|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Countable Discreet Data types |
| Results of rolling a dice | Countable Discreet Data types |
| Weight of a person | Continuous Data types |
| Weight of Gold | Continuous Data types |
| Distance between two places | Continuous Data types |
| Length of a leaf | Continuous Data types |
| Dog's weight | Continuous Data types |
| Blue Color | Classification Discreet Data types (Nominal Data Type) |
| Number of kids | Countable Discreet Data types |
| Number of tickets in Indian railways | Countable Discreet Data types |
| Number of times married | Countable Discreet Data types |
| Gender (Male or Female) | Classification Discreet Data types (Nominal Data Type) |

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 Data Type |
| High School Class Ranking | Ordinal Data Type |
| Celsius Temperature | Interval Data Type |
| Weight | Ratio Data Type |
| Hair Color | Nominal Data Type |
| Socioeconomic Status | Ordinal Data Type |
| Fahrenheit Temperature | Interval Data Type |
| Height | Ratio Data Type |
| Type of living accommodation | Nominal Data Type |
| Level of Agreement | Ordinal Data Type |
| IQ(Intelligence Scale) | Interval Data Type |
| Sales Figures | Ratio Data Type |
| Blood Group | Nominal Data Type |
| Time Of Day | Ordinal Data Type |
| Time on a Clock with Hands | Interval Data Type |
| Number of Children | Nominal Data Type |
| Religious Preference | Nominal Data Type |
| Barometer Pressure | Interval Data Type |
| SAT Scores | Ratio Data Type |
| Years of Education | Ordinal Data Type |

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

The tossing of three at a time. These are the number of possible outcomes {HHH,THH,TTH,TTT,HTT,HHT,THT,HTH}. the probability of getting two heads and one tail is 3/8 = 0.37

Q4) Two Dice are rolled, find the probability that sum is

1. **Equal to 1** : Probability is 0 because we are rolled two dice. So min sum of dice is 2.
2. **Less than or equal to 4:**

we rolled two dies then total outcomes possibility is 6\*6 = 36

Outcomes = 1 Outcomes is 0 then probability of P(1)=0

Outcomes =2 Outcomes is (1,1) then probability of P(2)= 1/36 = 0.03

Outcomes=3 Outcomes is(1,2),(2,1) then probability of P(3)=2/36 =0.05

Outcomes =4 Outcomes is (1,3)(3,1),(2,2) then probability of P(4)= 3/36= 0.08

1. **Sum is divisible by 2 and 3**

6and 12 both numbers are divisible by 2 and3.

Outcomes = 6 (1,5) (2,4), (3,3), (4,2),( 5,1) then probability of P(6)=5/36=0.14

Outcomes =12 (6,6) then probability of P(12)=1/36 =0.03

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?

Let Total number of balls be S= (2 + 3 + 2) = 7

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

n(S)=7C2 =(7×6)/(2×1) =21

Let Total number of non blue balls be E= (2 + 3) = 5

n(E) = Number of ways of drawing 2 balls out of 5

n(E)=5C2 =(5×4)/(2×1) =10

The probability that none of the balls drawn is blue : n(E)/n(S) = 10/21= 0.37

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

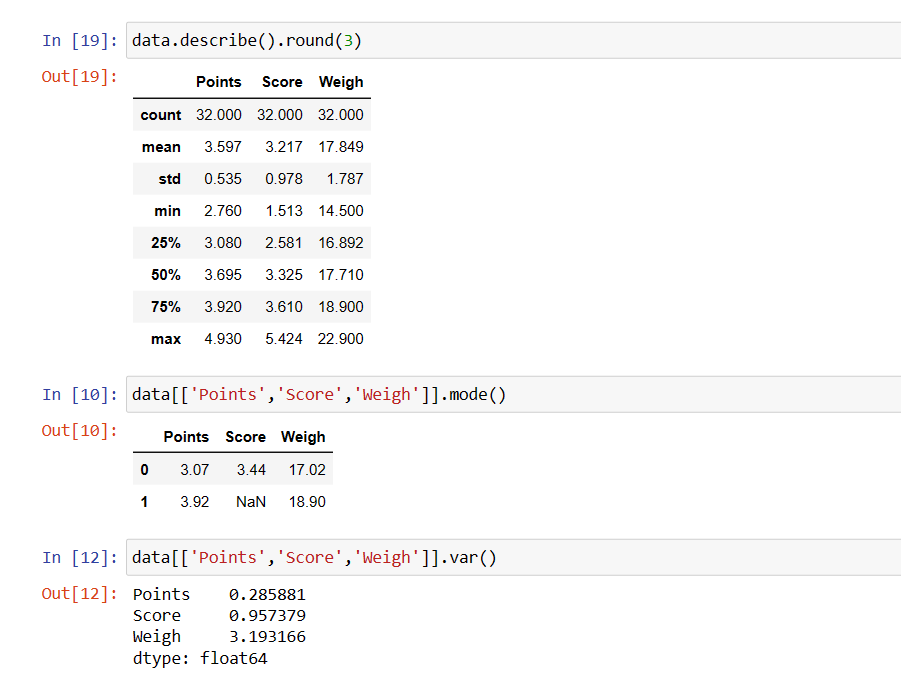
= 1 \* 0.015 + 4\*0.20 + 3 \*0.65 + 5\*0.005 + 6 \*0.01 + 2 \* 0.12

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24 **= 3.09**

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.



