|  |  |
| --- | --- |
| 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 | Ordinal |
| Fahrenheit Temperature | Interval |
| Height | Ratio |
| Type of living accommodation | Ordinal |
| 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 | Ratio |
| Years of Education | Interval |

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

Answer : We know that,

**Probability of an event (E) = Number of favourable outcomes / Total number of outcomes**

Let, H = Heads, T = Tails

Possible outcomes:

(H,H,H), (H,H,T), (H,T,H), (H,T,T), (T,H,H), (T,H,T), (T,T,H), (T,T,T)

Total number of outcomes = 8

Number of outcomes that gives two heads and one tail = 3

Such that, (H,H,T), (H,T,H), (T,H,H)

Thus, number of favourable outcomes = 3

Probability of getting two heads and one tail = Number of favourable outcomes / Total number of outcomes

= 3/8

**Thus, the probability of getting two heads and one tail on tossing three coins at once is equal to 3/8**

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

Answer : When two dices are rolled then,

Sample Space(S) = {(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)} = 36

1. P(Sum=1) = 0/36 = 0
2. P(Sum<=4) = {(1,1),(1,2),(1,3),(2,1),(2,2),(3,1)} = 6/36 = 1/6
3. P(Sum divisible by 2 and 3) = {(1,5),(2,4),(3,3),(4,2),(5,1),(6,6)} = 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?

Answer : Total number of balls = (2 + 3 + 2) = 7

Let S be the sample space

Then, n(S) = Number of ways of drawing 2 balls out of 7

n(S)=7*C*2

n(S)=(7×6)/(2×1)

n(S)=21

Let E = Event of 2 balls, none of which is blue  
n(E) = Number of ways of drawing 2 balls out of (2 + 3) balls

n(E)=5*C*2

n(E)=(5×4)/(2×1)

n(E)=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

Answer : Expected number 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

**Expected number of candies for a randomly selected child  = 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**

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?

Answer : The weights (X) of patients at a clinic (in pounds), are 108, 110, 123, 134, 135, 145, 167, 187, 199

one of the patients is chosen at random.

**To Find :** Expected Value

**Solution:**

Expected Value  =  ∑ ( probability  \* Value )

 ∑ 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  1/9   1/9  1/9   1/9   1/9   1/9   1/9  1/9

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

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

= (1/9)  (  1308)

= 145.33

Expected Value of the Weight of that patient = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**SP and Weight(WT)**

**Use Q9\_b.csv**

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

Answer :

Inferences for Histogram :- The most of the data concerted in the 50-150 with frequency 200. We can say that long tail on right side and data concentrated on left side.

Inferences for Boxplot :- In Boxplot, boxplot suggests that the distribution has lots of outliers towards upper extreme.



**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?

Answer : In this question we have not given population standard deviation, then we will use t distribution to Calculate 94%,98%,96% confidence interval .

Given Population (N)= 3,000,000 ,sample size( n)=2,000 ,Sample Mean (x̄)=200, sample standard deviation (s)=30, Degree of freedom=n-1=2000-1

1999, Standard Error=s/=30/30/44.7213=0.671

1**.For 94% Confidence interval**

α =1-0.94=0.06

(t-value for 1999 by using t-table)

t 1-α,n-1=t 0.94,1999=,

+ t 0.94,1999\*s/=200+1.881861\*0.671=200+1.26272873=201.262729

- t 0.94,1999\*s/=200- 1.881861\*0.671=200-1.26272873=198.737271

[198.737271, 201.104466]

2**.For 98% confidence interval**

α =1-0.98=0.02

(t-value for 1999 by using t-table at 98%)

t 1-α,n-1=t 0.98,1999= 2.328215

+ t 0.98,1999\*s/=200+2.328215\*0.671=200+1.56223227=201.562232

- t 0.98,1999\*s/=200-2.328215\*0.671=200-1.56223227=198.437768

[198.437768, 201.562232]

