|  |  |
| --- | --- |
| **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 | **Nominal** |
| Fahrenheit Temperature | **Interval** |
| Height | **Ratio** |
| Type of living accommodation | **Nominal** |
| Level of Agreement | **Ordina**l |
| 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**?

There are eight possibilities when a coin is tossed,but we have to find the probability that two heads and one tail are obtained .So,

**HHT + HTH + THH**

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

**=3/8**

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

* **Equal to 1**
* **Less than or equal to 4**
* **Sum is divisible by 2 and 3**

=>**Equal to 1**

If two dices are rolled then total possible cases = 36

Total favourable cases(Having sum = 1) = 0

As minimum sum is 2 for outcome(1,1). **Hence probability is 0.**

=>**Less than or equal to 4.**

If two dices are rolled then total possible cases = 36

possible outcomes: 6. They are:-

(1,1) (1,2) (2,1) (1,3) (3,1) (2,2)

Total favourable cases(Having sum Less than or equal to 4) = 6.

**the probability is 6/36 = 0.16**

=>**Sum is divisible by 2 and 3.**

If two dices are rolled then total possible cases = 36

possible outcomes:- 6 .They are:

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

Total favourable cases(Having sum divisible by 2 and 3) = 6

**the probability is 6/36 = 0.16**

**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?**

Probability of drawing non-blue balls = Probability of drawing red or green balls.

P(draw first red or green ball) = total red or green balls = 5/7

total balls

P(draw second red or green ball) = Balance red or green balls = 4/6

balance balls

P(draw both balls either red or green) = 5/7 \* 4/6 = 10/21

**P(draw both non-blue balls) = 10/21 = 0.47 = 47.62 %** .

**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 number of candies for a randomly selected child**

= 1\*0.015 + 4\*0.20 + 3\*0.65 + 5\*0.005 + 6\*0.01 + 2\*0.012

= 0.015 + 0.8 + 1.95 + 0.025 + 0.06 + 0.24

= 3.090 = **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:- POINTS :**

**Mean** : 3.5965625

**Median** : 3.6950000000000003

**Mode** : 0 3.07

1 3.92

**Standard deviation**: 0.5346787360709716

**Variance** : 0.28588135080645166

max 4.93

min 2.76

**range** : 2.17

**In POINTS, Median > Mean - NEGATIVELY SKEWED.**

**SCORE :**

**Mean** : 3.2172500000000004

**Median** : 3.325

**Mode** : 0 3.44

**Standard deviation** : 0.9784574429896967

**Variance** : 0.9573789677419356

max 5.424

min 1.513

**range** : 3.9110000000000005

**In SCORE, Median > Mean - NEGATIVELY SKEWED.**

**WEIGH:**

**Mean** : 17.848750000000003

**Median** : 17.71

**Mode** : 0 17.02

1 18.90

**Standard deviation**: 1.7869432360968431

**Variance** : 3.193166129032258

max 22.9

min 14.5

**range** : 8.399999999999999

**In WEIGH, Median < Mean - POSITIVELY SKEWED**

**Q8) Calculate Expected Value for the problem below**

* **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?**

**ANSWER:-**

**Expected value** = (108+110+123+134+135+145+167+187+199)/9

= **145.333**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS:-** speed is negatively skewed and dist is positively skewed. speed is having negative peakedness and dist is having positive peakedness**.**

**skewedness of speed : -0.11750986144663393**

**kurtosis of speed: -0.5089944204057617**

**skewedness of dist: 0.8068949601674215**

**kurtosis of dist: 0.4050525816795765**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**ANS:-** SP is POSITIVELY skewed and WT is NEGATIVELY skewed. SP is having HIGH peakedness and WT is having POSITIVE peakedness.

**skewedness of SP : 1.6114501961773586**

**kurtosis of SP: 2.9773289437871835**

**skewedness of WT: -0.6147533255357768**

**kurtosis of WT: 0.9502914910300326**

**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.

The boxplot 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:- 94%** confidence interval: (**198.738325292158, 201.261674707842)**

**96%** confidence interval: (**198.62230334813333, 201.37769665186667)**

**98%** confidence interval: (**198.43943840429978, 201.56056159570022)**

**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**

* **Find mean, median, variance, standard deviation**.

