|  |  |
| --- | --- |
| 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 | Qualitative |
| Number of kids | discrete |
| Number of tickets in Indian railways | discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Discrete |

***B.R.NIVEDHA***

***ASSIGNMENT-1***

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 | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ordinal |
| Years of Education | Ratio |

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

Sol: Event of getting 2 head and one tail

=3/8

3= Interested no of events

8=Total no of events

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

Sol. Rolling a dice twice

Sample space ={(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)

(2,1)(2,2)(2,3)(2,4)(2,5)(2,6)

(3,1)(3,2)(3,3)(3,4)(3,5)(3,6)

(4,1)(4,2)(4,3)(4,4)(4,5)(4,6)

(5,1)(5,2)(5,3)(5,4)(5,5)(5,6)

(6,1)(6,2)(6,3)(6,4)(6,5)(6,6)}

1. P(E1) = No. of events/no of samples

= 0/36

1. P(E2) = 6/36

S ={(1,1)(1,2)(1,3)(2,1)(2,2)(3,1)}

1. P(E3)= sum divisible by 2 and 3 are[6,12]

{(1,5)(2,4)(3,3)(4,2)(5,1)(6,6)}

=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?

Sol: Bag contains (2R+3G+2B) = 7 balls

No. of ways of getting 2 balls out of 7 is **7 c2** ways.

7!/(7-2)! \*2! =21 ways

E1 be the event of drawing two balls, None of which is Blue, No of ways of getting 2 Balls out of (2R+3G)Balls= 5C2 ways

5!/(5-2)!\*2! = 10 ways

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

Sol: n=6

The probability count of candies for children

=1(0.015)+4(0.20)+3(0.65)+5(0.005)+6(0.01)+2(0.120)

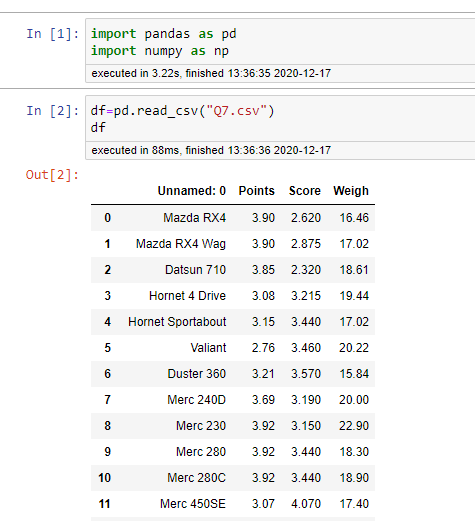
The expected number of candies for a randomly selected child is=3.09

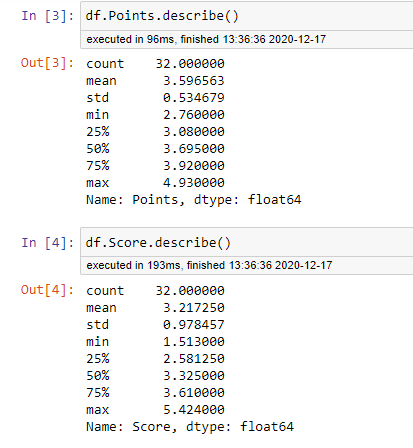
Q7) Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

