|  |  |
| --- | --- |
| 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) | **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?

Solution:

S **= { HHH, HHT, HTH, THH, HTT, THT, TTH, TTT }**

**Probability = 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

Solution : **When two dice are rolled, sample space is given as:**

**{(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)}**

**Total number of possible outcomes = 6 x 6 = 36**

1. **Equal to 1 = When the sum is equal to '1' the Probability will be Zero. P(E) = n(E) / n(S) = 0 / 36 = 0**
2. **Less than or equal to 4** = {(1,1) , (1,2), (1,3), (2,1), (2,2), (3,1)}

= **6/36 = 0.166**

**Favorable outcomes = sum is divisible by 2 and 3**

**Sum should be divisible by both 2 and 3**

**Favorable outcomes = (1,5) , (3,3) , (4,2) , (5,1) , (6,6)**

**Number of favorable outcomes = 5**

**Thus the probability that sum is divisible by 2 and 3 is 5/36 or 0.1388**

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?

Solution :

**2 RED: R1 R2**

**3 GREEN: G1 G2 G3**

**2 BLUE: B1 B2**

**Possible ways for (R1 , R2 , G1 , G2 , G3, B1 , B2 ) are (6, 5 ,4, 3, 2, 1, 0)**

**So the total outcomes will be (6+5+4+3+2+1) = 21**

**Probability that none of the balls drawn is blue is (R1 , R2 , G1 , G2 , G3) is (4,3,2,1) = (4+3+2+1) = 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

Solution: **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 \*0.12**

**= 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,Weight>

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

Solution:

****

**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?

Solution:

**Mean = (108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199)/9 = 145.333.**

**It is not mandatory for the expected value to be present in sample space. As calculated above Expected Value of 145.33 is not present in our sample space**

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

Cars speed and distance

Use Q9\_a.csv

Solution:

|  |  |  |
| --- | --- | --- |
|  | **Speed** | **Distance** |
| **Skewness** | **-0.113** | **0.782** |
| **Kurtosis** | **2.422** | **3.248** |

**For Speed the data concentration is on right side where as for Distance data concentration is on left side.**

**Kurtoses for Speed & Distance are positive means there is higher peaks.**

SP and Weight(WT)

Use Q9\_b.csv

|  |  |  |
| --- | --- | --- |
|  | **SP** | **WT** |
| **Skewness** | **1.58** | **-0.60** |
| **Kurtosis** | **5.72** | **3.81** |

**For Speed the data concentration is on left side where as for Weight data concentration is on Right side.**

**Kurtoses for Speed & Weight are positive means there is higher peaks.**

Q10) Draw inferences about the following boxplot & histogram



**Histogram:-**

1. **Chick weight data is right skewed or positively skewed.**
2. **More than 50% Chick Weight is between 50 to 150.**
3. **Most of the chick weight is between 50 to 100.**



**Boxplot:**

1. **The data is right skewed.**
2. **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?

Solution: **As n(the sample size)>Std deviation of sample (2000 > 30);**

**So the sample is normal and is a simple random sample.**

**Range of Confidence interval = Sample statistic + Margin of Error**

1. **For confidence interval 94%**

* **Compute alpha (α): α = 1 - (confidence level / 100) = 0.06**
* **Find the critical probability (p\*): p\* = 1 - α/2 = 1 - 0.06/2 = 0.97**
* **Find the degrees of freedom (df): df = n - 1 = 2000 - 1 = 1999**
* **The critical value is the t score having 1999 degrees of freedom and a cumulative probability equal to 0.97. From the t distribution we find that the critical value is 1.882.**

**Compute margin of error (ME): ME = critical value \* standard error = 1.88 \* 0.94 = 1.7672**

* **Specify the confidence interval. The range of the confidence interval is defined by the *sample statistic* + *margin of error*. And the uncertainty is denoted by the confidence level. Therefore, this 94% confidence interval is 200 + 1.76.**

**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?

Solution :

|  |  |
| --- | --- |
| **Mean** | **41** |
| **Median** | **40.5** |
| **Mode** | **41** |
| **Variance** | **25.52941** |
| **Std dev** | **5.052664** |

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

Solution : **Skewness = 0, Symmetric**

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

Solution : **Right Skewed**

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

Solution : **Left Skewed**

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

Solution : **Maximum peakedness or Sharp peak, Thick tails**

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

Solution : **Minimum peakedness or Broad peak. Thin tails**

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



What can we say about the distribution of the data?

Solution : **Data is Not Normally distributed.**

What is nature of skewness of the data?

Solution: **Left Skewed**

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

Solution: 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.

**Both are Normally Distributed**

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) = **0.35**
  2. P(MPG<40) = **0.73**
  3. c. P (20<MPG<50) = **0.898**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Solution: **MPG of the cars follows normal distribution**

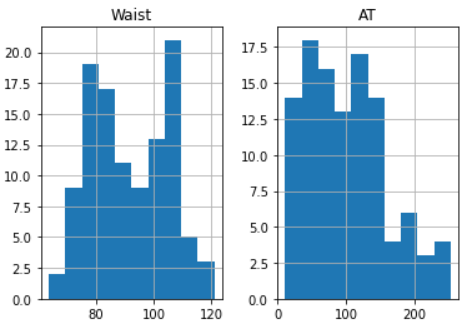
1. Check Whether the Adipose Tissue (AT) and Waist Circumference(Waist) from wc-at data set follows Normal Distribution

Dataset: wc-at.csv

Solution:

**Adipose Tissue (AT) follows Normal Distribution**

**Waist Circumference (Waist) does not follow Normal Distribution**



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

Solution:

* 1. **Z score of 90% Confidence Interval is 1.96**
  2. **Z score of 94% Confidence Interval is 1.88**
  3. **Z score of 60% Confidence Interval is 0.84**

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

Solution:

1. **t score of 95% Confidence Interval is 2.06**
2. **t score of 96% Confidence Interval is 2.17**
3. **t score of 99% Confidence Interval is 2.8**

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

Solution:

**X(bar ) = 260 , Mu = 270 , n = 18 , s = 90t = x(bar) - mu / (s/(root n))t = 260 - 270 / (90/4.24)t = -10 / 21.23t = -0.47Pt(-0.47,17) = 0.322**

**Hence, there is 32.2% probability that average lives of 18 selected bulbs are not more than 260 days.**