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

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

**Solution:**

Number of coins= 3

Total number of outcomes= 8 (HHH, HHT, HTH, THH, TTH, TTT, THT, HTT)

Probability= (favorable outcomes/total outcomes)

Probability of exact two heads and one tail= 3/8= 37.5%

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

**Solution**:

Total Possible outcomes:{(1,1),(1,2),(1,3),(1,4),(1,5),(1,6),(2,1),(2,2),(2,3),(2,4),(2,5),(2,6),(3,1),(3,2),(3,3),(3,4),(3,5),(3,6),(4,1),(4,2),(4,3),(4,4),(4,5),(4,6),(5,1),(5,2),(5,3),(5,4),(5,5),(5,6),(6,1),(6,2),(6,3),(6,4),(6,5),(6,6)}

1. Equal to 1

P(E1)= 0/36

=0

1. Less than or equal to 4

P(E2)= 6/36

=1/6

1. Sum is divisible by 2 and 3

(1,5)(2,4)(3,3)(4,2)(5,1)(6,6)

P(E3)=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?

Total No of balls= 2+3+2= 7

Let S be the sample space

Then, n(S)= No of ways of drawing 2 balls out of 7

= 7C2

=(7\*6)/(2\*1)

=21

Let E= Event of drawing 2 balls, none of which is blue.

n(E)= Number of ways drawing 2 balls out of (2+3) balls.

=5C2

=(5\*4)/(2\*1)

10

P(E)= n(E)/n(S)

=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

**Solution:**

Expected No of candies for a randomly selected child

=1\*0.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.12

=0.015+0.8+1.95+0.025+0.06+0.24

=3.090

=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**

**Solution:**

File name- Statistics 1(Q7)

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?

**Solution**:

Expected Value= Σ(Probability\*value)

=Σ(P(x)\*E(x))

There are 9 patients

Probability of selecting each patient=1/9

E(x)=108, 110, 123, 134, 135, 145, 167, 187, 199

P(x)= 1/9

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

=1/9(1308)

=145.33

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

Cars speed and distance

Use Q9\_a.csv

Solution:

File name: Statistics 1(Q9\_a)

SP and Weight(WT)

Use Q9\_b.csv

File name: Statistics 1(Q9\_b)

**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?

**Solution**:

Population N= 3000000

Sample size n= 2000

Standard deviation σ= 30

CI 94% confidence level

From Z- table values of z (c): 1.88

=200± 1.88(30/√2000)

=201.26 or 198.74

CI 98% confidence level

From Z- table values of z (c): 2.33

=200±2.33(30/√2000)

=201.56 or 198.43

CI 96% confidence level

From Z- table values of z (c): 2.05

=200±2.05(30/√2000)

=201.37 or 198.62

**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?

Solution:

File name= Statistics1 (Q12)

|  |  |
| --- | --- |
| Mean | 41 |
| Median | 40.5 |
| Variance | 25.52 |
| Standard deviation | 5.05664 |

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

Answer: Distribution has zero skewness

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

Answer: Distribution is positively skewed

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

Answer: Distribution is negatively skewed

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

 Answer:Distribution with positive kurtosis value indicates that the distribution has heavier tails and a sharper peak than the normal distribution.

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

Answer: Distribution with negative kurtosis value indicates that the distribution has lighter tails than the normal distribution.

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



What can we say about the distribution of the data?

Answer: Not normally distributed

What is nature of skewness of the data?

Answer: Negative skiwness

What will be the IQR of the data (approximately)?   
Answer: IQR of data 10-18

Q19) Comment on the below Boxplot visualizations?



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

Answer:

1. Have Same median
2. Normally Distributed

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)

**Solution**:

File name- Statistics1 (Q20)

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

File Name: Statistics1 (Q21\_a)

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

Dataset: wc-at.csv

File Name: Statistics 1(Q21\_b)

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

**Solution:**

|  |  |
| --- | --- |
| **Confidence Interval** | **Z Scores** |
| 60% | 0.84 |
| 90% | 1.65 |
| 94% | 1.88 |

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% | 2.06 |
| 96% | 2.17 |
| 99% | 2.78 |

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

****Solution:****

t - statistics for the data is given as follows:

t= (x-µ)/(s/√n)

x = mean of the sample of bulbs =  260

μ = population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

t = (260-270)/(90/√18)

t = (-10)/(90/3√2)

t = (-10)/(30/√2)

t =( -1\*√2)/3

t = -0.471

The degrees of freedom is n-1 so, 17

**t < - 0.471 with 17 degrees of freedom** assuming the population mean is true.

The probability of the bulbs lasting less than 260 days on average of **0.3218** assuming the mean life of the bulbs is 300 days.