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

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 | Interval |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Sol:

Three Coins are tossed {HHH, HHT, HTH, THH, HTT, THT, TTH, TTT}

No of sample n(s) = 8

The probability that two heads and one tail {HHT, HTH, THH}

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 2and 3

Sol :

{(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. 0/36=0

n(E1)= {(1,1),(1,2),(1,3), (2,1),(2,2), (3,1)}

1. 6/36=1/6

n(E1)= {(1,6),(2,4),(3,3),(4,2),(5,1),(6,6)}

1. 5/36

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:

5c2/7c2=10/21

=0.4761

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 |

Sol:

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

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

The expected no 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**

**Mean Median Mode Variance Standard Deviation Range**

Points 3.5965 3.695 3.07 0.2858 0.5346 2.17

Score 3.2172 3.325 0.9573 0.984 3.91

Weigh 17.848 17.710 3.1931 1.7869 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?

1/9(108+110+123+134+135+145+167+187+199)

The expected value of the weight of the patient is **145.333**

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

**Cars speed and distance**

**Use Q9\_a.csv**

Sol:

Skewness of Speed= -0.117510

Skewness of distance= 0.806895

**Since the Skewness of Speed is negative that indicates its skewed left**

**Since the Skewness of distance is positive that indicates its skewed right**

Kurtosis of Speed= -0.508994

Kurtosis of distance= 0.405053

**Kurtosis of speed has lighter tail**

**Kurtosis of distance has heavier tail**

**SP and Weight(WT)**

**Use Q9\_b.csv**

Sol:

Skewness of SP= 1.611450

Skewness of Weigh= -0.614753

**Since the Skewness of SP is negative that indicates its skewed left**

**Since the Skewness of Weigh is positive that indicates its skewed right**

Kurtosis of SP = 2.977329

Kurtosis of distance = 0.950291

**Kurtosis of SP has heavier tail**

**Kurtosis of Weigh has heavier**

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



Sol:

This histogram represents ‘ Unimodal distribution’ right significant right ‘skewed Positively skewness’ The distribution of the data follows ‘normal distribution’ Mode of the above frequency distribution is 200



Sol:

The boxplot with many outliers falling above upper whisker Lower whisker is smaller than the upper whisker 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?

From the above data

N=2000

Mean=200

Standard deviation=30

Standard Error=30/sqrt(2000)

=0.67082

**For 94%**

Alpha=1- (Cl/100)=0.06

Critical Probability=1- (0.03/2)=0.97

Degree of freedom (df)=2000-1=1999

The critical value is the t score having 1999 degrees of freedom and probability is 0.97.From the t chart the Critical value is 1.8816

Margin of error (ME): ME = critical value \* standard error

=1.8816 \* 0.67082

= 1.26

94 % confident interval of the population mean falls within the interval 200 +- 1.26

**For 98%**

Alpha=1-(Cl/100)=0.02

Critical Probability=1-(0.02/2)=0.99

Degree of freedom(df)=2000-1=1999

The critical value is the t score having 1999 degrees of freedom and probability is 0.99.From the t chart the Critical value is 2.3282

Margin of error (ME): ME = critical value \* standard error

=2.3282 \* 0.67082

= 1.5618

98% confident interval of the population mean falls within the interval 200 +- 1.5618

**For 96%**

Alpha=1-(Cl/100)=0.04

Critical Probability=1-(0.04/2)=0.98

Degree of freedom (df)=2000-1=1999

The critical value is the t score having 1999 degrees of freedom and probability is 0.98.From the t chart the Critical value is 2.0550

Margin of error (ME): ME = critical value \* standard error

=2.0550 \* 0.67082

= 1.3784

96 % confident interval of the population mean falls within the interval 200 +- 1.3784

**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?

Sol: 1 .Mean= 41

Median=40.5

Variance=25.529

Standard deviation=5.052

2.The average marks is 41 scored by student in a test

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

Symmetric

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

Positive Skew

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

Negative Skew

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

Distribution is peaked and has thick tail

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

Distribution is flat and has thin tail

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



What can we say about the distribution of the data?

Sol: The boxplot’s median is 15

Lower quartile Q1=10

Upper quartile Q3=18

Minimum value=2

Maximum value=18

Above maximum and below minimum values are Outliers

**What is nature of skewness of the data?**

Sol: Negative Skewness( Left skewed)

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

IQR= Q3 –Q1

IQR =8

Q19) Comment on the below Boxplot visualizations?



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

Sol:

Box 1 & Box 2 both are follow Normal distribution Variance of Box plot 1 is less than Box plot 2 Inter Quartile range of Box 1 = 25 & Box 2 = 100

Median Q2 of Both Box are same

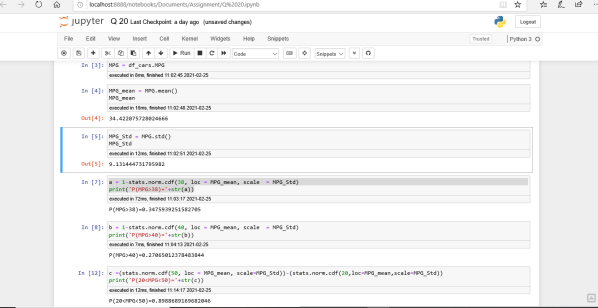
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

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

MPG<- Cars$MPG

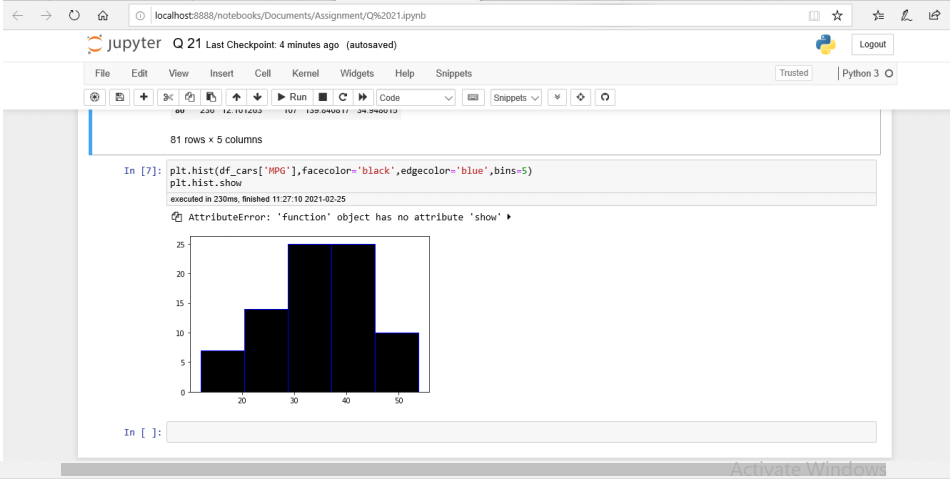
* 1. P(MPG>38)
  2. P(MPG<40)
  3. P (20<MPG<50)



Q 21) Check whether the data follows normal distribution

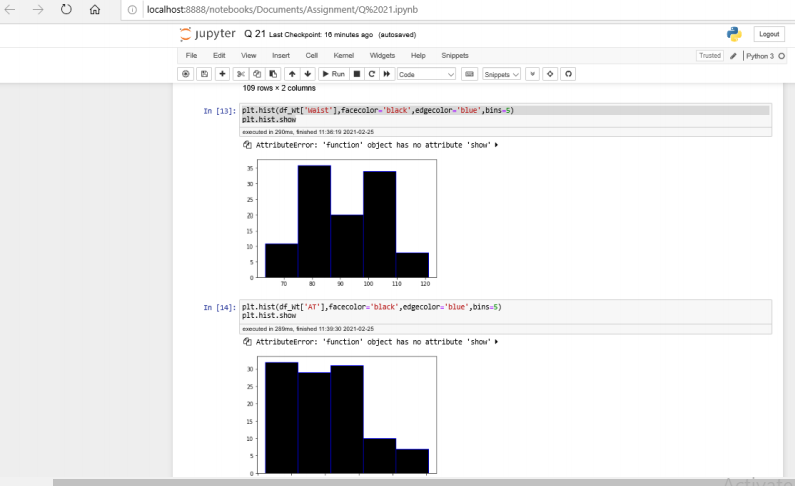
1. Check whether the MPG of Cars follows Normal Distribution

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

Sol:

**For 90%**

Z score =stats.norm.ppf(0.95)

**=**1.6448

**For 94%**

Z score =stats.norm.ppf(0.97)

**=**1.8807

**For 60%**

Z score =stats.norm.ppf(0.8)

**=**0.8416

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

**For 95%**

T score=stats.t.ppf( 0.975 ,df=24)

= 2.0638

**For 96%**

T score=stats.t.ppf( 0.98 ,df=24)

= 2.1715

**For 99%**

T score=stats.t.ppf( 0.995 ,df=24)

= 2.7969

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))

t [1]-0.4714045 Prob=pt(t,df=17) Prob [1]0.3216725

The probability that t<-0.471 with 17 degrees of freedom assuming the population mean is true ,the t-value is less than the t-value obtained with 17 Degrees of freedom and a t score of -0.471,the probability of the bulbs lasting Less than 260 days on average of 0.3218 assuming the mean life of the bulbs Is 300 -