|  |  |
| --- | --- |
| 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 | 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 | 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: sample s= {H, T}, {H, T}, {H, T}

2 HEADS ans 1 TAIL E= {HHT,HTH,THH}

P(X=2) = P(HHT) +P(HTH) +P(THH)

=1/2 .1/2 .1/2 + 1/2 .1/2.1/2 +1/2. 1/2 . 1/2

=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

ANS:

a) If two dices were rolled, then total possible cases =36

Total Favourable cases (Having sum =1) = 0

As minimum sum is 2 for outcome (1,1).

Hence, probability is 0.

b)possibilities of sum is less than or equal to 4 are (1,1) (1,2) (1,3) (2,2) (2,3)

probability(<=4) = (1,1)+ (1,2)+ (1,3)+ (2,2) +(2,3)

=1/36+ 1/36+ 1/36+ 1/36 +1/36

=5/36

c) Favorable outcomes = sum is divisible by 2 and 3

Sum should be divisible by both 2 and 3

Favorable outcomes = (1 , 5) , (3 , 3) , (4 , 2) , (5 , 1) , (6 , 6)

Therefore,

Number of favorable outcomes = 5

Thus the probability that sum is divisible by 2 and 3 is 5/36.

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 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

= 7C2

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

=21

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

∴n(E)= Number of ways of 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

Ans: 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

= 3.09

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?

Ans : 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**



HISTOGRAM: 1. It is right skewness

2. mean is higher than the median

3. standard deviation is 50.

4. there is no outlier.



BOXPLOT: 1. Median is located at centre.

2. it is right skewness.

3. there are multiple outliers.

**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: n=2000

Sample mean(x)= 200 pounds

Sample std(s)= 30 ponds

Here, n>30 therefore go for z-value

Confidence interval= x ±z(α-1) s/√n

1. confidence interval at 94% or 0.94

CI=x±z×s/√n

= 200±1.8808×30/√2000

= 200±1.262

= [198.738-201.262]

2. CI for 98%

CI= x±z×s/√n

= 200±2.3263×30/√2000

= 200±1.561

=[198.439-201.561]

3. CI for 96%

CI= x±z×s/√n

= 200±2.0537×30/√2000

= 200±1.378

= 198.622-201.378]

**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: 2. Mean = median

Must follow normal distribution.

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

Ans : if mean=median, skewness is zero.

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

Ans: if mean>median, distribution is positively skewed.

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

Ans: if median>mean, distribution is negatively skewed.

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

Ans: positive kurtosis value indicates that distribution of data is peak and thick tails.

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

Ans: negative kurtosis value indicates that distribution of data is flat and thin tails.

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



What can we say about the distribution of the data?

Ans: Distribution of data is unsymmetrical, because median is not middle of the box.

What is nature of skewness of the data?

Ans: It is left skewed of the data.

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

Ans: Q1= 10

Q3= 18

Q2(IQR)= 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: a) box 1 and box 2 are different.

1. Median: both the medians of the box 1 and box 2 are lies at middle of the box.
2. Whiskers: box 1 has shorter whisker that means distribution of data is

compare to box 1 so distribution of the data is wider.

1. There are no outliers in box 1 and box 2.

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)



Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv



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

Dataset: wc-at.csv

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

Ans:



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

Ans:



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: population mean/avg µ= 270

No. of random bulbs n = 18

Sample mean/avg x = 260

Standard deviation s = 90

Probability (x<260) =?

n<30, therefore conduct t-test

i.e t=(x-µ)

s/√n

=(260-270)

90/√18

=-10

21.226

t=-0.4711

p(x<260)=.