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

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

Answer 🡪 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 🡪 29/36

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?

Probability of drawing 2 blue balls =2/7 \* 1/6=2/42

Probability that none of the ball is blue= 1-2/42=40/42

20/21

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

Expected Value = ∑ (probability \* Value )

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

Expected value=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.

**Points**

Mean: - 3.596563, median:- 3.695,Mode:- 3.07 3.92,variance:- 0.2858814,SD:- 0.5346787,Range:- 2.76 4.93

Min value=2.76 and max= 4.93

There is very less deviation since SD value is approaching 0

There are 2 values (3.07 ,3.92) which are repeating max num of times

**Score**

Mean: - 3.21725, median:- 3.325,Mode:- 3.44,var:- 0.957379,SD:- 0.9784574,Range:- 1.513 5.424

**Weigh**

Mean: -17.84875, median:- 17.71,Mode:- 17.02 18.90,Var:- 3.193166,SD:- 1.786943,Range:- 14.5 22.9

**Use Q7.csv file**

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?

Expected Value  =  ∑ ( probability  \* Value )

 ∑ 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

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

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer:-

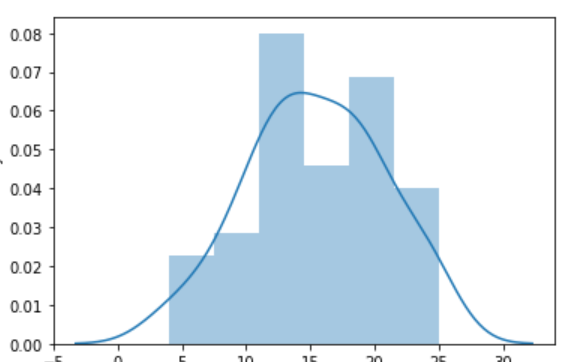
For Speed

Skewness value 🡪 -0.11750986144663393

Speed is left skewed

Kurtosis value 🡪 0.8068949601674215

Speed has lighter tails than normal distribution since value is less than 3



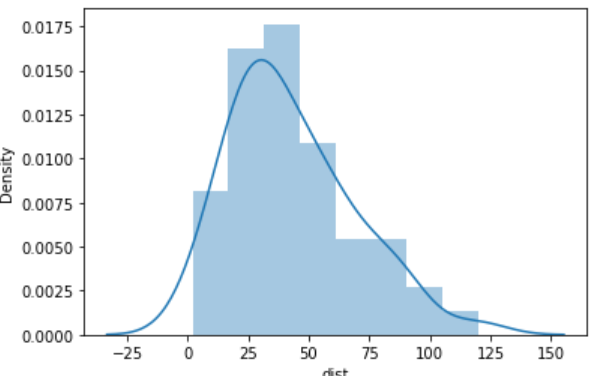
For distance

Skewness value 🡪0.8068949601674215

Distance is right skewed

Kurtosis value 🡪 -0.5089944204057617

Distance has lighter tails than a normal distribution since value is less than 3



**SP and Weight(WT)**

**Use Q9\_b.csv**

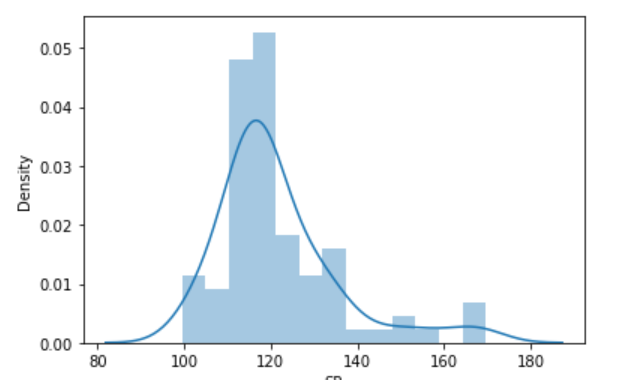
**SP**

Skewness:- 1.6114501961773555

Sp is right skewed

Kurtosis:- 2.9773289437871764

Tails are bit heavier



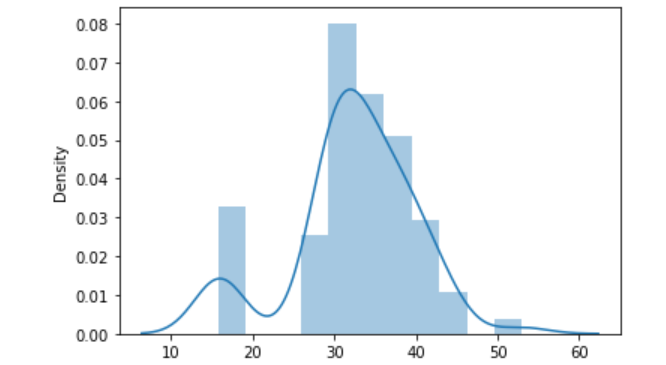
**Weight**

Skewness :- -0.6147533255357768

Weight is left skewed

Kurtosis:- 0.9502914910300326

Tails are lighter than that of normal distribution.



