|  |  |
| --- | --- |
| 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) | Ratio |
| 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 | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

Ans: Total no. of outcomes = 2^3 = 8 ( HHH, THH, HTH, HHT, HTT, THT, TTH, TTT)

Favorable outcomes = 3 ( THH, HTH, HHT )

Probability = Favorable outcomes/ Total no. of outcomes

P(two heads and one tail) = 3/8

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

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

Ans: Total outcomes = 6^2 = 36

a) 0 (There is no chance of getting zero because dice have no negative values)

b) favourable outcomes= 6 ((1, 1), (1, 2), (1, 3), (2, 1), (2, 2), (3, 1))

P = 6/36 = 1/6

c) favourable outcomes = 6((1,5), (2,4), (3,3), (4,2), (5,1), (6,6))

P = 6/36 = 1/6

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: Total Outcomes = 7c2 = 7!/5!2! = 21

Favourable outcomes = 5c2 = 5!/3!2! = 10 ( So the two balls drawn are not blue so outcomes decrease from 7 to 5)

Probability = 7c2/5c2 = 10/21

Q6) Calculate the Expected number of candies for a randomly selected child

Below are the probabilities of the count of candies for children (ignoring the nature of the child-generalised 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: Expected Value = Sum of the product of candies count to probability

E = 1(0.015) + 4(0.20) + 3(0.65) + 5(0.005) + 6(0.01) + 2(0.120) = 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 | MEDIAN | MODE | Variance | Std. Dev | Range |
| Points | 3.59 | 3.70 | 3.92 | 0.29 | 0.53 | 2.17 |
| Score | 3.22 | 3.33 | 3.44 | 0.96 | 0.98 | 3.91 |
| Weigh | 17.85 | 17.71 | 17.02 | 3.19 | 1.79 | 8.40 |

* We can observe after plotting the graphs for those there are outliers for ‘score’ and ‘weigh’

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 randodfm. What is the Expected Value of the Weight of that patient?

Ans: Expected Value = Mean = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans**: Skewness Kurtosis

Speed -0.11 2.42

Dist 0.76 3.25

* As we can observe from graphs ‘speed’ is negatively skewed because the more values are concentrated on right side, whereas ‘dist’ is positively skewed because the values are concentrated on left side.
* Both dist and speed are positive kurtosis

**SP and Weight(WT)**

**Use Q9\_b.csv**

Skewness Kurtosis

SP 1.55 5.72

WT -0.59 3.82

* “SP” is positively skewed where as “WT” is negatively skewed. Thus SP has distribution of data concentrated on the right whereas WT has distribution on the left. As seen in the graph
* Both WT and SP has positive Kurtosis

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



* The histogram is a negatively skewed because the tail is longer on left side.
* That means more chicks are present in the range of 50 -150 weight.
* As we can also observe that going to right side there are very few chicks are present with more weight.
* So we can also divide the chicks into 3 categories i.e., unhealthy(<50), healthy(50-150), fatty(>150).



Ans:

* Data has some outliers.
* It is a positively skewed

**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% | 98% | 96% |
| Upper | 201.04 | 201.38 | 201.17 |
| Lower | 198.96 | 198.62 | 198.83 |

**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.

ANS:

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.50 |
| Variance | 25.53 |
| Std Deviation | 5.05 |

1. What can we say about the student marks?

* By observing the graph, we can find some outliers in the data.
* It is not normally distributed.
* Majority of students are scored in between 35-45 marks.

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

Ans: Skewness=0, It is Symmetrical in nature.

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

Ans: Skewness=Positive. In this the right tail is long relative to left.

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

Ans: Skewness=Negative. In this the left tail is long relative to right.

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

Ans: High and narrow peak on central part of the data Or Leptokurtic

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

Ans: wider peak on central part of the data Or Platykurtic

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



What can we say about the distribution of the data?

Ans: It is not symmetric, It is more distributed on left side.

What is nature of skewness of the data?

Ans: It is negative because the median is greater than mean.

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

Ans: IQR = 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.

Ans:

* If we clearly observe Boxplot 1, the data is distributed light more on the upper side whereas in Boxplot 2 it is relatively symmetrical.
* By careful observation the median of the Two boxplots are almost same.
* There are no outliers for both Boxplots.
* The boxplot 1 is less distributed and the boxplot is more distributed.
* The IQR in boxplot 1 is less when compared to Boxplot 2.

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) = 0.4074074
  2. P(MPG<40) = 0.7530864
  3. P (20<MPG<50) = 0.8518519

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Ans: It follows Normal Distribution the p-value >0.05

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

Ans: Both are not following Normal Distribution because the p-value < 0.05.

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

Ans:

Confidence interval 90% 94% 60%

Z – Score 1.6 1.8 0.84

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

Ans:

95% 96% 99%

± 2.05 ± 2.16 ± 2.79

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: T-Score = -0.4714

Degrees of freedom = 17

pt(tscore, df) = pt(-0.4714,17) = 0.3216