|  |  |
| --- | --- |
| 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 | 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?

Ans: A coin has two sides and here 3 coins are tossed at the same time. The number of possibilities can be calculated by 23 = 8. The possibilities would be following TTH,THT,HTT,HHH,TTT,THH,HTH,HHT. Out of which ,2 heads and 1 tail are obtained three times.

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

ANS :When 2 dice are rolled the total probabiltity of outcomes is 36.

1. Nothing can be equal to zero : hence = 0
2. The possible outcomes where the sum is </= 4 are

(1,1),(1,2),(1,3),(2,1),(2,2),(3,1) = 6 times

Hence 6/36 =0.16

c)There are only two possible outcomes of the two dice where their sum is divisible by both 2 and 3,those are 6 and 12

out of all the outcomes 6 comes 5 times and 12 comes only once.

Hence in total there are 6 outcomes , 6/36 =0.16

----------------------------------------------------------------------------------------------------

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.

Ans: Total number of balls in the bag = 7

Number of balls excluding blue colour are 5.

Total probability of outcomes of balls 7\*6= 42

Probability of outcomes of drawing only 2 balls = 42/2 = 21

No of Outcomes with no blue outcomes (no of red and green balls to the power of 2) = 5 2 =25.

Outcomes with 2 balls out of 5 balls is 25/2 = 10.

Hence final outcome is 10/21 = 0.47

----------------------------------------------------------------------------------------------------

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

Ans:Expected number of candies = A(Probability\*Number)+B(P\*N)+C(P\*N)+D(P\*N)+E(P\*N)+F(P\*N)

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

=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**

|  |  |  |  |
| --- | --- | --- | --- |
|  | points | Score | Weight |
| Mean | 3.5965625 | 3.21725 | 17.84875 |
| MEDIAN | 3.69 | 3.215 | 17.6 |
| STDEV | 0.534678736 | 0.978457443 | 1.786943236 |
| MODE | 3.92 | 3.44 | 17.02 |
| VARIANCE | 0.28588 | 0.957378968 | 318.57787 |
| RANGE | 2.17 | 3.911 | 8.4 |

By the evaluation of the given data the car MERC 240 tends to be the best car in terms of points,score and also weight.

**-------------------------------------------------------------------------------------------------------------**

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?

Sum of (probability\*value)

Sum of[p(x)\*E(x)]

Probability of 1 in 9 = 1/9

Expected value = (1/9)\*(108+110+123+134+135+145+167+187+199)

**=145.318**

-------------------------------------------------------------------------------------------------------------

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Kurtosis =**

speed -0.508994

dist 0.405053

**SKEWNESS =**

speed -0.117510

dist 0.806895

SP and Weight(WT)

Use Q9\_b.csv

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



* It is a positively skewed histogram.
* Highest number of chicks weigh between 50-100.
* Outliers exist at the maximum fence.
* Best possible Mean = 176.25 (summation of mini/N)



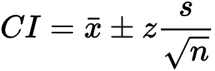
ANS:

* It is a positively skewed boxplot.
* More amount of data is present in the Q3.
* Outliers are present over the maximum.
* Median is towards minimum.

--------------------------------------------------------------------------------------

**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?

Ans: confidence interval formula =



Sample size = 2000

Sample avg weight/ mean = 200

SD of sample = 30

Confidence level value of

94%=1.88

98%=2.326

96%=2.05

[Confidence level value can be found by =

Find the AUC value = 1+Confidence level in decimals/2

And the given AUC should be found in the standard normal distribution table and both the column and row number if added gives the confidence level value.

For eg :94% = 1+0.94/2 = 0.97(it is AUC)

Finding it in the standard ND table gives the Z score as – 1.88]

a) Confidence interval **94%** = 200+- 1.88\*300/(root of 2000)

= **201.25 / 198.74.**

b)Confidence interval **98%** = 200+- 2.326\*30/(root of 2000)

= **201.55 / 198.442**.

c) Confidence interval **96%**= 200+-2.05\*30/(root of 2000)

= **201.373 / 198.627**.

-----------------------------------------------------------------------------------

**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?

Ans:

import pandas as pd

import numpy as np

import statistics as stats

data = [34,36,36,38,38,39,39,40,40,41,41,41,41,42,42,45,49,56]

print(stats.mean(data)).

Output: 41

print(stats.median(data))

40.5

print(stats.variance(data))

25.529411764705884

print(stats.stdev(data))

5.05266382858645

2Ans)On an average , marks of the students is 41.

-------------------------------------------------------------------------------------------------------------

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

It means that the data is symmetric and have no deviations.

-------------------------------------------------------------------------------------------------------------

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

It means the data is positively skewed which means the graph is tilted to the left side.

-------------------------------------------------------------------------------------------------------------

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

It means the data is negatively skewed which means the graph is tilted to the right side.

-------------------------------------------------------------------------------------------------------------

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

It means the data is highly peaked and the data is away from the mean and has thick tails.

-------------------------------------------------------------------------------------------------------------

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

It means the data has less peak and data is surrounded near the mean and has thin tails which suggests it closeness to a normal distribution.

-------------------------------------------------------------------------------------------------------------

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



What can we say about the distribution of the data?

=More amount of data is present in q1.

What is nature of skewness of the data?

=It is a negatively skewed boxplot.

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

=IQR = Q3-Q1 = 8  
------------------------------------------------------------------------------------------------------------

Q19) Comment on the below Boxplot visualizations?



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

ANS: Both the boxplots has median at the same point whereas the first plot is slightly positively skewed and the second plot is normally distributed.

The first plot has very less iqr when compared to the second one.

-------------------------------------------------------------------------------------------------------------

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)

a)33/81 = 0.407

b)61/81 = 0.753

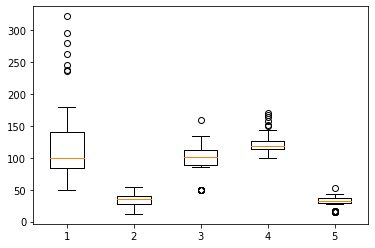
c)69/81 = 0.851.

-------------------------------------------------------------------------------------------------------------

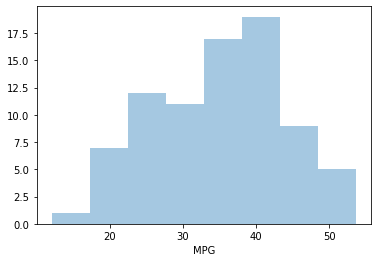
Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



Ans: No the data is not normaly distributed.



ANS: MPG data is slightly negative skewed.

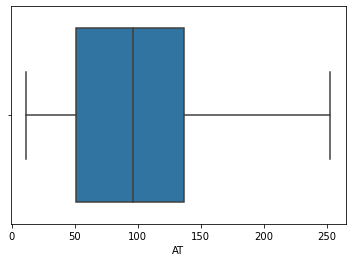
----------------------------------------------------------------------------------------------------

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

Dataset: wc-at.csv

sns.boxplot(df['AT'])

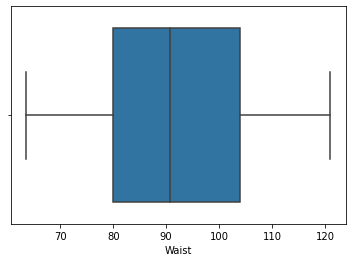
plt.show()



It is positively skewed boxplot(Adipose Tissue).

sns.boxplot(df['Waist'])

plt.show()



Both the whiskers are of same length and the median is slightly shifted to the left side. It is slightly positive skewed but almost it is normally distributed.

----------------------------------------------------------------------------------------------------

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

ANS:

a)90% CI: formula is stats.norm.ppf(1-alpha)

from scipy import stats

from scipy.stats import norm

#z score - 90% CI.

stats.norm.ppf(0.95)

**= 1.645**

b)94% CI.

from scipy import stats

from scipy.stats import norm

#z score - 94% CI.

stats.norm.ppf(0.97)

**= 1.880**

c)60% CI:

from scipy import stats

from scipy.stats import norm

#z score - 60% CI.

stats.norm.ppf(0.8)

**= 0.842**

----------------------------------------------------------------------------------------------------

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

Sample Size = 25

Degrees of freedom = N – 1 = 25-1 = 24

From the t- table:

95% = 2.064

96% = 2.172

99% = 2.797

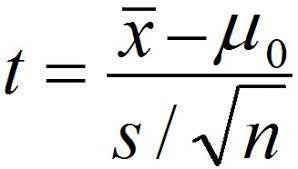
Q 24**)** A Government company claims that an average light bulb lasts 270 days. A researcher randomly selects 18 bulbs for testing. The sampled bulbs last an average of 260 days, with a standard deviation of 90 days. If the CEO's claim were true, what is the probability that 18 randomly selected bulbs would have an average life of no more than 260 days

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom

T test Formula:



Population mean = 270

Sample size = 18

Sample mean = 260

Standard deviation = 90

Degrees of freedom = 18-1 = 17

= 260-270/90/(root of 18)

= -0.471

The above t test score is checked with the degrees of freedom that is 17 in the Ttest table = the probability is 0.3218.0.