|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Continuos |
| Weight of Gold | Continuos |
| Distance between two places | Continuos |
| Length of a leaf | Continuos |
| Dog's weight | Continuos |
| 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 | Ratio |
| Weight | Interval |
| Hair Color | Nominal |
| Socioeconomic Status | Nominal |
| Fahrenheit Temperature | Ratio |
| Height | Interval |
| Type of living accommodation | Ordinal |
| Level of Agreement | Nominal |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Nominal |
| Number of Children | Nominal |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Ans:

Three coins are tossed means total set = {(H,H,H),(H,H,T),(H,T,T),(T,T,T),(T,H,H),(T,H,T),(T,T,H),(H,T,H)} = 8

Probability for two heads and one tail are obtained is lets say set A= {(H,H,T),(T,H,H),(H,T,H)} = 3

P (A) = Elements of A / Total no of outcomes = 3/8

So P(A) = 3/8

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:

Total outcomes for the two dice are rolled are: {(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)}

Elements of total set = N = 36

1. Lets say set A = {sum is equal to 1} = {0}

So Elements of A i.e. A = 0

P(A) = A/N = 0/36 = 0

1. Lets say B = {Sum is less than or equal to 4} = {(1,1),(1,2),(1,3), (2,1),(2,2), (3,1)}

So Elements of B i.e. B = 6

P(B) = B/N = 6/36 = 1/6

1. Lets say C = {Sum is divisible by 2 and 3} = {(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)}

So Elements of C i.e. C = 6

P(C) = C/N = 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?

Ans:

Total probability of drawing two balls is 7C2 = 7! /(7-2)!\*2! = 7\*6\*5! /5!\*2! = 21

None of the balls drawn is blue means we are drawing red and green balls out of total. So at time we are drawing two balls so probability of drawing two balls which consists of red and green is 5C2 = (5! /(5-2)!\*2!) = (5\*4\*3!/3!\*2!) = 20/2 =10

So probability of that none of balls is blue = 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

Answer:

Probability for the randomly selected child means it is selecting either A or B or C or D or E or F = P(AUBUCUDUEUF) …. As these are independent events.

So P(AUBUCUDUEUF) = P(A)+ P(B)+ P(C)+ P(D)+ P(E)+ P(F) = (1\*0.015)+(4\*0.20)+(3\*0.65)+(5\*0.005)+(6\*0.01)+(2\*0.120) = 3.09

Expected number of candies for a randomly selected child = 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:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | 3.5965625 | 3.21725 | 17.84875 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 3.07 and 3.92 | 3.44 | 18.9 and 17.2 |
| **Variance** | 0.285881351 | 0.957379 | 3.193166129 |
| **standard deviation** | 0.534678736 | 0.978457 | 1.786943236 |
| **Range** | 2.76,4.93 | 1.513,5.424 | 14.5,22.9 |

(refer assignment.ipynb file Q-7)

Comment: There is not much difference in mean and median of the points, variance is also low so this is much steady data with less variation. Same thing as like points is applied to score only slight variance is there compared with points. Weigh has larger variation compared with points and score.

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?

Answer:

Expected Value of the Weight of that patient = mean of (108, 110, 123, 134, 135, 145, 167, 187, 199) = 145.3333

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answer:

|  |  |  |
| --- | --- | --- |
| Parameter | Cars Speed | Cars distance |
| Skewness | -0.1139 | 0.7824 |
| Kurtosis | 2.422 | 3.248 |

Refer assignments.ipynb Q9A

Inference: Cars speed skewness is negative it implies that the mass is distributed on the right side whereas Cars distance skewness is positive it implies that the mass is distributed on the left side. Cars speed and cars distance kurtosis is positive which implies that it is thinner at peak and wider at the tails compared with the normal distribution.

**SP and Weight(WT)**

**Use Q9\_b.csv**

|  |  |  |
| --- | --- | --- |
| Parameter | SP | WT |
| Skewness | 1.5814 | -0.6033 |
| Kurtosis | 5.723 | 3.8194 |

Refer assignments.ipynb Q9B

Inference: SP skewness is positive it implies that the mass is distributed on the left side. WT skewness is negative it implies that the mass is distributed on the right side. SP and WT kurtosis is positive which implies that it is thinner at peak and wider at the tails compared with the normal distribution.

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



Answer:

From histogram it is observed that the most of the data lie between the 0 to 200 and very few data lie after 200. On boxplot it is observed that some of data is lying outside the limits i.e. some outliers are there.

**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:

|  |  |  |
| --- | --- | --- |
| **Confidence Interval** | **Z value** | **Range** |
| 94% | 1.88 | 198.74, 201.26 |
| 96% | 2.06 | 198.62, 201.38 |
| 98% | 2.33 | 197.43, 201.56 |

(refer assignment.ipynb file Q-11)

**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:

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Mode | 41 |
| Variance | 25.52941 |
| Standard deviation | 5.052664 |

Refer Q12 excel file for explanation.

Average marks obtained by the students are 41 and 40.5 divides both 50% data on either side. There is large variation between lowest and highest marks.

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

Ans: There is no skewness in the data

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

Ans: When mean is greater than the median then mass is concentrated on the right side and having negative skewness.

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

When mean is greater than the median then mass is concentrated on the right side and having negative skewness.

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

Ans: Positive kurtosis value indicates thinner peak and wider tails.

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

Ans: Negative kurtosis value indicates wider peak and thinner tails.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

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

Answer:

More data is distributed on the right side of the box plot so not normally distributed

Skewness is negative

IQR of data is 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.

Answer:

Boxplot 1 covers data with very less variation and boxplot 2 covers data with high variation compared with boxplot 1.

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:

1. P(MPG>38) = 0.3475939251582705 = 34.75%
2. P(MPG<40) = 0.7293498762151616 = 72.93%
3. P (20<MPG<50) = 0.8988689169682046 = 89.88%

Refer file assignments.ipynb Q-20

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

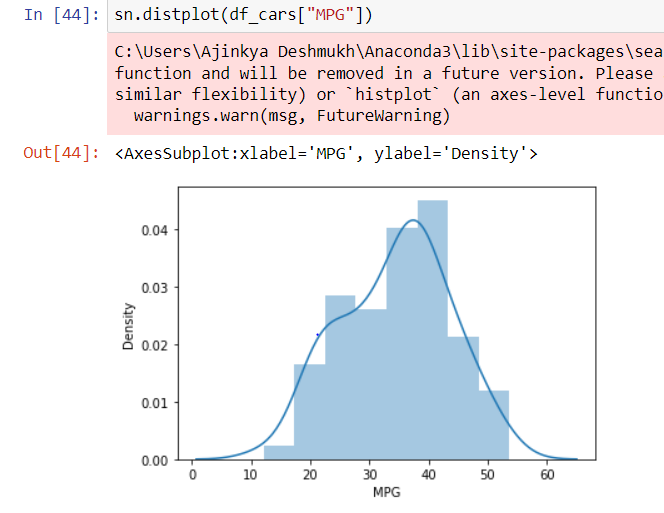
Dataset: Cars.csv

Answer:

Refer assignments.ipynb Q-21a

Skewness of MPG of cars: -0.1764

It means MPG of the cars is having negative skewness with more mass distributed on the right side so it does not follows normal distribution



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

Dataset: wc-at.csv

Answer:

Refer assignments.ipynb Q-21b

Skewness of Waist = 0.1322

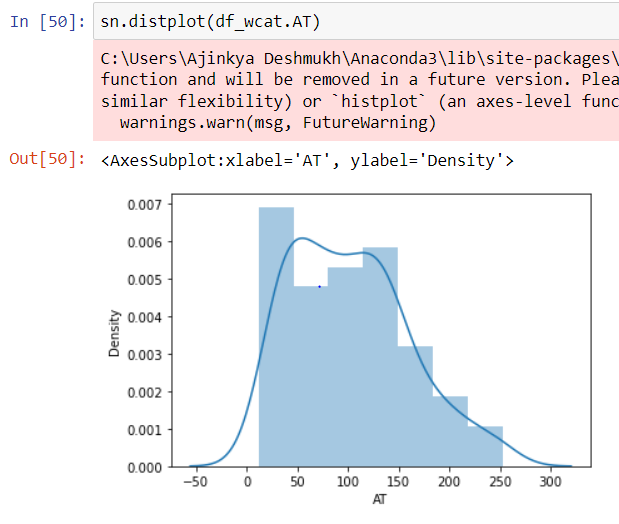
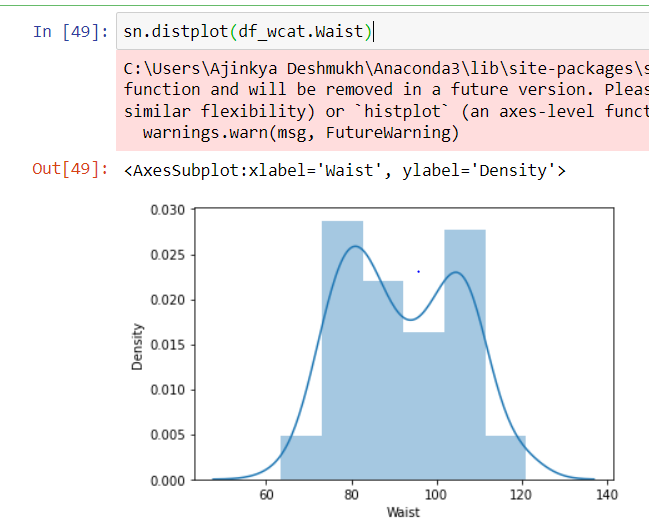
It means Waist is having positive skewness with more mass distributed on the left side so it does not follows normal distribution (it is also clear from its shape from the python file)

Skewness of AT = 0.5767

It means AT is having positive skewness with more mass distributed on the left side so it does not follows normal distribution (it is also clear from its shape from the python file)

Waist Circumference(Waist) from wc-at data does not follows normal distribution and Adipose Tissue (AT) also not follows normal distribution as both are not symmetrical about the axis.

Refer below screenshots and refer file assignments.ipynb.



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

Ans:

|  |  |
| --- | --- |
| Confidence Interval | Z score |
| 60% | 0.8416 |
| 90% | 1.6448 |
| 94% | 1.8807 |

refer file assignments.ipynb – Q22

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

Ans:

|  |  |
| --- | --- |
| Confidence Interval | t score |
| 95% | 2.0638 |
| 96% | 2.1715 |
| 99% | 2.7969 |

refer file assignments.ipynb – Q23

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: n= 18

u=270 days, s(sample deviation) = 90 days, xnbar = 260

P(u<=260) = 0.3216 or 32.16%

refer file assignments.ipynb – Q24