|  |  |
| --- | --- |
| 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 | Ratio |
| Socioeconomic Status | Interval |
| Fahrenheit Temperature | Ratio |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| Level of Agreement | Interval |
| IQ(Intelligence Scale) | Interval |
| Sales Figures | Interval |
| Blood Group | Ratio |
| Time Of Day | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Interval |
| Religious Preference | Ratio |
| Barometer Pressure | Interval |
| SAT Scores | Ratio |
| Years of Education | Nominal |

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

**Ans:** C={HHH,HHT,HTH,THH,TTH,THT,HTT,TTT}

P =

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

**Ans:**

N(S)=36

1. P = 0
2. P = 6/36=1/6
3. 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:** Probability of getting no 2 blue balls = 2 red + 3 Green = 5

= /

= 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 |

**Ans:**

Child A – probability of having 1 candy = 0.015.

Child B – probability of having 4 candies = 0.20

Ans: P= Candies count \* propablity

= (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.

**Ans:**

Mean ,median and mode values lies close to each other of three Points, Score and weigh.

If drawn a box-plot median lies closer to upper quartile.

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:**

= 1/9 \* (108+110+123+134+135+145+167+187+199)

= 1308/9

= 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans:** For Speed both skewness & kurtosis are negative.

For distance both skewness & kurtosis are positive.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**  For SP both skewness & kurtosis are Positive

For WT skewness is negative and kurtosis is positive

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



Ans: Histogram is positively skewed. Mean is greater than median.



Ans: It is Positive skewed. It has many outliers.

Median is closer to lower quartile.

**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% CI- (143.57619175546247, 256.42380824453755)

For 98% CI-(130.2095637787748, 269.7904362212252)

For 96% CI-(138.38753268104531, 261.61246731895466)

**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=24.11, Standard deviation = 4.91

Avg. marks students got is 41. Mean, median and mode lies close to each other.

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

**Ans:** Skewness=0

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

**Ans:** Right skewed

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

**Ans:** Left skewed

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

**Ans:** Sharp peak

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

**Ans:** Flat peak

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



What can we say about the distribution of the data?

**Ans:** This box plot distribution. Median is close to upper quartile.

What is nature of skewness of the data?

**Ans:** It is left skewed.

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

**Ans:** Let’s assume Q1=10, Q3=18

IQR= Q3-Q1=18-10

= 8

Q19) Comment on the below Boxplot visualizations?



**Ans:** 1) For 1st box-plot median is close to lower quartile. Hence it is positively skewed.

There are no outliers.

2) For 2nd box-plot median is in middle. Hence it is symmetrical distributed.

There are no outliers

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

**Ans:** Inter quartile range for 2nd diagram is greater than 1st diagram.

2nd diagram has wide distribution than 1st diagram.

IQR1= 275-250=24

IQR2= 310-225=85

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

a. P(MPG>38)- 0.3475939251582705

b. P(MPG<40)- 0.7293498762151616

c. P (20<MPG<50)- 0.8988689169682046

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Ans:** MPG of cars does not follow normal distribution as mean is not equal to median.

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

Dataset: wc-at.csv

**Ans:** AT and Waist does not follow normal distribution as mean is not equal median .

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

Ans: 90 % CI-1.6448536269514722

94% CI-1.8807936081512509

60 % CI-0.8416212335729143

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

Ans: 95% CI- 2.0638985616280205

96% CI- 2.1715446760080677

99% CI-2.796939504772804

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:** tstats =x-µ/(s/sqrt(n))

=260-270/(90/4.24)

= -0.4711



P=0.3218