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

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

**Answer-**

We get Set of Combinations {HHH, HHT, HTH, THH, TTH, TTT, HHT}

Two Heads and One Tail Combinations {HHT, HTH, HHT} i.e. 3

Probability = 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

**Answer-**

We get Total 36 Combination’s i.e. {(1,1), (1,2), …… (6,6)}

1. The sum is equal to one is zero because they start with (1,1) i.e. Sum equal to 2
2. The sum is equal to 4 possible outcomes are {(1,1), (1,2), (1,3), (2,1), (2,2), (3,1)}

So, n(b)= 6/36 =1/6

1. Sum is Divisible by 2 and 3 possible outcomes are {(1,1), (1,2), (1,3), (1,5), (2,1), (2,2), (2,4), (2,6), (3,1), (3,3), (3,5), (3,6), (4,2), (4,4), (4,5), (4,6), (5,1), (5,3), (5,4), (5,5), (6,2), (6,3), (6,4), (6,6)}

So n(c)= 24/36= 2/3

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?

**Answer-**

Total balls= 7

n(P)=Number of ways of drawing 2 balls out of 7 = 21

A= Event of drawing 2 balls, none of which is blue.

n(A)=Number of ways drawing 2 balls out of (2 +3) balls = 10

P(A)=n(A)/n(P)=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-**

Suppose,

Candies count=xi

Probability=P(xi)

Then Expected number of candies for a randomly selected child is

= 1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

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

= 3.09

Expected number of candies for a randomly selected child = 3.09

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

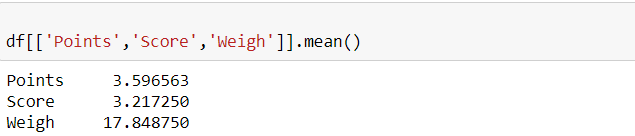
* For Points, Score, Weigh>

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

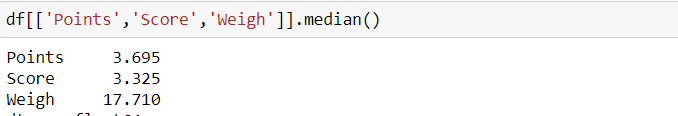
**Use Q7.csv file**

**Answer-**

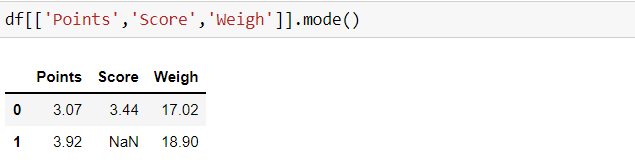
**Mean-**

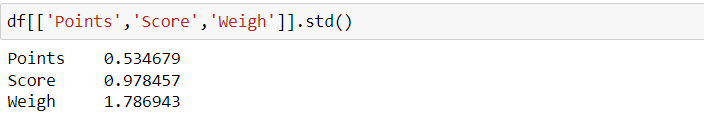
****

**Median-**

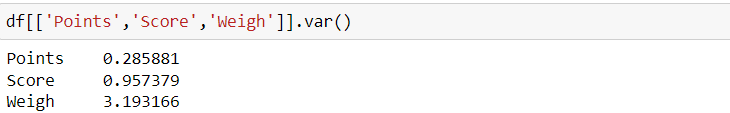
****

**Mode-**

****

**Standard Deviation-**

**Variance-**

****

**Range-**

****

**Answers by Excel-**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | 3.5965625 | 3.21725 | 17.84875 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 3.92 | 3.44 | 17.02 |
| **Variance** | 0.276947559 | 0.927460875 | 3.093379688 |
| **Standard Deviation** | 0.526258072 | 0.963047701 | 1.758800639 |
| **Range** | 2.17 | 3.911 | 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-**

