|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Integer |
| Results of rolling a dice | Integer |
| Weight of a person | Float |
| Weight of Gold | Float |
| Distance between two places | Float |
| Length of a leaf | Float |
| Dog's weight | Float |
| Blue Color | String |
| Number of kids | Integer |
| Number of tickets in Indian railways | Integer |
| Number of times married | Integer |
| Gender (Male or Female) | String |

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 | Ordinal |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Ordinal |
| Years of Education | Interval |

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

**ANS: -**

Total Number of probabilities: 8 (HHH, HHT, HTH, THH, TTH, THT, HTT, TTT.)

Number of chances that two head and one tail is obtain: 3(HHT, THH, HTH)

The Probability of getting two heads and one tail is 3/8 or 0.375.

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

1. **Equal to 1**: - The probability of getting a sum of 1 is 0 because its impossible to getting a sum as 1 if we have numbers between 1 to 6.

1. **Less than or equal to 4: -** Total number of outcomes = 6^2 = 36

Number less than or equal to 4 = 6 {(1,2), (2,1), (1,3), (3,1), (2,2), (1,1)} = 6/36 = 1/6

1. Sum is divisible by 2 and 3: -

6 = {(6, 6), (6, 5), (5, 6), (4, 6), (6, 4), (6, 3), (3, 6)}

Probability of sum divisible by 2 and 3 = 6/36

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?

**ANS: -**  Total number of Balls = 5 (2Red + 3green + 2blue)

Total number of outcomes = 7c2

Total no of outcomes where they are none of the balls drawn is blue 5c2.

Probability = 5c2/7c2 = 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

**ANS: -**

The expected no of candies for a randomly selected child are

E(x)= 3.125

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: - we have used df.describe() function to calculate all the values**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **points** | **Score** | **Weigh** |
| **count** | **32** | **32** | **32** |
| **mean** | **3.596563** | **3.21725** | **17.84875** |
| **std** | **0.534679** | **0.978457** | **1.786943** |
| **min** | **2.76** | **1.513** | **14.5** |
| **25%** | **3.08** | **2.58125** | **16.8925** |
| **50%** | **3.695** | **3.325** | **17.71** |
| **75%** | **3.92** | **3.61** | **18.9** |
| **max** | **4.93** | **5.424** | **22.9** |
| **Variance** | **0.285881** | **0.957379** | **3.193166** |

**Note: - I will attached the ipnyb file for the reference…**

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 of all weights)/(Total number of patients)

Sum of all weights= 108+110+123+134+135+145+167+187+199 = 1308

Total number of patients = 9

Expected value 1308/9 = 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**

**ANS: -**  Answer are available in ipynb file, under Q9.

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



**ANS: -** The histogram is positively skewed, and the graph is about the frequency vs weight. The number of people from 0 to 100 is more than the rest of them 7.

**ANS: -** By looking at the boxplot, we can see that there are some outliers present, we also observe that the median is less than mean as there are less data points between Q1 and below points range, hence we can consider it as the right skewed.



**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:** - Answer is in ipynb file.

**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 Meadian: 40.5 Standard deviation: 4.910306620885412Variance: 24.11111111111111

1. What can we say about the student marks?

**ANS: -** Overall, based on these statistics, we can say that the student's marks are consistent and centered around 40.

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

**ANS**: - When the Mean and Median of the data is equal it indicates the data is symmetrical in that case the nature of the skewness is zero or close to zero.

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

**ANS:** - when the mean is greater than median the data will be positively skewed.

Positively skewed means the data is distributed to the right side.

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

ANS: - when the median is greater than mean the data will be negatively skewed.

Negatively skewed means the data is distributed to the left side.

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

**Ans: -** A positive kurtosis value indicates that the peak of the probability distribution of a dataset is sharper (more peaked) and the tails are heavier (more outliers) than the normal distribution.

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

**Ans:** - A negative kurtosis value indicates that the peak of the probability distribution of a dataset is flatter, and the tails are lighter (fewer outliers) than the normal distribution.

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



1. What can we say about the distribution of data?

Ans: - As we can see from the box plot visualization, we can easily say that the data is distributed between 10 and 18 most of the data is right side of box plot graph so it negatively skewed.

1. What is the nature of skewness of the data?

Ans: - In the boxplot, most of the data was lying in the range of 10-18, hence it will be considered as the left skewed, as the data values in upper quadrant contains more than the low quadrant.

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

Ans: - IQR (Inter quartile range) = Q3 - Q1 = 18.3 - 10.3 = 8 approx

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

ANS: - In the above given box plots, we can observe that mean and median are equal, even though the whisker level is high in box plot 2, hence we can consider the distribution is symmetrical.

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 🡪 pt(tscore,df)

df 🡪 degrees of freedom