**For 96% confidence interval**

α =1-0.96=0.04

(t-value for 1999 by using t-table at 96%)

t 1-α,n-1=t 0.96,1999=2.05509

+ t 0.96,1999\*s/=200+2.05509\*0.671=200+1.37896539=201.378965

- t 0.96,1999\*s/=200-2.05509\*0.671=200-1.37896539=198.621035

[198.621035, 201.378965]

**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?

Answer : Mean=41, Median= 40.5, Variance=25.52941, Standard Deviation=5.052664

We say that the scores obtained by a student in tests in increasing order and most Occurring value is 41 and the smallest marks is 34 and highest marks is 56.

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

Answer : No skewness because skewness is the measurement of asymmetry in the distribution . but when Mean, Median of data are equal so the distribution is symmetric.

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

Answer : Positive Skewness

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

Answer : Negative Skewness

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

Answer : 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 : negative kurtosis value indicates that the distribution has lighter tails and a flatter peak 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 : Let’s assume above box plot is about age’s of the students in a school.in Box Plot, Box contain middle of the 50% data of the student’s age above 10yrs.and upper whisker contain 25% data of the student’s age above 18yrs,and lower whisker contain 25% data of the student’s age less 10yrs. Upper Quartile(Q3)=18, Lower Quartile(Q1)= 10

What is nature of skewness of the data?

Answer : Left Skewed, Data concentrate on right side. Median is greater than Mean.

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

Answer : Upper Quartile(Q3)=18, Lower Quartile(Q1)= 10

(**Inter Quartile Range**)**IQR = Q3-Q1**

**IQR =** 18-10

**IQR =**8

Q19) Comment on the below Boxplot visualizations?



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

Answer : Here we assume they are two salesmen who sold mobiles of different companies to customers in the last one year.

The Box contain 50% of the data and the two Whisker contains 25%,25% data.

Plot-1:Upper Quartile(Q3)=280.5,Lower Quartile(Q1)=256.5, IQR(inter Quartile Range)=Q3-Q1=280.5-256.5=24

Major Outlier on upper extreme =Q3+3(IQR)=352.5>then major outlier

Minor Outlier on upper extreme= Q3+1.5(IQR)=316.5

Major Outlier on lower extreme=Q1-3(IQR)=184

Minor outlier on lower extreme=Q1-1.5(IQR)=220

Plot-2: Upper Quartile(Q3)=315.5, Lower Quartile(Q1)=223.6,

IQR(inter Quartile Range)=Q3-Q1=315.5-223.6=91.9

Major Outlier on upper extreme =Q3+3(IQR)=591.2

Minor Outlier on upper extreme= Q3+1.5(IQR)=453.35

Major Outlier on lower extreme=Q1-3(IQR)=-52.1

Minor outlier on lower extreme=Q1-1.5(IQR)=85.75

1. There is no outlier for upper quartile and lower quartile.
2. Symetrical/Normal Distribution of data.
3. Both the Box plots median are same.

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) = 26/81 = 0.3209
  2. P(MPG<40) = 61/81 = 0.75308

c. P (20<MPG<50) = 69/81 = 0.85185

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Answer : We can say that data is fairly symmetrical, i.e fairly normally distributed.

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

Dataset: wc-at.csv

Answer : waist is multimodal, AT is bimodal data. Mean > median, right whisker is larger than left whisker, data is positively skewed. Mean > median, both the whisker are of same length, median is slightly shifted towards left. Data is fairly symmetrically distributed.

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

Answer : Z scores of 90% confidence interval =1.29

Z scores of 94% confidence interval=1.56

Z scores of ,60% confidence interval=0.26

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

Answer : t scores of 95% confidence interval for sample size of 25 = 1.7108820799094275

t scores of 96% confidence interval for sample size of 25= 1.828

t scores of 99% confidence interval for sample size of 25 = 2.492

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

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

tscore= = = -10/21.23= -0.47

> pt(-0.47,17)

[1] 0.3221639

Required probability = 0.32=32%