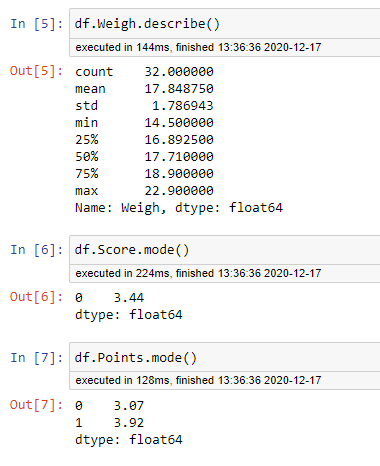
* For Points,Score,Weigh>

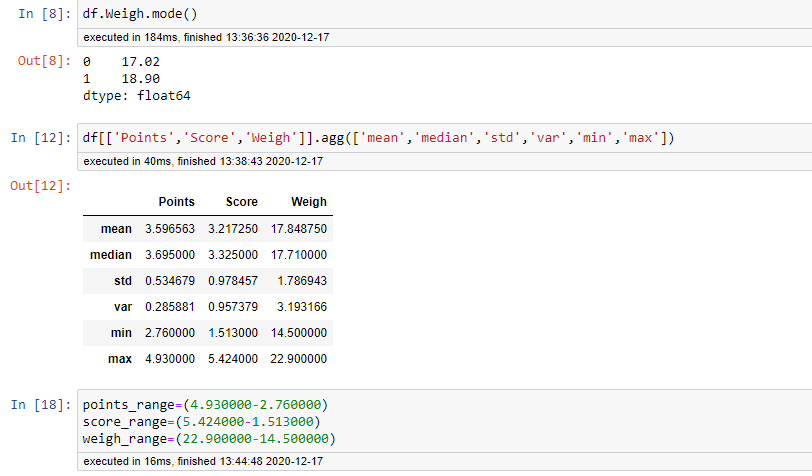
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

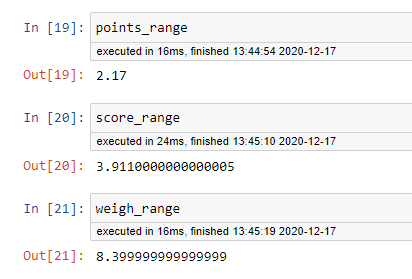
**Use Q7.csv file**











**Inference:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Values** | **Points** | **Score** | **Weigh** |
| Mean | 3.596563 | 3.217250 | 170848750 |
| Median | 3.695000 | 3.325000 | 17.710000 |
| Mode | 3.44 | 3.07 | 17.02 |
| Standard deviation | 0.534679 | 0.978457 | 1.786943 |
| Variance | 0.285881 | 0.957379 | 3.193166 |
| Range | 2.17 | 3.911000 | 8.39 |

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?

Sol: The Mean is : Expected value

= ( 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199 )

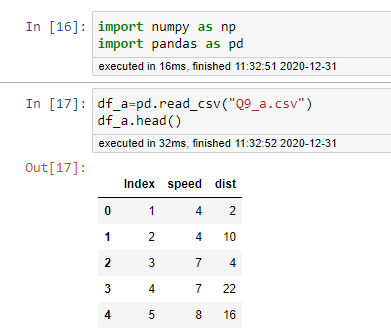
9

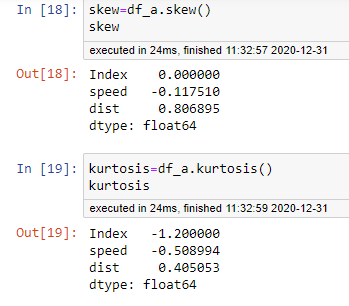
= 145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

****

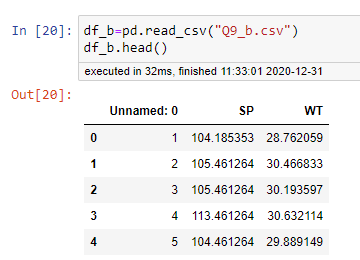
****

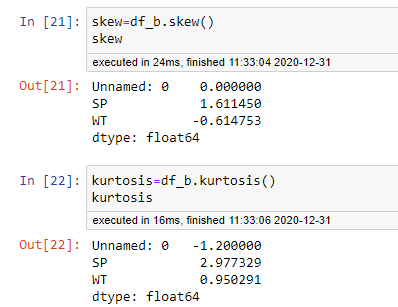
*Inference****:* Speed is Negatively Skewed and Plati Kurtoic**

**Distance is Positively Skewed and Lepto Kurtoic**

**SP and Weight(WT)**

**Use Q9\_b.csv**

****

****

*Inference****:* Sp is Positively Skewed and Lepto kurtoic**

**Weight is Negatively Skewed and Lepto kurtoic**

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



This histogram represents “Unimodal distribution” right significant right skewed

Positively skewed.

The distribution of the data follows ‘Normal distribution’

Mode of the above frequency distribution is 200



The boxplot with many outliers falling above Upper whisker

Lower whisker is smaller than Upper whisker

Box plot is Right side positively 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?

