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

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

ANS: the Probability of getting two heads and one tails in the toss of three coins simultaneously is defined as P (Two heads and One Tail) = Number of desired outcomes is 3/8 or 0.375.

Q4) Two Dice are rolled, find the probability that sum is

1. Equal to 1

ANS: n(s) = 36, n(E)=0, P(E) = n(E) / n(s), Probability of get sum of 1= 0/36 = 0.

1. Less than or equal to 4

ANS: (1,3),(2,2),(3,1),(1,1),(1,2),(2,1) = 6 Outcomes,

= 6/36 i.e. 1/6

1. Sum is divisible by 2 and 3

ANS: E = {(5,1) (4,2) (3,3) (2,4) (1,5) (6,6)}

P(E) = n(E) / n(S) = 6/36 =1/36

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?

1. ANS: 2/7+3/7+2/7= 12/21

2 balls are already drawn at random so ans = 10/12

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:

= 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

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: Points :- Mean = 3.596, Median = 3.659, Mode = 3.92,

Variance =0.285, Standard deviation =0.534, Range=2.1

Score :- Mean = 3.217, Median = 3.325, Mode = 3.44,

Variance =0.957, Standard deviation =0.978, Range=3.911

Weigh :- Mean = 17.848, Median = 17.710, Mode = 18.90,

Variance =3.197, Standard deviation =1.78, Range=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?

ANS: = Probability of selecting each patient = 1/9

(1/9) X (1308)

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

ANS: Skewness:

Speed= -0.117510

Dist = 0.806895

Kurtosis:

speed = -0.508994

dist = 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

ANS: Skewness: SP= 1.611450

WT= -0.614753

Kurtosis: SP= 2.977329

WT= 0.950291

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



ANS:  **histogram=** The data is right skewed.

**boxplot** - There are outliers at upper side

**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:n=2000,x̄= 200,Std= 30

Confidence Interval Estimate= x̄± z 𝑠/√𝑛 => 200 ± Z 30√2000

94% Confidence: 200 ± 1.88\* 30√2000 =198.74 -201.26

96% confidence: 200 ± 2.05\* 30√2000 = 198.62-201.38

98% Confidence: 200 ± 2.33\* 30√2000 =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=41.0, median=40.5, variance=24.111, std=4.91

1. What can we say about the student marks?

ANS: Mean is greater than Median

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

ANS: It is perfectly symmetrical.

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

ANS: It is right skewed

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

ANS: It is left skewed.

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

ANS: distribution is peaked and possess thick tails

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

ANS: distribution is flat and has thin tails

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



What can we say about the distribution of the data?

ANS: We can say that there is a big difference between upper quartile and upper extreme compared to the lower quartile and lower extreme.

What is nature of skewness of the data?

ANS: Left Skewed

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

ANS: IQR=Q3-Q1 = IQR=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: Both are having normal distribution

Mean and median are equal so there is no skewness

Outliers are not found in this plot

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

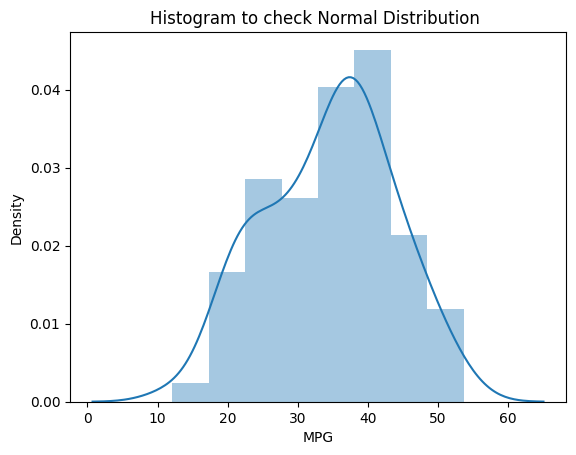
1. P(MPG>38) = 0.347
2. P(MPG<40) = 0.729
3. P (20<MPG<50) = 0.013

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

ANS :

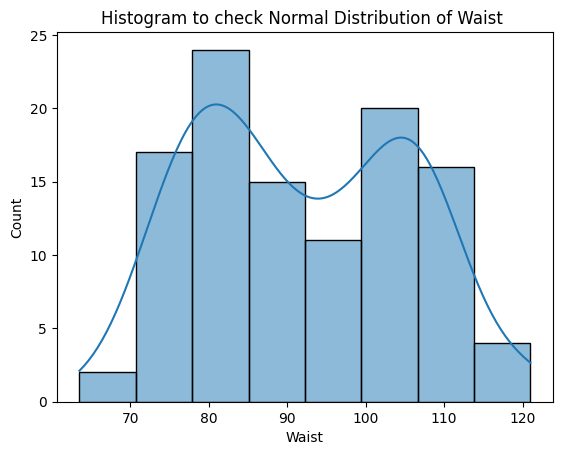
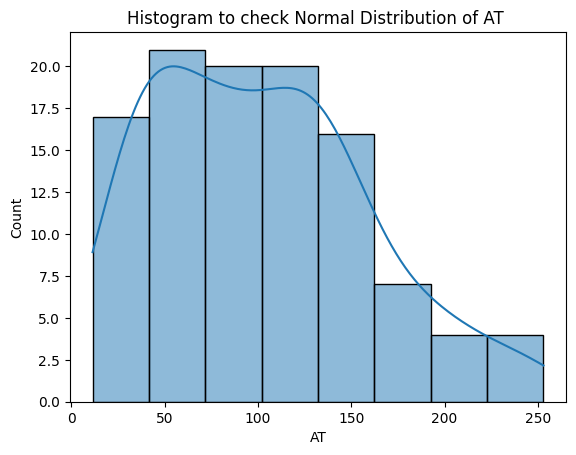


=>Here the data follows normal distribution

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

Dataset: wc-at.csv

ANS :



=> waist and AT follow Normal Distribution

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

ANS: Z scores of:

90% = 1.6448536269514722

94% = 1.8807936081512509

60% = 0.8416212335729143

Q 23) Calculate the t scores of 95% confidence interval, 96% confidence interval,

99% confidence interval for sample size of 25

ANS: T-Score:

95% = 2.0638985616280205

96% = 2.1715446760080677

99% = 2.796939504772804

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 à pt(tscore,df)

df à degrees of freedom

ANS:

df=n-1=18-1= 17

t score= -0.4714

probability = t.cdf(t\_score, df)

Required probability = 0.3216 = 32%