|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continous |
| Weight of Gold | Continous |
| Distance between two places | Continous |
| Length of a leaf | Continous |
| Dog's weight | Continous |
| 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 | Interval scale |
| Celsius Temperature | Ratio |
| Weight | Ratio |
| Hair Color | Nominal |
| Socioeconomic Status | Interval |
| Fahrenheit Temperature | Ratio scale |
| Height | Ratio scale |
| Type of living accommodation | Ordinal |
| Level of Agreement | Interval Scale |
| IQ(Intelligence Scale) | Interval scale |
| Sales Figures | Ratio scale |
| Blood Group | Nominal |
| Time Of Day | Interval scale |
| Time on a Clock with Hands | Ratio scale |
| Number of Children | Ratio |
| Religious Preference | Ordinal |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Ratio |

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

Solution:

Number of outcomes: 8

Getting two heads and one tail: {H,H,T],{H,T,H},{T,H,H}

P(Two heads and one tail) =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:

Total number of possible outcomes: 36

1. Equal to 1

Fsavorable outcomes={}

P(sum equal to 1)=0

b) Less than or equal to

Favorable outcomes:{1,1},{1,2},{1,3},{2,1},{2,2},{3,1}

P(<=4)=6/36 = 0.167

c) Sum is divisible by 2 and 3

Possibe outcomes:{1,5},{2,4},{3,3},{4,2},{5,1}

P(Sum is divisible by 2 and 3) = 5/36 = 0.138

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  
Let S be the sample space.  
Then, n(S) = Number of ways of drawing 2 balls out of 7  
=7C2​  
=(7×6)/(2x1)​  
=21  
Let E = Event of drawing 2 balls, none of which is blue.  
n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls.  
=5C2​  
=(5×4)​(2x1)  
=10  
∴P(E)=n(S).n(E)​=2110​

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:

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

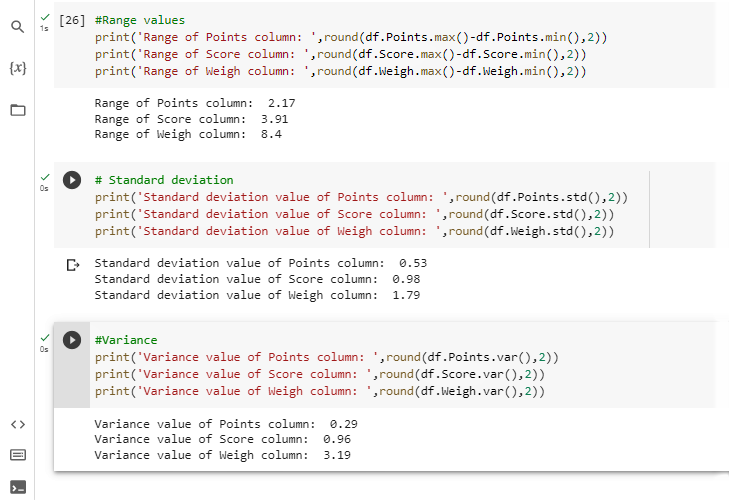
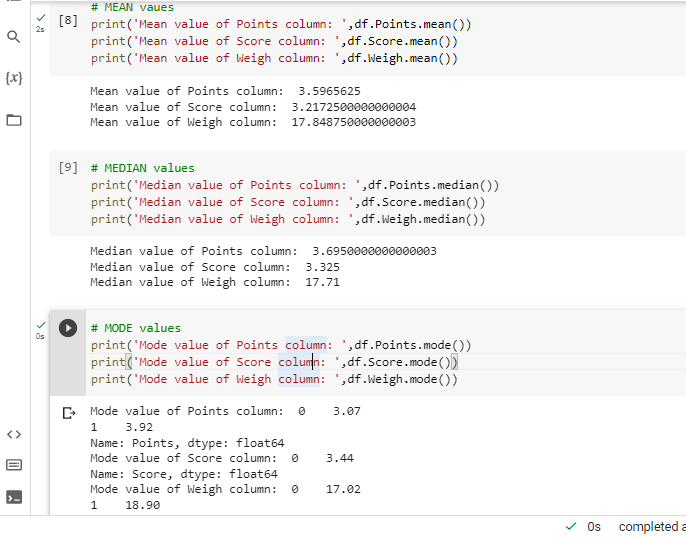
= 3.09

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.**Use Q7.csv file**

Solution:



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:

Expected Value  =  ∑ ( probability  \* Value )

  ∑ P(x).E(x)

n(S)=9

Probability of selecting each patient = 1/9

E(x)  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

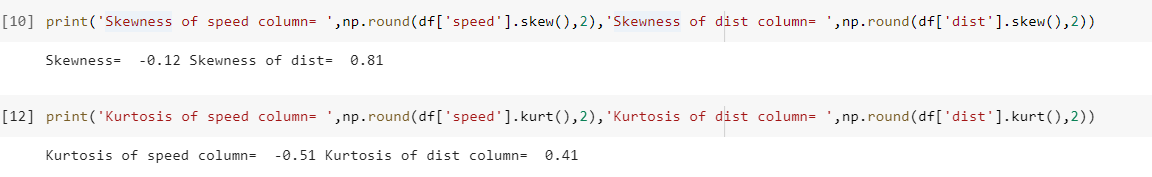
Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

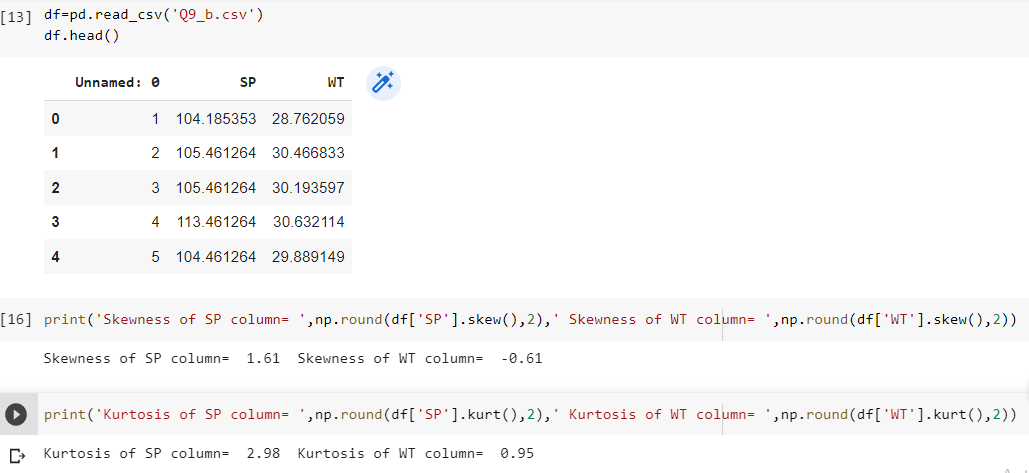
Solution:

****

**SP and Weight(WT)**

**Use Q9\_b.csv**

Solution:

****

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



Solution:

* From the above graph we can observe that the chicks usually weigh less than 200g.
* Maximum number of chicks weigh from 50-100g.
* Here the data is more right skewed.
* There are 8 outliers on the higher side of data,specifies that the weigh of few chicks varies significantly than others.

**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?

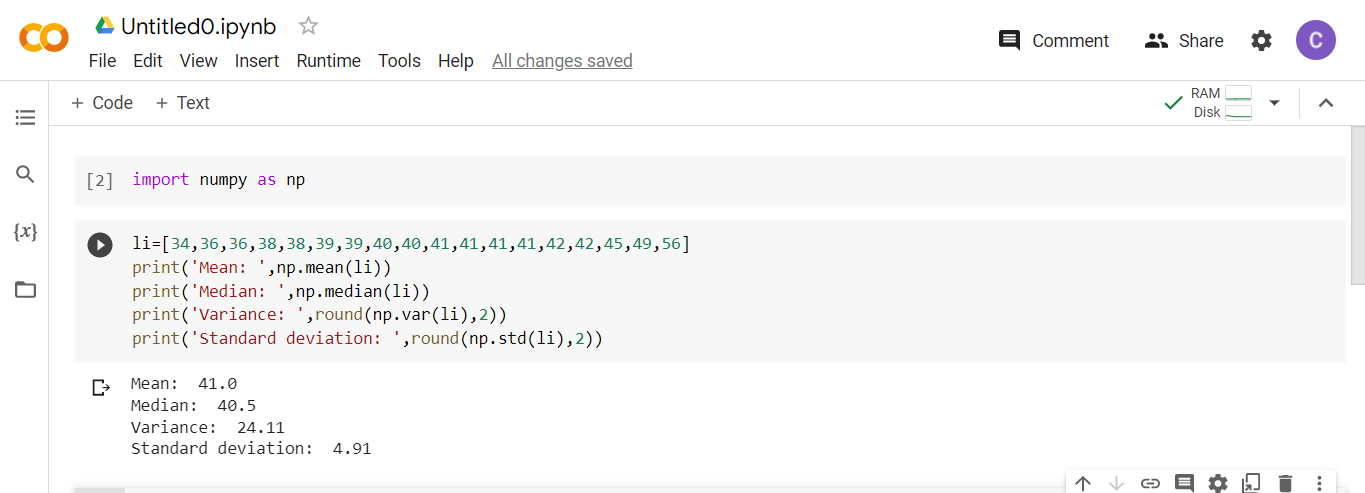
Solution:



**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?

Solution: 

2)

* The average marks scored by the student is 41 which is deviated by approximately 5.
* Here the data is normally distributed as the mean is almost equal to the median.

