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

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

Sample space= {HHH,HHT,HTH,HTT,TTT,TTH,THT,THH}=8outcomes

P(two heads and one tail)=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=6/36=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 = 7

Number of ways 2 balls drawn out of 7 is 7 = 21

Number of balls which are not blue = 5

Number of ways 2 balls drawn from this 5 balls is 5 = 10

P(none of the balls drawn is blue)=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.

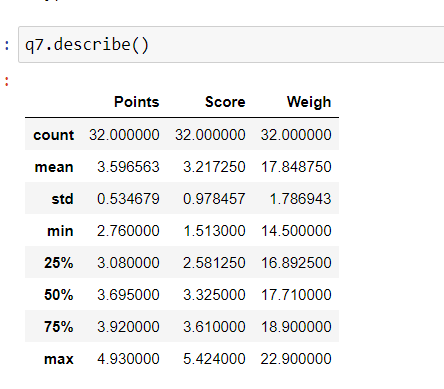
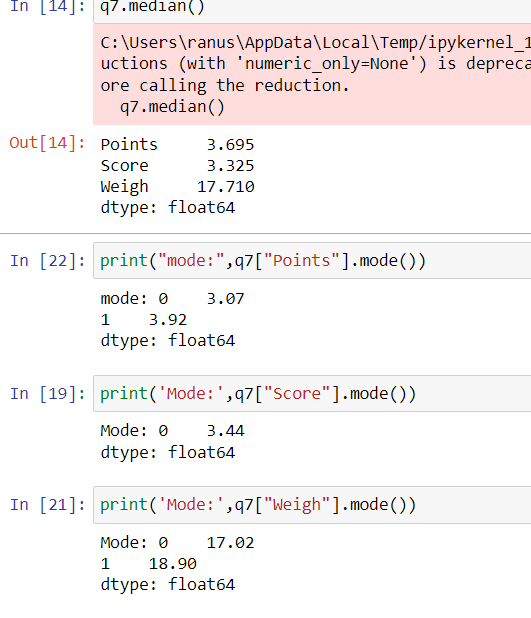
Child B – probability of having 4 candies = 0.20

Answer: 1\*0.015+4\*20+3\*0.65+5\*0.005+6\*0.01+2\*0.120=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.

****

**Most of the observations lie nearest to the mean.**

**There are possibly no outliers when mean and median is compared .**

**Variance of ‘Weigh’ are more that means data points are much far from the mean value while in ‘Score’ data points are less scattered and in Points all the data points are nearer to the mean value.**

**Range = Max value- Min value**

**Points=2.17 ,Score=3.911,Weigh=8.4**

Q8) Calculate Expected Value for the problem below

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

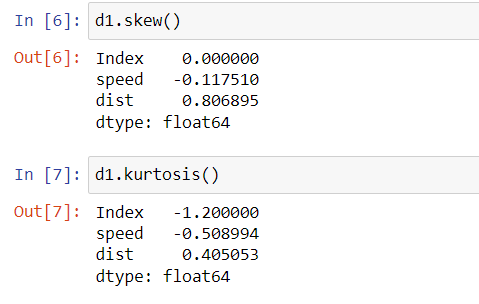
108, 110, 123, 134, 135, 145, 167, 187, 199

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

**Answer: 145.33 pounds**

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

Cars speed and distance

****

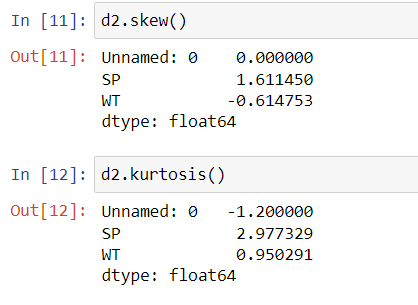
**Speed of the cars are negatively skewed and Distance are positively skewed**

