|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete (because it can’t be divided or it is in fractional form) |
| Results of rolling a dice | Discrete |
| Weight of a person | Continue (it is a quantitative or it is in decimal) |
| Weight of Gold | Continue |
| Distance between two places | Continue |
| Length of a leaf | Continue |
| Dog's weight | Continue |
| Blue Color | Discrete |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Continue |
| 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( doesn’t related to each other) |
| High School Class Ranking | Ordinal(using in feedback, related to each other) |
| Celsius Temperature | Ratio |
| Weight | Ratio |
| Hair Color | Ordinal |
| Socioeconomic Status | Ordinal |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Ordinal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ration |
| Religious Preference | Ordinal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Interval |

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

Ans : 1 coin (s)=(h, t)

For 3 coin=23

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

SO {HHT,HTH,THH} are having 2 head and 1 tail so

P=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 are (1,2,3,4,5,6) = 6 for 1 dice

For 2 dice sample are (1,1)(1,2)(1,3)(1,4)(1,5)(1,6)

. .

. .

. .

. .

(6,1)(6,2)(6,3)(6,4)(6,5)(6,6) = 36

1. p1=0
2. (1,1)(1,2)(1,3)(2,1)(2,2)(3,1) =p2=6/36
3. P3=6/36

=1/6

=0.166

**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 are =2+3+2=7

So there are 2 blue

* p1=7c2 the way of taking 2 balls

p1=21

* e1=5c2  the event of taking 2 balls

e1=10

* P=e1/p1

P=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\*0.120

=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**

Ans:

Mean for points= 3.596, score=3.217, weight=17.848

Median for points= 3.695, score=3.325, weight=17.71

Mode for points= 3.92, score=3.44, weight=17.02

Variance for points= 0.285, score=0.957, weight=3.193

Standard deviation for points= 0.534, score=0.978, weight=1.786

Range for points= 2.17, score=3.911, weight=8.4

**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

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

**Cars speed and distance**

**Use Q9\_a.csv**

Ans : speed for skewness is -0.12 and for kurtosis is -0.51

Distance for skewness is 0.81 and for kurtosis is 0.41(code in notebook)

**SP and Weight(WT)**

**Use Q9\_b.csv**

Ans: sp skewness is 1.61 and kurtosis is 2.98

Wt for skewness is -0.61 and kurtosis is 0.95 **(code in notebook)**

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





Ans : fig 1.1

Ans: histogram peak has max frequency at the weight of 50-100 and min at 350-400 and it has tail on right side .

Boxplot have max outlier left tail , min outlier in right side (fig 1.1)

**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 for 94% =(198.737-201.262)

Confidence interval for 96%=(198.621-201.378)

Confidence interval for 98%=(198.438-201.561) **(code in notebook)**

**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:

Mean for marks is 41.00

Median for marks is 40.5

Variance for marks is 25.5

Standard deviation is 5.05.

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

Ans: No skewness is present we have a symmetrical distribution

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

Ans: Skewness is towards right side and also tail.

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

Ans: Skewness is toward left side .

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

Ans: Positive kurtosis means the curve is more peaked.

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

Ans: Negative kurtosis means the curve is flatter and broder.

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



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

Ans: The boxplot is not normally distributed as the mean is toward lower vale.

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

Ans: The data is skewed towards left and the range of whisker of minimum values is greater than maximum values.

**What will be the IQR of the data (approximately)?**Ans: The inner quartile range

=Q3-Q1

=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.**

* Boxplot 1 and 2, both are normally distributed
* There is no skewness at the maximum or minimum range
* There is no outliers in both the boxplot
* Both have the same median value from 275-225

**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)**

Ans: P(MPG>38) = 0.348

P(MPG<40) = 0.729

P(MPG>20) = 0.057

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**

Ans: yes its 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**

Ans: no it is not follows normal distribution.

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

Ans: z score for 60% is -1.6449

Z score for 90% is -1.8808

Z score for 94% is -0.8416

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

Ans: t scores of 95% is -2.0639

t score of 96% is -1.974

t score of 99% is -2.7969

**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 freed**

ans: df=n-1

=18-1

df=17

t score=-0.47

probability=32%