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

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

**HHH, HHT, HTH, THH, TTH, THT, HTT, TTT are the combination when 3 coins are tossed.**

**Therefore, The number of combinations which have two heads and one tail are:**

**HHT, HTH, TTH which makes them 3 in number. -> 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

Answers:

1. **Zero probability that sum is equal to 1**
2. **Possibilities are (1,1);(1,2); (1,3); (2,1); (2,2); (3,1) when two dice are roll, Therefore, 6/36 => 1/6 -> 0.1667**
3. ***When two dice are rolled, sample space is given as:***

**(1,1) (1,2) (1,3) (1,4) (1,5) (1,6) -> (1,5)**

**(2,1) (2,2) (2,3) (2,4) (2,5) (2,6) -> (2,4)**

**(3,1) (3,2) (3,3) (3,4) (3,5) (3,6) -> (3,3)**

**(4,1) (4,2) (4,3) (4,4) (4,5) (4,6) -> (4,2)**

**(5,1) (5,2) (5,3) (5,4) (5,5) (5,6) -> (5,1)**

**(6,1) (6,2) (6,3) (6,4) (6,5) (6,6) -> (6,6)**

**Therefore, Total number of possible outcomes = 36**

**6/36=> 1/6 -> 0.1667**

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?

Answers:

**Total no. of chances to draw 2 balls at random from 7coloured balls, n(s)=7C2=21**

**Let E = Event of drawing 2 balls, none of which is blue ball.**

**n(E)= Number of ways of drawing 2 balls out of (2 + 3) balls= 5C2=10**

**p(E)=n(E)/n(s)=10/21**

**so, probability=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 the3

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

Answers:

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

**= 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0 .24 -> 3.09**

**So, therefore 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**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | **3.59** | **3.21** | **17.84** |
| **Median** | **3.695** | **3.325** | **17.710** |
| **Mode** | **3.07**  **3.92** | **3.44** | **17.02**  **18.90** |
| **Variance** | **0.2858** | **0.9573** | **3.2130** |
| **Standard Deviation** | **0.5346** | **0.9784** | **1.7869** |
| **Range** | **2.17** | **3.911** | **8.399** |

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?

Answers:

**Mean/average = sum of all observation / Total number of observation**

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

**= 145.333**

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

**Cars speed and distance**

**Use Q9\_a.csv**

Answers:

|  |  |  |
| --- | --- | --- |
|  | **Cars Speed** | **Cars Distance** |
| **Skewness** | **-0.117510** | **0.806895** |
| **Kurtosis** | **-0.508994** | **0.405053** |

**SP and Weight(WT)**

**Use Q9\_b.csv**

Answers:

|  |  |  |
| --- | --- | --- |
|  | **SP** | **Weight (WT)** |
| **Skewness** | **1.611450** | **0.614753** |
| **Kurtosis** | **2.977329** | **0.950291** |

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



Answers:

**This histogram is Right side skewed which mean it’s a positive skewed**

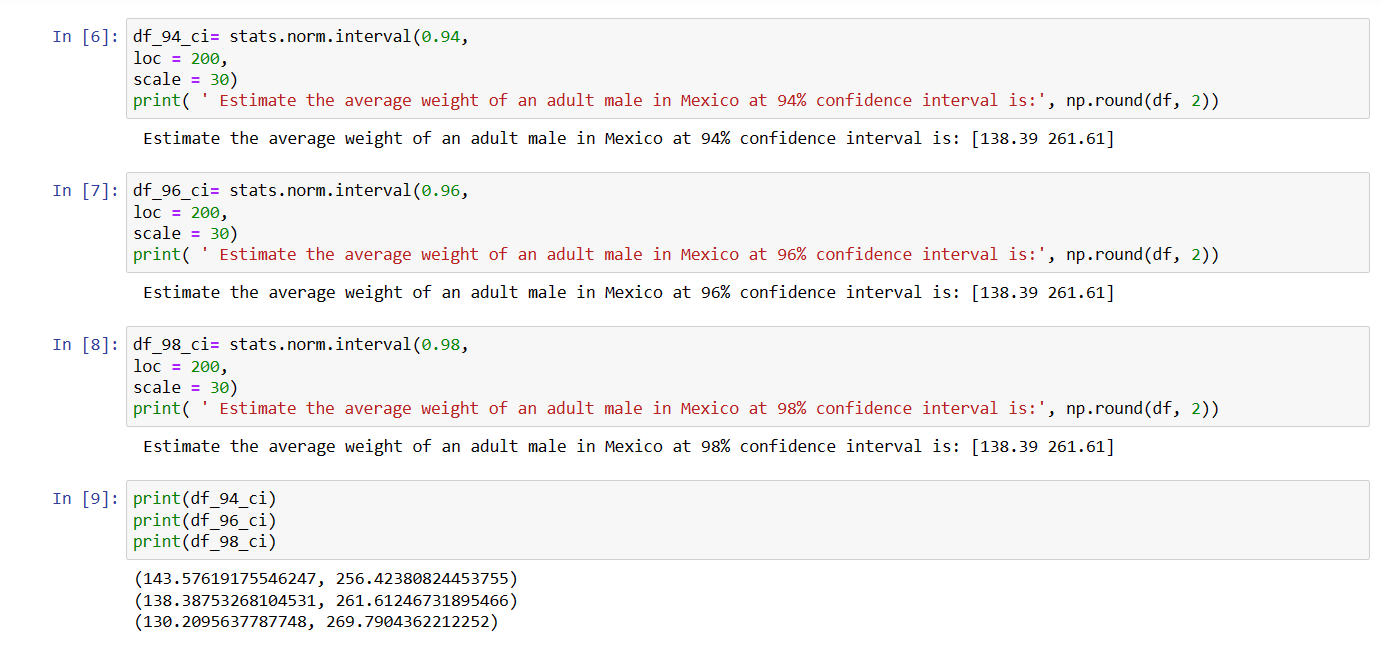


Answers:

**The distribution for this boxplot is considered “positively skewed” mean > median. It means the data constitute higher frequency of high valued scores.**

**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?

Answers:



**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?

Answers:

1.

|  |  |
| --- | --- |
| Mean | **41.00** |
| Median | **40.5** |
| Variance | **25.52** |
| Standard Deviation | **5.052** |

2**. Average scores obtained by a student in tests is 41 with a standard deviation of 5.05**

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

Answers:

**Then the distribution is symmetric.**

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

Answers:

**Then the distribution is positively skewed.**

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

Answers:

**Then the distribution is negatively skewed.**

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

Answers:

**Thinner peak and longer tails.**

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

Answers:

**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?

Answers:

**The distribution of above visualized boxplot left side tail is longer than right side of the visualized boxplot. Hence, it is not normally distributed.**

What is nature of skewness of the data?

Answers:

**The distribution of above visualized boxplot is negatively skewed, because the boxplot shows the median closer to the upper quartile or top quartile.**

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

**Q1 -> 10**

**Q3 -> 18**

**Therefore, IQR = Q3 - Q1**

**-> 18 – 10 = 8   
-> IQR\*1.5**

**->8\*1.5 = 12**

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect 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)
  2. P(MPG<40)

c. P (20<MPG<50)

Answers:

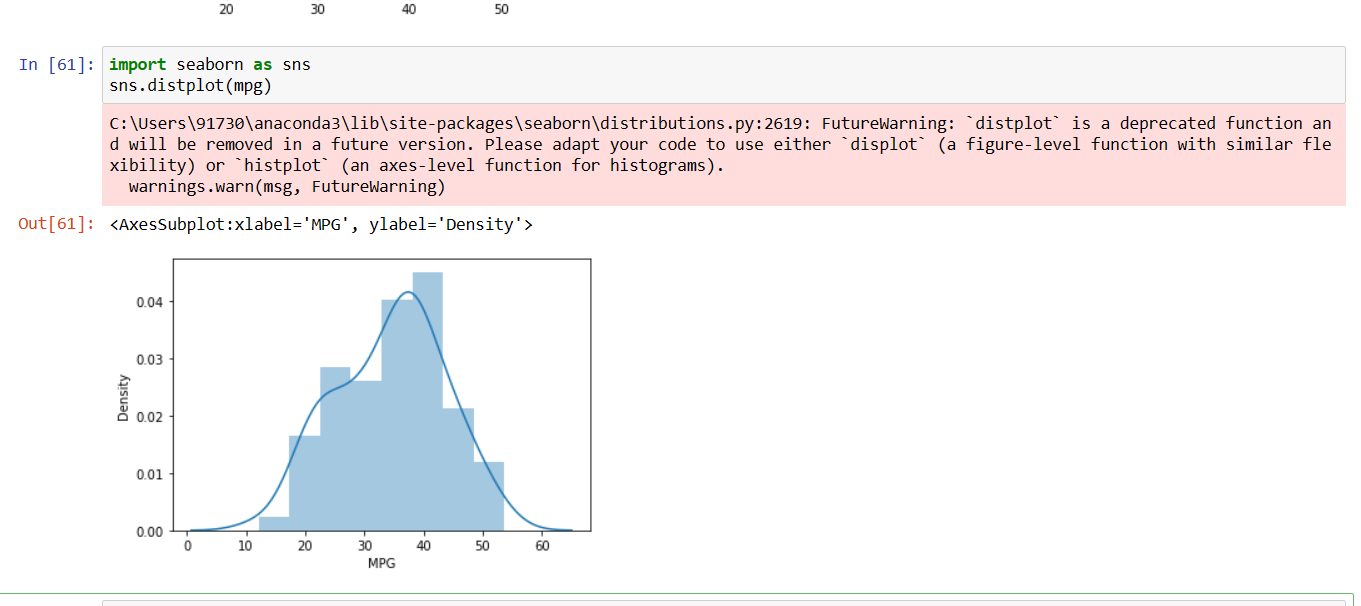


Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answers:

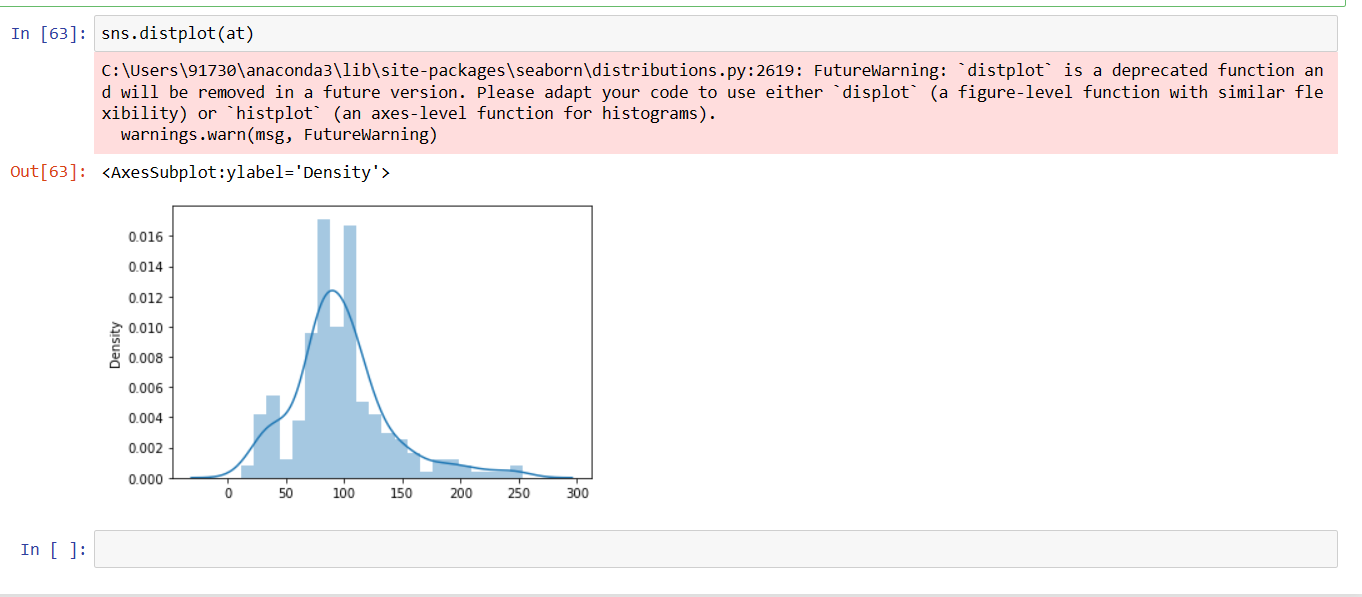


**Normally Distributed.**

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

Dataset: wc-at.csv

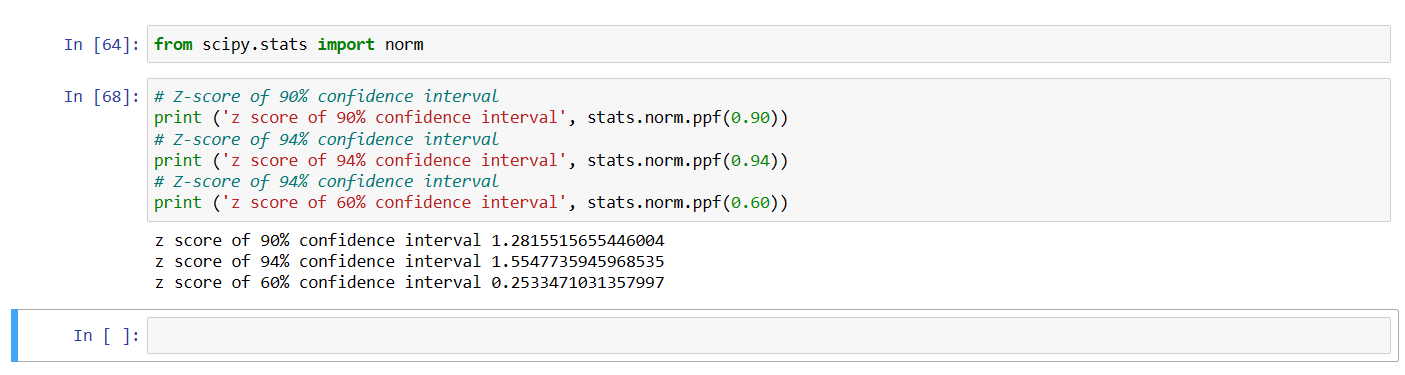
Answers:



**Normally Distributed.**

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

Answers:



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

Answers:

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

Answers:

