|  |  |
| --- | --- |
| 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 | Ordinal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Ordinal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

Ans: When three coins are tossed, the number of outcomes is 8 that is

S= {HHH, HHT, HTH, HTT, TTT, TTH, THT, THH}.

So, total number of possible outcomes = 8

Now, the possible outcomes to get two heads and one tail is

{HHT, HTH, THH} =3.

So, no of favorable outcome = 3

The probability of getting two heads and one tail =

The probability of getting two heads and one tail = 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: Two dice are rolled, total possible outcome =.

1. Favorable outcome (sum equal to 1) is 0 i.e. not possible that sum always exceed to 1.
2. Favorable outcome (sum is less than equal to 4) is 6 i.e. (1,1), (1,2), (1,3), (2,1), (2,2), (3,1).
3. Favorable outcomes (sum is divisible by 2 and 3) is 6 i.e.

(1, 5), (2 ,4), (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: **-** Total number of balls = 2 + 3 + 2 = 7

Let S be the sample space.

Then, n(S) = Number of ways of drawing 2 balls out of 7

= 7C2

=

= 21

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

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

=5C2

=

= 10

⸫ P(E) = = = 0.4762

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

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

= 3.09

Expected number of candies for randomly selected child is 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:** for points:

Mean = 3.5965, Median = 3.965, Mode = 3.07, Variance = 0.2858,

Standard deviation = 0.5346, Range = 2.17

The dataset is skewed to the left, meaning that most of the values are

clustered around the higher end of the range, and there are some

outliers on the lower end. This is because the mean is lower than the

median which indicates that the distribution is not symmetrical.

The dataset has a moderate variability, meaning that the values are

not to spread out or too close together. This is because the

standard deviation is 0.5346, which is about 14.9% of the mean.

A rule of thumb is that a standard deviation that is less than 15%

of the mean indicates a low variability, while a standard deviation

that is more than 30% of the mean indicates a high variability.

For score:

Mean = 3.2172, Median = 3.325, Mode = 3.44, Variance=0.9573,

Standard deviation = 0.9784, Range= 3.911

The dataset is skewed to the left, meaning that most of the values are

clustered around the higher end of the range, and there are some

outliers on the lower end. This is because the mean is lower than the

median which indicates that the distribution is not symmetrical.

The dataset has a moderate variability, meaning that the values are

not to spread out. This is because the standard deviation is 0.9784,

which is about 29% of the mean. A rule of thumb is that a standard

deviation that is less than 15% of the mean indicates a low variability,

while a standard deviation that is more than 30% of the mean

indicates a high variability.

For weigh:

Mean= 17.8487, Median = 17.71, Mode= 17.02, Variance = 3.1931,

Standard deviation = 1.7869, Range = 8.4

The dataset is skewed to the right, meaning that most of the values are

clustered around the lower end of the range, and there are some

outliers on the upper end. This is because the mean is greater than the

median which indicates that the distribution is not symmetrical.

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: There are 9 patients in the clinic, so probability of selecting each patient

is 1/9.

Expected value =

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | 108 | 110 | 123 | 134 | 135 | 145 | 167 | 187 | 199 |
|  | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 | 1/9 |

The expected value of the weight of the patient

= 1/9\*108+ 1/9\*110+ 1/9\*123+ 1/9\* 134+ 1/9\* 135+ 1/9\*145+ 1/9\*167+

1/9\*187+ 1/9\* 199

= 145.33

Expected value of the weight of patient is 145.33 pounds.

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight (WT)**

**Use Q9\_b.csv**

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



Ans: According to histogram the is positively skewed. It means outliers are at upper extreme. Maximum data lie between 50 to 100 approximately and mean is also in between 50 to 100.

According to boxplot the data is concentrated downward and Q2-Q1 value is less than Q3-Q2 which shows positive skewness and has outliers above Q4.

**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: Confidence interval for 94% is in range 198.7383 to 201.2617

Confidence interval for 98% is in range 198.4394 to 201.5605

Confidence interval for 96% is in range 198.6223 to 201.3777

**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 = 25.529411

Standard deviation = 5.052663

1. What can we say about the student marks?

Ans: we don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

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

Ans: When the mean, median of data are equal then the distribution is symmetric and the skewness of the distribution is zero that is normalized skewness.