**ANS:- Mean:** 41

**Median:** 40.5

**Variance:** 25.529411764705884

**Standard Deviation:** 5.05266382858645

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

**ANS**:-

The distribution of scores appears to be **positively skewed**, as indicated by the fact that the mean is greater than the median.

**Average** of student marks is **41**.

The student marks range from **34 to 56.**

Mode is **41**.

Most of students **scored between 36 to 42**.

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

**ANS**:- When the mean and median of a dataset are equal, it typically indicates that the distribution of the data is symmetric.

When the mean and median are equal, the skewness of the data is zero.

A skewness of zero indicates a symmetric distribution.

If mean = median = mode, the distribution is perfectly symmetrical.

So, when the mean and median are equal, the skewness is zero, indicating symmetry in the distribution of the data**.**



**Mean = Median = Mode**

**No Skewedness.**

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

**ANS:-**When the mean is greater than the median, it indicates that the distribution is skewed to the **right**, also known **as positively skewed or right-skewed.**

**In summary, when mean > median:**

The distribution is **positively skewed.**



**POSITIVELY SKEWED / RIGHT SKEWED**

**Mode < Median < Mean**

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

**ANS:-**When the median is greater than the mean, it indicates that the distribution is negatively skewed, also known as **left-skewed.**

In summary, when the median is greater than the mean, it suggests that the distribution is **negatively skewed.**



**NEGATIVELY SKEWED / RIGHT SKEWED**

**Mode >Median > Mean**

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

**ANS:-**A positive kurtosis value indicates that a dataset has heavier tails and a sharper peak compared to the normal distribution.

In simpler terms, **positive kurtosis** suggests that the dataset has more extreme values and is **more "peaked" or "tailed" than a normal distribution**. It may indicate the presence of outliers or a more pronounced central peak in the data.

A **LEPTOKURTIC DISTRIBUTION**  is **fat-tailed**, meaning that there are a lot of **outliers**.

**Leptokurtic distributions** are **more kurtotic** than a normal distribution. They have:

**A kurtosis of more than 3**

**An excess kurtosis of more than 0**

**Leptokurtosis** is sometimes called **positive kurtosis**, since the **excess** **kurtosis is positive**.



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

**ANS:-A negative kurtosis value indicates that a dataset has lighter tails and a flatter peak compared to the normal distribution.**

In simpler terms, **negative kurtosis** suggests that the dataset has fewer extreme values and is **less peaked** or has a **flatter peak** than a normal distribution. It may indicate that the data is more spread out or lacks extreme values compared to a normal distribution.

A **platykurtic distribution** is **thin-tailed**, meaning that **outliers are infrequent.**

**Platykurtic** distributions have **less kurtosis** than a normal distribution. In other words, platykurtic distributions have:

**A kurtosis of less than 3**

**An excess kurtosis of less than 0**

Platykurtosis is sometimes called negative kurtosis, since the excess kurtosis is negative.



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



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

**ANS:- NOT NORMALLY DISTRIBUTED.**The above Boxplot is not normally distributed the median is towards the higher value

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

**ANS:- NEGATIVE SKEWNESS.**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)?**

**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:-** First there are **no outliers**. Second both the box plots share the **same median** that is approximately in a range between 275 to 250 and 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**

* P(MPG>38) = **0.348**
* P(MPG<40) **= 0.729**

c. P (20<MPG<50) **= 0.013**

**Q 21) Check whether the data follows normal distribution**

* **Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv**

**ANS: a**.) MPG of cars **follows normal distribution**

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

**Dataset: wc-at.csv**

**ANS**:- Adipose Tissue (AT) and Waist does **not follow Normal Distribution .**

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

**ANS:-** Z score of **90%** confidence interval: **1.6448536269514722**

Z score of **94%** confidence interval: **1.8807936081512509**

Z score of **60%** confidence interval: **0.8416212335729143**

**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%** confidence interval: **2.0638985616280205**

t score of **96%** confidence interval: **2.1715446760080677**

t score of **99%** confidence interval: **2.796939504772804**

**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:-** Probability that 18 randomly selected bulbs would have an average life of no more than 260 days: **0.32167253567098364**