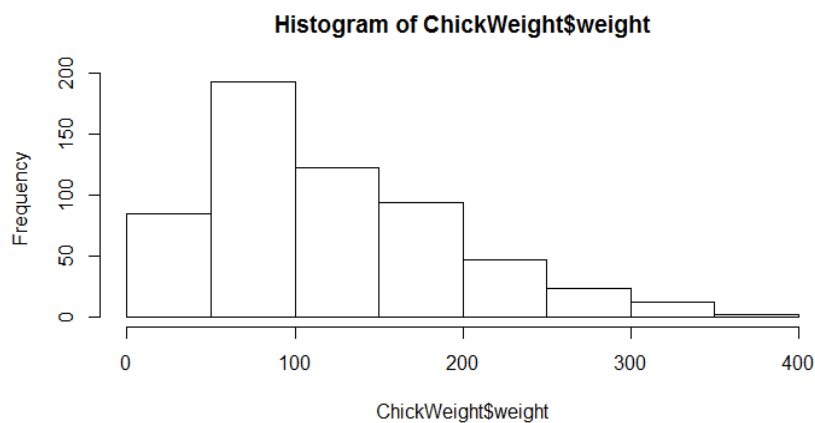
Q9_b.csv

- a) Skewness of speed (SP) = 1.611
- b) Kurtosis of speed (SP) = 2.977
- c) Skewness of weight (WT) = - 0.615
- d) Kurtosis of weight (WT) = 0.95

The speed (SP) has positive skewness and positive kurtosis.

The weight (WT) has negative skewness and positive kurtosis.

Q10) Draw inferences about the following



boxplot & histogram



ANS :- For Histogram :-

The histogram has positive skewness.

For box-Plot :-

The box-plot has outlier above the “upper fence “.

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 :-

`stats.norm.interval (0.94, loc = 200, scale = 30/np.sqrt(2000))`

1. confidence interval for 94% = 198.73833 , 201.2616

`stats.norm.interval (0.98, loc = 200, scale = 30/np.sqrt(2000))`

2. confidence interval for 98% = 198.4394 , 201.5605

`stats.norm.interval (0.96, loc = 200, scale = 30/np.sqrt(2000))`

3. confidence interval for 96% = 198.6223 , 201.3776

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 :-

- a) Mean = 41
- b) Median = 40
- c) Variance = 27.125
- d) Std. devn = 5.208

This students has most repeated marks in test is 41.

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

ANS :- If the values of mean and mode are equal then, the skewness will be zero (0).

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

ANS :- If the mean is greater than median then the skewness will be positive.

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

ANS :- If the mean is less than median then the skewness will be negative.

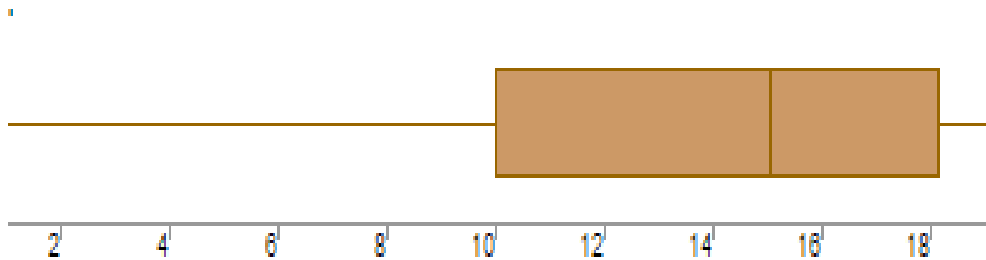
Q16) What does positive kurtosis value indicates for a data ?

ANS :- If the probability distrubution has positive kurtosis then it indicates that it has higher peak and has thick tails.

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

ANS :- If the probability distrubution has positive kurtosis then it indicates that it has flat distortion and has thin tails.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

ANS :- 1.)

a) Upper quartile of the above box-plot (Q3) = 10

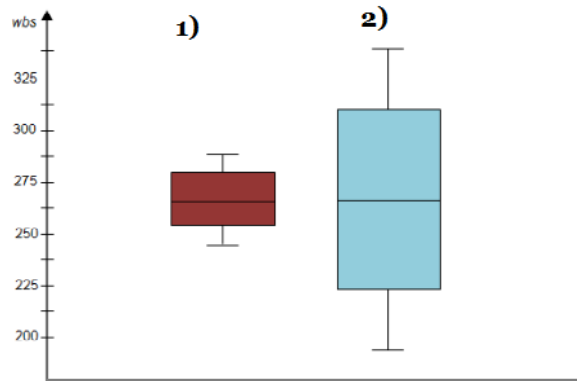
b) Lower quartile of the above box-plot (Q1) = 18

$$\begin{aligned}\text{IQR} &= Q3 - Q1 \\ &= 10 - 18 \\ &= -8\end{aligned}$$

2.) From the boxplot we can say that the nature of skewness is negative.

3.) The whisker for the above boxplot is less at the left side it is negative skew also the median is closer to the the right side.

Q19) Comment on the below Boxplot visualizations?



