|  |  |
| --- | --- |
| 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) | Ratio |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ratio |
| Time on a Clock with Hands | Interval |
| Number of Children | Nominal |
| 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-

when two coins are tossed the total number of outcomes is 2^3=8.

Possible combinaton is{ HHH,HHT,HTH,THH,TTH,THT,HTT,TTT.}

There are three possible ways of two head and one tail{HHT,THH,HTH}=3

The probability is 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 combination for the above event is =6^2=36.

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

Total favourable cases = 0

As minimum sum is 2. Hence, probability is **0.**

1. sum of Less than or equal to 4

Total no. of outcomes = 36

Expected outcomes = 6

Probability = 6/36 = **1/6=0.16=16.66%.**

1. Sum is divisible by 2 and 3

Total no. of outcomes = 36

Expected outcomes = 6

Probability = 6/36 = **1/6=0.16=16.66%.**

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 = 7

Let S be the sample space

Then, n(S)= No. of ways of drawing balls out of 7

n(S) = 7\*6/2\*1

n(S) = 21

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

n(E) = No. of ways of drawing 2 balls out of (2+5) balls

n(E) = 5\*4/2\*1

n(E) = 10

Hence probability = 10/21 = **0.4761**

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

Expected no. of candies for a randomly selected child = **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?

Ans:

Expected value=sum(X\*Propbabilty 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)+ (1/9)(135)=**145.33**

**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**: In given histogram peak has right skew and tail is an right, mean>median. It have outliers on the higher side.

**Ans**: In given boxplot median is less than mean, and there are some outliers on the uperside of the boxplot.



**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:

**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?

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

**Ans**: If the mean and median of data are equal, this means that the distribution is symmetric. And such distribution has zero skewness.

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

**Ans**: If mean is greater than median, then the nature will be positively skewed.

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

**Ans**: If median is greater than mean, then the nature will be negatively skewed.

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

**Ans**: Positive values of kurtosis indicates that distribution is peaked and has thick tail.

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

**Ans**: Negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

**Ans**: The above boxplot is following normal distribution, the median is towards higher value.

What is nature of skewness of the data?

**Ans**: The data is left skewed, the whisker range of minimum value is greater than maximum value.

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

**Ans**: IQR=Q3-Q1(Upper quantile – Lower quantile)

IQR=18-10

**IQR=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 the boxplot has no outliers.
* Both the boxplots are normally distributed, hence no skewness can be seen.
* Both the boxplots shares the same median.

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

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

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:** import numpy as np Import scipy as stats t\_score = (x - pop mean) / (sample standard daviation / square root of sample size) (260-270)/90/np.sqrt(18)) t\_score = -0.471 stats.t.cdf(t\_score, df = 17) 0.32 = 32%