|  |  |
| --- | --- |
| 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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Nominal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Ordinal |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
|  |  |

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

Answer :-

## When three coins are tossed together, the total number of outcomes =8

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

Let E be the event of getting exactly two heads

Then,no.of favourable events,n(E)=(HHH,HHT,THH)

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

Answer :-

1. Equal to 1

## If two dices were rolled, then total possible cases =36

Total Favourable cases (Having sum =1) = 0

As minimum sum is 2 for outcome (1,1).

Hence, probability is 0.

1. Less than or equal to 4

## When two dice are rolled, the total number of outcomes =36

Total possible outcomes when rolling two dices: 6faces on first die \* 6 faces on second die=36

Number of Favourable outcomes =3 +3(2,3 and 4)=6

1. Sum is divisible by 2 and 3

When two dice are rolled, the total number of outcomes =36

Number of Favourable outcome is 3+3=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?

Answer :-

Total number of balls =2+3+2=7

Number of not blue balls=2red+3green=5

So,probability of non blue balls is 5/7

Now find chance of picking non blue ball on second draw but this time considering that first was not blue.After taking 1 ball out of them remainings are 6 in bag now in bag there are 4 balls that are non blue so the chance of this is 4/6 which simplies 2/3

To get overall probability that none of the ball drawn is blue,we have to multiply the both probability(First draw and second draw)

Probability = Probability of First draw \* Probability of Second Draw

Probability = (5/7)\*(2/3)

Probability = 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

Answer :-The Expected number of candies for a 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

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**

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?

Answer:- Expected Value of X

S(108, 110, 123, 134, 135, 145, 167, 187, 199)

n(S) = 9

P(X) = n(X)/n(S)

P(X) = 1308/9

P(X) = 145.333333

**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**



Answer:-

10-a

(1) The distribution of data is right skewed or positive skewed and most of the data lie on the right side

(2) Majority of chick are having weight less than 200

(3) The most of the data points are connected in the range 50-100 and lest range in 0-10 are in 400

(4) The man in the case of distribution will be greater than the median

10-b

1. The above boxplot we can see that are outliers beyond the upper extreme.
2. The boxplot is skewed right and most of the data is concentrade in the lower quartike

**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?

Answer:-

X+/-(Z1- α. σ/sqrt(n)

Degrees of freedom= 2000-1= 1999

Confidence interval= 94%

(1- σ/2)= 1-0.03) =0.97

for confidene interval for 94% is 1.882

Confidence interval for 98%= 2.33

Confidence interval for 96% = 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.
2. What can we say about the student marks?

Answer:-

1)

|  |  |
| --- | --- |
| mean | 41.00 |
| median | 40.50 |
| variance | 25.5294117647059 |
| Standard deviation | 5.05266382858645 |

1. The average marks obtained by student in test is 41 and we can say that student is maintaining their marks around 41 as deviation is very less

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

Answer :- Symmetric Distribution

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

Answer :- Right Skewed Distribution

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

Answer :- Left Skewed Distribution

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

Answer :-A distribution with a positive kurtosis value indicates that the distribution has heavier tails than the normal distribution.

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

Answer:- A distribution with a negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



a)What can we say about the distribution of the data.s

b)What is nature of skewness of the data?

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

Ans:-a) We can say that the distribution of data is not symmetrical. In other words, most of the data lies on the right side or in the upper quartile.

1. -Nature: Data is left skewed or negative skewed
2. Approximately =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. There are no outliers in both boxplots
2. Both boxplots are uniform distribution
3. The range of data of boxplot 1 is smaller than the range of data of boxplot 2.
4. The IQR for boxplot 1 is lower than the IQR for 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)
  2. P(MPG<40)

c. P (20<MPG<50)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

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

Dataset: wc-at.csv

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

Ans:-

Z score of 90% confidence interval is 1.644854

Z score of 94% confidence interval is 1.880794

Z score of 60% confidence interval is 0.8416212

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

Ans:-

t scores of 95% confidence interval is 1.959964

t scores of 96% confidence interval is 2.053749

t scores of 25% confidence interval is 0.3186394

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

Answer:-

Mean = 270 days = u

Sample size = 18 = n

Smaple mean = 260 = x

Deviation sample = 90 days =s

Sol:-

IMG_256

t = 260-270/90/sqrt18

t = -10/90/3^2

t = -10/30/2^

t = -1\*2^/3

t = 0.417