|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete Data |
| Results of rolling a dice | Discrete Data |
| Weight of a person | Continuous Data |
| Weight of Gold | Continuous Data |
| Distance between two places | Continuous Data |
| Length of a leaf | Continuous Data |
| Dog's weight | Continuous Data |
| Blue Color | Discrete Data |
| Number of kids | Discrete Data |
| Number of tickets in Indian railways | Discrete Data |
| Number of times married | Discrete Data |
| Gender (Male or Female) | Discrete Data |

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

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

**Ans:**

Total number of possible combinations = **2**^3 as below

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

P(HHT)+(HTH)+(THH)

=1/8+1/8+1/8

=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:

1. Minimum number in dice is 1 so there is no probability of getting sum=1 while two dices are rolled. Total no. of combinations =6^6=36

P=0/36=0

1. Probability of getting total 4 are (1,3), (2,2), (3,1) = 3 outcomes, 3/36=1/12.
2. Probability of getting total divisible by 6 are (1,5), (5,1), (2,4), (4,2), (3,3), (3,3) = 6 outcomes, 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 No. of balls= 2R,3G,2B= 7 balls

No. of ways picking 2 balls at a time out of 7= (7\*6)/2=21

Total no. of balls is not Blue= 7-2=5

No. of ways picking 2 balls out of 5 non-Blue balls= (5\*4)/2=10

Looking at favorable outcomes 10 out of total 21 outcomes, **probability is 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 randomly selected child = 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 Comment about the values/ Draw some inferences.

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weigh** |
| **Mean** | 3.596563 | 3.21725 | 17.84875 |
| **Median** | 3.695 | 3.325 | 17.71 |
| **Mode** | 3.07 | 3.44 | 17.02 |
| **Variance** | 0.285881 | 0.957379 | 3.193166 |
| **Standard Deviation** | 0.534679 | 0.978457 | 1.786943 |
| **Range** | 2.17 | 3.911 | 8.4 |

**Note:** Mean & Median are almost equal.

**Use Q7.csv file**

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:**

P(X)= 1/9

Expected Value E(X) = ∑ (X)\*P(X)

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

= (1308) \*1/9

=145.33

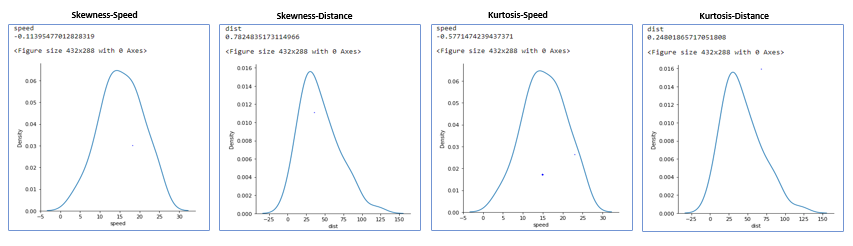
**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data**

**Cars speed and distance**

**Ans:** If the skewness is between -0.5 & 0.5, the data are nearly symmetrical.

If the skewness is between -1 & -0.5 (negative skewed) or between 0.5 & 1(positive skewed), the data are slightly skewed.

If the skewness is lower than -1 (negative skewed) or greater than 1 (positive skewed), the data are extremely skewed.



|  |  |  |
| --- | --- | --- |
|  | **Speed** | **Distance** |
| Mean | 15.4 | 42.98 |
| Median | 15 | 36 |
| Mode | 20 | 26 |

We can see that Mode<Median<Mean for Distance, so it is showing positive skewness/Kurtosis while Skewness and Kurtosis for Speed is bit negative.

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



**Ans:**

The most of the data points are concentrated in the range 50-100 with frequency 200 while least range of weight is between 300-400 somewhere around 0-10.

So, the expected value the above distribution is 75. Also, observed Positive Skewness.



**Ans:** Median is less than mean right skewed and we have outlier on the upper

side of box plot and there is less data points between Q1 and bottom point.

**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:**

CL @94%=Alpha at 94% = (1-94/100) = 0.06

Z(alpha/2) or Critical value=0.0300

Looking at Z Table **Confidence interval for 94% is 1.882**

CL @98%=Alpha at 98% = (1-98/100) = 0.02

Z(alpha/2) or Critical value=0.0100

Looking at Z Table **Confidence interval for 98% is 2.32**

CL @96%=Alpha at 96% = (1-96/100) = 0.04

Z(alpha/2) or Critical value=0.0200

Looking at Z Table **Confidence interval for 96% is 2.05**

**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 | 24.11111 |
| Std Dev | 4.910307 |

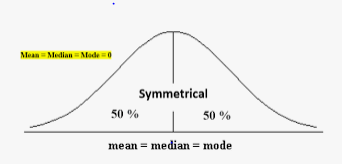
1. What can we say about the student marks?

