Q1) Identify the Data type for the Following:

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

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

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

Probaility=desired outcome / total number of outcome

Three Coins

*Number of favorable outcomes* =2 head n 1 Tail=0

**0**

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

1. Equal to 1

*Number of favorable outcomes* =0

*Total number of possible outcomes*=36

*P(E)*= *Number of favorable outcomes/Total number of possible outcomes*

P=0/36=0

**0**

1. Less than or equal to 4

*Number of favorable outcomes* =6

*Total number of possible outcomes*=36

*P(E)*= *Number of favorable outcomes/Total number of possible outcomes*

P=6/36=0.166

**16.6%**

1. Sum is divisible by 2 and 3

🡺(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)

*Number of favorable outcomes* =6

*Total number of possible outcomes*=36

*P(E)*= *Number of favorable outcomes/Total number of possible outcomes*

P=6/36=0.166

**16.6%**

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?

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) /(2×1) ​  
=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)​//(2×1) =10  
∴P(E)=n(E)/n(S)​=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

🡺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

=       3.090

=  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 🡺 BasicStatAss\_1.ipynb**

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?

🡺108, 110, 123, 134, 135, 145, 167, 187, 199

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

=1308/9= **145.333**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Answer: 🡪 Q9\_a.ipynb**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Answer: 🡪 Q9\_b.ipynb**

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



**Answer: 🡪 1) Most of the chick weights are between 50-150**

**2) RightSKew / +ve skew**

**1) Most of the chick weights are within 50-100**



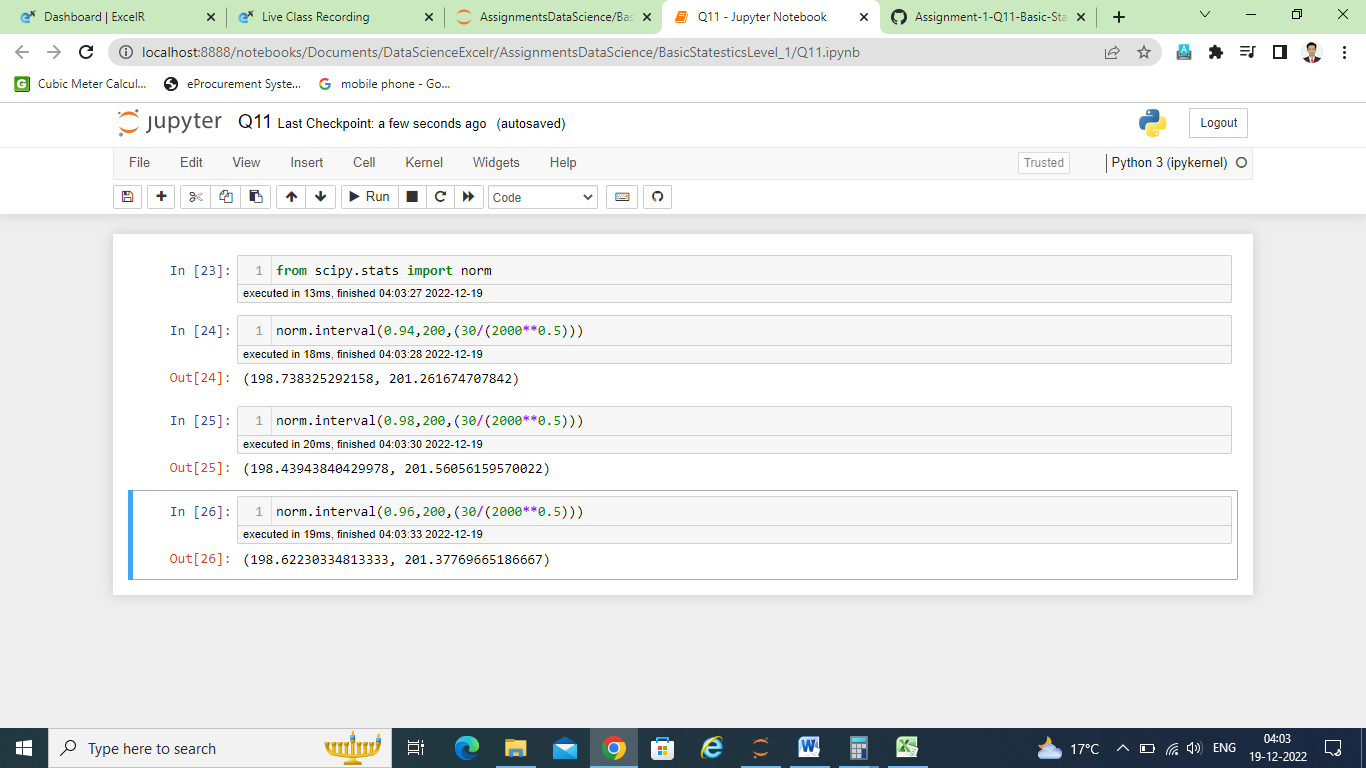
**Answer: 🡪 1) Outliers on upper side**

**2) RightSKew / +ve skew**

3)MEDIAN NEAR TO lower quartile

**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: 🡪 Q11.ipynb**



**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:** 🡪 **Q12.ipynb**

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

Mean=median=0 or Normal destribution

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

* +ve distribution or right skew

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

* -ve distribution or left skew

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

* Thinner Peak , most points r located around tail instead of mean

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

* Wider peak

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



What can we say about the distribution of the data?

* Left skewness or –ve skew
* Mean<median
* Most of the data live below lower quartile i.e, below 10
* Median is closer to upper quartile (18)
* Whisker range is min>max
* No outliers

What is nature of skewness of the data?

* Left skew or –ve skew

What will be the IQR of the data (approximately)?   
🡪upper\_Quartile-Lower\_Quartile🡺18-10 🡪  
IQR= 8

Q19) Comment on the below Boxplot visualizations?



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

* No outliers in both
* Both having Same median in range between 250-275
* Both are Normally distributed Mean=median=0
* No skewness present

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

Data \_set: Cars.csv

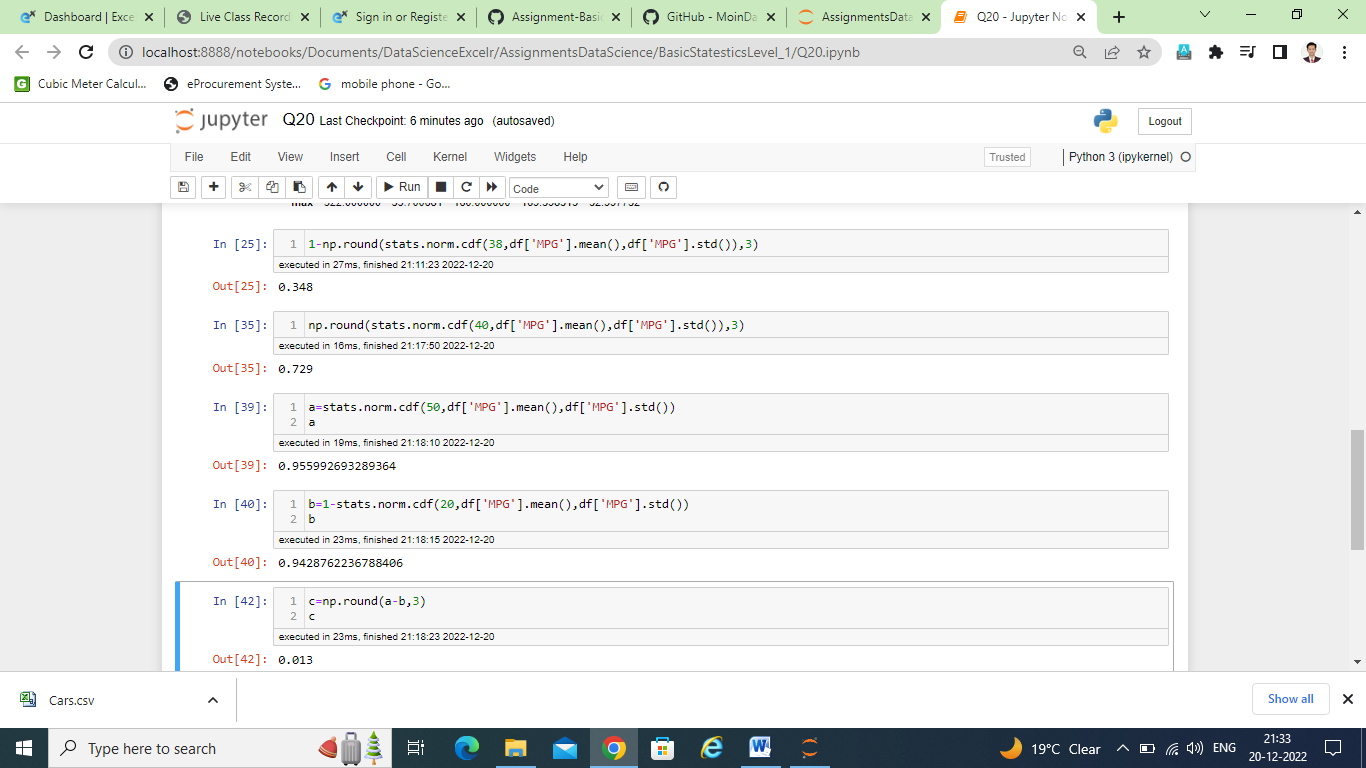
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

