|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete [Countable] |
| 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 | Ordinal |
| Celsius Temperature | Interval |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Ordinal |
| 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 | Nominal |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Ratio |
| Years of Education | Ordinal |

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

Let us assume ‘S’ is the sample space then, S = [H,T] [H,T][H,T]

When three coins are tossed the probability of getting two heads and one tail

P = [ H,H,T] [H,T,H] [T,H,H] The occurance of getting head is 2

P = [1\2 \* 1\2\*1\2]+[1\2\*1\2\*1\2] +[1\2\*1\2\*1\2] = 1\8+1\8+1\8 =3\8

So, the probability of getting two heads and one tail when three coins are tossed is 3 8

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

1. Equal to 1 = 0
2. Less than or equal to 4 = 1\6
3. Sum is divisible by 2 and 3=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?

## A : 🡪 n(x) is no.of ways of drawing 2 balls out of 2

## nCr = 7C2 = 7\*6/2\*1 = 42/2=21

## 🡪y event of drawing 2 balls , none of which is blue

## 🡪n(y) is no.of ways of drawing 2 balls out of 2+3 balls

## nCr = 5C2 =5\*4/2\*1 = 20/2 = 10

P(E) =n(y)/n(x)

= 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 =( 1\*0.015)= 0.015

# Child B – probability =( 4\*0.20) = 0.8

# Child C – probability =( 3\*0.65) = 1.95

# Child D – probability =( 5\*0.005) = 0.025

# Child E – probability =( 6\*0.01) = 0.06

# Child F – probability =( 2\* 0.120) = 0.24

## Sol: Expected number of candies for randomly selected child =

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

# = 3.9

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**

Mean:

## Points : 3.596563

## Score : 3.217250

## Weigh : 17.848750

Median:- median is middle most value is known as “median”

### Points : 3.695

## Score : 3.325

## Weigh :17.710

Mode:- most frequently occurring values is known as “mode”

## Points= 0 3.07

## 1 3.92

## Score = 0 3.44

## Weigh= 0 17.02

## 1 18.90

Variance:- variance is variety of information more information more variance

## Points : 0.285881

## Score : 0.957379

## Weigh : 3.19316

## 

## 

## 

## Standard deviation:- It is square root of variance is known as “standard deviation”

## 

## Points : 0.534679

## Score : 0.978457

## Weigh : 1.786943

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 :

## The above given , the weights(x) of patients at a clinic are

## (x) = 108,110,123,134,135,145,167,187,199

## 

## P(x) = No. of events/ total no. of events

## = 108+110+123+134+135+145+167+187+199/9

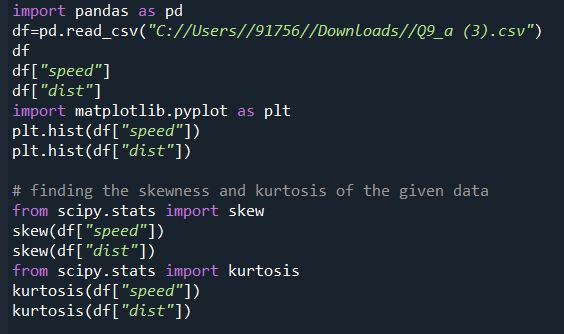
## = 1308/9

## =145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

****

**a): Skewness:**

## speed : -0.11751

## distance: 0.806895

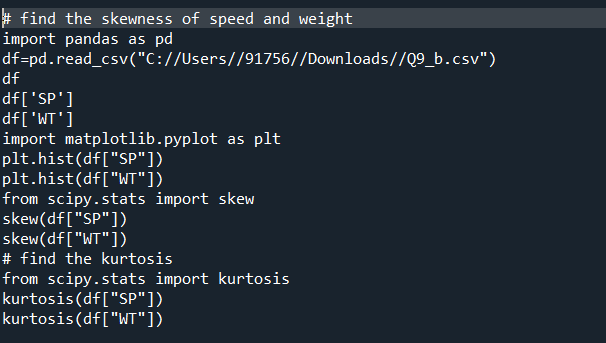
**Kurtosis:**

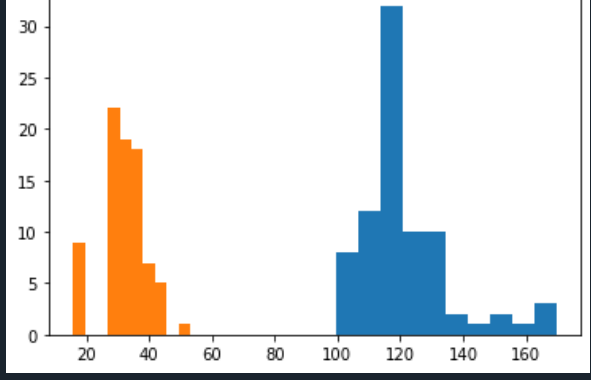
## Speed: -0.508994

## Distance: 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

****

Skewness:

## SP : 1.611450

## WT : -0.64753

Kurtosis:

## SP : 2.97732

## WT : 0.950291

CONCLUSION :

## Skewness of speed= -0.1139548, skewness value is negative so it is left skewed. Hence magnitude is slightly greater than 0 it is sightly left skewed.