Range= Max-Min

for Points =4.930-2.760= 2.17

for Scores =5.424 - 1.513= 3.91

for Weigh=22.900-14.500= 8.34

For Points :1. The median (50th percentile) point score is 3.70, which is close to the mean(3.29), fairly symmetric distribution.

2. The interquartile range (IQR) between Q1 and Q3 (3.08 to 3.92) suggests that the middle 50% of data is spread across this range.

for Scores:1. The median (50th percentile) score of 3.33 is greater than the mean (3.23), left-skewed distribution

2. The interquartile range (IQR) between Q1 and Q3 (2.58 to 3.61) suggests that the middle 50% of data is spread across this range.

for Weigh: 1.The median (50th percentile) weigh of 17.71 is close to the mean (17.85), fairly symmetric distribution.

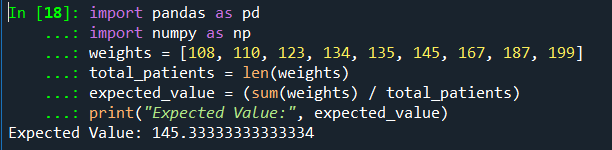
2. The interquartile range (IQR) between Q1 and Q3 (16.89 to 18.90) suggests that the middle 50% of data is spread across this range.

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?

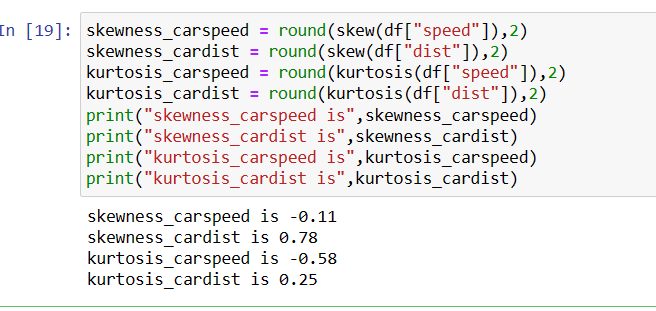


The Expected Value of the Weight of that patient is 145.33 Pounds.

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

**Cars speed and distance**

**Use Q9\_a.csv**

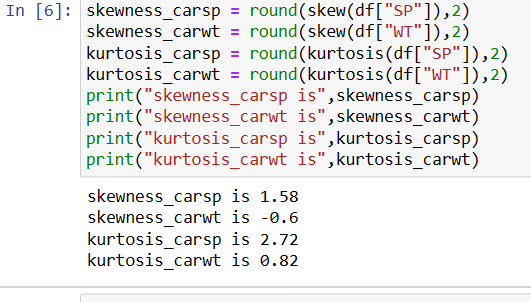
****

Skewness\_carspeed is -0.11 means it’s form a symmetrical distribution. Kurtosis\_carspeed is -0.58 i.e its form a platykurtic (short tailed distribution) means possibility of most of the data lies around the mean.

Skewness\_cardist is -0.78 means it’s form a slightly skewed distribution. Kurtosis\_cardist is 0.25 i.e its form a platykurtic (short tailed distribution) means possibility of most of the data lies around the mean.

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

Skewness\_carsp is 1.58 means it’s forma extremely symmetrical distribution. Kurtosis\_carsp is 2.72 i.e its form a platykurtic (short tailed distribution) means possibility of most of the data lies around the mean.

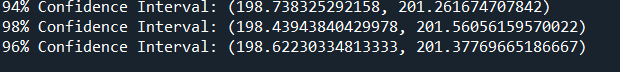
Skewness\_carwt is -0.6 means it’s form a slightly negative skewed distribution. Kurtosis\_carwt is 0.82 i.e its form a platykurtic (short tailed distribution) means possibility of most of the data lies around the mean.

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



Negative skewed Distribution and have few outliers which may be impact on mean.

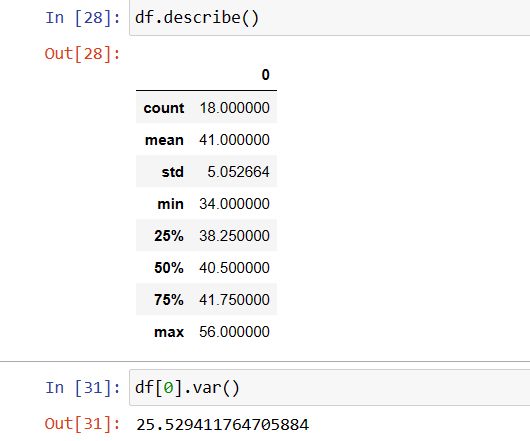
**Q11)** Suppose we want to estimate the average weight of an adult male in Mexico. We draw a random sample of 2,000 men from a population of 3,000,000 men and weigh them. We find that the average person in our sample weighs 200 pounds, and the standard deviation of the sample is 30 pounds. Calculate 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.