Expected Value = ∑ (probability \* Value) =∑ P(x). E(x)

there are 9 patients so Probability of selecting each patient = 1/9

E(x) = 108, 110, 123, 134, 135, 145, 167, 187, 199

P(x) = 1/9 of each

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**

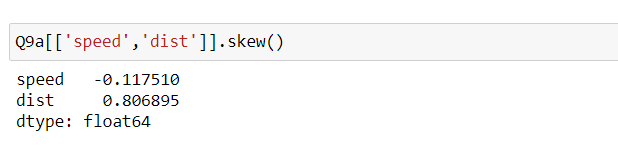
**Use Q9\_a.csv**

**SP and Weight (WT)**

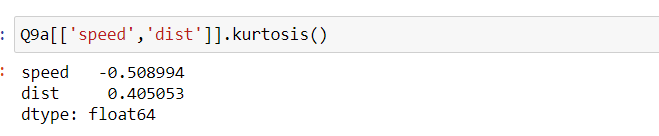
**Use Q9\_b.csv**

**Answer-** Q9\_a

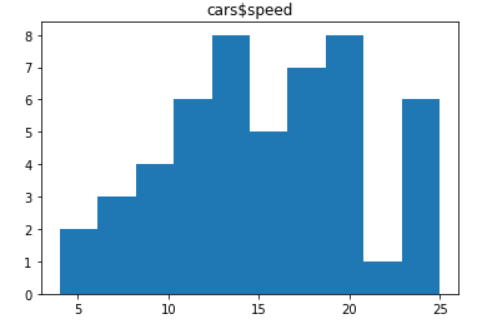
**Skewness: speed and distance-**

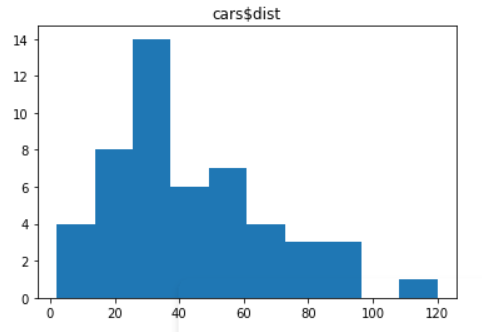


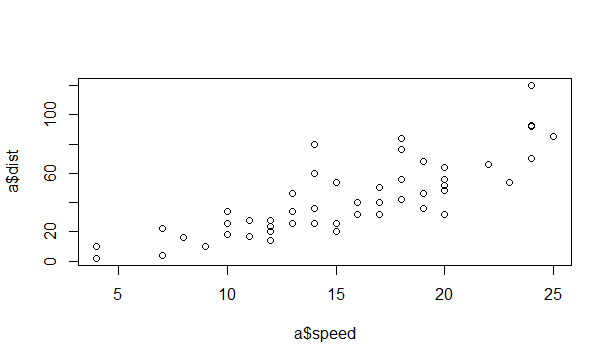
**Kurtosis: speed and distance-**



**Draw inferences: speed and distance –**

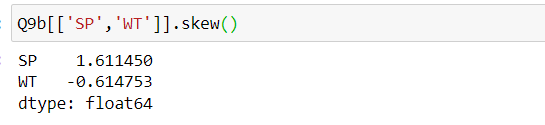
****

****

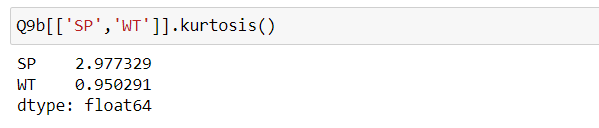


**Answer-** Q9\_b

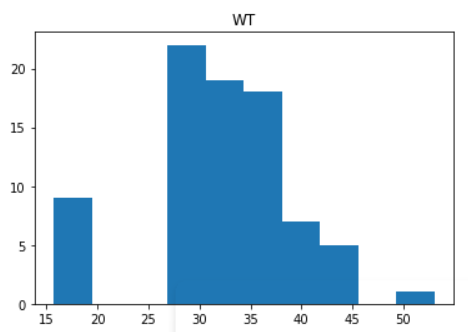
**Skewness of SP and Weight (WT):**

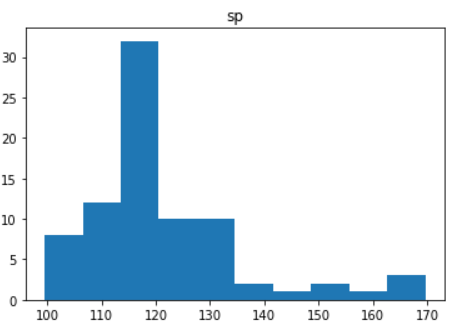
****

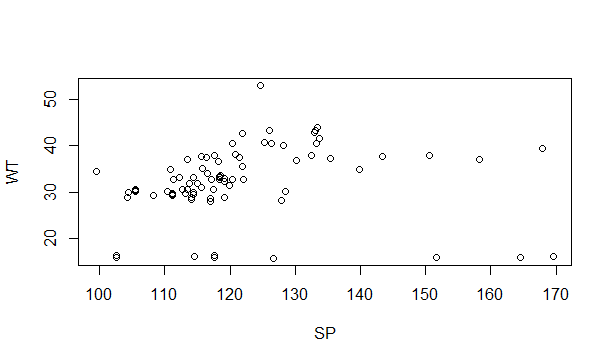
**Kurtosis of SP and Weight (WT):**

****

**Draw inferences: SP and Weight (WT)-**

****

****



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



**Answer-**

|  |  |
| --- | --- |
| Chick Weight | Frequency (Approximately) |
| 0-50 | 80 |
| 50-100 | 200 |
| 100-150 | 125 |
| 150-200 | 100 |
| 200-250 | 50 |
| 250-300 | 25 |
| 300-350 | 12.5 |
| 350-400 | 10 |

* 50-100 weight having highest frequency of around 200
* 350-400 weight having lowest frequency of less than 10
* Data is right skewed
* 0-50 weight having frequency of 80
* 100-150 weight having frequency of 125
* 7 outliers are present in above box plot
* Q1 is smaller than the Q3

**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?

**Answer-**

Standard deviation of population is not given So,

