Q.1

|  |  |
| --- | --- |
| ACTIVITY | DATA TYPE |
| Number of beatings from wife | Discrete data |
| Results of rolling a dice | Discrete data |
| Weight of a person | Continuous data |
| Weight of gold | Continuous data |
| Distance between two places | Continuous data |
| Length of leaf | Continuous data |
| Dogs weight | Continuous data |
| Blue color | Discrete data |
| Number of kids | Discrete data |
| No.of tickets in Indian railways | Discrete data |
| No. of times married | Discrete data |
| Gender (male or female) | Nominal data |

Q.2 identify the data types , which were among the following

Nominal,ordinal, interval, ratio

|  |  |
| --- | --- |
| DATA | DATA TYPE |
| Gender | Nominal data |
| High school class ranking | Ordinal data |
| Celcius temperature | Interval data |
| Weight | Ratio data |
| Hair colour | Nominal data |
| Socioeconomic status | Ordinal data |
| Fahrenheit temperature | Interval data |
| Height | Interval data |
| Type of living accommodation | Nominal data |
| Level of agreement | Ordinal data |
| IQ(intelligence scale) | Interval data |
| Sales figures | Ratio data |
| Blood group | Nominal data |
| Time of day | Ratio data |
| Time on clock with Hands | Interval data |
| Number of children | Ratio data |
| Religious preference | Ordinal data |
| Barometer pressure | Ratio data |
| SAT score | Interval data |
| Years of education | Ratio data |
|  |  |

Q.3 Three coins are tossed ,find the probability that two heads and one tail are obtained?

ANS:

when three coins are tossed the total no.of possible combinations are 2^3=8.these combinations are HHH,HHT,HTH,THH,TTH,THT,HTT,TTT. The number of combinations which have to heads and one tail are : HHT,HTH,TTH which makes then 3 in number . therefore the probability of getting two heads and one tails in the toss of three coins simultaneously is defined as: P(two heads and one tail)=no.of desired outcomes

=3=0.375.

Q.4 ) Two dice are rolled find the probability that sum is,

a)equal to 1

b)less than or equal

c)sum is divisible by 2 and 3

ANS:

total possible outcome =6^2=36

a) The probability thatr the sum is equal to 1 is given by

Total cases : 36 favourable cases : OP=0/36

The probability that the sum is equal to 1 is zero.

b)the probability that the sum is less than or equal to 4 is given by:

total cases: 36 favourable cases:3P=3/36

the probability that the sum is less tan or equal to 4 is 8.33%.

c) the probability that the sum is divisible by 2 and 3 is given by:

total cases :36 =36/2=18 and 36/3=12 .so 36 is divisible by 2 and 3.

Q.5) 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:

Total number of balls = (2+3+2)=7

Let s be the sample space .

Then ,n(s)= number of ways of drawing 2 balls out of 7

=7C2

= (7\*6)/(2\*1)

=21

Let E =event of drawing 2 balls ,none of which is blue .

Therefore,n(E)= number of ways of drawing 2 balls out of (2+3)balls.

=5C2

=(5\*4)/(2\*1)

= 10

Therefore, P(E )=n(E) /n(S)= 10/21.

Q.6) 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 -generalied 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 :

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

=0.015+ 0.8+ 1.95 + 0.025 + 0.06 + 0.24

=3.090

=3.09

Expected number of candies for a randomly seleted child = 3.09.