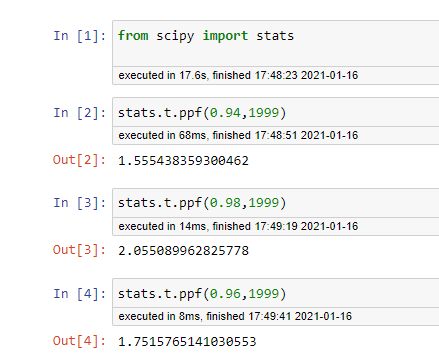
Sample mean =X=200

Sample standard deviation=s=30

Sample size =n=2000

Confidence interval=94%,98%,96%

|  |  |  |
| --- | --- | --- |
| ɑ=94 | ɑ=98 | ɑ=96 |
| 1-ɑ=0.94 | 1-ɑ=0.98 | 1-ɑ=0.96 |



X±tc\*S/√n

At 94% CI Tc = 1.55

[200-1055\*30/√2000,200+1.55\*30/√2000]

[198.9569,201.04]

At 98% CI Tc = 2.055

[200-2.055\*30/√2000,200+2.055\*30/√2000]

[198.621,201.37]

At 96% CI Tc = 1.7515

[200-1.7515\*30/√2000,200+1.7515\*30/√2000]

[198.824,201.175]

**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?

>score=(34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56)

A)ƩX/N

= (34+36+36+38+38+39+39+40+40+41,41,41+41+42+42+45+49+56)

18

Mean=41

Median=no.of observations=18(even)

=(N/2+1)th  Term

=(18/2+1)th  Term

=(9+1)th Term

=10th Term

Median:-41

σ2 = (34-41)+(36-41)+(36-41)+(38-41)+(38-41)+(39-41)+(39-41)+(40-41)+(40-41)+(41-41)+(41-41)+(41-41)+(41- 41)(42-41)+(42-41)+(45-41)(49-41)+(56-41) .

18

Variance (σ2):-24.21

Standard deviation(σ) = 4.910193

Student marks are uniformly distributed and the range of marks is 22

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

1. Symmetric or 0 Skewness

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

1. Positively skewed

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

1. Negatively skewed

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

1. The curve which is more peaked then normal curve is called **lepto kurtoic** curve, β(2)>3,γ(2)>2

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

1. The curve which is flatter than normal curve is called **Plati kurtoic** curve, β(2)<3,γ(2)<2

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



What can we say about the distribution of the data?

This represents **Left significant, Left skewness, Negatively skewed data.**

What is nature of skewness of the data?

The nature of skewness is **Negatively skewed.**

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

**IQR =Q3-Q1**

**=18-10**

Q19) Comment on the below Boxplot visualizations?



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

Box plot 1 and 2 both follows normal distribution

Variance of Boxplot 1 is less than Boxplot2

IQR of Boxplot 1 is 25

IQR of boxplot 2 is100

Q2 Median of both Boxplots are same

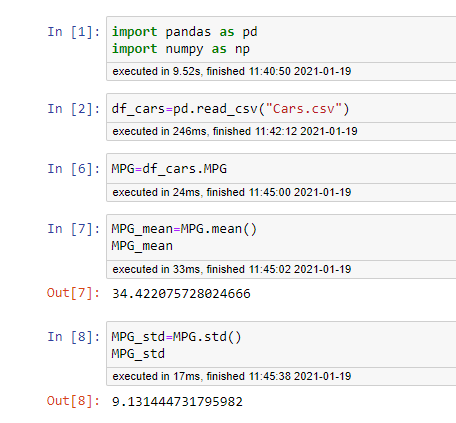
Q 20) Calculate probability from the given dataset for the below cases

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)