**Speed of the cars data are negatively skewed**

**Speed of cars distribution is not too flat looks somewhat like normal distribution . distance is leptokurtic.**

**+**

**SP and Weight(WT)**

****

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

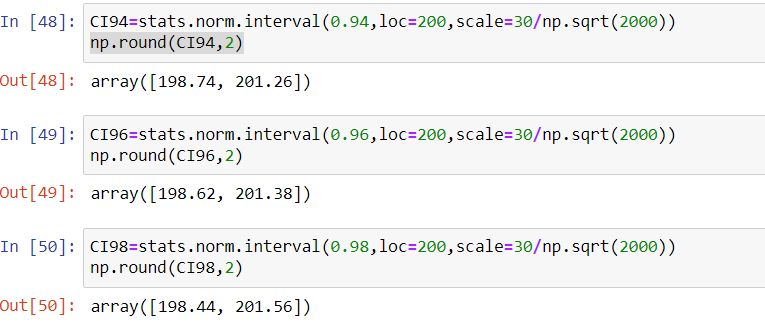


**Here in the given histogram Data is right skewed . Mean median will lie to the right side of the graph**



**In the given box plot there are outliers and the data is right skewed**

**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?



**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?

**Mean =41 ,median=40.5 , mode=41**

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

**The data is normal distributed**

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

Right skewed

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

**Left skewed**

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

**Leptokurtic i.e highly peaked distribution and flat tails**

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

**Platykurtic i.e flatter than the normal distribution and thin tails**

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



What can we say about the distribution of the data?

**The data is not normally distributed and no outliers present**

What is nature of skewness of the data?

**It is left skewed data**

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

**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.

**Both Boxplot 1 and Box plot2 are normaly distributed. Median of both boxplot is same. There are no outliers. Interquartile region of boxplot1 is less than boxplot2 which implies that data is more spread in 2 compared to 1.**

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

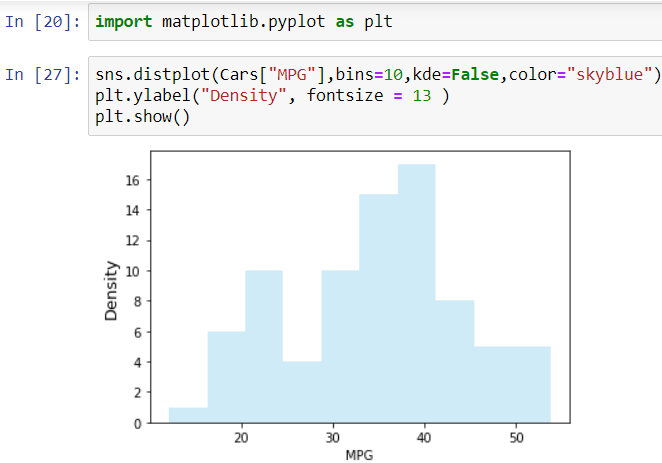
* 1. P(MPG>38)
  2. P(MPG<40)

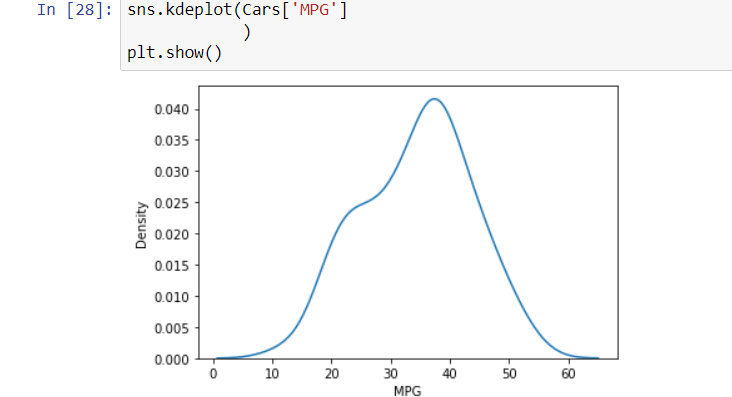
c. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

