|  |  |
| --- | --- |
| 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 | 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 | 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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| 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**

Total outcome={HHH,HHT,HTH,THH,TTT,TTH,THT,HTT}

=8

1. find the probability that two heads and one tail

P(A)=(3/8)

=========0.375

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

**Solution**

1. Less than or equal to 4

Total Outcome =

{(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. Equal to 1

=P(A)=(0/8)

=0

1. Less than or equal to 4

{(1,1)(1,2)(1,3)(2,1)(2,2)(3.1)}

=6

P(B)=(6/36)

=0.1666

1. Sum is divisible by 2 and 3

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

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

P(C)=(24/36)

=0.667

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**

2=red

3=green

2=blue

Total ball is =7C2

Total Outcome=21

1. What is the probability that none of the balls drawn is blue?

2C2+3C2

=1+3

=4

P(A)=(4/7)

=0.571

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)

**Solution**

|  |  |  |
| --- | --- | --- |
| 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

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Child** | **Candies Count** | **Probability** | **Expected Number** |
|  | A | 1 | 0.015 | 0.015 |
|  | B | 4 | 0.2 | 0.8 |
|  | C | 3 | 0.65 | 1.95 |
|  | D | 5 | 0.005 | 0.025 |
|  | E | 6 | 0.01 | 0.06 |
|  | F | 2 | 0.12 | 0.24 |
| **Total Number** |  |  |  | **3.09** |

1. The Expected number of Candies for randomly selected for child= 3.09

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

**Solution**

* For Points,Score,Weigh>

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

1. Mean,Standard.Deviation, Range

data.describe()



1. Variance
2. data.var()



1. Median

data.median()



1. mode

data.mode()



**Use Q7.csv file**

Q8) Calculate Expected Value for the problem below

**Solution**

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. What is the Expected Value of the Weight of that patient=∑xi/n

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

= 145.3333333

What is the Expected Value of the Weight of that patient=145.333

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Solution**

1. **skewness**

**data.skew()**



1. **Kurtosis**

**data.kurt()**



**SP and Weight(WT)**

**Use Q9\_b.csv**

1. **skewness**

**data.skew()**



1. **kurtosis**

**data.kurt()**



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

**Solution**



Conclusion:

1. From above the Histogram,we conclude that Chick Weight data is right skewed or Positively Skewed
2. More than 50% ChickWeight data is between 50 to 150



Conclusion:

1. From above the boxplot We conclude that there are outliers is the upper side

**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:**

N=300000, n=2000, =200,=30

Confidence Interval Estimate= N=3,000,000 , n=2000 , x̅ = 200 , s=30

Confidence Interval Estimate= Z => 200 Z

94% Confidence :

((1-0.94)/2)+0.94)=(0.97)=1.880794=Z

2001.88\*

Lower limit= 198.7389

Upper limit=201.2611

98% Confidence :

Confidence Interval Estimate= Z => 200 Z

**98% Confidence:** ((1-0.98)/2)+0.98) = (0.99) = [1.880794=Z]

200 2.33\*

Lower limit =198.44

Upper limit = 201.56

**96% Confidence:** ((1-0.96)/2)+0.96) = (0.98) = [1.880794=Z]

200 2.05\*

Lower limit =198.62

Upper limit = 201.38

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

**Solution**

**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. df.describe()

|  |
| --- |
| count 18.000000 |
| mean 41.000000 |
| std 5.052664 |
| min 34.000000 |
| 25% 38.250000 |
| 50% 40.500000 |
| 75% 41.750000 |
| max 56.000000 |
| dtype: float64 |

1. variance

df.var()

25.529411764705884

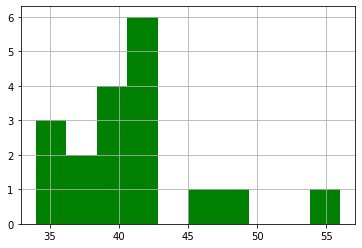
In [42]:

1. What can we say about the student marks?

plt.hist(df,color='g')

plt.grid()

plt.show()



Conclusion:

From above Histogram we can see that most of the data falling near the mean

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

* There is no skewness, and distribution is symmetric

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

* Right skewed distribution (tail on the right side)

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

* Left Skewed distribution (tail on the left side).

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

* Positive values of kurtosis indicate that distribution is more peaked and possesses thick tails.

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

* negative values of kurtosis indicate that distribution is less peaked and possesses thin tails.

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



What can we say about the distribution of the data?

* Negative skewed distribution

What is nature of skewness of the data?