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

Solution: If mean is equal to median then the data is distributed normally,For normally distributed data skewness will be equal to zero.

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

Solution: The distribution will be skewed to “Right”, if mean>median.

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

Solution: The distribution will be skewed to “Left”, if median>mean.

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

Solution: Positive values of kurtosis indicate that distribution is peaked and possesses thick tails

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

Solution: A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution

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



What can we say about the distribution of the data?

Solution: The data here is not normally distributed here as the median line is not at the centre of the box but slightly towards right.

What is nature of skewness of the data?

Solution: The data is left skewed in the above plot.

What will be the IQR of the data (approximately)?   
Solution: IQR=Upper limit – Lower limit

Hence,IQR=18-10  
 IQR=08

Q19) Comment on the below Boxplot visualizations?



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

Solution:

* Here in both the plot the median is almost same and both are

distributed normally.

* The variance of of data in graph-1 is lesser than the graph-2 as the spread of data in graph-2 in more.

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

Data \_set: Cars.csv

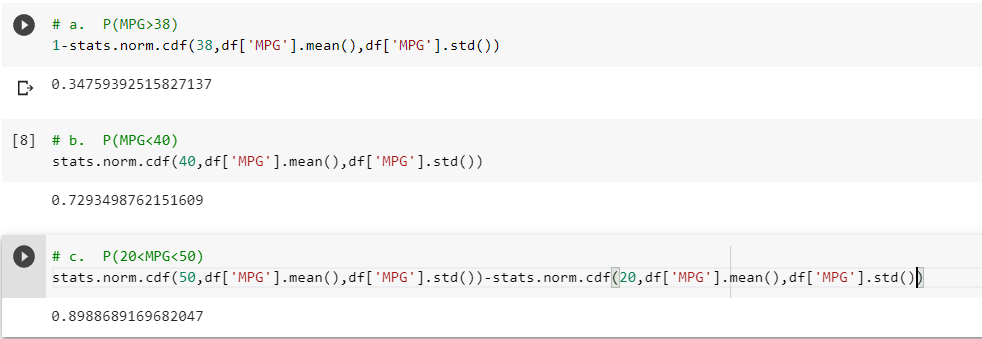
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

Solution:



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Solution: The data here is 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

Solution: The data here is not normally distributed.

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

Solution:

For 90% confidence interval

A=(1+Confidence interval)/2

A=(1+0.9)/2

A=0.95

Z score,By using the z-score table

Z=1.645(By taking the average of 1.64 and 1.65,as 0.95 is close to both the values 1.64 and 1.65)

For 60% confidence interval

A=(1+Confidence interval)/2

A=(1+0.6)/2

A=0.8

Z score,By using the z-score table

Z=0.85

For 94% confidence interval

A=(1+Confidence interval)/2

A=(1+0.94)/2

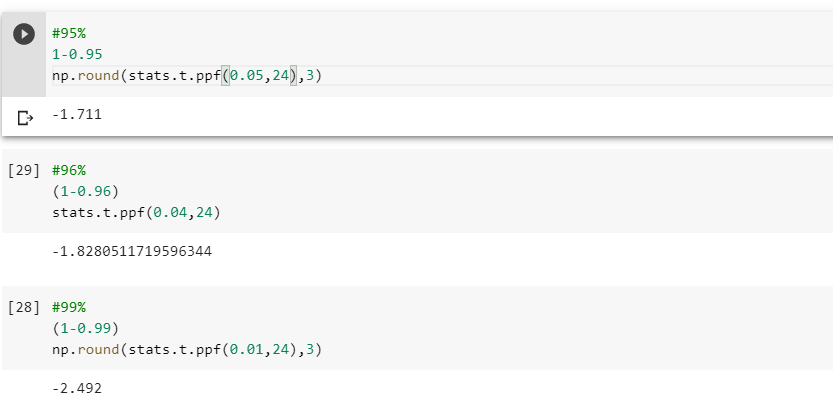
A=0.97

Z score,By using the z-score table

Z=1.88

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

Solution



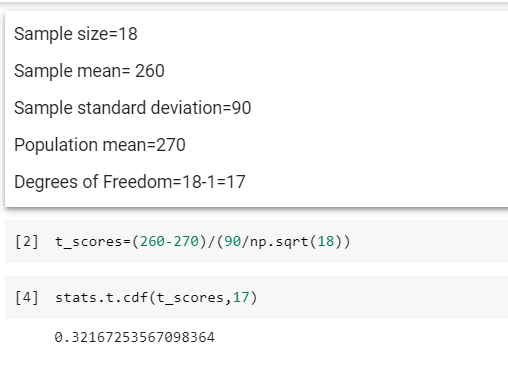
Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

Solution



Therefore the probability is 32%