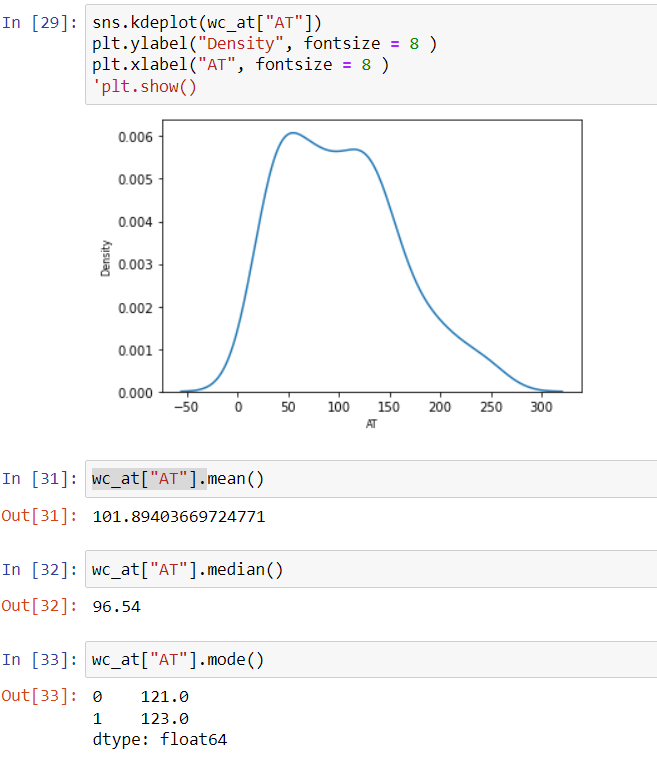


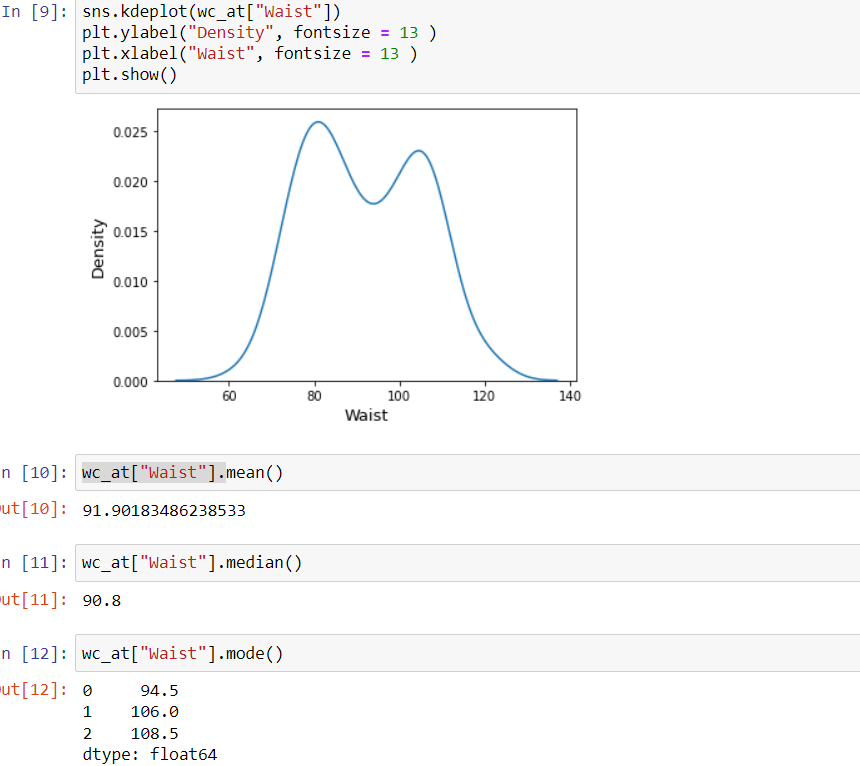


**Not normally distributed**

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

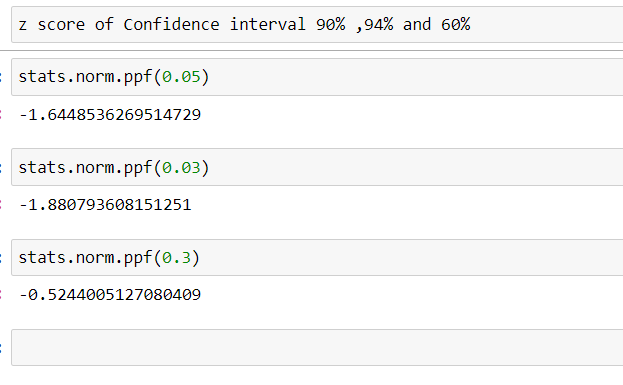
Dataset: wc-at.csv



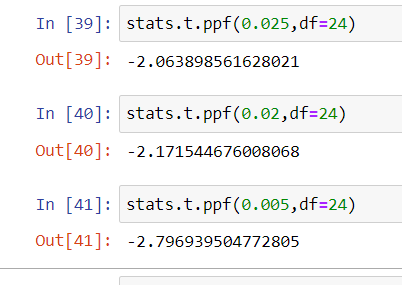


**The above distributions are not normal. Mean median and mode are not equal.**

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



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



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

Mean of Population =µ=270

n=18

df= n-1=17

Sample mean ==260

Sample std = s= 90

t score = [260-270)/(90/√18)]= -0.4714

the probability that 18 randomly selected bulbs would have an average life of no more than 260 days=

