|  |  |
| --- | --- |
| 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 | Qualitative |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Qualitative |

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

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

Total number of samples when tossing 3 ={TTT,TTH,THT,THH,HTT,HTH,HHT,HHH}

1. P(E) = 3/8 =0.375=37.5%

2. P(E) = 3/8 =0.375=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
4. P(E)=0

b) P(E)=6/36 = 0.1666 =16.66%

1. P(E) = 5/36 = 0.1388 =13.88 %

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 number of balls= 2R+3G+2B =7 Balls

2 balls are drawn randomly who are not blue.

5 are not blue color.

P (E) = 5c2/ 7c2= 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: 1.P(A) = Expected number of candies(A)/ Candies count(A)

0.015= Expected number of candies(A)/1

Expected number of candies(A)= 0.015

2) P(B) = Expected number of candies(B)/ Candies count(B)

0.20= Expected number of candies(B)/4

Expected number of candies(B)= 0.8

3)Expected number of candies(C)/ Candies count(C)

0.65= Expected number of candies(C)/3

Expected number of candies(B)= 0.65\*3=1.95

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

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  = 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**



1. **Ans**:Histogram of ChickWeight$weight it is observer that it is positive skewness because data is skewed at positive side.



1. **Ans**:From above boxplot it is identified that there are Outliers in the data.

**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**: Here t test is used because data is given related to samples.

a)For 94% confidence interval:

1-=94%

=1-0.94=0.06

200(2.067)\*44.2735

(108.49, 291.51)

b)For 98% confidence interval:

1-=98%

=1-0.98=0.02

200(2.34)\*44.2735

(96.4,303.96)

c) )For 96% confidence interval:

1-=96%

=1-0.96=0.04

200(1.89)\*44.2735

(116.33, 283.67)

**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?
3. Mean=41

Median=41

Mode=41

Variance=41

Std. Deviation=41

1. The average marks of the students is same.

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

Ans :- Skewness can be positive, negative or zero. when mean, median of data are equal then there is no skewness.

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

Ans:-When mean is greater than median then it is called as positive skewness. Because more data is distributed to positive side.

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

Ans:-When median is greater than mean then it is called as negative skewness. Because more data is distributed to negative side.

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

Ans:-In positive kurtosis data set is an indicator that data has heavy tails or outliers. If there is a high kurtosis, then, we need to investigate why do we have so many outliers

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

Ans:- A distribution with a 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?

Ans:-Data is distributed (Q3-Q2)<(Q2-Q1) so it is negative.

What is nature of skewness of the data?

Ans:-Negative Skewness.

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

Ans:-IQR is near about 35%

Q19) Comment on the below Boxplot visualizations?



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

Ans:-Range of 1 is less than range of 2 so IQR of boxplot 1 is less than 2.

1 & 2 boxplot both follows same normal distribution.

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)

c. 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:= For 90% confidence interval:

1-=90%

=1-0.90=0.1

0.05 value will be going towards left & right side.

Z score for 90%=(-1.64,1.64)

For 94% confidence interval:

1-=94%

=1-0.94=0.06

0.03 value will be going towards left & right side.

Z score for 90%=(-1.79,1.79)

For 60% confidence interval:

1-=60%

=1-0.60=0.4

0.03 value will be going towards left & right side.

Z score for 90%=(-0.84,0.84)

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

For 95% confidence interval:

1-=95%

=1-0.95=0.05

0.025 value will be going towards left & right side.

t score for 95%=2.064

For 96% confidence interval:

1-=96%

=1-0.96=0.04

0.02 value will be going towards left & right side.

t score for 96%=2.1715

For 99% confidence interval:

1-=99%

=1-0.99=0.01

0.005 value will be going towards left & right side.

t score for 99%=2.797

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

t - statistics for the data is given as follows:

https://tex.z-dn.net/?f=t%3D%5Cdfrac%7Bx-%5Cmu%7D%7B%5Cfrac%7Bs%7D%7B%5Csqrt%20n%7D%7D

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

https://tex.z-dn.net/?f=t%3D%5Cdfrac%7B260-270%7D%7B%5Cfrac%7B90%7D%7B%5Csqrt%2018%7D%7D

https://tex.z-dn.net/?f=t%20%3D%20%5Cdfrac%7B-10%7D%7B%5Cfrac%7B90%7D%7B3%5Csqrt%202%7D%7D

https://tex.z-dn.net/?f=t%20%3D%20%5Cdfrac%7B-10%7D%7B%5Cfrac%7B30%7D%7B%5Csqrt%202%7D%7D

https://tex.z-dn.net/?f=t%20%3D%20%5Cdfrac%7B-1%20%5Ctimes%20%5Csqrt%202%7D%7B3%7D

t=-0.4714

from scipy import stats

stats.t.cdf(t-valaue,df)

stats.t.cdf(-0.4714,17)