|  |  |
| --- | --- |
| 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 | Discrete |
| 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 | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ordinal |
| 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?

Solution:

H-Head

T-Tail

Possible outcomes are: { HHH , HHT, HTH, HTT, THH, THT, TTH, TTT}

To get 2 heads and 1 tail the possible probability would be -> HHT,THH,HTH P(E)= Number of favorable outcomes/ Total number of possible outcome

Therefore, 3/8=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

Solution:

P(E)= Number of favorable outcomes/ Total number of possible outcome

Possible outcome are 36.

1. When 2 dice are rolled sum can’t be equal to zero. Minimum sum will be 2. So, 0/36 i.e. Probability is zero.
2. Event of getting sum Less than or equal to 4 are [(1,1) (1,2) (1,3) (2,1) (2,2) (3,1)]

Therefore, P(E)= 6/36=0.1666

1. Event of getting sum is divisible by 2 and 3 are (E) ->[( 1,5) (2,4) (3,3) (4,2) (5,1) (6,6) ]

Therefore, P(E)=6/36=1/6=0.166

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?

Solution:

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

n(S) = Number of ways of drawing 2 balls out of 7

n(S)=7C2

n(S)=(7×6) / (2×1)

n(S)=21

Let E = Event of 2 balls

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

P( E) =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

Solution:

E(X) = ΣX \* P(X)

Expected number of candies for randomly selected child = 1\*0.015+ 4\*0.20+ 3\*0.65+ 5\*0.005+ 6\*0.01 +2\*0.120

E( X) =3.09

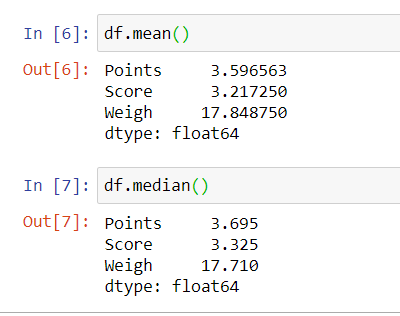
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

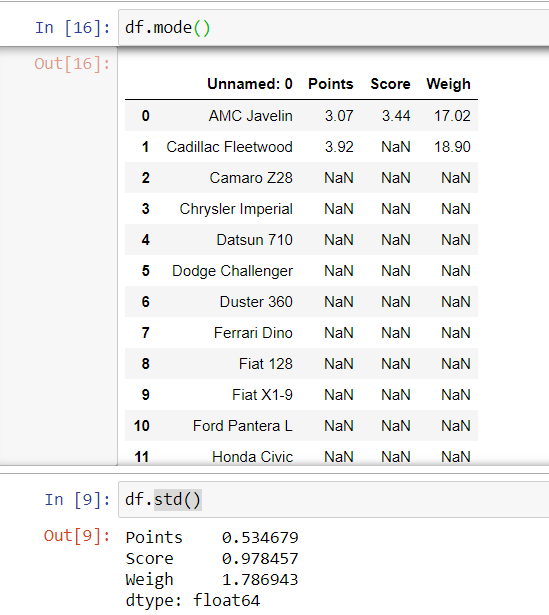
* For Points,Score,Weigh>

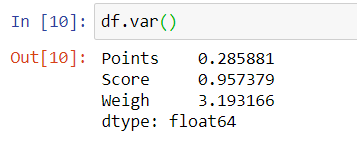
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

**Solution:**







**Standard deviation value is smaller which indicates that values are closer to the mean.**

**----------------------------------------------------------------------------------------------------------**

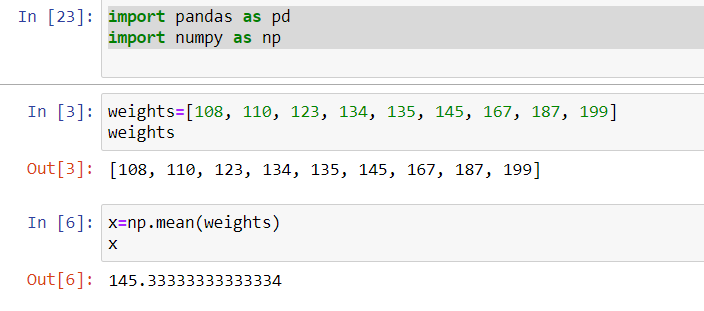
Q8) Calculate Expected Value for the problem below

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?

