|  |  |
| --- | --- |
| `Activity | Data Type |
| Number of beatings from Wife | Numeric (continuous) |
| Results of rolling a dice | Numeric (discrete) |
| Weight of a person | Numeric (continuous) |
| Weight of Gold | Floating value(continuous) |
| Distance between two places | Numeric (discrete) |
| Length of a leaf | Numeric (discrete) |
| Dog's weight | Floating (continuous) |
| Blue Color | characteristics |
| Number of kids | Numeric (discrete) |
| Number of tickets in Indian railways | Numeric(continuous) |
| Number of times married | Numeric (discrete) |
| Gender (Male or Female) | Characteristics |

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

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

Answer:-These combinations are{ (HHH), (HHT), (HTH), (THH), (TTH), (THT), (HTT),( TTT)}.

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

HHT, HTH, THH which makes them 3 in number

2^3=8 we can take 3/8

Probability of getting two heads one tails 3/8

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

Answer: A 0 outcome

B There are 6 possibility outcome’s 6/36=1/6

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

(2,1) (2,2) (2,3) (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) ]

The possibility outcomes are[(1,5) (2,4)(3,3)(4,2) (5,1)(6,6)]

The probability outcomes for sum is divisible by 2 and 3 are 6

6/36=1\6

The probability of sum is divisible by 2 and 3 is 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?

**Answer**:- Total number of balls =(2+3+2)=7

Lets s be the sample space

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

7c2 =

Formula (n!)

-----------

r!(n-r)!

therefore 7!

------------

2!(7-2)!

7!

----------- = 21 ways

2!.5!

Let E be the event of drawing two balls , non of the balls are blue balls

5c2=

(n!)

Formula -----------

r!(n-r)!

therefore 5!

------------

2!(5-2)!

5!

----------- = 10 ways

2!.3!

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

Answer:- 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 \* 0.12

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

=       3.090

=  3.09

Child A – probability of having 1 candy = 0.015= 1\*0.015=0.015

Child B – probability of having 4 candies = 0.20=0.8

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**

**Answer** :- Mean for point=3.596563

Mean for score = 3.21725

Mean for weigh =17.84875

Median for point = 3.695

Median for score = 3.325

Median for weigh = 17.71

Standard deviation for points = 0.094519

Standard deviation for score = 0.172968

Standard deviation for weigh = 1.786943

Variance for points = 0.285881

Variance for score = 0.957379

Variance for weigh = 3.193166

Range for points = 2.17

Range for score = 3.911

Range for weigh = 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?

Answer:- (1/9)\*(108)+(1/9)\*(110)+(1/9)\*(123)+(1/9)\*(134)+(1/9)\*(145)+(1/9)\*(167)+(1/9)(187)+(1/9)\*(199)

=(1/9)\*(1308)

=145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer :-**  skewness for carspeed is = -0.11751

Skewness for distance = 0.806895

Kurtosis for speed = -0.50899

Kurtosis for distance =0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

skewness for sp is = 1.61145

Skewness for weight= -0.61475

Kurtosis for sp= 2.977329

Kurtosis for weight= 0.950291

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



Answer:- In the histogram data where check weight is on x axis and frequency is located on y axis there are some points are occurred in histogram which represents on x axis which indicates the weights is calculated in frequency is distributed , from 0 to 100 weights is occurred in between 50 to 100 , from 100 to 200 there are some weights are on x axis the frequency the weights are disturbuted x between in between from 100 to 200 , from 200 to 300 there are some weights in the frequency the data points are fall between 0 to 100, from 300 to 400 there are some points are in y axis from 0 to 100

(ii) In the above boxplot diagram minimum point of the box plot is very near to the quantile 1 , the median of the of the box blot is far from quantile 3 maximum is very far from the quantile 3 there are some suspected outliers

**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?

Answer:- z=x-u

------

Sd

Where x= 2000

u = 200pounds

standard deviation = 30

2000-200

----------- = 13.33it is positive

30

At 96% of confidence interval = 1+0.96

--------=0.98= +/-2.06

2

At 94% of confidence interval 1+0.94

--------=0.97=+/-1.89

2

At 98% of confidence interval 1+0.98

--------=0.99 =+/- 2.33

2

**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?

Answer: mean = 34+36+36+38+38+40+40+41+41+41+41+42+45+49+56

------------------------------------------------------------------------- =41

18

Median = 40+40 80

------------- ------- = 40

2 2

Mode = 41

Variance = 31.35294

Standard deviation = 1.3580

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

Answer :- If the distribution is symmetric, then the mean is equal to the median, and the distribution has zero skewness. If the distribution is both symmetric and unimodal, then the mean = median = mode.

