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

**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 | Ordinal |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

* Sample Space = [HHH, THH, HTH, HHT, TTH, HTT, THT, TTT ]

P(Two heads and one tail) = 3/8

**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**

* Sample Space = [(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),

(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),

(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),

(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),

(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),

(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)]

1. P(Equal to 1) = 0/36
2. P(less than or equal to 4) = 6/36 = 1/6
3. P(sum is divisible by 2 and 3 = 24/36 = 3/4

**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?**

* Total number of balls = 2 (red) + 3 (green) + 2 (blue) = 7 balls

Probability of the first ball is not blue = 5/7

Probability of the second ball is not blue = 4/6

Probability of both events happening = (5/7) \* (4/6)

= 20/42 = 0.476

So, the probability that none of the balls drawn is blue is 47.6%

**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**

* Expected value = ∑ probability \* candies

= (1\*0.015) + (4\*0.20) + (3\*0.65) + (5\*0.005) + (6\*0.01)+(2\*0.120)

= 3.09

The Expected number of candies for a 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**

* solution in Basic statistics 1.ipynb (I have attached python file to email)

**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?**

* Expected value = ∑ probability \* weights

= (1/9) \* [108+110+123+134+135+145+167+187+199)

= 1308/9

The Expected Value of the Weight of that patient is 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**

* solution in Basic statistics 1.ipynb (I have attached python file to email)

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



* The peak and tail of the histogram are skewed to the right. Mean > Median. On the higher side, there are outliers.



* Based on the above box plot, it is apparent that outliers are present above the upper whisker. Additionally, the mean is greater than the median, as indicated by the median line's positioning toward the lower quartile of the data.

**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?**

* For 94% confidence interval Range is [ 198.73 – 201.26]

For 98% confidence interval range is [198.43 – 201.56]

For 96% confidence interval range is [198.62 – 201.37]

* code in Basic statistics 1.ipynb (I have attached python file to email)

**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?**

* Mean =41, Median =40.5, Variance =25.52 and Standard Deviation =5.05

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

* solution in Basic statistics 1.ipynb (I have attached python file to email)

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

* The distribution has zero skewness and the mean and median are equal in a symmetric distribution.

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

* Positive skewness is present in the distribution when the mean is greater then median.

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

* Negative skewness is present in the distribution when the median is greater then mean

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

* A distribution with thick tails and a peak is indicated by positive kurtosis values.

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

* When a distribution has a lower tail than the normal distribution, it is said to have a negative kurtosis value.

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



**What can we say about the distribution of the data?**

* The above Boxplot is not normally distributed the median is towards the higher value

**What is nature of skewness of the data?**

* The data is a skewed towards left. The whisker range of minimum value is greater than maximum

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

* Q1 = 10

Q5 = 18

Inter Quartile Range = (Q3 – Q1)

=(18 - 10 )

=8(approx)

**Q19) Comment on the below Boxplot visualizations?**



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

* Here when we compare box plot 1 with box plot 2 we can say that the data in boxplot 1 is widely spread. Here the main inference is that since the data range varieshigh in box plot 2 it is hard to make a prediction in box plot 2. The median in the 2box plots are equal. And the data spread in both of them are 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)**
  3. **P (20<MPG<50)**

=>a. P(MPG>38)= 0.348

b. P(MPG<40)= 0.729

c. P(20<MPG<50)= 0.899

code in Basic statistics 1.ipynb (I have attached python file to email)

**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**

* solution in Basic statistics 1.ipynb (I have attached python file to email)

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

=>Z score of 90% confidence interval= 1.645

Z score of 94% confidence interval= 1.880

Z score of 60% confidence interval= 0.842

code in Basic statistics 1.ipynb (I have attached python file to email)

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

=>t score of 95% confidence interval= 2.064

t score of 96% confidence interval= 2.172

t score of 99% confidence interval= 2.797

code in Basic statistics 1.ipynb (I have attached python file to email)

**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**

* The probability of getting an average life of no more than 260 days are 32.181

code in Basic statistics 1.ipynb (I have attached python file to email)