Solution:

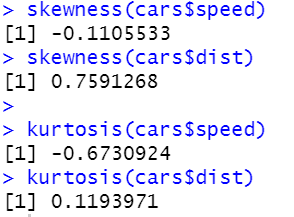


Therefore, mean is: 145.33

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

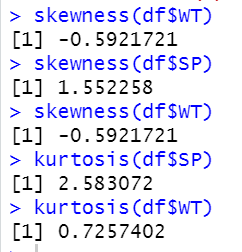
**Cars speed and distance**

**Use Q9\_a.csv**



**SP and Weight(WT)**

**Use Q9\_b.csv**



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



Solution:

* The least range of weight is 400 somewere around 0-10.
* Tthe expected value the above distribution is 75.
* Skewness- we can notice a long tail towards right so it is heavily right skewed.

Solution:

* It is right skewed .
* Outlier are on the upperside of box plot.
* There exist less data points between Q1 and bottom point

( To observe the result and comment on it)

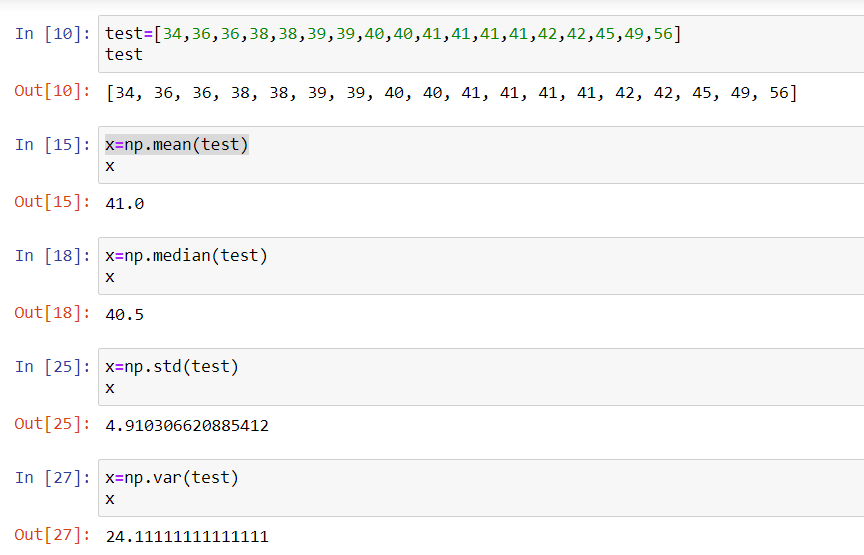
**Q11)** Below are the scores obtained by a student in tests

**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?

Solution:

I.



ii.

* The average score of students is 41
* Therefore, resulting std deviation as 4.91 test scores are between range of [41+4.910, 41-4.910] which is[36.09, 45.91] which indicates the range of the students test score. & also tells the data lies in the first deviation which is 68%.

Q12) What is the nature of skewness when mean, median of data are equal?

Solution: The curve is symmetric

Q13) What is the nature of skewness when mean > median ?

Solution: The curve is positively skewed.

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

Solution: The curve is negatively skewed.

Q15) What does positive kurtosis value indicates for a data ?

Solution: It indicates that peak is sharper and tails are heavier than the normal distribution.

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

Solution: It indicates that peak is flattened tails are light than the normal distribution.

Q17) Answer the below questions using the below boxplot visualization.



What can we say about the distribution of the data?

Solution:

Given box plot tells us about the age’s of the students in a school.

50% of the people are above 10 yrs old and remaining are less.

What is nature of skewness of the data?

Solution: Left skewed, as median is greater.

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

Solution: IQR=Q3-Q1=18-10=8

Q18) Comment on the below Boxplot visualizations?