**Ans:** Majority of the students are falling above the mean, shows Right skewed.

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

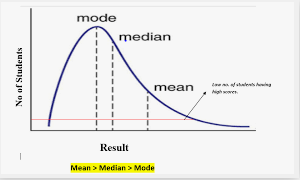
**Ans:**

It would be symmetrical.



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

**Ans:** It will show as Right Skewed.



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

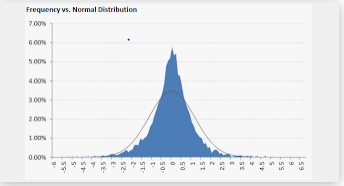
**Ans:** It will show Left Skewed.

Diagram

Description automatically generated with medium confidence

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

**Ans:** Positive values of kurtosis indicate that distribution is peaked and possesses thick tails. Data is distributed normally, and Kurtosis value is “0”.



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

**Ans:** A negative kurtosis means that your distribution is flatter than a normal curve with the same mean and standard deviation

Diagram

Description automatically generated

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



What can we say about the distribution of the data?

**Ans:** 50% of the volume is above 10 unit of the objects & 40% of the volume is between 15-20.

What is nature of skewness of the data?

**Ans:** It is Left Skewed as Median>Mean.

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

**Ans:** IQR= Q3-Q1=10-18=-8

Q19) Comment on the below Boxplot visualizations?



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

**Ans:** Mean & Median are same for both the Box Plot, hence distribution is symmetrical. Comparing to Boxplot 1, whiskers are showing higher side in Boxplot 2.

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

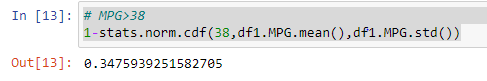
Data \_set: Cars.csv

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

MPG <- Cars$MPG

* 1. P(MPG>38)

**Ans:**



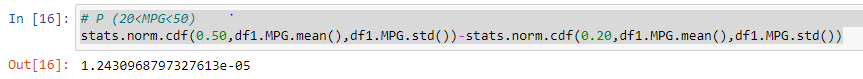
* 1. P(MPG<40)

**Ans:**

Graphical user interface, text, application, Word

Description automatically generated

* 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:** As skewness of the MPG is showing symmetrical, so data is Normal Data Set.

Chart, line chart

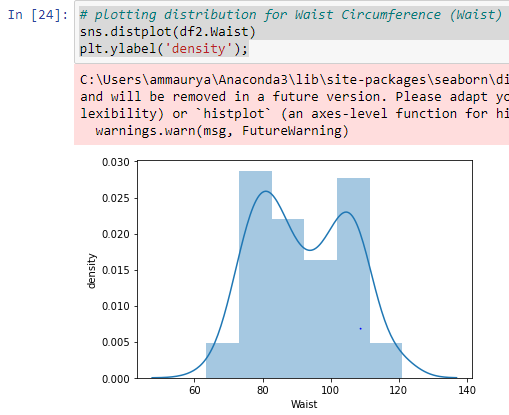
Description automatically generated

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **HP** | **MPG** | **VOL** | **SP** | **WT** |
| Mean | 117.4691 | 34.42208 | 98.76543 | 121.5403 | 32.41258 |
| Median | 100 | 35.15273 | 101 | 118.2087 | 32.73452 |

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

Dataset: wc-at.csv

**Ans:** More than 68% of data set lying within symmetrical part, dataset is following normal distribution.



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

**Ans:**

|  |
| --- |
| CL @90%=Alpha at 90% = (1-90/100) = 0.10 |
| Z(alpha/2) or Critical value=0.0500 |
| Looking at Z Table **Confidence interval for 90% is 1.645** |

CL @94%=Alpha at 94% = (1-94/100) = 0.06

Z(alpha/2) or Critical value=0.0300

Looking at Z Table **Confidence interval for 94% is 1.882**

|  |
| --- |
| CL @60%=Alpha at 60% = (1-60/100) = 0.40 |
| Z(alpha/2) or Critical value=0.2000 |
| Looking at Z Table **Confidence interval for 60% is 2.005** |

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

**Ans:**

n=25, df(degree of freedom)=n-1=24

following t\_table;

t\_value at 95% confidence level=2.064

t\_value at 96% confidence level=2.172

t\_value at 99% confidence level=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

**Ans:**

**t-scores at x=260; t=(s\_mean-P\_mean)/(s\_SD/sqrt(n))**

**(260-270)/{(90/18)\*0.5)}=0.4714**

Graphical user interface, text, application, chat or text message

Description automatically generated