we have use T-Distribution to calculate confidence interval

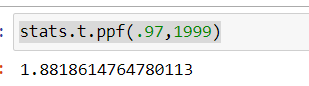
Sample mean (X̄)=200 pounds

Sample standard deviation (S)=30

Sample (n)=2000

CI=

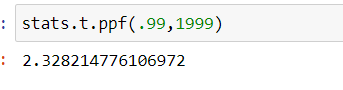
Value of t for confidence level 94% -



Put this value in above equation

**Confidence interval for 94% is [198,201]**

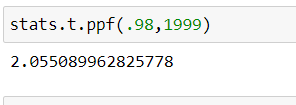
Value of t for confidence level 98%-



Put this value in above equation

**Confidence interval for 98% is [198.4,201.4]**

Value of t for confidence level 96% -

  
Put this value in above equation

**Confidence interval for 96% is [198.6,201.3]**

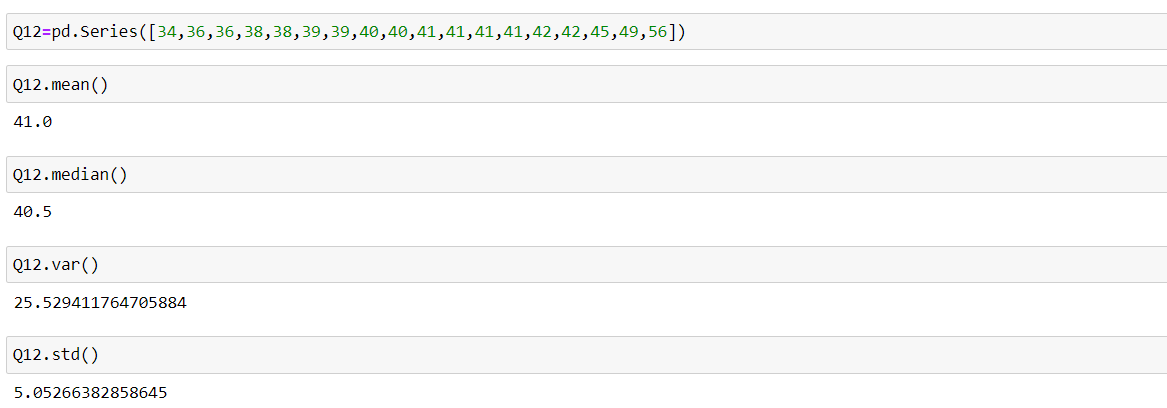
**Q12)** Below are the scores obtained by a student in tests

**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?

**Answer-**

|  |  |  |  |
| --- | --- | --- | --- |
| mean | median | variance | standard deviation |
| 41 | 40.5 | 25.52941 | 5.052664 |

****

Mean > Median, this implies that the distribution is slightly skewed towards right. No outliers are present.

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

**Answer-**

the skewness is zero, Symmetric

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

**Answer-**

Positively skewed

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

**Answer-**

Negatively skewed

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

**Answer-**

Distribution is peaked (Sharp Pick) and possesses thick tails

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

**Answer-**

Distribution is flat (Broad Pick) and has thin tails

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



What can we say about the distribution of the data?

