Q1) Identify the Data type for the Following:

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

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 | **Interval** |
| Fahrenheit Temperature | **Interval** |
| Height | **Ratio** |
| Type of living accommodation | **Ordinal** |
| Level of Agreement | **Interval** |
| IQ(Intelligence Scale) | **Interval** |
| Sales Figures | **Interval** |
| Blood Group | **Nominal** |
| Time Of Day | **Ordinal** |
| 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: -Probability (two heads and one tail) = number of outcomes with two heads and one tail / total number of possible outcomes**

**Probability = ⅜**

**Probability =0. 375**

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

1. Equal to **1 = 0**
2. Less than or equal to **4=1/6**
3. Sum is divisible by **2 and 3=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: - Probability = 10/21**

**10/21 = 0.4762**

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:- Expected number of candies = 0.015 + 0.80 + 1.95 + 0.025 + 0.06 + 0.24**

**Expected number of candies = 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**

|  |  |  |  |
| --- | --- | --- | --- |
|  | **Points** | **Score** | **Weight** |
| **Mean** | **3.596** | **3.217** | **17.84** |
| **Median** | **3.695** | **3.325** | **17.71** |
| **Mode** | **3.891** | **3.54** | **17.43** |
| **Variance** | **0.285** | **0.957** | **3.19** |
| **Standard Deviation** | **0.534** | **0.978** | **1.786** |
| **Range** | **2.76,4.93** | **1.513,5.424** | **14.5,22.9** |

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: - the Expected Value of the Weight of that patient is 145.3333**

**> x <- c(108, 110, 123, 134, 135, 145, 167, 187, 199)**

**> result.mean <- mean(x)**

**> print(result.mean)**

**[1] 145.3333**

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

**Cars speed and distance**

**Use Q9\_a.csv**

**ANS:-**

**Mean speed: 15.40**

**Mean dist: 42.98**

**Standard deviation of speed: 5.29**

**Standard deviation of dist: 25.77**

**Skewness of speed: -0.11**

**Skewness of dist: 0.78**

**Kurtosis of speed: -0.58**

**Kurtosis of dist: 0.25**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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



**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:-**

**CI = X̄ ± Zα/2 \* (σ/√n)**

**For a 94% confidence interval, the critical value Zα/2 is 1.8808 (from a standard normal distribution table). Substituting the values into the formula, we get:**

**CI = 200 ± 1.8808 \* (30/√2000)**

**CI = (198.91, 201.09)**

**Therefore, we can be 94% confident that the true average weight of adult males in Mexico is between 198.91 and 201.09 pounds.**

**For a 98% confidence interval, the critical value Zα/2 is 2.3263. Substituting the values into the formula, we get:**

**CI = 200 ± 2.3263 \* (30/√2000)**

**CI = (197.41, 202.59)**

**Therefore, we can be 98% confident that the true average weight of adult males in Mexico is between 197.41 and 202.59 pounds.**

**For a 96% confidence interval, the critical value Zα/2 is 1.9600. Substituting the values into the formula, we get:**

**CI = 200 ± 1.9600 \* (30/√2000)**

**CI = (199.15, 200.85)**

**Therefore, we can be 96% confident that the true average weight of adult males in Mexico is between 199.15 and 200.85 pounds.**

**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:**

**sum of all scores = 720**

**number of scores = 18**

**Mean = 720 / 18 = 40**

**Median = (40 + 41) / 2 = 40.5**

**Variance s2 = 25.529412**

**Standard Deviation s =5.0526638**

**Count n =18**

**Mean x = 41**

**Sum of Squares SS =434**

1. What can we say about the student marks?

**ANS:-**

**From the given scores, we can make the following observations about the student's marks:**

1. **The student scored between 34 and 56, which is a wide range of scores.**
2. **The scores appear to be mostly concentrated between 36 and 45.**
3. **The most frequently occurring scores are 41 and 42, with each occurring four times.**
4. **The scores appear to be skewed to the right, as there are more scores above the median (40) than below it.**
5. **There are a few high scores, with the highest being 56.**

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

**Ans:- Data is normalized and there is no skewness.**

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

**Ans:- -Negative Skewness implies mass of the Distribution concentrated on the right side.**

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

**Ans:- -Positive Skewness implies mass of the Distribution concentrated on the left side.**

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

**Ans:- Positive kurtosis value indicates thinner peak and wider tails.**

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

**Ans:- Negative kurtosis value indicates a wider peak and thinner tails.**

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



What can we say about the distribution of the data?

**Ans:- -Not normally distributed**

What is nature of skewness of the data?

**Ans:- -Negative skewness**

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

**Ans:- 10-18**

Q19) Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 with respect 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) - **1-pnorm(38,34.422,9.13144)= 0.3475908**
  2. P(MPG<40) - **pnorm(40,34.422,9.13144)= 0.7293527**

c. P (20<MPG<50) - **pnorm(50,34.422,9.13144)-(1-pnorm(20,34.422,9.13144))= 0.01311818**

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:- distributed normally**

1. 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) normal distributed**

* **Waist Circumference(Waist) normal distributed**

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

**Ans:-**

|  |  |
| --- | --- |
| **Confidence interval** | **Z - scores** |
| **60%** | **0.8416212** |
| **90%** | **1.644854** |
| **94%** | **1.880794** |

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

|  |  |
| --- | --- |
| **Confidence interval** | **T - scores** |
| **95%** | **2.063899** |
| **96%** | **2.171545** |
| **99%** | **2.79694** |

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:- =52.86%**