|  |  |
| --- | --- |
| 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 ( Categorical - Nominal )** |
| Number of kids | **Discrete** |
| Number of tickets in Indian railways | **Discrete** |
| Number of times married | **Discrete** |
| Gender (Male or Female) | **Discrete ( Categorical - Nominal )** |

**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 | **Interval** |
| Time on a Clock with Hands | **Interval** |
| Number of Children | **Ratio** |
| 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 : set ( HHH,HHT,HTH,THH,HTT,THT,TTH,TTT)**

**Desired out put set – HHT,HTH,THH**

**Probability : 3/8**

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

1. Equal to 1 – **Ans 0**
2. Less than or equal to 4 : **Ans 1/6** {( 1,1) (1,2) (1,3) (2,1) (2,2) (3,1)}
3. Sum is divisible by 2 and 3 : **Ans 1/6 {(1,5),(2,4),(3,3),(4,2),(5,1)(6,6)}**

**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 10/21 = 0.47**

**(R1,R2),(R1,G1),(R1,G2), (R1,G3), (R2,G1), (R2,G2), (R2,G3), (G1,G2),(G1,G3),(G2,G3)**

Combination:

Total possibilities = selecting 2 from 7 balls = C27 = 21

Selecting red & green balls = selecting 2 from 2red+3 green balls = C25 = 10

**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 + 0.8  + 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, Weigh>**

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

**Use Q7.csv file**

**Calculated – Refer python code 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: Average = 145.33**

**Calculated – Refer python code file**

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

**Cars speed and distance (Use Q9\_a.csv)**

|  |  |  |
| --- | --- | --- |
|  | **Cars speed** | **distance** |
| **Skewness** | **-0.117510** | **0.806895** |
| **Kurtosis** | **-0.508994** | **0.405053** |

**SP and Weight(WT) (Use Q9\_b.csv)**

|  |  |  |
| --- | --- | --- |
|  | **SP** | **Weight** |
| **Skewness** | **1.611450** | **-0.614753** |
| **Kurtosis** | **2.977329** | **0.950291** |

**Calculated – Refer python code file**

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



This Histogram is having Positive Skewness (Tail to the Right), outliers are on right side. Data is concentrated on right of graph. Average Chick weight is approx. around 120-150 Kg.



box plot tells us about the distribution of data in less space, this box plot shows that data is not symmetrical and outliers are present. Outliers at top side.

**Both Graphs shows the Asymmetric nature & outliers present in 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?

**Ans** : n=2000, N=3,000,000, x bar =200 & S = 30, xbar +/- (z\*S/sqrt n)

Confidence Interval for

94% - (198.738325292158, 201.261674707842)

98% - (198.43943840429978, 201.56056159570022)

96% - (198.62230334813333, 201.37769665186667)

**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: 41.0

Median: 40.5

**Variance:**

When calculated using – numpy lib - 24.11111111111111

When calculated using – pandas lib - 25.529411764705884

**Standard Deviation:**

When calculated using – numpy lib - 4.910306620885412

When calculated using – pandas lib - 5.05266382858645

2. Mean>median. Marks of students are positively skewed (distribution of marks).

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

Ans: Zero Skew / No Skew / Symmetric / Data is concentrated in middle

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

Ans: Positive Skewness / Right Skew / (Tail) Out liars on Right / Data is concentrated at left

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

Ans: Negative Skewness / Left Skew / (Tail) Out liars on Left / Data is concentrated at Right

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

Ans : Kurtosis tell us about the peak, Thin Peak & wider Tails ( less variation)

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

Ans : Kurtosis tell us about the peak, Thick / Wider Peak and Thin Tails (high variation)

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



What can we say about the distribution of the data? Ans:- **Asymmetric -** Skewed to left - Not normal Distribution  
What is nature of skewness of the data? Ans:- Left Skew (Negative Skewness)

What will be the IQR of the data (approximately)? Ans:- 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.

* In this box plots, box plot 1 shows the data is tightly grouped than box plot 2. (i.e., Variations are more for 2)
* Also, the median of both the Box plots are approx. same i.e. 260
* No outliers in both the box plots
* Both box plots are symmetric, no skewness (length of 4 parts equal)

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 v$ MPG

1. P(MPG>38) - 0.3475939251582705 (34.76%)
2. P(MPG<40) - 0.7293498762151616 (72.93%)
3. P (20<MPG<50) - 0.8988689169682046 (89.89%)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

* Follows normal distribution approximately as mean and median are approx. same.

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

Dataset: wc-at.csv

* Waist data is fairly symmetrical & AT data is positively skewed.

- Plots are plotted in python file, kindly refer python file

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

stats.norm.ppf()

Z Scores of 90% CI - stats.norm.ppf(0.95) - 1.6448536269514722

Z Scores of 94% CI - stats.norm.ppf(0.97) - 1.8807936081512509

Z Scores of 60% CI - stats.norm.ppf(0.80) - 0.8416212335729143

stats.norm.interval(CI,0,1)

Z Scores of 90% CI – stats.norm.interval(0.90,0,1)

(-1.6448536269514729,1.6448536269514722)

Z Scores of 94% CI – stats.norm.interval(0.94,0,1)

(-1.8807936081512509, 1.8807936081512509)

Z Scores of 60% CI – stats.norm.interval(0.60,0,1)

(-0.8416212335729142, 0.8416212335729143)

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

By ppf

T Scores of 95% CI - stats.t.ppf(0.975,df=24) = 2.0638985616280205

T Scores of 96% CI - stats.t.ppf(0.98,df=24) = 2.1715446760080677

T Scores of 99% CI - stats.t.ppf(0.995,df=24) = 2.796939504772804

**stats.t.interval(CI,0,1)**

T Scores of 95% CI - stats.t.interval(0.95,25,loc=0,scale=1) =

(-2.059538552753294, 2.059538552753294)

T Scores of 96% CI - stats.t.interval(0.96,25,loc=0,scale=1) =

(-2.1665866344527562, 2.1665866344527562)

T Scores of 99% CI - = stats.t.interval(0.99,25,loc=0,scale=1) = (-2.787435813675851, 2.787435813675851)

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

Ans:

Population Mean = 270 days

Sample size = n= 18 bulbs

Sample mean = 260 days

Sample Std Deviation = 90 days

t(n-1)=(260-270)/(90/sqrt 18) = -0.471

with df = 17

probability = 32.167%

stats.t.cdf(-0.471,df=17)

Note: In mail attached python file a