Q.7 ) calculate mean ,median,mode ,variance,standard deviation ,range and 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 .

|  |  |  |
| --- | --- | --- |
| Points | Score | Weight> |
| 3.9 | 2.62 | 16.46 |
| 3.9 | 2.875 | 17.02 |
| 3.85  3.00  3.15 | 2.32  3.210  3.44 | 18.61  19.44  17.02 |
| 2.76 | 3.46 | 20.22 |
| 3.21 | 3.57 | 15.84 |
| 3.69 | 3.19 | 20 |
| 3.92 | 3.15 | 22.9 |
| 3.92 | 3.44 | 18.3 |
| 3.92 | 4.07 | 18.9 |
| 3.07 | 3.73 | 17.4 |
| 3.07 | 3.73 | 17.6 |
| 3.07 | 3.78 | 18 |
| 2.90  3 | 5.25  5.424 | 17.90  17.02 |
| 3.23 | 5.345 | 17.42 |
| 4.08 | 2.2 | 19.47 |
| 4.93 | 1.615 | 18.52 |
| 4.22 | 1.835 | 19.9 |
| 3.7 | 2.465 | 20.01 |
| 2.76 | 3.52 | 16.87 |
| 3.15 | 3.435 | 17.3 |
| 3.73 | 3.84 | 15.41 |
| 3.08 | 3.845 | 17.05 |
| 1.08  4.40 | 1.235  2.14 | 18.0  16.7 |
| 3.77 | 1.513 | 16.9 |
| 4.22 | 3.17 | 14.5 |
| 3.62 | 2.77 | 15.5 |
| 3.54 | 3.57 | 14.6 |
| 4.11 | 2.78 | 18.6 |

ANS :

MEAN :

Mean = (sum of all the values)/(number of values )

So,the mean for points ,score,and weight are :

Mean of points

=3.590625

Mean of score

=3.21725

Mean of weight

Median of score = (3.29 +3.44)/2=3.365

Median of weight = (18.805 +18.755)/2=18.78

=19.8478125

Therefore, the mean of points is 3.590625,the mean of score is 3.21725, nad the mean of weight is 19.8478125.

Median :

To calculate the median ,we need to arrange the values in ascending order and find the middle value.

The median for points ,score ,and weight are :

Median of points =(3.07+3.15)/2=3.11

Mode:

To calculate the mode ,we need to find the value that appears the most number of times in the dataset . if there are multiple values that appear with the same frequency ,then there is no mode .

The mode for points ,score ,and weight are :

Mode of points =3.92

Mode of score =3.44

Mode of weight=no mode.

Q.8) 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 :

The weight (X) is expcted value

Expected value = sum of (probability \*value ) sum of P(X).E(X)

There are 9 patients

Probability of selecting each patient =1/9

Ex. 108,110,123,134,135,,145,167,187,199

P(X) 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9 1/9

Expected value =(1/9)(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)(108+110+123+134+135+145+167+187+199)

=(1/9)(1308)

=145.33

Expected value of the weight of that patient =145.33.

Q.9)calculate the skewness ,Kurtosis ,and draw inferences of following data cars speed and distance.

|  |  |
| --- | --- |
| Speed | Distance |
| 4 | 2 |
| 4 | 10 |
| 7 | 4 |
| 7 | 22 |
| 8 | 16 |
| 9 | 10 |
| 10 | 18 |
| 10 | 26 |
| 10 | 34 |
| 11 | 17 |
| 11 | 28 |
| 12 | 14 |
| 12 | 20 |
| 12 | 24 |
| 12 | 28 |
| 13 | 26 |
| 13 | 34 |
| 13 | 34 |
| 13 | 46 |
| 14 | 26 |
| 14 | 36 |
| 14 | 60 |
| 14 | 80 |
| 15 | 20 |
| 15 | 26 |
| 15 | 54 |
| 16 | 32 |

ANS :

Given data:

Import stastics

Import numy

Data =[ [4,2],[4,10],[7,4],[7,22],[8,16],[9,10],[10,18],[10,26],[10,34],[11,17],[11,28],[12,14],[12,20],[12,24],[12,28],[13,26],[13,34],[13,34],[13,46],[14,26],[14,36],[14,60],[14,80],[15,20],[15,26],[15,54],[16,32]]

#extract first column as x\_values and second column as y \_values

X\_values =[x[0]for x in data]

Y\_values =[x[1] for x in data]

#calculate mean

Mean\_x=stastics .mean (x\_values)

Mean\_y= stastics mean(y\_values)