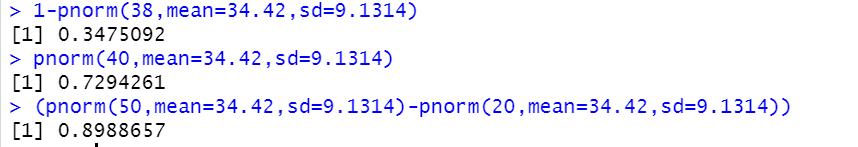
Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Q 19) 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)= 0.7294
  3. P (20<MPG<50)= 0.89866



Q 20) Check whether the data follows normal distribution (check beml code)

1. Check whether the MPG of Cars follows Normal Distribution

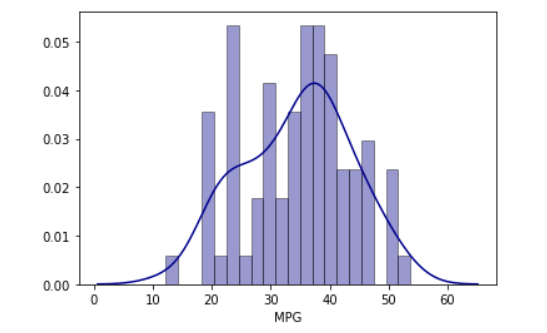
Dataset: Cars.csv

No it doesn’t follows normal distribution. By looking at the graph we can understand it doesn’t follow normal bell curve and skewness can be noticed. And also from the from theory a curve follows normal distribution when mean=median=mode which in this case its not.

Mean: 34.42

Median: 35.14

Std dev: 9.074

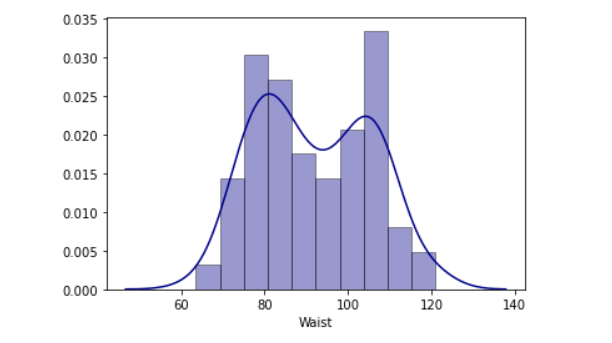


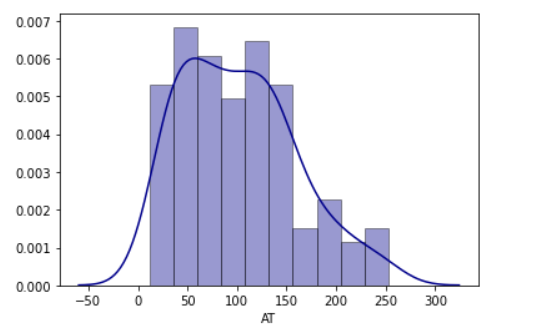
1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

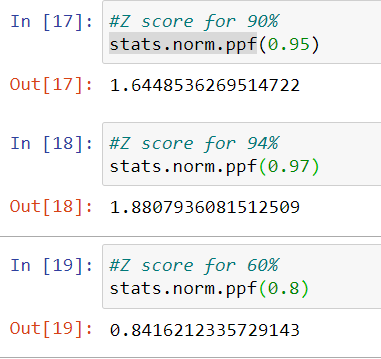
Solution:

Both don’t follow the normal distribution.



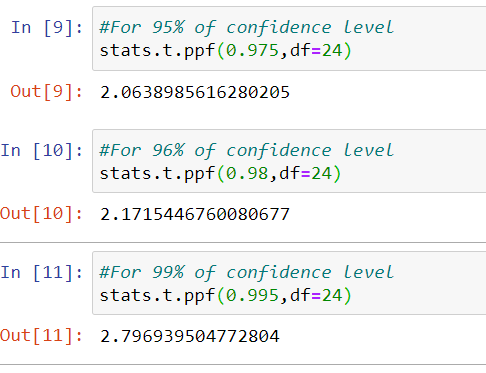


Q 21) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



Q 22) Calculate the t scores of 95% confidence interval, 96% confidence interval, 99% confidence interval for sample size of 25

Solution:



Q 23**)** 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

Solution:

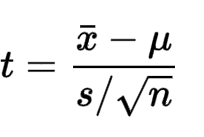
Population mean : 270 days

Sample size ->n= 18

Sample mean: 260

Sample Std deviation:90

By using the below formula:



Subsitute the values:

t=260-270/(90/sqrt(18) )

t= -0.4714

when applied on r using the command we get,

Degrees of freedom🡪n-1=18-1=17



**Therefore, 32.18 % is the that 18 randomly selected bulbs would have an average life of no more than 260 days**