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

Answer:- 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. If the mean is less than the median, the distribution is negatively skewed

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

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

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

Answer:- Positive values of kurtosis indicate that a distribution is peaked and possess thick tails.

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

Answer:- Negative values of kurtosis indicate that a distribution is flat and has thin tails.A platykurtic distribution is flatter (less peaked) when compared with the normal distribution, with fewer values in its shorter

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)?   
  
Answer:-

(iii) As per the given above boxplot Q1 is 10,median is 15 & Q3 is 18

Q19) Comment on the below Boxplot visualizations?



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

Answer :- By above given boxplot 1 as we can see that min value is 225 q1 is

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)

A . P(MPG>38)

Ans :- there are 33 possible outcomes out of 82 entries which are grater than 38

B. P(MPG<40)

ANS :- there are 61 possible outcomes, out of 82 entries which are less than 40

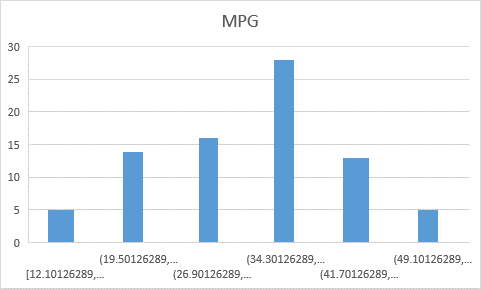
C. P (20<MPG<50)

Ans:- there are 76 possible outcomes which lies between 20 and 50

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



Mean = 34.42208

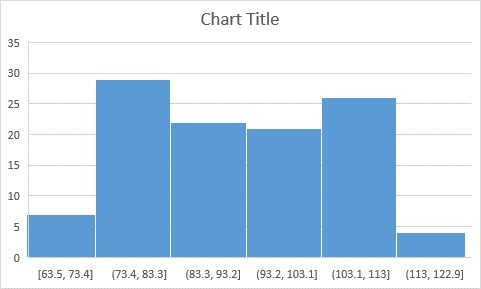
Median = 35.1523

Mode = 29.629

Hence median >mean>mode

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

Dataset: wc-at.csv

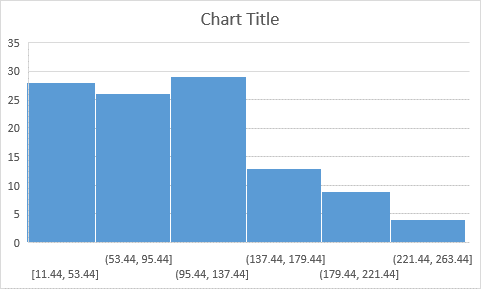
Waist Circumference(Waist):- 

Mean = 91.90183

Median =90.08

Mode = 94.5

Hence mode>mean>median

Adipose Tissue (AT)

Mean = 101.894

Median = 96.5

Mode= 121

Hence mode>mean>median

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

Answer:

Formula 1+confidence interval

------------------------------

2

At the confidence interval for 90 % is = 1+0.9

----------- = 0.95

2

At the confidence interval for 94% is = 1+0.94

----------- = 0.97

2

At the confidence interval for 94% is = 1+0.6

----------- = 0.8

2

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

Sample size = 25

Degree of freedom = n-1 = 24

At the 95% of confidence interval = 2.06390

At the 99% of confidence interval = 3.7454

At the 96% of confidence interval = 2.492

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 random

ly selected bulbs would have an average life of no more than 260 days

Answer :-

Formla Is :- x-

--------

S

-------

Square root(n)

X = mean of the sample bulbs =260

µ = population of bulbs = 270

s = standard deviation is 90 days

n = number of items in the sample 18

260-270

-------------

90

-----------

Square root of 18

-0.471

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom