|  |  |
| --- | --- |
| 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) | Nominal |

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) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| 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?

ANS: **3/8**

Explanation:

(HHH, HHT, HTH, THH, TTH, THT, HTT, TTT)

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: **a)0 b)1/6 c)4/9**

Explanation:

1. The probability of getting sum=1 is 0.

Hence, P(1) = 0

b) The probability of getting sum <= 4 is 1/6

When two dies are thrown, the probability less than or equal to 4 is

(1,1), (1,2), (1,3), (2,1), (3,1)

Number of favourable outcomes=6

Total number of possibilities=36

Probability= Number of favourable outcomes / Total number of possibilities

=6/36

=1/6

c) The probability of getting the sum which is divisible by 2 & 3 is 5/36

favourable outcomes = (1,5), (3,3), (4,2), (5,1), (6,6)

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: **10/21**

Probability of getting no blue balls is 10/21

Explanation:

Total number of balls = (R, R, G, G, G, B, B) = (2+3+2) =7.

Let S is the sample space.

Then, N(S)=number of ways of drawing 2 balls out of 7.

= 7 C 2 ​ = (5×4) / (2×1)​ = 10

Let “e” is the event of drawing 2 balls, none of which is blue.

n(e) = Number of ways of drawing 2 balls out of (2 + 3) balls.

= 5 C 2 ​ = (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

ANS:

**Expected number of candies for a randomly selected child is 3.09**

**Explanation:**

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

**=0.015+0.8+1.95+0.025+0.06+0.24**

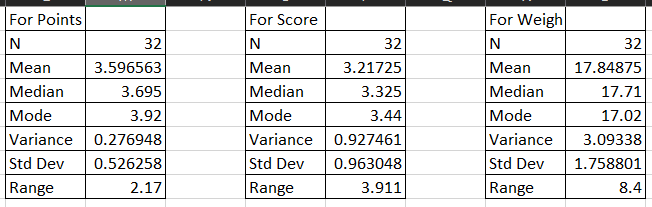
**=3.09**

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**

**ANS: **

\***Calculated these values from Excel**

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?

ANS: **Expected value of weight for a random selected patient =145.33**

Explanation:

Expected value = ∑(probability \*value)

Probability of selecting each patient = 1/9

Expected value

= (1/9)108+(1/9)110+(1/9)134+(1/9)135+(1/9)145+(1/9)167

+ (1/9)187+(1/9)199

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

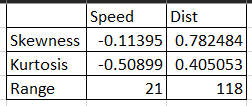
= (1/9) (1308)

=145.33

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

**Cars speed and distance**

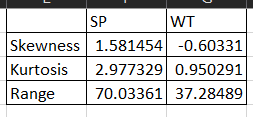
**Use Q9\_a.csv**

**ANS: **

\***Calculated these values from Excel**

**SP and Weight (WT)**

**Use Q9\_b.csv**

**ANS: **

\***Calculated these values from Excel**

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



1. **We can see that it is not normal distribution**
2. **It is right skewed distribution, we can also call it positive distribution**
3. **Mode is approximately 75**
4. **Outliers lies between 350 to 400**



1. **Given boxplot is not symmetric.**
2. **We can see outliers above extreme high**
3. **Boxplot is right skewed**
4. **We can also see 25%,50%,75% quartile.**
5. **50% quartile means where median lies**

**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: **1. 94% Confidence Interval - 198.74 pounds to 201.26 pounds**

**2. 98% Confidence Interval - 198.44 pounds to 201.56 pounds.**

**3. 96% Confidence Interval - 198.83 pounds to 201.17 pounds.**

**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.

ANS: **Mean = 41, Median = 40.5, Variance = 40.11, Std Dev = 6.34**

1. What can we say about the student marks?

ANS: **1. We can say that average marks obtained is 41**

**2. Min marks obtained by student is 34 and max marks is 56**

**3. We can not any any outlier in the data as far as now**

**4. Variance is 40.11 so we can see the spread of the data**

**5. Mode is 41 it means it is most frequent occurring data point**

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

ANS: **When mean and median of data are equal then nature of skewness is symmetrical. It is zero skewed and we can also say that data is normally distributed.**

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

ANS: **When mean is greater than median, data is distributed to the right so we can say that it is skewed in right side and it is also known as positive skewness in nature.**

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

ANS: **When median is greater than mean, data is distributed to the left so we can say that it is skewed in left side and it is also known as negative skewness in nature.**

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

ANS : **A positive kurtosis value indicates that a dataset has heavier tails and a more peaked distribution compared to a normal distribution. It suggests that the data has a higher concentration of values around the mean and more extreme values in the tails.**

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

ANS: **negative kurtosis indicates that the dataset has a lighter-tailed distribution, meaning that there are fewer values in the tails compared to a normal distribution. The data is more concentrated towards the mean, with a lower probability of observing values far away from the mean.**

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



What can we say about the distribution of the data?

ANS: **Dataset does not follow Normal/Gaussion Distribution**

**1. By looking at the boxplot, we can say that median is 15**

**2. Datapoint ‘10’ is at 25th quartile, datapoint 18 is at 75th quartile**

**3. frequency of datapoints is maximum between 10 to 18**

**4. We can also say that data is negatively skewed and datapoints 2 to 6 can be outliers**

What is nature of skewness of the data?

ANS: 1. Given data is negatively skewed

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

ANS: **The Interquartile Range (IQR) is a measure of statistical dispersion that quantifies the spread of a dataset. It is calculated by finding the difference between the upper quartile (Q3) and the lower quartile (Q1) of the data.**

**IQR = Q3 – Q1 = 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.

ANS: 1. **By looking at given boxplots we can say that they are symmetric and follow Gaussion Distribution**

**2. Also they have same median as we can see i.e. aprox 267**

**3. They do not contain outlier**

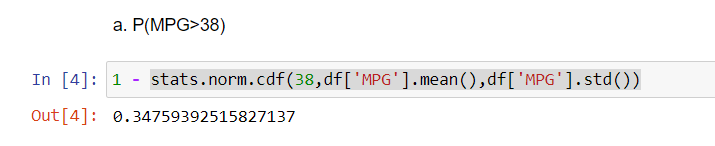
Q 20) Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

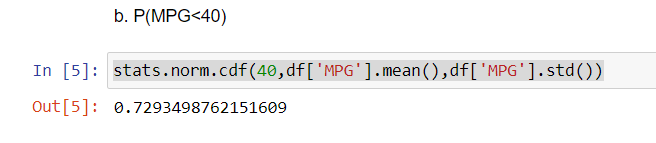
Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

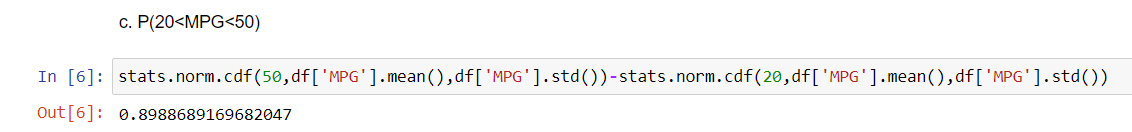
* 1. P(MPG>38)

ANS: 

* 1. P(MPG<40)

ANS: 

* 1. P (20<MPG<50)

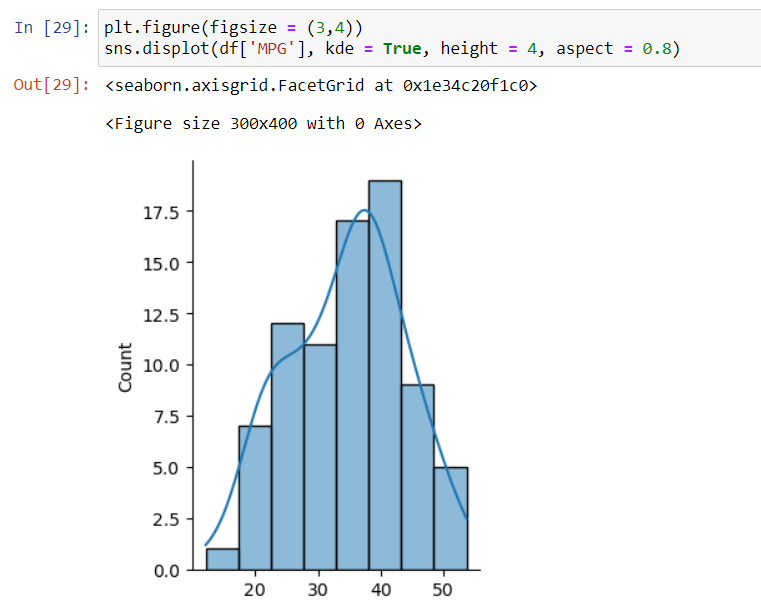
ANS: 