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

ANS :- From above we can say that both the box-plot has same median = 262.5

For box plot 1:- Upper quartile of the box-plot = 275
and Lower quartile of the box-plot = 250

$$\text{IQR} = 275 - 250 = 25$$

For box plot 2 :- Upper quartile of the box-plot = 300
and Lower quartile of the box-plot = 225

$$\text{IQR} = 300 - 225 = 75$$

As the median for both the box-plot is in the middle of and whisker are about the same in size on both side of box hence the distribution is symmetric.

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
```

a. $P(\text{MPG} > 38)$

b. $P(\text{MPG} < 40)$

$$c. \quad P(20 < \text{MPG} < 50)$$

ANS :-

```
cars = pd.read_csv("C:\\Users\\RUSHIKESH\\Downloads\\cars.csv")
```

MEAN = 34.42 AND STD = 9.13

a. $P(\text{MPG} > 38) :-$

$$Z = (38 - 34.42) / 9.13 = 0.39$$

$$P(\text{MPG} > 38) = 0.65173$$

b. $P(\text{MPG} < 40) :-$

$$Z = [(40 - 34.42) / 9.13] = 0.61$$

$$P(\text{MPG} < 40) = 1 - 0.729$$

c. $P(20 < \text{MPG} > 50) :-$

$$Z_1 = (50 - 34.42) / 9.13 = 1.706$$

$$P(\text{MPG} < 50) = 0.96080$$

$$Z_2 = (20 - 34.42) / 9.13 = -1.57$$

$$P(\text{MPG} < 20) = 0.5821$$

$$P(20 < \text{MPG} > 50) = P(\text{MPG} < 50) - P(\text{MPG} < 20)$$

$$= 0.9608 - 0.5821$$

$$= 0.3787$$

Q 21) Check whether the data follows normal distribution

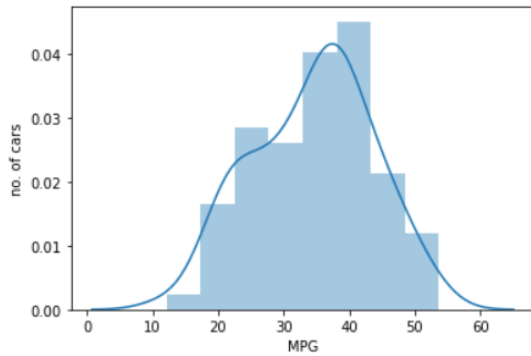
a) Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS :- YES

```
import seaborn as sn
```

```
sn.distplot(cars.MPG , )
```

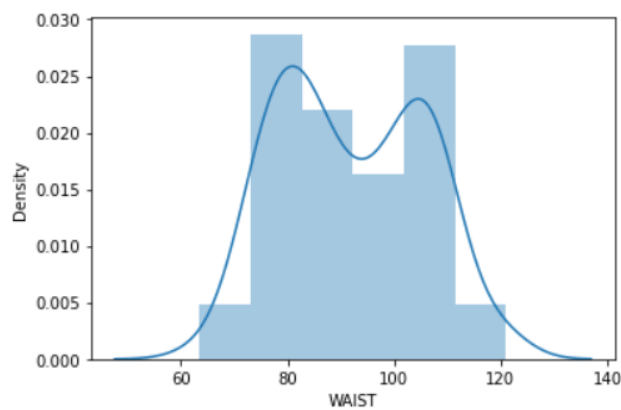


b) Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

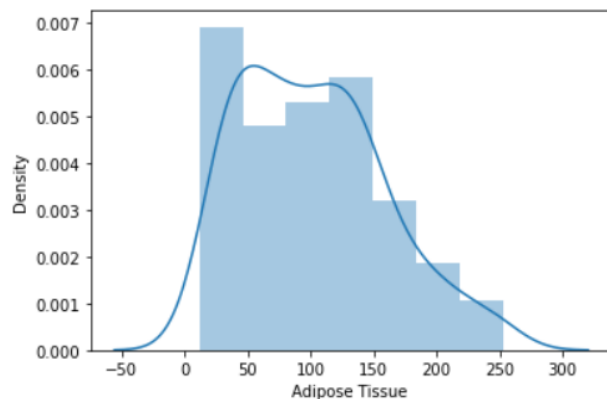
ANS :- 1.Waist circumference (waise) = NO

`sn.distplot(wc.Waist ,)`



2. Adipose Tissue (AT) = NO

`sn.distplot(wc.AT)`



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

ANS :-

- a. From scipy import stats
Stats.norm.ppf (0.95)
Z Score of 90% confidence interval = 1.645
- b. From scipy import stats
Stats.norm.ppf (0.97)
Z score of 94% confidence interval = 1.88
- c. From scipy import stats
Stats.norm.ppf (0.80)
Z score of 60% confidence interval = 0.841

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

ANS :- a. From scipy import stats

Stats.t.ppf (0.975 , 24)

t score of 95% confidence interval for sample size 25= 2.063

b. stats.t.ppf (0.98, 24)

t score of 96% confidence interval for sample size 25= 2.1715

c. `stats.t.ppf (0.995 , 24)`

t score of 96% confidence interval for sample size 25= 2.7969

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 t

Hint:

rcode \rightarrow `pt(tscore,df)`

df \rightarrow degrees of freedom

ANS :-

`Stats.t.cdf (260, 17, 270, 90)`

Where ,

X=260 Df = 17 , loc= 270, scale=90

The probability that 18 randomly selected bulbs would have an average life of no more than 260 days will be 0.456