#calculate median

Median \_x=stastics .median (X\_values )

Median \_y=stastics .median (Y\_values )

#calculate mode

mode \_x=stastics .mode(X\_values)

mode\_y=stastics .mode(Y\_values)

#calculate standard deviation

Stdev\_x=stastics .stdev(X\_values)

Stdev\_y=stastics .stdev(Y\_values)

#calculate skewness

Skewness\_x=(3\*(mean \_x-median \_x))/stdev\_x

Skewness\_y=(3\*(mean \_y-median \_y))/stdev\_y

#calculate kurtosis

Kurtosis\_x=((mean\_x-mode\_x)/stdev\_x)\*\*4

Kurtosis\_y=((mean\_y-mode\_y)/stdev\_y)\*\*4

Print (“skewness for x \_values is :”,skewness\_x)

Print(“skewness for y \_values is :”,skewness\_y)

Print (“kurtosis for x \_values is:”,kurtosis\_x)

Print (“kurtosis for y\_values is :”,kurtosis\_y)

To calculate the skewness and kurtosis ,you would need to use the formulas after calculating he mean,median,mode ,and standard deviation of the data .

Skewness:

Skewness is a measure of the asymmetry of the probability distribution of a real valued random variable about its means .

Skewness =(3(mean-median))/standard deviation

Kurtosis:

Kurtosis is measure of the “talledness”of the probability distribution of a real valued ramdom variable .

Kurtosis= ((mean-mode)/standard deviation )^4

Q.10) draw inferences about the following boxplot and histogram .

ANS :



Here we can see that the major chick weights fall in the category of 50 -100g (measure in x) as the maximum which is 200. The minimum weights have a frequency if less than or equal to 5.

The plot os right sqewed which show that there is lesser concentration of chick weights in the 300-400g category .

The expected value should be above 46.45

Median is less than mean right skewed and we have outlier on the upperside of box plot and there is less data points between Q1 and bottom point.

Q.11) suppose we want to estimate the average weight of an adult male in Mexico.we draw a random sample of 2,000men from a population of 3,000,000men and weight them.we find that the average person in our sample weights 200 pounds ,and the standard deviation of the sample is 30 pounds .calculate 94%,98%,96% confidence interval ?

ANS:

In which t is the critical value for the two tailed confidence interval.

Considering a 94% confidence level,using a calculator ,with 200-1=199 df,the d=critical value is t=1.8916,hence:

-t s/=200-1.8916 30/= 198.73

+t s/=200+1.8916 30/= 201.27

The 94% confidence interval is (198.73,201.27).

Considering a 98% confidence level ,using a calculator with 200-1=199 df,the critical value is t

=2.3452, hence :

-t s/=200 – 2.3452 30/= 198.43

+ s/=200+ 2.3452 30/=201.57

The 98% confidence interval level is (198.43,201.57).

Considering a 96% confidence level ,using a calculator ,with 200-1=199 df, the critical value is t =2.0673,hence:

– t s / =200-2.0673 30/ =198.61

+ t s/ = 200+2.0673 30 / =201.39

The 96% confidence interval level is (198.61,201.39)

Q.12) below are the scores obtained by a student in tests

34,36,36,38,38,39,39,40,40,41,41,41,42,42,45,49,56

Find mean ,median , mode ,variance, standard deviation .

ANS :

Mean =

(34+36+36+38+38+39+39+40+40+41+41+41+42+42+45+49+56)/18

=41

There are 18 elements .so middle elements are 9 th and 10 th elements ,that is 40 and 41.so median is=

(40+41)/2=40.5

In this series 41 is present in highest number of time that is four time .so mode =41

Variance =5.05

Standard deviation =25.529.

Q.13) what is the nature of skewness when mean ,median of data are equal ?

ANS:

Nature of skewness : skewness can be positive ,negative or zero, when the value of mean ,median and more are equal there is no skewness

Skewness is a measure of the asymmetry of probability distribution of real valued mean.

Q.14) what is the nature of skewness when mean>median?

ANS :

The mean ,mode and median can be used to figure out if you have a positively or negatively skewed distribution . if the mean is greater than the median ,the distribution is positively skewed .

Q.15) what is the nature of skewness when median >mean?

ANS :

Skewness and tail are towards left.

If the median is greater than the mean ,the distribution is negatively skewed.

Q.16) what does positive kurtosis value indicates for a data?

ANS :

Positive values of kurtosis indicates that distribution is peaked and possesses thick tails. An extremely positive kurtosis indicates a distribution where more numbers are located in the tails of the distribution instead of around the mean.

Q.17) what does negative kurtosis value indicates for a data?

ANS :

A distribution with anegative kurtosis value indiacres that the distribution has lighter tails than the normal distribution.

Negative values for the skewness indicate data that are skewed left and positive values for the skewness indicates data that are skewed right. By skewed left ,we mean that the left tails is long 1

Q.18) answer the below qutions using the below boxplot visualisation.

1)What can we say about distribution of the data?

2)What is the nature of skewness of the data?

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

ANS :

1. The data is not actually equally distributed across the plane .there might be outliers influencing the data.median of the data is 14.7

25 percent of the data lies between 0-10

50 percent of the data lies between 10-18

25 percent of the data lies after 18=20 approximately

1. The data will be left skewed since whisker length on the upper quadrent is higher than the data on the lower quadrant .median will be greater than the mean since data is left skewed.
2. IQR is the inter quartile range .

Here Q1=10

Q2=14.7

Q3=18

IQR = Q3-Q1=8(approx.)

Q.19) comment on the below boxplot visualization?

ANS :

Here there is a representation of 2 box plots in which box plot 2) is a highly distributed across the plane and 1) is slightly less distributed .

Whiskers in these diagrams also show this .100% of the data is spread across values from 350 in 2 whereas its spread in range 250-290 app x in 1)

Draw an inference from the distribution of data for box plot 1 with respect box plot 2.

Here when we compare box plot 1 with box plot 2 we can say that the data in box plot 1 is widely spread. Hare the main inference is that since the data range varies high in box plot 2 it is hard to make a prediction in box plot 2 , the median in the 2 box plots are equal .and the data spread in both of them are symmetrical.

Q.20) calculate probability from the given data set 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)

ANS :

P (MPG>38)

=mean (MPG) = 34.42208

=sd(MPG)=9.131445

=1-pnorm(38,mean (MPG),sd(MPG))

=0.330

=33%

P (MPG<40)

=pnorm (40,mean (MPG),sd(MPG))

=0.7293499

=72.3%

P (20<MPG<50)

=pnorm (50,mean (MPG),sd(MPG))-pnorm(20,mean(MPG),sd(MPG))

=0.955-0.057

=0.8988689

Q.21) check whether the data follows normal distribution

a) check whether the MPG of cars follows normal dataset : cars .csv

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. When we plot check the qqnorm and qqline we can almost get a straight line thus the data is normalized .
2. Majority of the data points lie on the qqline hence normal.

This data set is not normal because the data points follows an abnormal curve.

Q.22) calculate the z scores of 90% confidence interval ,94% confidence interval , 60% confidence interval .

ANS :

Z scores

=90%

=95+2.5

=97.5

=qnorm (0.975)

=1.96

94%

=94+4

=97

=qnorm(0.97)

=1.88

60%

=60+20

=80

=qnorm(0.80)

=0.841

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

ANS:

TSCORE CALCULATION

T((1,alpha ),(n-1))

Here n=25

n-1=24

hence t score values will be :

95%

=qt(0.975,24)

=2.063899

96%

=qt(0.98,24

=2.171545

99%

=qt(0.995,24)

=2.79694

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 :

R code -pt(score,df)

Df-degrees of freedom

ANS :

Sample size =18=n

Sample mean =260days =x

Sample standard deviation =s=90 days

=260-270/90/SQRT(18)

=-10/9.487

=-1.054.