|  |  |
| --- | --- |
| 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Interval |
| 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 | Interval |
| Years of Education | Ratio |

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

Ans: P (Two heads and one tail) = N (Two heads and one tail)/N(Three coins tossed) = 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: Number of possible outcomes for the above event is:

N (Two dice rolled) = 36

1. P (Sum is equal to 1) = 0
2. P (Sum is less than or equal to 4) = 6/36 =0.166
3. P (Sum is divisible by 2 and 3) = 6/36 = 0.166

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 = 7

Number of balls hat are not blue = 5 (2red+3green)

P (None of the balls drawn is blue) = 5C2/7C2 = 10/21= 0.476

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 = 0.015+0.20+0.65+0.005+0.01+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>

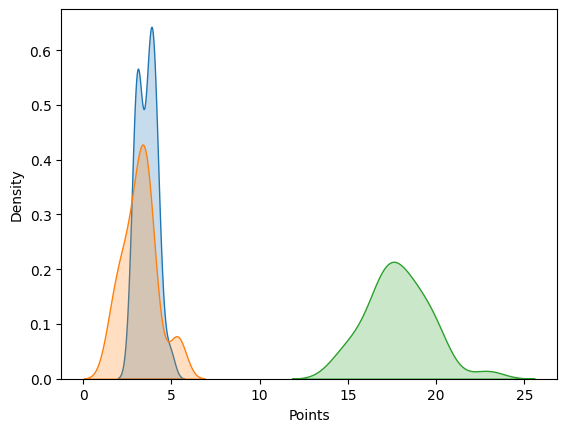
Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Ans:

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mean | 3.59 | 3.22 | 17.85 |
| Median | 3.69 | 3.33 | 17.71 |
| Mode | 3.07 | 3.44 | 17.02 |
| Variance | 0.29 | 0.96 | 3.19 |
| Std\_dev | 0.53 | 0.98 | 1.79 |
| Min | 2.76 | 1.51 | 14.50 |
| max | 4.93 | 5.42 | 22.90 |

Inferences:

Points and score variables have very close mean, median and mode values but score clearly has large variance and standard deviation. We can see score has larger range than points. In Weigh variable, mode is much smaller than mean and median thus the distribution of Weigh must be little positively-skewed.



You can clearly see the distribution of Points, Score and weigh. Points and Score are close together but Weigh has different range.

**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: Expected value = Sum (X\*Probability of X) = (1/9) (108) + (1/9) (110) + (1/9) (123) + (1/9) (134) + (1/9) (145) + (1/9) (167) + (1/9) (187) + (1/9) (199)

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans:

|  |  |  |
| --- | --- | --- |
|  | skewness | kurtosis |
| Car’s speed | -0.117 | -0.508 |
| Car’s distance | 0.806 | 0.405 |

Inferences:

Variable speed is left skewed and distance is right skewed. Negative kurtosis of speed shows that the data is flatten and positive kurtosis of distance shows that the distribution has sharp peak.

**SP and Weight (WT)**

**Use Q9\_b.csv**

Ans:

|  |  |  |
| --- | --- | --- |
|  | skewness | kurtosis |
| SP | 1.61 | 2.97 |
| Weight (WT) | -0.61 | 0.95 |

Inferences:

SP is right skewed and Weight is left skewed. Both the variable has positive kurtosis, so they have sharp peak as compare to the normal distribution. Comparing Weight and SP, SP has sharper peak than Weight.

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



Ans :



Ans:

* Data is right skewed. We have more data in the right side.
* Mode < Median < Mean
* There are outliers present.

**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: Total population = 3000000 Sample population = 2000 men

Sample mean = 200 pounds Sample std dev = 30 pounds

|  |  |
| --- | --- |
| Percentage | Confidence interval |
| 94% | (198.74, 201.26) |
| 96% | (198.62, 201.38) |
| 98% | (198.44, 201.56) |

**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 =24.11

Standard deviation = 4.91

1. What can we say about the student marks?

Ans: Mean is slightly greater than median, so data is slightly right skewed. We don’t have outliers.