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

c. P (20<MPG<50)

**Answer:** 🡪 **Q20.ipynb**



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Answer:** 🡪 **Q21\_A\_B.ipynb**

No outliers

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

**Answer:** 🡪 **Q21\_A\_B.ipynb**

Waist is normally distributed

AT is right skewed

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

* 1) Z score for 90%

(1+0.90)/2=1.90/2=0.95

Check in +ve Zscore table value 0.95

It will be 1.6+0.05=1.65

This left side CI 1.65

So the Z scores of 90% confidence interval 1.65

* 2) Z score for 94%

(1+0.94)/2=1.94/2=0.97

Check in +ve Zscore table value 0.97

It will be 1.8+0.08=1.88

So the Z scores of 94% confidence interval 1.88

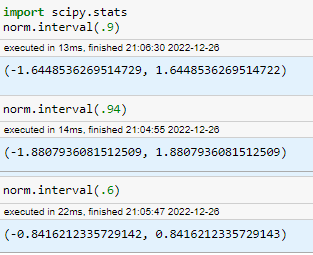
* 2) Z score for 60%

(1+0.60)/2=1.60/2=0.8

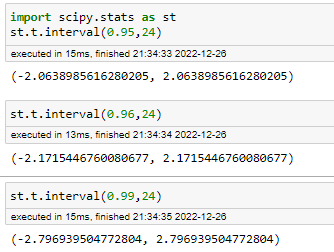
Check in +ve Zscore table value 0.8

It will be 0.8+0.04=0.84

So the Z scores of 60% confidence interval 0.84



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

🡪 **Answer:** 🡪 **Q24.ipynb**

Ho=(u(mue)>=260)=null hypothesis

Ha=(mue<260)=Alternate hypothesis

Alpha=1-95%=1-0.95=0.05=Confidence level

Population mean=U(mue) = 270

Sample mean=x=260

Standard deviation=s=90

Sample size=n=18

One tail=T test

t-statestics=T-test=

(x-u)/(s/sqroot(n))

=(260-270)/(90/sqroot(18))

=-10/(90/4.24)

=-10/21.23

=-0.471

a=abs(-0.4714)=0.4714

Degree of freeredom=df=1-n=1-18=17

import scipy.stats as st

p\_val=1-st.t.cdf(a,17)

p\_val=0.3216

or

p\_val=st.t.sf(a,17)

p\_val=0.3216

here p\_val(0.3216)< alpha(0.05)

So Reject null hypothesis(Ho) means accept Ha (alternate hypothesis)

Means mue < 260 sample