|  |  |
| --- | --- |
| 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 | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | 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 | Nominal |
| Level of Agreement | Ordinal |
| IQ(Intelligence Scale) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Interval |
| Time on a Clock with Hands | Nominal |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Interval |
| Years of Education | Interval |

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

**Answer** = 3/8

**Explanation -** The probability of getting two heads and one tail, which can occur in 3 of the 8 possible outcomes (THH, HTT, HTH).

Therefore, the probability of getting two heads and one tail when tossing three coins is **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

**Answers:**

**a). Sum equal to 1:** The probability is 0/36 = 0.

**Explanation:** There is no combination of dice rolls that results in a sum of 1.

**b). Less than or equal to 4:** The probability is 3/36 = 1/12 or 0.0833.

**Explanation:** These outcomes are possible only when both dice land on 1: (1, 1), (1, 2), and (1, 3**).**

**c). Sum divisible by 2 and 3:**

A sum is divisible by 2 if both dice rolls are even or both are odd. However, only certain even sums are divisible by 3: 6, 9, and 12.

* Even sums:
  + 2 + 2, 2 + 4, 2 + 6, 4 + 2, 4 + 4, 4 + 6, 6 + 2, 6 + 4, 6 + 6
* Even sums divisible by 3:
  + 3 + 3, 6 + 6

Therefore, there are 9 favorable outcomes for even sums and 2 for even sums divisible by 3, making a total of 11 favorable outcomes. **The probability is 11/36 = 0.3055 or roughly 30.5%.**

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?

**Answer:**

**Total number of ways to draw two balls:**

We have 7 balls (2 red + 3 green + 2 blue) and need to choose 2 without replacement. This can be done in ⁷C₂ ways, which is calculated as:

**⁷C₂ = (7 \* 6) / (2 \* 1) = 21**

**Number of ways to draw two balls without including any blue:**

We only need to consider red and green balls, so there are 5 total (2 red + 3 green). We again need to choose 2 without replacement, using ⁵C₂:

**⁵C₂ = (5 \* 4) / (2 \* 1) = 10**

**Probability:**

The probability is the number of favorable outcomes (drawing 2 non-blue balls) divided by the total number of possibilities (drawing any 2 balls):

Probability = (Number of ways without blue) / (Total number of ways) = 10 / 21

**Therefore, the probability of drawing two balls without any blue balls is 10/21. This can also be expressed as roughly 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

**Answer:** The expected number of candies is 3.09.

**Explanation:** The expected number of candies is calculated by summing the product of the number of candies and the probability of each child having that number of candies.

For example, the expected number of candies for child A is (1 candy) \* (0.015 probability) = 0.015 candies. Similarly, the expected number of candies for child B is (4 candies) \* (0.20 probability) = 0.80 candies.

By summing the expected number of candies for all children, we get the overall expected number of candies for a randomly selected child.

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.

**Answer:**

**Mean:**

Points 3.596563

Score 3.217250

Weigh 17.848750

**Median:**

Points 3.695

Score 3.325

Weigh 17.710

**Mode:**

Points 3.07, 3.92

Score 3.44

Weigh 17.02. 18.90

**Standard Deviation:**

Points 0.534679

Score 0.978457

Weigh 1.786943

**Range:**

Points 2.170

Score 3.911

Weigh 8.400

**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?

**Answer:** The expected value of the weight of a patient is: **145.33** pounds.

**Explanation:** we calculated the expected value (mean) of the weights by dividing the sum of weights by the total number of patients. This is the mathematical formula for calculating the expected value of a set of data points.

In this case, the set of data points is the weights of the patients at the clinic, and the expected value is the average weight of a patient at the clinic.

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

**Cars speed and distance**

**Answers:**

**Skewness:**

Speed -0.117510(**Negative Skewness**)

Distance 0.806895(**Symmetric**)

**Kurtosis:**

Speed -0.508994(**The Distribution is less peaked Planty Kurtic**)

Distance 0.405053(**The Distribution has Normal peak)**

**B .SP and Weight(WT)**

**Answers:**

**Skewness:**

SP : 1.611450(**Positive Skewness**)

WT : -0.614753(**Negative Skewness)**

**Kurtosis:**

SP 2.977329 (The distribution is more peaked **(leptokurtic**))

WT 0.950291 (Distribution has a normal peak (**mesokurtic))**

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



***Answer:***

In the above Histogram the values in the range of 60 to 100 has more frequency and The values in the range of 100 to 150 has the second highest frequency .So in conclusion the above histogram has **Positive Skewness**



**Answer:**

The above diagram is an boxplot with upper quartile ,lower quartile , Upper Extreme, lower extreme , so as we saw in the above boxplot the boxplot is containing **Outliers.**

**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?

**Answers:**

***94% confidence interval:*** (198.96 ,201.63)

***98% confidence interval:*** (198.62 ,201.37)

***96% confidence interval:*** (198.82 ,201.17)

**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?
3. ***Answers:***

Mean: 41.0

Median: 40.5

Variance : 25.529412

Standard deviation: 5.052664

1. ***Answer:***

From the above students marks **18** students are having average marks as **41.0** with the variance **25.52** and maintaining standard deviation **5.05**

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

Answer: The nature of skewness when the have same mean ,median is **Symmetric**

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

Answer: **Positive skewness**

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

Answer: **Negative skewness**

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

Answer: The Distribution has more peak(**Leptokurtic**)

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

Answer: The Distribution has less peak(**plantykurtic)**

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



What can we say about the distribution of the data?

Answer: As shown in the above, the boxplot is not normally distributed and the median is towards the higher value.

What is nature of skewness of the data?

Answer: from the above plot,The data is skewed towards the left ,The whisker range of minimum value is greater than the maximum.

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

Answer: The Inter Quartile Range=Upper Quartile(Q3)-Lower Quartile(Q1)

IQR=Q3-Q1

IQR=18-10=8

**IQR=8(approximately)**

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect Boxplot 2.

Answer: The two boxplots are not having outliers, and both boxplots are sharing same median that is approximately in the range of 250 and 275, They are normally distributed with zero to no skewness.

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)

Answers:

**a)P(MPG>38)=**0.3474

**b)P(MPG<40)=**0.7294

**c)P(20<MPG>50)=** 0.8989

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer: MPG of cars **is 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

Answer: Adipose Tissue (AT) and Waist Circumference(Waist) are **not follows Normal Distribution**

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

Answers:

**Z scores of 90%:** 1.2815

**Z scores of 94%:** 1.5547

**Z scores of 60%:**0.2533

Explanation: For Calculating Z-Scores we use **ppf** function from stats library.

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

Answers:

**t scores of 95%:** 1.71

**t scores of 96%:** 1.82

**t scores of 99%:**2.49

Explanation: As same as Z-Scores we use same library but in t score we need to mention T ,because by default it will take Z so for calculating t score we need to use t and also we need to give Degrees of freedom in the ppf function.

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

Answer:32%

From the Above data

T-score=-0.4714

Degrees of freedom:17

**P(t)=0.32**