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

Ans: No skewness.

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

Ans: Right skewed.

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

Ans: left skewed.

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

Ans: Sharp peak.

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

Ans: flatten or broader curve.

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



What can we say about the distribution of the data?

Ans: Data is not normally distributed. It is skewed towards left.

What is nature of skewness of the data?

Ans: Data is left skewed.

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

Ans: Inter Quartile range (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:

* There are no outliers.
* Both boxplot has same median between 250 and 275.
* IQR of 2nd plot is greater than 1st.
* Both are almost normally distributed.

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)

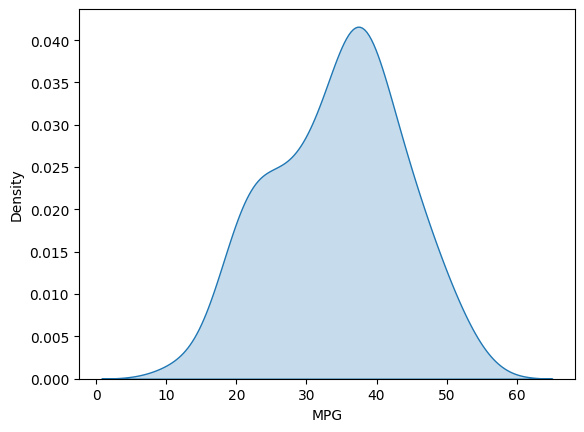
Ans: a) 0.35

1. 0.73
2. 0.89

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

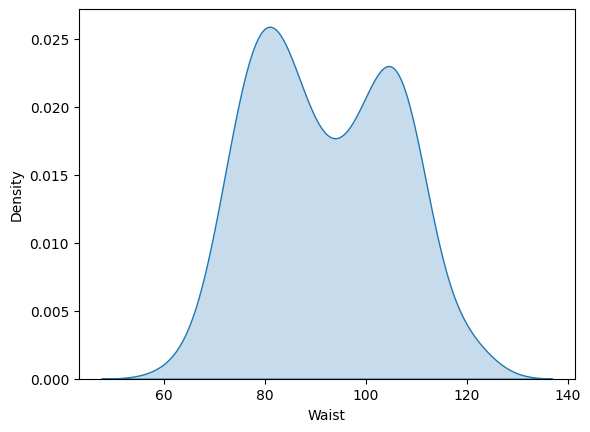
Dataset: Cars.csv

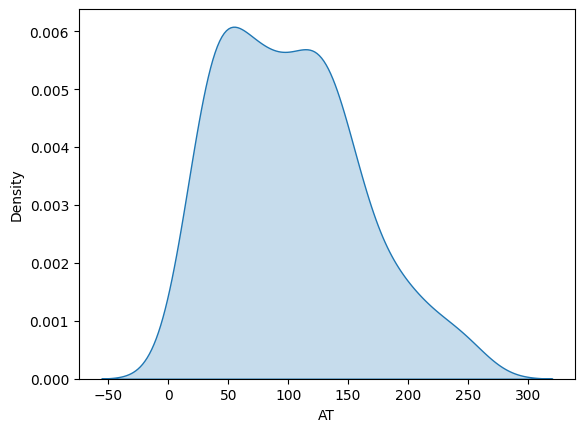
Ans: Yes, MPG of Cars follows Normal Distribution.

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

Dataset: wc-at.csv

Ans: Both AT and Waist doesn’t follow normal distribution.





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

Ans:

|  |  |
| --- | --- |
| Confidence | Z scores |
| 90% | -1.64 |
| 94% | -1.88 |
| 60% | -0.84 |

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

Ans:

|  |  |
| --- | --- |
| Confidence | T scores |
| 95% | -2.06 |
| 96% | -2.17 |
| 99% | -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: 0.32