Mean = 41

Median= 40.50

Std. = 5.05

Var=25.53

1. What can we say about the student marks?

Negative skew. And median and mean and little bit far .

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

Skew symmetrical

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

Positive skewed

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

Negative skewed

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

Greater than 3

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

Less than 3

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



What can we say about the distribution of the data?

50% of that falls between 10-18. And mean is around 15.

What is nature of skewness of the data?

Negative skewed

What will be the IQR of the data (approximately)?   
  
IQR= 60%

Q19) Comment on the below Boxplot visualizations?



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

Boxplot1: no outliers and range of data set is [250,275] where as Boxplot2 no outliers and data range is [225,315] both have same mean. And it’s a symmetrical distribution.

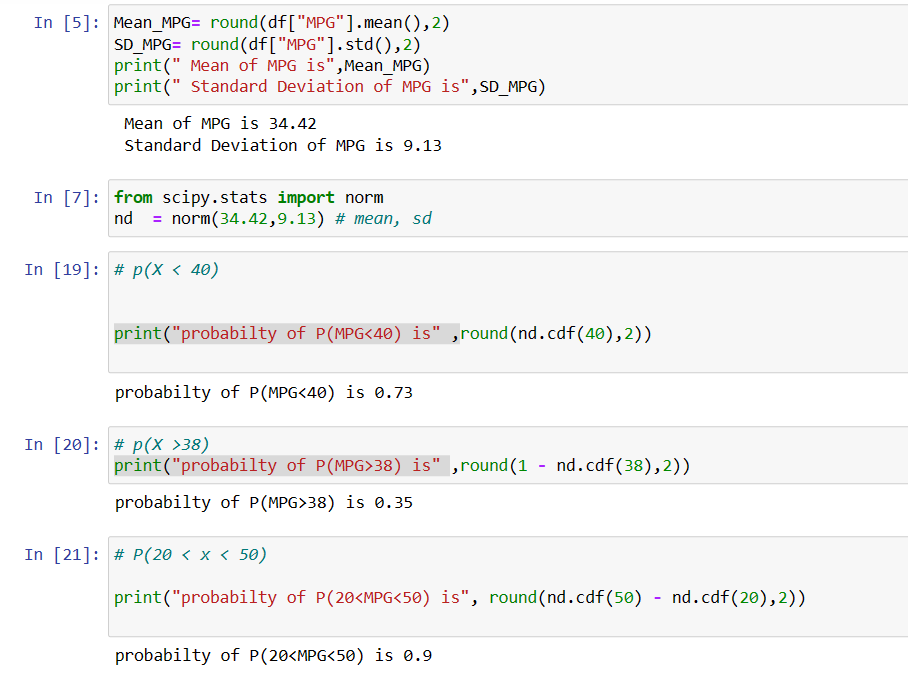
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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

MPG <- Cars$MPG

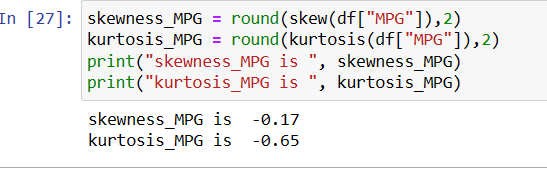
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

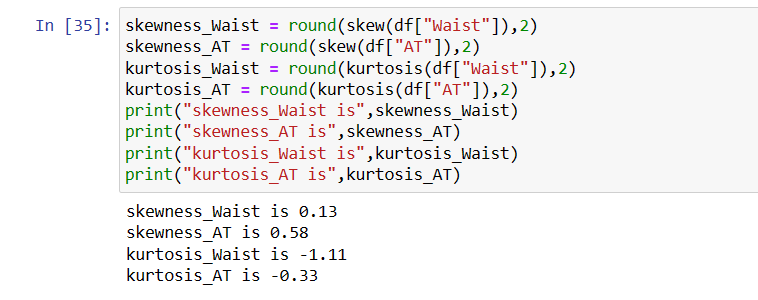
1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



Skewness of MPS is -0.17 which lies between -0.5 - +0.5. means it’s symmetrical data. And Kurtosis less than 3 mean less than normal distribution means median and mean are slightly near.

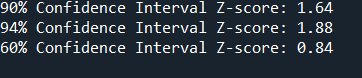
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv 

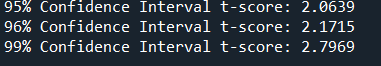
Skewness of Waist is 0.13 which lies between -0.5 - +0.5. means it’s s symmetrical data. And Kurtosis is -1. less than 3 mean less than normal distribution means median and mean are slightly equal.

Skewness of AT is 0.58 which lies between +0.5 -1 means it’s positive skewed data. And Kurtosis less than 3 mean less than normal distribution means median and mean are slightly 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

The probability that 18 randomly selected bulbs would have an average life of no more than 260 days: 0.3217