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

Ans: When mean > median there is right skewness.

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

Ans: when median > mean then there is left skewness.

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

Ans: Positive value of kurtosis indicates sharp peak in plot, less gap between tails to x- axis.

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

Ans: Negative kurtosis value indicate broader peak under the curves and more gap between tails and x-axis.

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



What can we say about the distribution of the data?

Ans: the data is assigned in de-assigned format.

What is nature of skewness of the data?

Ans: left skewness

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

Ans: 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: First there are no outliers. Second both the box plot shares the same median that is approximately in a range between 275 to 250 and they are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

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)

Ans: p(MPG>38) = 0.3476

P(MPG<40) = 0.7293

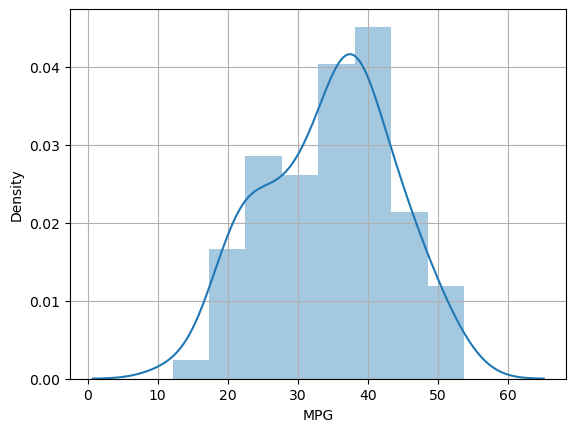
P(20<MPG<50) = 1.2430

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

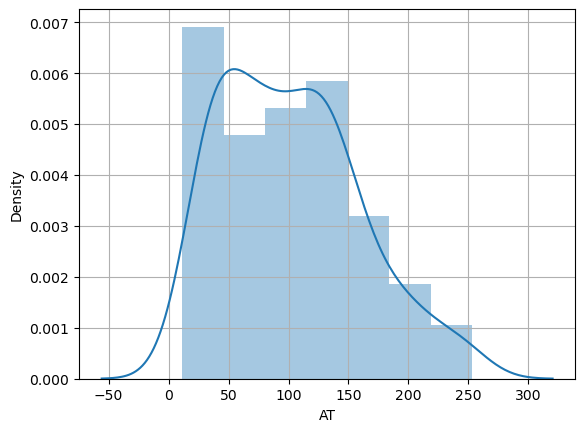
Ans: MPG of Cars follows Normal Distribution.

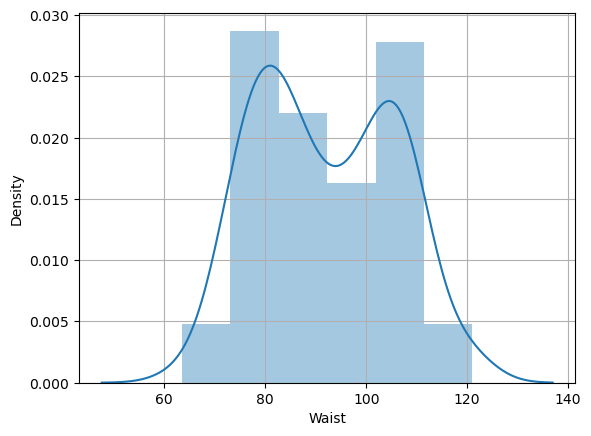


1. Check Whether the Adipose Tissue (AT) and Waist Circumference (Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Ans: Adipose Tissue (AT) and Waist circumference does not follow normal distribution.





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

Ans: Z-score of 90% confidence interval is 1.6448

Z-score of 94% confidence interval is 1.8807

Z-score of 60% confidence interval is 0.8416

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

Ans: for sample size of 25 t score of

a) 95% confidence interval is 2.0638

b) 96% confidence interval is 2.1715

c) 99% confidence interval is 2.7969

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: The probability that 18 randomly selected bulbs would have an average life of no more than 260 days is 32%.