**Answer-**

* Data is not normally distributed, No outliers.
* Most of data is ranges from 10 to 18, Median is between 15 to 16.

What is nature of skewness of the data?

**Answer-**

Negative Skewness

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

**Answer-**

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.

**Answer-**

* The median of the two boxplots is same approximately 260.
* The boxplots are not skewed in Positive or Negative direction.
* Outliers doesn’t exist in both of the boxplots.

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

Data \_set: Cars.csv

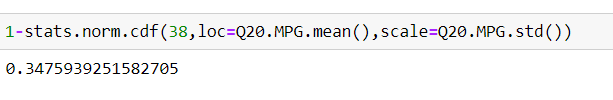
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

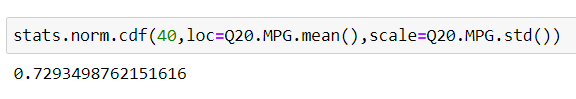
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)

**Answer-**

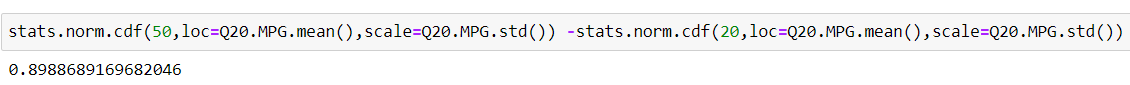
1. P(MPG>38)



1. P(MPG<40)



1. P (20<MPG<50)



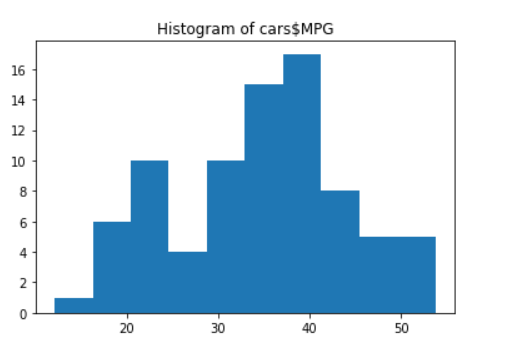
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Answer-**

MPG of cars follows normal distribution

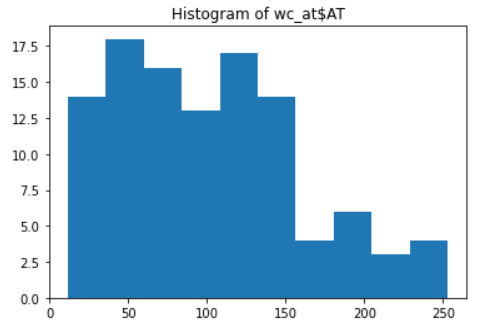


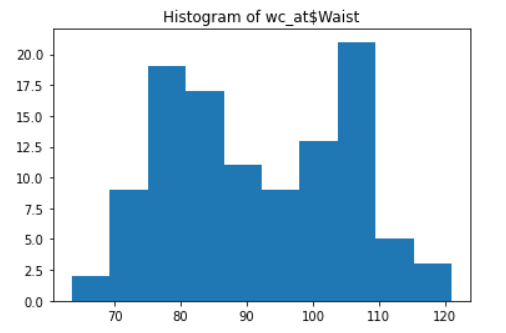
1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc.at data set follows Normal Distribution

Dataset: wc-at.csv

**Answer-**

Adipose Tissue (AT) and Waist Circumference (Waist) follows normal distribution

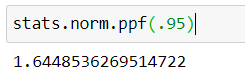




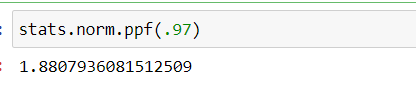
Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval.

**Answer-**

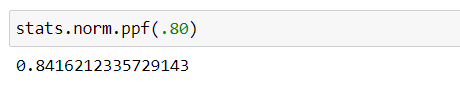
Z scores of 90% confidence interval-



Z scores of 94% confidence interval-



Z scores of 60% confidence interval-



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

**Answer-**

t scores of 95% confidence interval-



t scores of 96% confidence interval-



t scores of 99% confidence interval-



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

**Answer-**

µ=270

=260

SD=90

n=18

df= n-1

df=18-1

df= 17

