|  |  |
| --- | --- |
| 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 a leaf | Continuous data |
| Dog's weight | Continuous data |
| Blue Color | Discrete data |
| Number of kids | Discrete data |
| Number of tickets in Indian railways | Discrete data |
| Number of times married | Discrete data |
| Gender (Male or Female) | Discrete data |

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 | Nominal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ordinal |
| Years of Education | Nominal |

Q3) Three Coins are tossed, find the probability that two heads and one tail are obtained?  
=Probability of two heads and one tail / total possible combination

=3/8

Total number of events is = 8 (HHH/TTT/HTH/THH/HHT/TTH/HTT/THT)

No. of interested events for 2 heads and 1 tail is = 3 (HTH/THH/HHT)

Probability = No. of interested events/Total no of events

Probability = 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

a)

Since there are no outcomes which correspond to a sum equal to 1

Therefore in this case i.e, equal to 1 is Probability = No. of interested events/Total no of events Probability

= 0/36=0

b)

Less than or equal to 4 {(1,1) (1,2) (1,3) (2,1) (2,2) (3,1)}

Probability = No. of interested events/Total no of events Probability

= 6/36

=1/6 (0.16666)

c)

There are 5 rolls that produce 6, i.e., 1–5, 2–4, 3–3, 4–2, 5–1. This we have 6 of the 36 possible rolls that produce sums that are divisible by both 2 and 3.

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?

n(S)=7C2

n(S)=(7×6)/(2×1)

n(S)=21

Let E = Event of 2 balls, none of which is blue  
∴ n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(E)=5C2

n(E)=(5×4)/(2×1)

n(E)=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

=

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

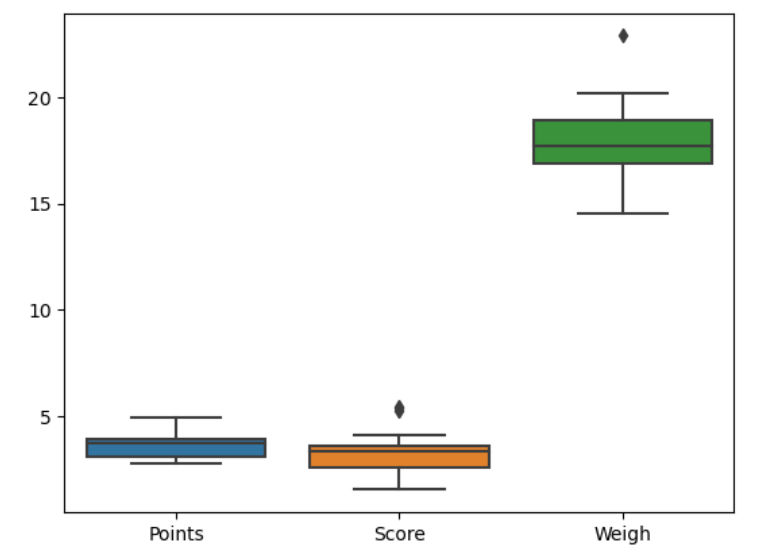
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* 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**

|  |  |  |  |
| --- | --- | --- | --- |
| **Results** | **Points** | **Score** | **Weigh** |
| **Mean** | **3.597** | **3.217** | **17.849** |
| **Median** | **3.695** | **3.325** | **17.710** |
| **Mode** | **3.07, 3.92** | **3.440** | **17.02, 18.90** |
| **Range** | **2.170** | **3.911** | **8.400** |
| **Std** | **0.535** | **0.978** | **1.787** |
| **Variance** | **0.286** | **0.957** | **3.193** |



Out layers :there are out layers present in score and weight data set

Upper and lower extreme points are closer to IQR in point and score as data points are less but in weight spread is more so large data point

Range : variation is high in weight as the value is 8.400 as compared to score and point data 2.170 , 3.911

Q8) Calculate Expected Value for the problem below

1. The weights (X) of patients at a clinic (in pounds), are

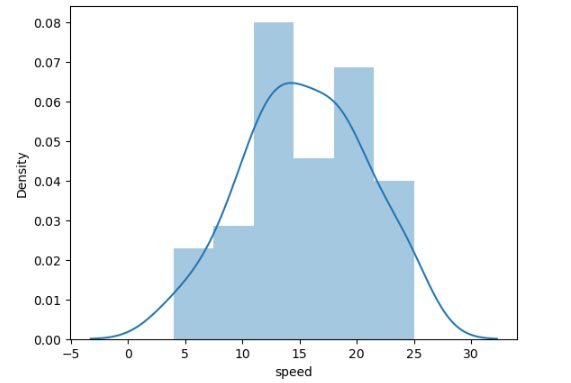
108, 110, 123, 134, 135, 145, 167, 187, 199

Assume one of the patients is chosen at random. What is the Expected Value of the Weight of that patient?