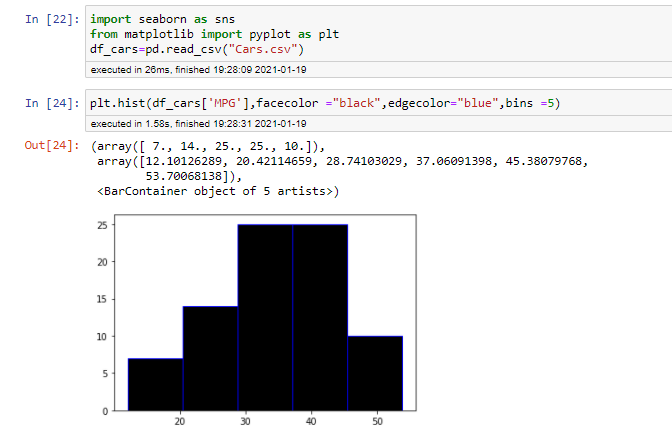




Q 21) Check whether the data follows normal distribution

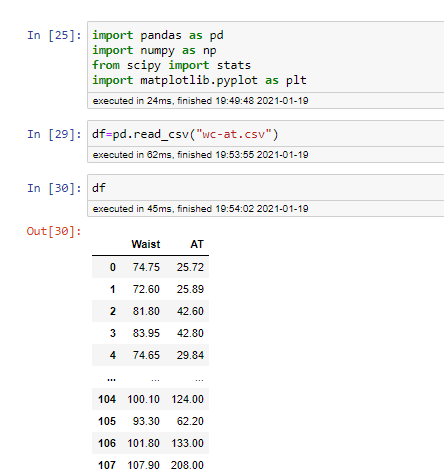
1. Check whether the MPG of Cars follows Normal Distribution

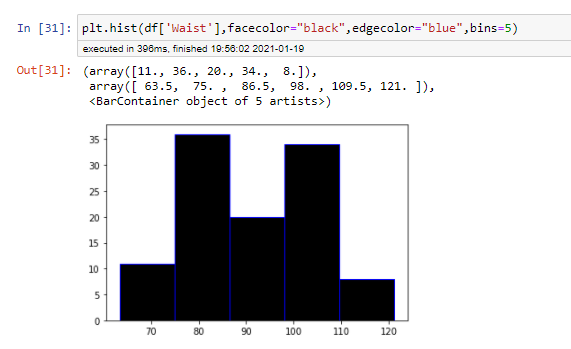
Dataset: Cars.csv

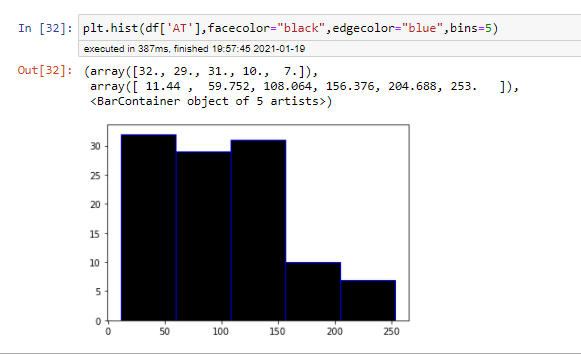


1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

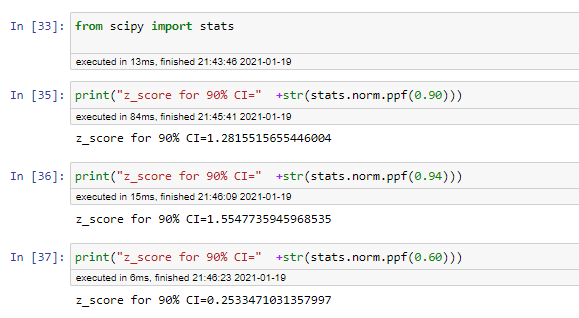
Dataset: wc-at.csv



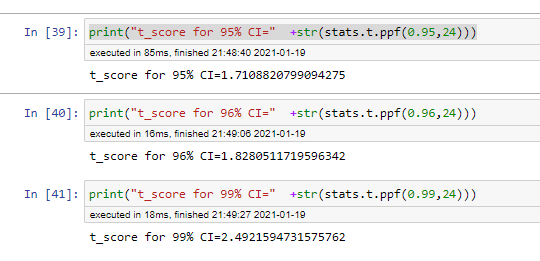




Q 22) Calculate the Z scores of 90% confidence interval,94% confidence interval, 60% confidence interval



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

Sol: x=260 , Mu=270 , s=90 , n=18

T=(x-Mu)/(s/sqrt(n))

**[1] -0.4714045**

>prob=pt(t,df=17)

>prob

**[1]0.3216724**

The probability that t < -0.4714 with 17 degrees of freedom assuming the population mean is true , the t-value is less that the t-value obtained with 17 degree of freedom and a t score of -0471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.