|  |  |
| --- | --- |
| 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) | Categorical |

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

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

Ans. = HHT + HTH + THH

= 1/8 + 1/8 + 1/8

= 3/8

=37.5%

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

1. **Equal to 1**

= **0**, As no one of the outcomes will be equal to one.

1. **Less than or equal to 4**

= (1,3), (2,2), (3,1), (1,1), (1,2), (2,1) = 6/36 = 1/6 =**16.66%**

1. **Sum is divisible by 2 and 3**

when we roll two dice there are 6 × 6 = 36 possibilities.

The only numbers divisible by 2 and 3 are 6,12

Where 12 can be made by 1 way (double 6) and 6 by 5 ways (1,5) (5,1 4,2) (2,4) (3,3)

Probability p(A) = The number of favorable outcomes / Total number of possibilities= (1+5)/36=6/36

**=**1/6 = **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 number of balls = (2 + 3 + 2) = 7.

Let s be the sample space.

Then, n(s) = Number of ways of drawing 2 balls out of 7=7C2 = 21

Let e = Event of drawing 2 balls, none of which is blue.

n(e) = Number of ways of drawing 2 balls out of (2 + 3)

balls =5C2 = 10

  Therefore, P(e) = n(e)/n(s)

**= 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 = 1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.12

= 0.015 + 0.80 + 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**

**ANS =** Refer file Q7.ipynb in same folder.

**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 = 108+110+123+134+135+145+167+187+199 /9

= 1308/9

= **145.33**

**Expected value of the weight of the 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**

ANS = Refer file Q9.ipynb in same folder.

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





**ANS =** The histograms peak has right skew and tail is on right. Mean > Median. We have outliers on the higher side.

According to the boxplot, it has outliers on the maximum 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**?

**ANS =** Confidence Interval=Sample Mean±(Critical Value×(Standard Deviation/√sample size))

Where:

* Sample Mean (*x*ˉ) = 200 pounds
* Standard Deviation (*σ*) = 30 pounds
* Sample Size (*n*) = 2,000 men

CI 94%:

α = 1 - (confidence level / 100) = 1- (94%/100) = 0.06

Critical probability (p\*) = 1 - α/2 = 1-0.03 = 0.97

Degrees of freedom =n-1=2000-1=1999

t-score =1.882

* For 94% confidence level:1.88
* For 98% confidence level: 2.33
* For 96% confidence level:2.05
* **For 94% Confidence Interval:**

= 200± (1.88 \* 30/√2000)

= 200± (1.88 \* 0.671)

= 200± (1.262)

**= (198.738,201.262)**

* **For 98% Confidence Interval:**

= 200± (2.33 \* 30/√2000)

= 200± (2.33 \* 0.671)

= 200± (1.563)

**= (198.437,201.563)**

* **For 96% Confidence Interval:**

= 200± (2.05 \* 30/√2000)

= 200± (2.05 \* 0.671)

= 200± (1.374)

**= (198.626,201.374)**

**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?**

**ANS =1)**

Mean= sum of all scores / number of scores

Mean(µ)= 34+36+36+38+38+39+39+40+40+41+41+41+41+42+42+45+49+56

​ 18

**Mean** =738/18 **=41**

**Median**: The median is the middle value when the data is sorted in ascending order.After sorting the data: 34, 36, 36, 38, 38, 39, 39, 40, 40, 41, 41, 41, 41, 42, 42, 45, 49, 56.

The median is the average of the two middle values (since there is an even number of data points), which are 40 and 41.

Therefore, the **median** is 40+41/2= **40.5**

**Variance**:

Variance=

=++………..+/18

=**25.23**

**Standard deviation:**

Standard deviation=

= **=5.022**

**2)What can we say about the student marks?**

Ans= We don’t have outliers and the data is slightly skewed towards right because mean is greater than median.

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

Ans= No skewness is present we have a perfect symmetrical distribution

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

Ans= Skewness and tail is towards Right

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

Ans= Skewness and tail is towards left.

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

Ans=Positive kurtosis means the curve is more peaked and it is Leptokurtic.

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

Ans= Negative Kurtosis means the curve will be flatter and broader.

**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 not normally distributed the median is towards the higher value.

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

Ans= The data is a skewed towards left.

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

Ans= The Inter Quantile Range = Q3 Upper quartile – Q1 Lower Quartile = 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: There are no outliers.

Both the box plot shares the same median.

They are normally distributed with zero to no skewness neither at the minimum or maximum whisker range.

**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)**

**ANS =** Refer file Q20.ipynb in same folder.

**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**

**ANS =** Refer file Q21.ipynb in same folder.

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

**ANS =** Z score of 90% = 1.65

Z score of 94% = 1.88

Z score of 60% = 0.84

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

**Ans=** T score of 95%= 2.064

T score of 94%= 1.974

T score 0f 99%= 2.797

**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**

**Ans=**

Given:

Population mean (claimed by the CEO): μ = 270 days

Sample size: n = 18 bulbs

Sample mean: x̄ = 260 days

Standard deviation of the population: σ = 90 days

First, let's calculate the standard error of the mean (SEM), which is the standard deviation of the sampling distribution of the sample mean:

SEM = σ / √n

Where:

σ is the population standard deviation

n is the sample size

SEM = 90 / √18 ≈ 21.213

Now, let's find the z-score for x̄ = 260 days:

z = (x̄ - μ) / SEM

z = (260 - 270) / 21.213 ≈ -0.471

Now, we find the probability corresponding to this z-score using a standard normal distribution table or calculator. The probability of finding a z-score less than or equal to -0.471 is approximately 0.3208.

Therefore, the probability that 18 randomly selected bulbs would have an average life of no more than 260 days, given the CEO's claim, is approximately 0.3208 or 32.08%.