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

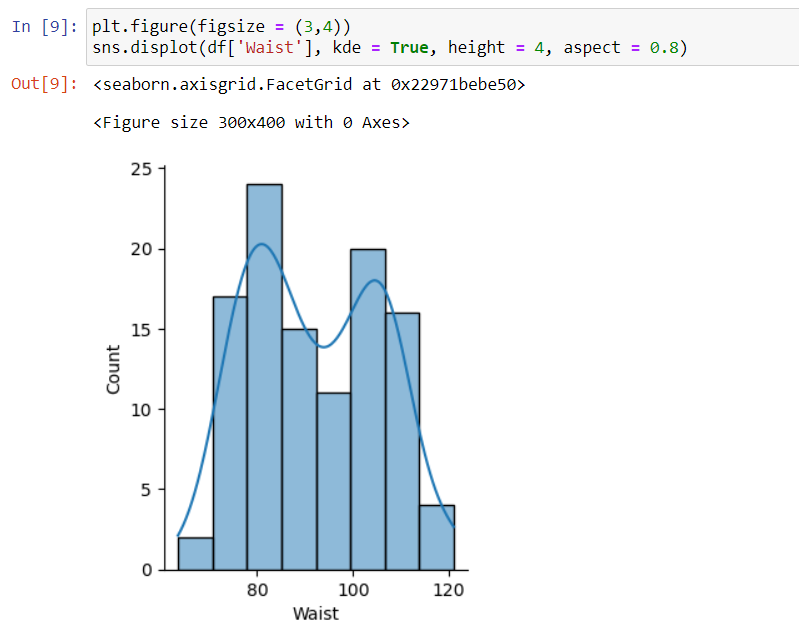
ANS: **By looking at the graph we can say that somehow it is following normal distribution, not in the pure form but we can see the bell-shaped curve. But there is -ve skewness**



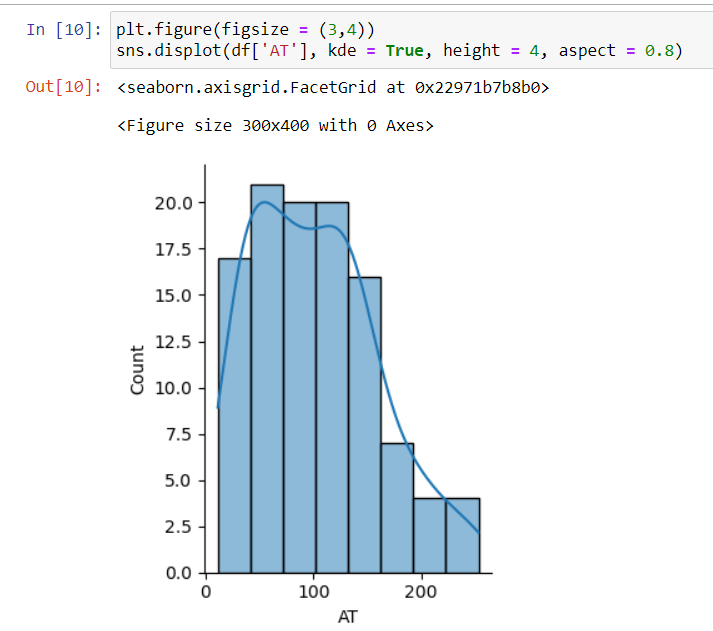
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

ANS: Waist Circumference does not follow normal distribution.



ANS: AT also does not follow normal distribution, but we can say that it is Right Skewed.



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

ANS: Z score of 60% CI = 0.84

Z score of 90% CI = 1.645

Z score of 94% CI = 1.881

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

ANS: **T Score of 95% CI = 2.06**

**T Score of 96% CI = 2.17**

**T Score of 99% CI = 2.79**

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

ANS:

For probability calculations, the number of degrees of freedom is n - 1, so here you need the t-distribution with 17 degrees of freedom.

The probability that t &lt; - 0.471 with 17 degrees of freedom assuming the population mean is true, the t-value is less than the t-value obtained with 17 degrees of freedom and a t score of - 0.471, the probability of the bulbs lasting less than 260 days on average of 0.3218 assuming the mean life of the bulbs is 300 days.