Expected Value= (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)/9 =145.333

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

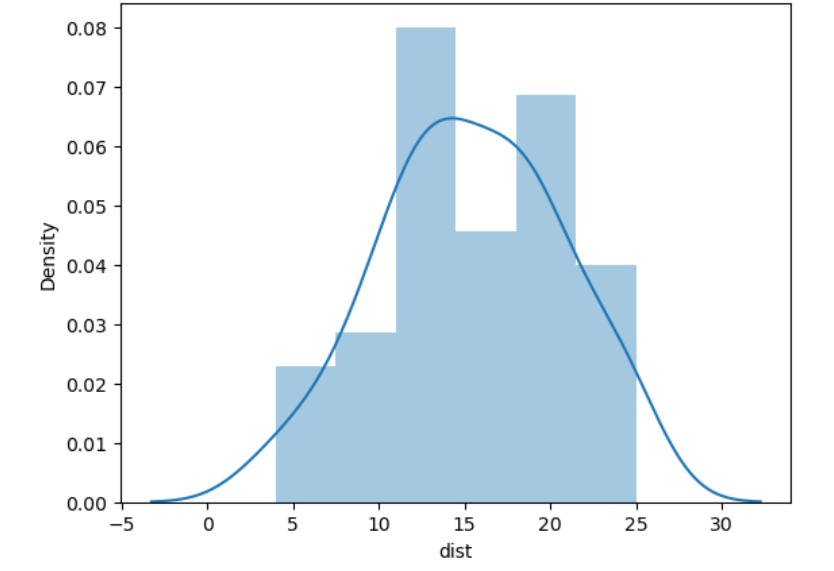
**Cars speed and distance**



Speed data is left skewed/negative as mean(15.4) < mode (20)

Skewness is -0.11395

Kurtosis is -0.57714



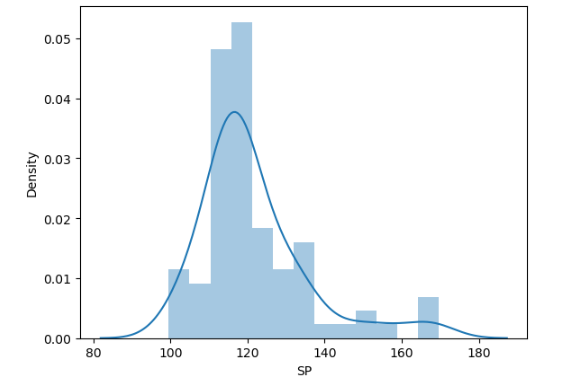
Distance data is Right skewed /positive as mean (42.98) > mode (26)

Skewness is 0.782483517

Kurtosis is 0.24801865

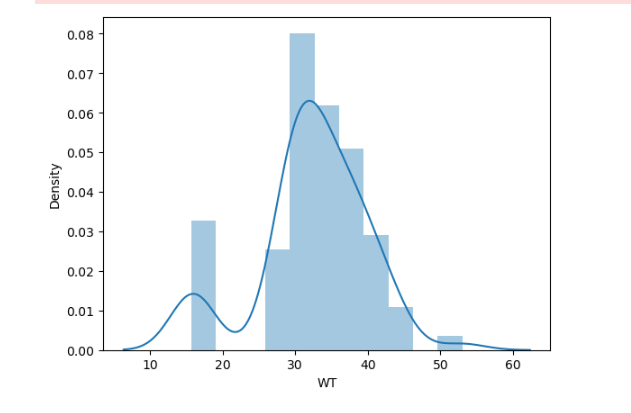
**SP and Weight(WT)**

**Use Q9\_b.csv**



SP data is Right skewed /positive as mean (121.54) > mode (118.28)  
Skewness is 1.5814

Kurtosis is 2.723



WT data is Right skewed /positive as mean (32.41) > mode (15.71)  
Skewness is -0.6033

Kurtosis is 0.8194

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



Histogram represents distribution is normal or not

According to above histogram it is not normal also there are out layers we can see

From box plot

**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?

N = 2000 , σ = 30, x = 200

|  |  |  |
| --- | --- | --- |
| Confidence Level | Z value | CI |
| 94% | 1.88 | (198.738 , 201.261) |
| 96% | 2.053 | (198.622 , 201.377) |
| 98% | 2.326 | (198.440 , 201.560) |

**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?

|  |  |
| --- | --- |
| Result | Values |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.52 |
| std. deviation | 5.05 |

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

When mean median are equal , the curve is standard normal distribution

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

when mean > median the degree of asymmetry or skewness is positive/skewed right

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

when median > mean the skewness is negative/skewed Left

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

A positive value for the kurtosis indicates a distribution more peaked than normal.

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

In contrast, a negative kurtosis indicates a shape flatter than normal

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



What can we say about the distribution of the data?

Ans : not normally distributed

As there is large difference between min point and first quartile

What is nature of skewness of the data?

Negatively 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.

There is no out layer

Median is same for both

But the data points are more in boxplot 2

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

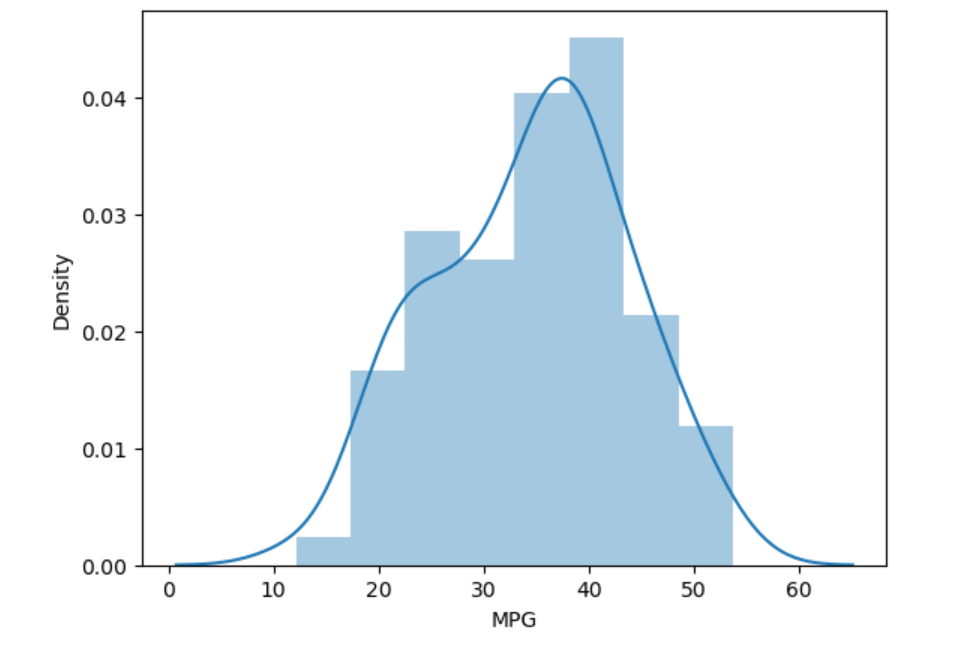
* 1. P(MPG>38) = 0.3475
  2. P(MPG<40) = 0.7293
  3. P (20<MPG<50) = 0.8988

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

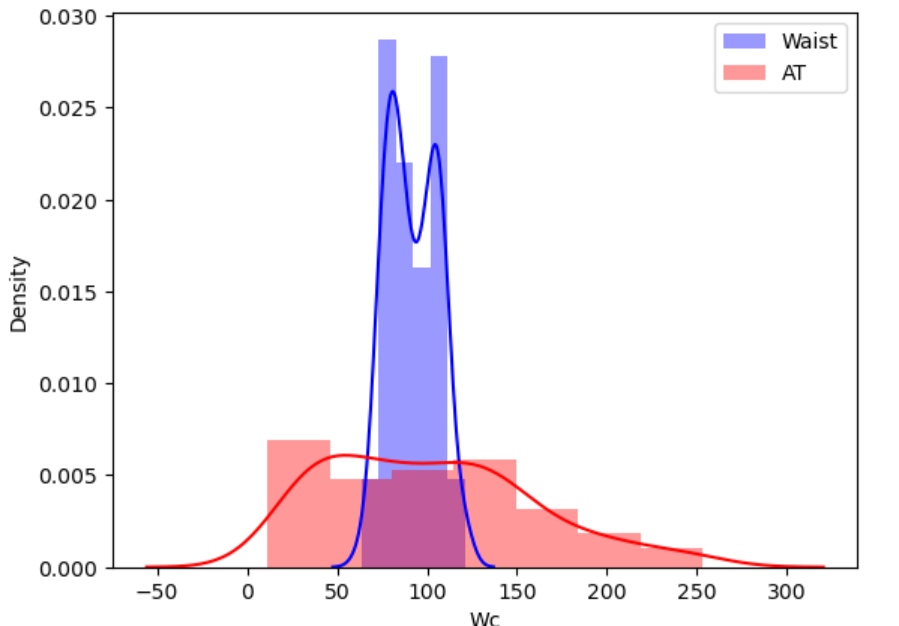
MPG data set of cars does follow Normal distribution



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

Dataset: wc-at.csv

Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follow Normal distribution



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

|  |  |  |
| --- | --- | --- |
| Confidence Level | P value | Z value |
| 90% | 0.05 | 1.644 |
| 94% | 0.03 | 1.88 |
| 60% | 0.2 | 0.841 |

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

|  |  |
| --- | --- |
| Confidence Level | t value |
| 95% | 2.063 |
| 96% | 2.171 |
| 99% | 2.796 |

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

Probability:

P= 0.3216725 (but we have 1 sample mean and another one is given mean also here H0 is >< so its one tail test )

P= P/2 =0.1608