|  |  |
| --- | --- |
| 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 | interval |
| 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 | interval |
| Number of Children | nominal |
| Religious Preference | nominal |
| Barometer Pressure | ratio |
| SAT Scores | interval |
| Years of Education | ordinal |

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

Ans=1/4

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

1. Equal to 1 ans=0
2. Less than or equal to 4 ans=1/6
3. Sum is divisible by 2 and 3 ans=2/3

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=6/49.

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= 3.13 candies.

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.

Points score weight

mean: 3.596563 mean: 3.21725 mean: 17.84875

median: 3.695 median: 3.325 median: 17.71

mode: 3.92 mode: 3.44 mode: 17.02

variance: 0.276948 variance: 0.927461 variance: 3.09338

std: 0.526258 std: 0.963048 std: 1.758801

range: 2.17 range: 3.911 range: 8.4

**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?

Ans=E(x)=145.3

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Speed=**

**Skew:** -0.11751

**Kurtosis:** -0.50899

**distance=**

**Skew:** 0.806895

**Kurtosis:** 0.405053

**SP and Weight(WT)**

**Use Q9\_b.csv**

**sp=**

**Skew:** 1.61145

**Kurtosis:** 2.977329

**wt=**

**Skew** :-0.61475

**Kurtosis:** 0.950291

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

**Ans=**

**Histogram is represents the frequency distribution**

**Histogram is positively skewed**



This graph shows the distribution of data

It shows the outliers of data

**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, the critical value is 1.645

98% confidence interval, the critical value is 2.33

96% confidence interval, the critical value is 1.96

**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:40.5, variance:24.55, std=4.95

1. What can we say about the student marks?

Ans=41

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

Ans=symmetrical

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

Ans=positively skewed

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

Ans=negatively skewed

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

Ans=Peaked and thick tails

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

Ans=wider peak and thinner tail

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



What can we say about the distribution of the data?

Ans=left skewed

What is nature of skewness of the data?

Ans=negatively skewed

What will be the IQR of the data (approximately)?   
ans=8

Q19) Comment on the below Boxplot visualizations?



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

Ans=

In 1st boxplot majority of the data is in 200 to 275 and it is kurtosis

In 2nd boxplot is normal distribution and the majority of data

is in range 250 to 310

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)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans= it is little bit left skewed but we can say it is 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=

Wc follows normal distribution

At is right skewd it means it is not following normal distribution

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

90= 1.645

94= 1.88

60= 0.25

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

95= 1.697

96= 1.672

99= 2.500

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

Using a t-distribution table or calculator with 17 degrees of freedom and a one-tailed test, we can find the probability of getting a t-score less than -2.22. This probability is approximately 0.0157.

the probability that 18 randomly selected bulbs would have an average life of no more than 260 days if the company's claim were true is approximately 0.0157 or 1.57%. This probability is quite low, which suggests that the company's claim is not true or there's some other factors that affecting the average life of the bulb

Hint:

rcode 🡪 pt(tscore,df)

df 🡪 degrees of freedom