* Left skewed distribution

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

* Inter Quartile Range

=Upper Quartile- Lower Quartile

=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.

1. The median of the two boxplots are same approximately 262.
2. Both follows normal distribution
3. Outliers doesn’t exist in both of the boxplots.

(iiii) Here ,range will be less in boxplot 1 than in boxplot 2

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)

1-stats.norm.cdf(38,loc=Cars.MPG.mean(),scale=Cars.MPG.std())

= 0.3475939251582705

* 1. P(MPG<40)

stats.norm.cdf(40,loc=cars.MPG.mean(),scale=cars.MPG.std())

= 0.7293498762151616

c. P (20<MPG<50)

(i) Cars.MPG.mean()

= 34.422075728024666

(ii)Cars.MPG.std()

= 9.131444731795982

(**iii)stats.norm.cdf(50,loc=34.422076,scale=9.1314 45)-stats.norm.cd**

**f(20,loc =34.422076,scale=9.131445)**

**=** 0.8988689076273199

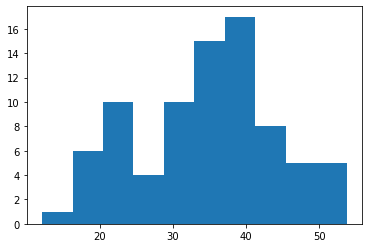
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Solution**

plt.hist(data.MPG)



Conclusion: From above Histogram We can seen that datas are does not normal

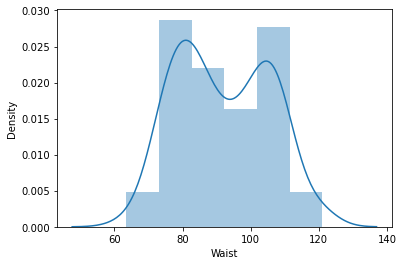
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

1. Waist

sns.distplot(data['Waist'])

plt.show()



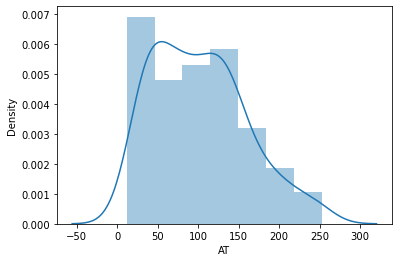
Conclusion:

Data dose not follow normalty

(ii)Adipose Tissue

sns.distplot(data['AT'])

plt.show()



Conclusion:

Data dose not follow normalty

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

**Solution:**

from scipy import stats

from scipy.stats import norm

import pandas as pd

import numpy as np

data=pd.read\_csv("C:\\Users\\Hp\\Desktop\\wc-at.csv")

**(i) for the 90% Confidence interval**

data=stats.norm.interval(0.90,loc=data.mean(),scale=data.std())

print('Mean at 90% Confidence interval is :',np.round(data,4))

Mean at 90% Confidence interval is : [[ 69.5991 7.6525]

[114.2046 196.1355]]

**(ii) for the 94% Confidence interval**

1. data=stats.norm.interval(0.94,loc=data.mean(),scale=data.std())
2. print('Mean at 94% Confidence interval is :',np.round(data,4))

Mean at 94% Confidence interval is : [[ 66.3999 -5.8656]

[117.4037 209.6537]]

**(ii) for the 60% Confidence interval**

1. data=stats.norm.interval(0.60,loc=data.mean(),scale=data.std())
2. print('Mean at 60% Confidence interval is :',np.round(data,4))

Mean at 60% Confidence interval is : [[ 80.4902 53.6735]

[103.3135 150.1145]]

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

**Solution**

(i)For the 95% of confidence interval

#t score of 95% confidence interval for sample size of 25

stats.t.ppf(0.975,24) # df= n-1

=2.0638985616280205

In [4]:

1. For the 96% of confidence interval

# t score of 96% confidence interval for sample size of 25

stats.t.ppf(0.98,24) # df= n-1

=2.1715446760080677

1. For the 99% of confidence interval

# t score of 99% confidence interval for sample size of 25

stats.t.ppf(0.995,24) # df= n-1

= 2.796939504772804

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:**

H0 : The average life of bulbs are260 days

H1 : The average life of bulbs are not more than 260 days

µ = 270,

n = 18,

= 260,

s = 90

tscore=

1. tscore=(260-270)/(90/math.sqrt(18))

tscore

= -0.4714045207910317

1. stats.t.cdf(t,17)

=0.32167253567098364

Conclusion:

Here P-value is greater than 0.05.so we do not reject the that and

Conclude the average life of bulls are 260 days