|  |  |
| --- | --- |
| `1Activity | 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 | Continuous |
| 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 | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Nominal |
| Level of Agreement | Ordinal |
| 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 | 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: Three coins tossed = 2^3=8

The possible outcomes are HHH,HHT,HTH, THH,TTH,THT,HTT,TTT.

The probability that two heads and one tail are obtained= 3/8

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: Total no of balls= (2+3+2)=7

Number of ways of drawing 2 balls out of 7= 21

Number of ways of drawing 2 balls out of (2+3)balls= 10

Probability that none of the balls drawn is blue= 10/21.

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- 3.09

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.

Ans: Mean= Points 3.596563

Score 3.217250

Weigh 17.848750

Median= Points 3.695

Score 3.325

Weigh 17.710

Mode= 3.07, 3.44, 17.02

Variance= Points 0.285881

Score 0.957379

Weigh 3.193166

Standard deviation= Points 0.534679

Score 0.978457

Weigh 1.786943

**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: 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 skewness=** -0.11750986144663393

**Speed kurtosis=** -0.5089944204057617

Distance skewness = 0.8068949601674215

Distance Kurtosis=0.4050525816795765

**SP and Weight(WT)**

**SP skewness=** 1.6114501961773586

**SP Kurtosis=** 2.9773289437871835

**Weight Skewness=** -0.6147533255357768

**Weight Kurtosis=** 0.9502914910300326

**Use Q9\_b.csv**

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

**Histogram:**

**Chick weight data is right skewed or positively skewed**

**More than 50% chick weight is between 50 -150**

**Most of the chick weight is between 50-100**



Ans: the data is right skewed

There are outliers at upper 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: For 94 % confidence interval = (198.738325292158, 201.261674707842)

For 98% confidence interval= (198.43943840429978, 201.5605615957002

For 96% confidence interval=(198.62230334813333, 201.3776966518666

**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?

Ans: Mean= 41

Median= 40.5

Variance= 25.52

Standard Deviation=5.05

Q13) What is the nature of skewness when mean, median of data are equal? The distribution has zero skewness, normalized skewness

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

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

Q16) What does positive kurtosis value indicates for a data ? peaked and possess thick tails, heavy tailed.

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

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



What can we say about the distribution of the data?

Deassigned Format.

What is nature of skewness of the data? Left side skewed

What will be the IQR of the data (approximately)? 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: 1) Box plot 1 designed with range =3

1. Box plot 2 range = 1.5
2. Outliers doesn’t exist in both of the boxplots.

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)
  3. P (20<MPG<50)

Ans: mean of mpg = 34.422076

Standard deviation = 9.131445

a.P(MPG>38) = 0.6524060595854699

P(MPG>38)=1-P(MPG<38)

1-0.65

=0.35

b. P(MPG<40) = 0.7293498604157946

c. P (20<MPG<50) = 0.8988689076273199

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: 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

Ans: yes, waist and Adipose Tissue follows normal distribution by normal Q-Q plot.

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

Ans: z scores for 90%, 94%, 60%

1.2815515655446004

1.5547735945968535

0.2533471031357997

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

Ans: T-scores of 95%, 96%, 99% confidence Interval 1.7108820799094275

1.8280511719596342

2.4921594731575762

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: µ=270, =260, SD=90, n=18, df=n-1=18-1= 17

tscore= = = -10/21.23= -0.47

Required probability = 0.32=32%