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

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

ANS:- when three coins are tossed the total number of possible combinations are 8

No. of combination which have two head and one tail= {HHT,HTH,THH}

P(X=2)= probability of occurrence of two head and one tail

=P(HHT)+P(HTH)+P(THH)

= 1/8 + 1/8 + 1/8

= 3/8 or 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:- two dice, there are ( 6 \* 6 = 36 ) possible combinations of numbers.

The minimum sum possible for the two dice thrown is (1, 1) = a sum of (2 )

The maximum sum possible for the two dice thrown is (6, 6) = a sum of (12).

1. Sum = 1

The minimum possible sum is (1, 1) = ( 2 ).

Therefore P( 1 ) = 0 / 36 = 0

1. sum less than or equal to 4

The possible combination of sum is less than or equal to 4 = (1,1) ,(1,2), (1,3), (2,1), (2,2), (3,1)

P(Sum is less than or equal to 4) = 6 / 36 = 1 / 6

1. Sum is divisible by 2 and 3

The possible combination of sum is divisible by 2 and 3 = (1 , 5) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)

P(Sum is divisible by 2 and 3) = 5/36

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:- 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.  
n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls.  
=5C2

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

=10

P(E)= n(E)​ / n(S) = 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 :- 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

= 3.09

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:- For Points:-

Mean = 3.596563

Median = 3.695

Mode = 3.92

Variance = 0.276948

Standard Deviation = 0.526258

Range = 2.17

For Score:-

Mean = 3.21725

Median = 3.325

Mode = 3.44

Variance = 0.927461

Standard Deviation = 0.963048

Range = 3.911

For Weight:

Mean = 17.84875

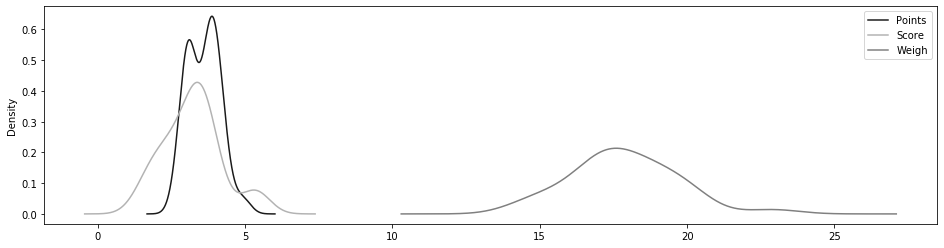
Median = 17.71

Mode = 17.02

Variance = 3.09338

Standard Deviation = 1.758801

Range = 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?

ANS:- Expected Value  =  ∑ ( probability  \* Value )

  = ∑ P(x)\*E(x)

there are 9 patients

Probability of selecting each patient = 1/9

E(x):-  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

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 pounds

Expected Value of the Weight of that patient = 145.33 pounds

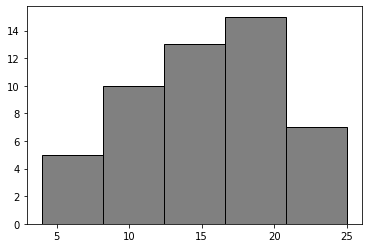
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**ANS:-** FOR CARS SPEED

Skewness:- -0.11751

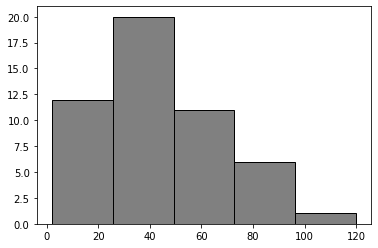
Kurtosis :- -0.50899



FOR CARS DISTANCE

Skewness:- 0.806895

Kurtosis :- 0.405053

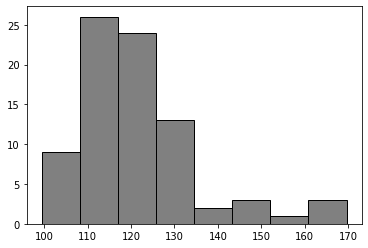


* SP and Weight(WT)

FOR CARS SP

Skewness:- 1.61145

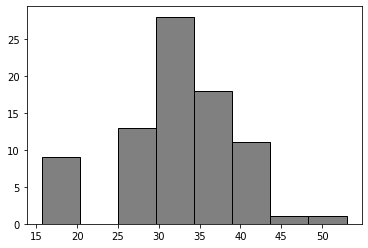
Kurtosis :- 2.977329



FOR CARS WEIGHT

Skewness:- -0.61475

Kurtosis :- 0.950291



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



ANS:-

Chick weight data is right skewed or positively skewed.