## And for DISTANCE = 0.7824 835 , right skewed i.e Positive ,slight magnitude is right skewed.

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



## 🡪Histogram is a 3rd moment of business Decision.

## 🡪 in histogram, data is continuous and shape of the distribution line is positively symmetry .

## 🡪in histogram intervals are present .

## 🡪 Mode<median<mean

## 🡪the most of the data points are in range 50 – 100 with frequency 200.

## 🡪so the expected value the above distribution is 75.

## 🡪skewness- we can observe a long tail towards right ,hence it is right skewed.

S





BOX PLOT :

## It is also called as box and “ whisker plot”. It is used for detecting the outliers .

## Boxplot is having 5 piece of information .

## I)the Minimum

## 2) Q1

## 3) The median

## 4)Q3

## 5)The maximum

## 🡪 Outliers are appeared above boxplot.

## 🡪in box plot ,upper whisker length is greater than the lower whisker length

## 🡪Median is less than the mean ,right skewed and outliers on the upper side of the boxplot and there is less data points between Q1 and bottom point.

**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

## Import numpy as np

## df\_ci = stats.norm.interval(0.94,loc=200.scale=30)

## print(“94% confidence interval is:” ,np.round(df\_ci,4)

## df\_ci = stats.norm.interval(0.98,loc=200.scale=30)

## print(“98% confidence interval is:” ,np.round(df\_ci,4)

## df\_ci = stats.norm.interval(0.96,loc=200.scale=30)

## print(“96% confidence interval is:” ,np.round(df\_ci,4)

Confidence interval values are:

## 94% confidenece interval is : [143.5762 265.4238]

## 98% confidenece interval is : [130.2096 269.7904]

## 94% confidenece interval is : [138.3875 261.6125]

**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.

Mean = 41.05,

Median = 40.5

Variance = 24.111

1. What can we say about the student marks?

ANS: Students marks are given as sequence order like ascending order

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

ANS: Symmetry

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

ANS: Right skewed

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

ANS: Left skewed

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

ANS: Positive kurtosis value indicate that the distribution is peaked possesses the thick tails. It is normally distributed and kurtosis is 0

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

ANS: Negative kurtosis value of the data indicates that the distribution is flat and having thin nails.

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



What can we say about the distribution of the data?

ANS : Asymmetric(negative or left skewness of the data)

What is nature of skewness of the data?

# ANS: Negative skewness of the data, median is greater than the mean

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

# ANS: IQR = Upper Quartile – Lower Quartile

= 18-10 = 8(approximately)

Q19) Comment on the below Boxplot visualizations?



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

# ANS:

## 🡪 In both plots, whisker’s level is high in boxplot 2.

## 🡪mean and median are equal hence distribution is symmetrical.

## 🡪there is 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:

## MPG which are greater than 38.

## P(MPG>38) = 0.347593

## (B) MPG which are lesser then 40.

## P(MPG<40) = 0.7293

## (C) Between 20 to 50 observations are find out.

## P(20<MPG<50) = 0.898868

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

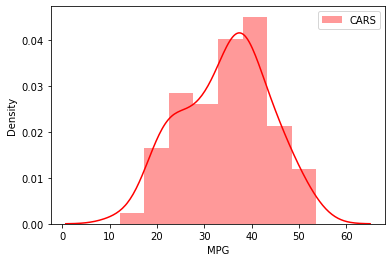
Dataset: Cars.csv

## Mean = 34.422075728024666

## Median = 35.152

## Skewness= -0.17796474

## Kurtosis = -0.611678655.



Normal distribution Properties :

## 🡪The shape of the normal distribution always symmetrical and having bell shaped curve .

## 🡪Mean = Median.

## 🡪Skewness = 0.

## 🡪Kurtosis = 3.

Conclusion :

## To calculate the given Cars$MPG data is Normal Distribution ,to satisfy the normal distribution properties .but solved the problem don’t follow normal distribution properties.

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

Dataset: wc-at.csv

Waist :

## Mean = 91.9018

## Median = 90.8

## Skewness = 0.13405

## Kurtosis = -1.102666

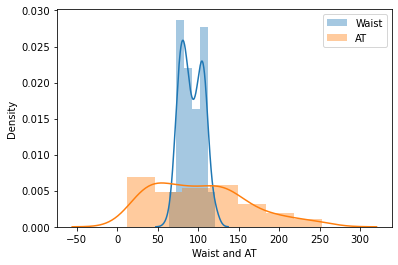
AT:

## Mean = 10.89403669

## Median = 96.54

## Skewness = 0.5848

## Kurtosis = -0.2855756750458445



Conclusion :

## The above Waist and AT do not satisfy the normal distribution properties so we can decide the both are not follow normal distribution .

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

# ANS:

# Z scores of 90% confidence interval is 1.65

# Z scores of 94% confidence interval is 1.55

# Z scores of 60% confidence interval is 0.8

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

# ANS:

# t scores of 95% confidence interval is 1.96

# t scores of 96% confidence interval is 2.5

t scores of 99% confidence interval is 2.47

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 :

## Avg of light bulbs lasts 270 days

## Sample size(n)= 18

## Sample avg xbar (s) = 260

## Sample standard deviation (S) = 90

## True and selected bulbs would have an avg of life no more than 260 days

## t-statistic/t-score = [xbar-MU)/(S)/sqrt root of n]

## = [260-270/90/sqrt root of 18]

## = [-10/90\*sqrt root of 18]

## = [-0.4714]

# Probability of p-value :

## Degree of Freedom (df) = n-1

## = 18-1

## = 17

## left of third test t = (-0.4714) and ( df)

# p-value = (0.3372)

# Hypothesis Testing :

## 🡪 If p-value >0.05

## Accept the null-hypothesis (H0)

# 🡪If p-value<0.05

## Accept the Alternate Hypothesis (H1)

## To compare the p-value of hypothesis testing >=20 so accept the Null hypothesis test .

## The probability that 18 randomly selected bulbs would have an average life of no more than 260 days .