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



Maximum of chicks weigh between the interval 0 and 200

There are very less number of chicks between 200 and 400

200 chicks weigh between 50-100 which is maximum amongst all other intervals

Almost 1-2 chicks weigh between 350-400



There are outliers which lie above upper extreme which are as follows:-

318 331 327 341 332 361 373 321 322

More values lie in upper whisker

Q3=163.8 Q1=63

**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?

198.738325292158, 201.261674707842------- 94%

198.43943840429978, 201.56056159570022-------98%

198.62230334813333, 201.37769665186667-----96%

**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 ,median-40.5,SD-5.052664

1. What can we say about the student marks?

Deviation is not much and students have scored almost alike.

56 is an outlier.

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

Zero skewed- symmetric

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

Positive skew

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

Negative skew

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

It means sharp peak and tails almost touch the axes

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

It means broad peak and the tails are much above the axes

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



What can we say about the distribution of the data?

More values in lower whisker.

No outliers

What is nature of skewness of the data? Left skewed

What will be the IQR of the data (approximately)? 8

Q19) Comment on the below Boxplot visualizations?



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

* Median value in case of boxplot 1 is equal to the median for boxplot 2 (262.5)
* Data in Boxplot 1 is less widely spread as compared to boxplot 2.
* Both Boxplot 1 and boxplot 2 are normally distributed.
* None of them have any outliers.
* Q1 and Q3 values for boxplot 1 are 250 and 275 whereas those for boxplot 2 are
* 225 and 312.5 respectively.
* IQR for boxplot 1 is 25 and IQR for boxplot 2 is 86.5

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) --🡪-0.34
  2. P(MPG<40) -🡪0.73

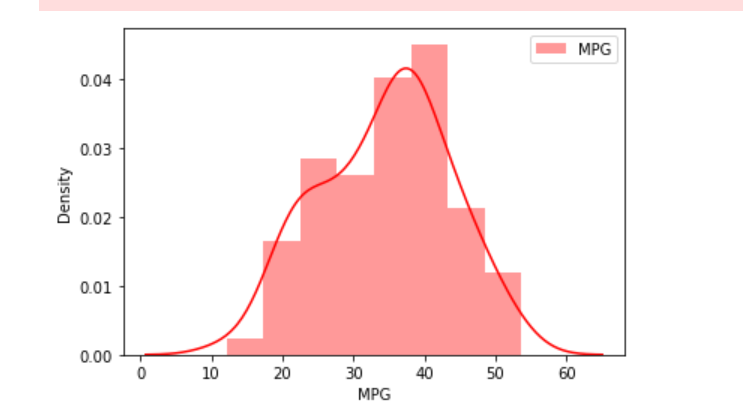
c. P (20<MPG<50)🡪0.90

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

No,it is not normally distributed since it can be seen in the graph that bell shaped curve is not formed and also In our case mean,median and mode are not equal which is a condition that should hold true for normal distribution

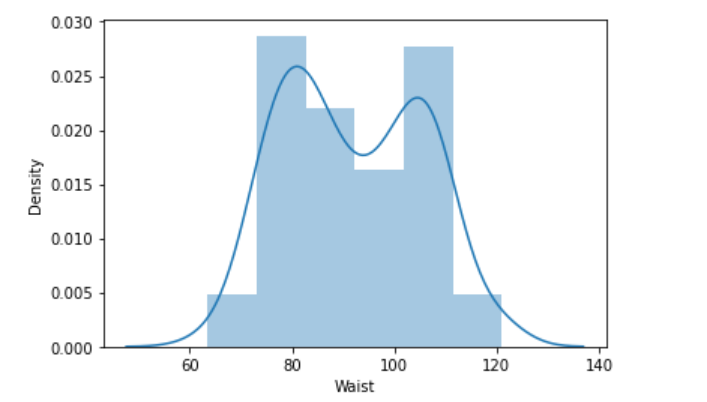


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

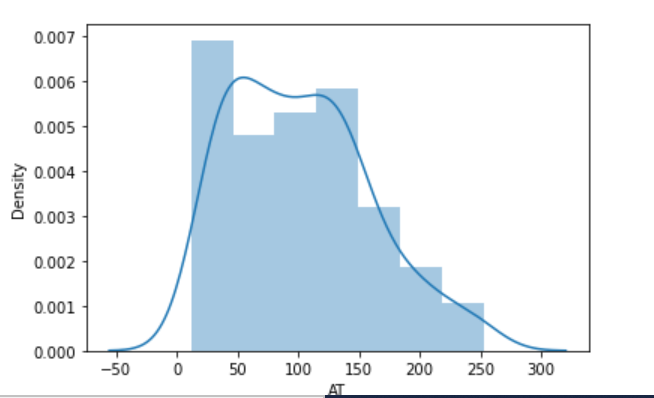
Waist

No,it is not normally distributed since it can be seen in the graph that bell shaped curve is not formed and also In our case mean,median and mode are not equal which is a condition that should hold true for normal distribution



Adipose Tissue (AT)

No, it is not normally distributed since it can be seen in the graph that bell shaped curve is not formed and also In our case mean, median and mode are not equal which is a condition that should hold true for normal distribution



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

90%--- 1.64

94%- 1.88

60%-- 0.84

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

95% -- 2.064

96%-- 2.172

99%-- 2.797

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

Answer :- (260-270) /90/np.sqrt(18))=-0.4714

🡪 0.321