0-50 chick weight having frequency 80.

50-100 Chick Weight having more frequency 180.

100-150 chick weight having frequency 120.

150-200 chick weight having frequency 100.

200-400 chick weight having very less frequency below 50.

Data is not a normal distribution.



ANS:-

7 outliers are present in above box plot.

Positive skewness.

Q1 is smaller than the Q3.

**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:- x̄ = 200 pounds, S = 30 pounds, n = 2000

We don’t have standard deviation for population. So we have to use the T-distribution to determine the confidence interval.

The confidence interval for 94% is [198, 201].

The confidence interval for 98% is [198.6, 201.3].

The confidence interval for 96% is [198.4, 201.4].

**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= 41

Median= 40.5

Variance = 25.5294

Standard deviation = 5.05266

2)Avg of student marks 41.

The student markes range from 34 to 56.

Mode is 41

Most of students score is bw 35 to 42

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

ANS:- when the mean, median of data are equal, there is no skewness also you can say the data is in normal distribution.

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

ANS:- If the mean is greater than the median, the distribution is positively skewed.

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

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

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

ANS:- A distribution with a Positive values of kurtosis indicate that a distribution has heavier tails than the normal distribution.

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

ANS:- A Negative kurtosis means that your distribution is flatter than a normal curve with the same mean and standard deviation.

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



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

ANS:-

* No outliers.
* Q1 greater than Q3.
* Median is between 15 to 16.
* Most of data present in range of 10 to 18.
* Not follower normal distribution.

1. What is nature of skewness of the data?

ANS:- Left skewness of data.

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

ANS:- IQR = 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:-

* Both the plots infer that their data is normally distributed.
* We can say that box plot 1 is for sample distribution and box plot 2 is for population or a sample with larger size.
* No outliers.
* Q1 is 25%, Q3=75%. IQR is 50% for both the box plots. So we can say both the distribution s follow normal distribution ( mean=median=mode).

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

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

MPG <- Cars$MPG

* 1. P(MPG>38)

ANS:- 0.3475939

* 1. P(MPG<40)

ANS:- 0.7293499

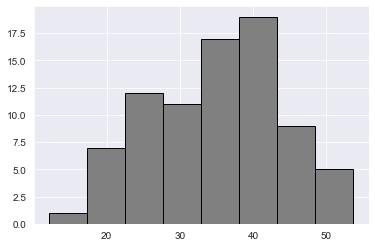
* 1. P (20<MPG<50)

ANS:- 0.8988689

Q 21) Check whether the data follows normal distribution

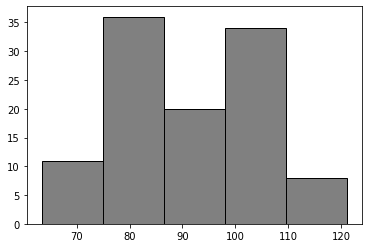
1. Check whether the MPG of Cars follows Normal Distribution

ANS:- MPG of cars not following normal distribution ,



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

ANS:- Waist & AT does 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 = 1.6448536269514722

Z scores of 94% confidence interval = 1.8807936081512509

Z scores of 90% confidence interval = 0.8416212335729143

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

ANS:-

the t scores of 95% confidence interval= 2.063899

the t scores of 95% confidence interval= 2.171545

the t scores of 95% confidence interval= 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:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

ANS:- t - statistics for the data is given as follows:



x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

t= {(260-270) / (90/√18)}

t=(-1 \* √2) / 3

t= - 0.471

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.