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

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

Ans: When three coins are tossed the total numbers of possible combinations are

N(S) = {HHH, HHT, HTH, TTH, THH, THT, HTT, TTT}

= 8

The number of combinations which have two heads and one tail are

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

= 3

Probability of getting two head and one tail (P) = N(P)/N(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: When two dices are rolled sample space is

S = { (1,1) (1,2) (1,3) (1,4) (1,5) (1,6)

(2,1) (2,2) (2,3)S (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) }

1. The number of combinations which have sum equal to 1 are :

P(a) = 0/36 = 0

1. The number of combinations which have sum less than or equal to 4 are 6

P(b) = 6/36 = 1/6 = 0.167

1. The number of combinations which have sum divisible by 2 and 3

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

Hence probability is 6/36 = 1/6

P(c) =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: There are 7 balls with 2 of them blue so probability of the first ball not being blue is 5/7, then 6 balls left with 2 blue balls.

The probability of second ball not being blue is by considering first ball not blue is = 4/6

The probability of neither ball drawn is blue = (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

Ans: Suppose,

Candies Count = xi

Probability = P(xi)

Expected number = sum of [xi\*P(xi)]

={(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.

Ans:

1. Mean

df[['Points','Score','Weigh']].mean()

Points 3.596563

Score 3.217250

Weigh 17.848750

dtype: float64

1. Median

df[['Points','Score','Weigh']].mean()

Points 3.695

Score 3.325

Weigh 17.710

dtype: float64

1. Mode

df[['Points','Score','Weigh']].mode()

| **Points** | **Score** | **Weigh** |
| --- | --- | --- |
| **0** | 3.07 | 3.44 | 17.02 |
| **1** | 3.92 | NaN | 18.90 |

1. Standard Deviation & Variance

df[['Points','Score','Weigh']].std()

Points 0.534679

Score 0.978457

Weigh 1.786943

dtype: float64

df[['Points','Score','Weigh']].var()

Points 0.285881

Score 0.957379

Weigh 3.193166

dtype: float64

1. Range

df[['Points','Score','Weigh']].min()

Points 2.760

Score 1.513

Weigh 14.500

dtype: float64

df[['Points','Score','Weigh']].max()

Points 4.930

Score 5.424

Weigh 22.900

dtype: float64

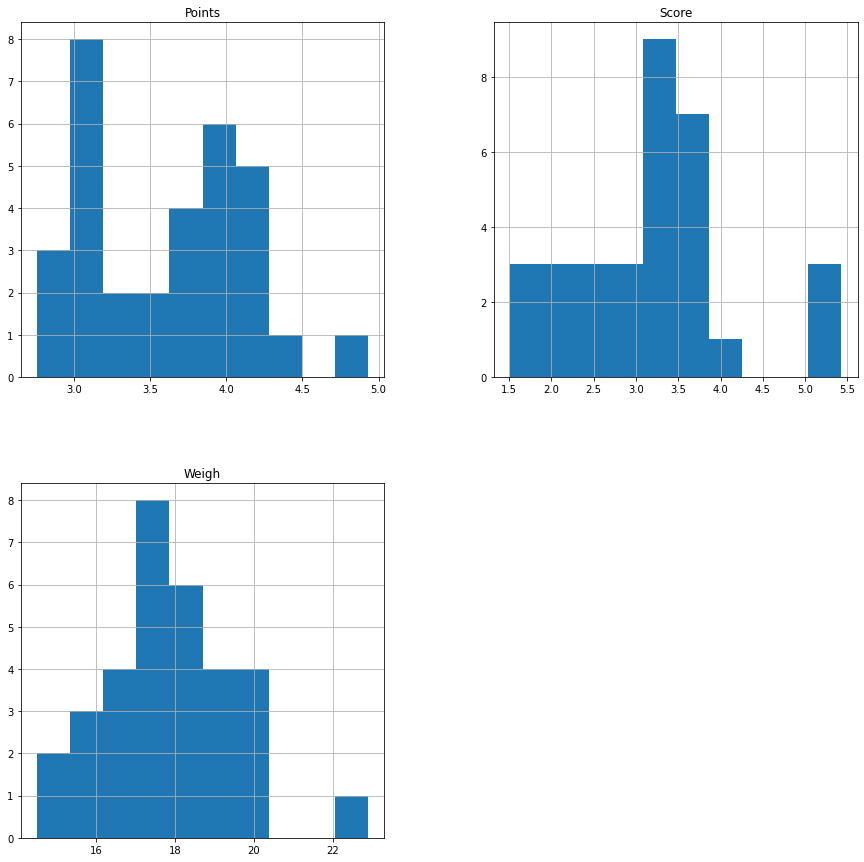
Here, Range is max - min

Points = 4.93 – 2.76

Score = 5.42 – 1.51

Weigh = 22.90 – 14.50

1. Inferences



Comments :

* It is a positively/right-skewed distribution.
* Points dataset is a bimodal because it is having two modes and Score, weight are unimodal because it having single mode.

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:

There are 9 patients

Probability of selecting each patient =1/9

Xi = 108,110,123,134,135,145,167,187,199

P(x) = 1/9

Expected Value = {(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\*1/9)}

= 145.33

Expected value of the weight of the patient is 145.33 kg.

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

**Cars speed and distance**

**Ans:**

1. **Skewness:**

**df[['speed','dist']].skew()**

speed -0.117510

dist 0.806895

dtype: float64

1. **Kurtosis:**

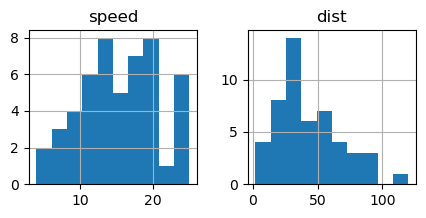
**df[['speed','dist']].kurtosis()**

speed -0.508994

dist 0.405053

dtype: float64

1. **Inferences:**

****

**Comments**:

* Speed data has high kurtosis
* Dist data is positively skewed and has high kurtosis

**SP and Weight(WT)**

**Use Q9\_b.csv**

1. **Skewness:**

**df[[‘SP,'WT']].skew()**

SP 1.611450

WT -0.614753

dtype: float64

1. **Kurtosis:**

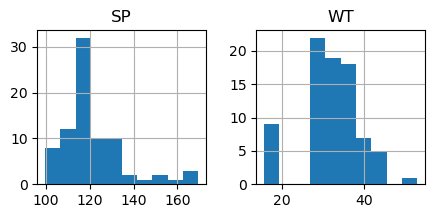
**df[[‘SP,'WT']].kurtosis()**

SP 2.977329

WT 0.950291

dtype: float64

1. **Inferences:**

****

**Comments**:

* SP data is positively skewed
* WT data is negatively skewed

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



**Ans:**

**Inferences**:

#**Histogram**:

The data has asymentry and it positively skewed.

The data has small kurtosis.

# **Boxplot:**

There is outliers on the Higher side of the boxplot

**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:**

from scipy import stats

print('Confidence Interval for 94% is:', stats.norm.interval(0.94,200,30/(2000\*\*0.5)))

print('Confidence Interval for 98% is:', stats.norm.interval(0.98,200,30/(2000\*\*0.5)))

print('Confidence Interval for 96% is:', stats.norm.interval(0.96,200,30/(2000\*\*0.5)))

Confidence Interval for 94% is: (198.738325292158, 201.261674707842)

Confidence Interval for 98% is: (198.43943840429978, 201.56056159570022)

Confidence Interval for 96% is: (198.62230334813333, 201.37769665186667)

In [ ]:

**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:**

1. **Mean, Median, Variance & Standard Deviation:-**

a=pd.DataFrame([34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56])

print(type(a))

#Mean

print('Mean:',a.mean())

#Median

print('\nMedian:',a.median())

#Variance

print('\nVariance',a.var())

#Standard Deviation

print('\nStandard Deviation',a.std())

O/p: <class 'pandas.core.frame.DataFrame'>

Mean: 0 41.0

dtype: float64

Median: 0 40.5

dtype: float64

Variance 0 25.529412

dtype: float64

Standard Deviation 0 5.052664

dtype: float64

1. Average marks of eighteen students are 41. Maximum marks obtain by student are 56 and minimum marks obtain by student are 34.

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

Ans: If mean & median of the data are equal that means the data is symmetric, and distribution has zero skewness.

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

Ans: If mean is greater than median then the distribution of data is positively skewed.

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

Ans: If median is greater than mean than the data distribution is negatively skewed.

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

Ans: Positive Value of kurtosis indicate that the distribution of data is at peak and has thick tails.

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

Ans: Negative value of kurtosis indicates that the distribution of data 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 the range of -2 and 30 and distributed on right side.

What is nature of skewness of the data?

Ans: The data is right/positively skewed.

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

Here, Q1 = 10, Q3 = 18, median = 18

IQR = 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.

Ans: Boxplot1 and Boxplot2 has same median value, 100% data of Boxplot1 is less than 50% of the data of Boxplot2, the range of Boxplot1 is less than range of Boxplot2. There are no outliers in both boxplots.

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

Data \_set: Cars.csv

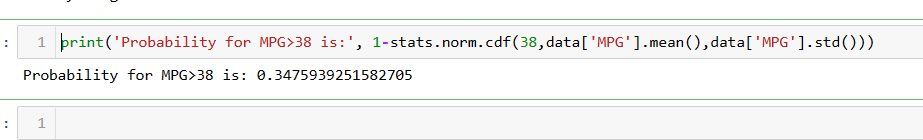
Calculate the probability of MPG of Cars for the below cases.

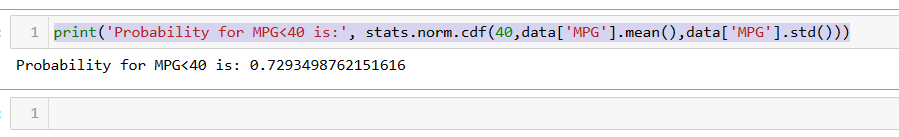
MPG <- Cars$MPG

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

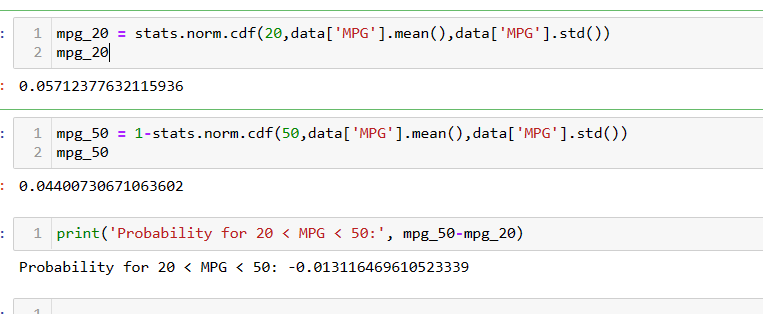
c. P (20<MPG<50)

Ans:

1. 



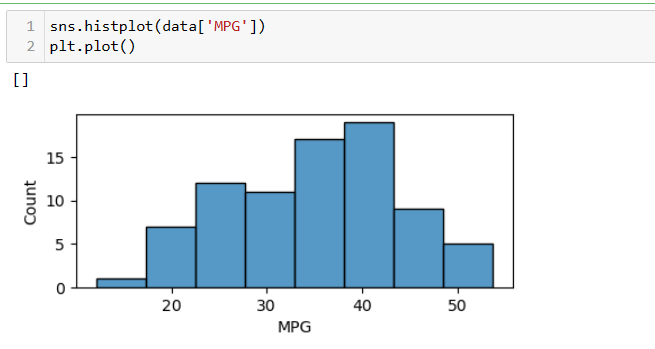




Q 21) Check whether the data follows normal distribution

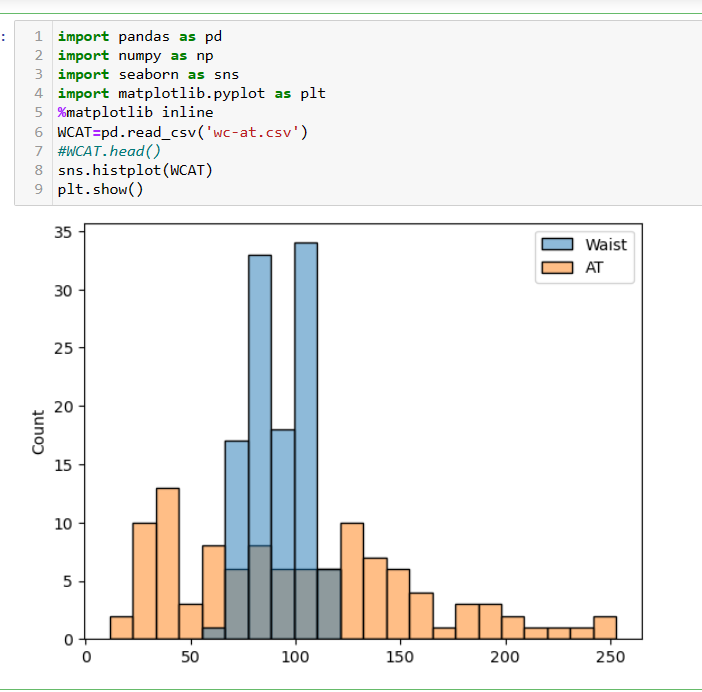
1. Check whether MPG of Cars follows Normal Distribution:

Dataset: Cars.csv



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

Dataset: wc-at.csv



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

Ans:

Z score for 90% CI:



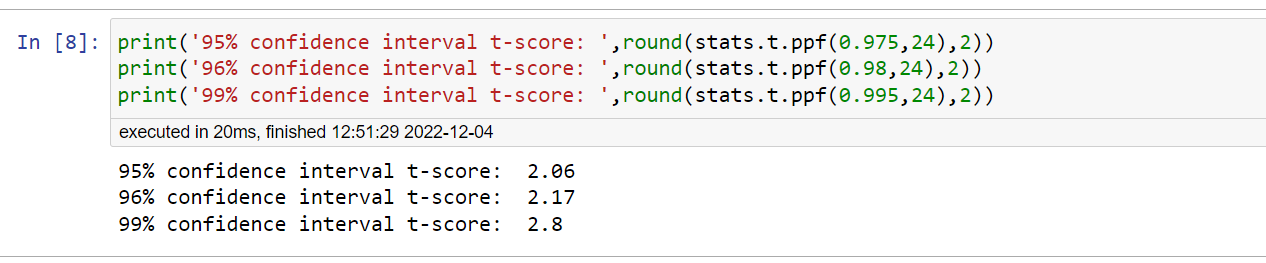
Z score for 94% CI:



Z score for 60% CI:



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

Ans:

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

