|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | Discrete |
| Weight of a person | Ratio |
| Weight of Gold | Ratio |
| Distance between two places | Continuous |
| Length of a leaf | Ratio |
| Dog's weight | Ratio |
| Blue Color | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Nominal |

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 | Interval |
| Time on a Clock with Hands | Interval |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Interval |
| SAT Scores | Ordinal |
| Years of Education | Ratio |

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

Total No. of outcomes :

[H,H,H T,T,T, H,T,T H,H,T H,T,H T,H,H T,T,H T,H,T]

8

No. of favorable outcomes : [ H,H,T H,T,H T,H,H ]

Probability of Event Happening P(E): No of favorable outcomes/Total no. of outcomes

P(E): 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

Total No. of Outcome: 36

1. 0
2. No. of favorable outcomes: [1,3 2,2 3,1] = 3

Probability = 3/36 = 1/12

1. No. of Favorable outcome: [1,5 2,4 3,3 4,2 5,1 6,6] = 6

Probability = 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 Let the balls in the bag be denoted by r1,r2,g1,g2,g3,b1,b2

The Total No. of outcomes if 2 balls are drawn at random

r1r2,g1g2,g1g3,g2g3,b1b2,r1g1,r1g2,r1g3,r1b1,r1b2,r2g1,r2g2,r2g3,r2b1,r2b2,g1b1,g1b2,g2b1,g2b2,g3b1,g3b2 = 21

Outcomes In which none of the balls drawn is blue =r1r2,g1g2,g1g3,g2g3,r1g1,r1g2,,r1g3,r2g1,r2g2,r2g3 = 10

Probability That none of the balls drawn is blue = 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 Let X denote the Count of Candies for a Randomly Selected Child

So, X is a Random Variable & can take Values 1,4,3,5,6,2

Let Probability Distribution of X is P(X)

Then,

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| X | 1 | 4 | 3 | 5 | 6 | 2 |
| P(X) | 0.015 | 0.20 | 0.65 | 0.005 | 0.01 | 0.120 |

Expected No. of Candies For a randomly selected child = 1\*0.015 + 4\*0.2 + 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.

**Use Q7.csv file**

Ans

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Mean | Median | Mode | Variance | STD | Range |
| **POINTS** | |  | | --- | | 3.596563 | | |  | | --- | | 3.695000 | | |  | | --- | | 3.07 | | 0.285881 | |  |  | | --- | --- | | 0.534679 |  | | 2.170 |
| **SCORE** | |  | | --- | | 3.217250 | | |  | | --- | | 3.325000 | | |  | | --- | | 3.44 | | 0.957379 | 0.978457 | 3.911 |
| **WEIGH** | 17.848750 | 17.71000 | 17.02 | 3.193166 | 1.786943 | 8.400 |

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?

Let P(X) be the Probability of choosing one random patient

So, P(X) = 1/9

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 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 of the Weight of the Patient = 108/9 + 110/9 + 123/9 + 134/9 +135/9 + 145/9 + 167/9 + 187/9 + 199/9 =1308/9 = 145.33

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Ans**

|  |  |  |
| --- | --- | --- |
|  | Speed | Distance |
| Skewness | -0.117510 | 0.806895 |
| Kurtosis | -0.508994 | 0.405053 |

As from the Graph we can observe that Speed is Moderately skewed to the left And Distance is Moderately skewed to the right.

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| Skewness | 1.611450 | -0.614753 |
| Kurtosis | 2.977329 | 0.950291 |

SP is Highly Skewed and WT is moderately skewed.

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



Ans- For the above Histogram It can be clearly said that the dataset is negatively skewed & Considering the Boxplot its Observant that there are outliers present at the 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?

Ans- import numpy as np

from scipy import stats

#94% C.I

stats.t.interval(.94,df=1999,loc = 200, scale = 30/np.sqrt(2000))

(198.7376089443071, 201.2623910556929)

#98% C.I

stats.t.interval(.98,df=1999,loc = 200, scale = 30/np.sqrt(2000))

(198.4381860483216, 201.5618139516784)

#96% C.I

stats.t.interval(.96,df=1999,loc = 200, scale = 30/np.sqrt(2000))

(198.6214037429732, 201.3785962570268)

**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.

Ans -

|  |  |  |
| --- | --- | --- |
| **Scores** | Xi- |  |
| 34 | -7 | 49 |
| 36 | -5 | 25 |
| 36 | -5 | 25 |
| 38 | -3 | 9 |
| 38 | -3 | 9 |
| 39 | -2 | 4 |
| 39 | -2 | 4 |
| 40 | -1 | 1 |
| 40 | -1 | 1 |
| 41 | 0 | 0 |
| 41 | 0 | 0 |
| 41 | 0 | 0 |
| 41 | 0 | 0 |
| 42 | 1 | 1 |
| 42 | 1 | 1 |
| 45 | 4 | 16 |
| 49 | 8 | 64 |
| 56 | 15 | 225 |
| **Mean**-738/18=**41** |  | **Variance –**  *=434/(18-1) =* ***24.11*** |

Standard Deviation = = √24.11 = 4.91

1. What can we say about the student marks?

Ans- The Range of the Scores is 56-34 = 22, Knowing just this metric, the teacher of the class can gain a quick understanding of the spread of the data of values in exam scores among all of the students.

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

Ans – Skewness is 0

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

Ans- Skewed to the right or positively skewed.

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

Ans- - Skewed to the left or negatively skewed.

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

Ans- Positive values of kurtosis indicate that distribution is peaked and possesses thick tails.

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

Ans- Negative values of kurtosis indicate that distribution is Flatter and possesses thin tails.

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



What can we say about the distribution of the data?

Ans – Since most of the data points are falling to the right side

Mean < Median

What is nature of skewness of the data?

Ans- negatively/left skewness

What will be the IQR of the data (approximately)?   
Ans- Let Q3 be Upper quartile = 18

And Q1 be The Lower Quartile =10

Therefore, IQR = (Q3-Q1) = 18-10 = 8(approx..)

Q19) Comment on the below Boxplot visualizations?



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

Ans- From the above Visualization we can say that distribution of data in both boxplot1 and 2 is normal. Hence there is no skewness and the median for both the dataset is similar, which approximately lies at 262.5.

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)

Ans- a. P(MPG>38) = .34759

b.P(MPG>40) = .7293

c.P(20<MPG<50) = .898

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans- No it Does not follow a normal distribution. There is some data points concentrated at an approximate range between 12-30.

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

Dataset: wc-at.csv

Ans- No the dataset does not follow normal distribution.

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

Ans- For 90%

Z1-α = Z0.90

α/2 = .1/2=.05

AUC = .95

Value for .95 in Z table = 1.64

Z.94 = 1.88

Z.6 = .841

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

Ans- C.L for 95%

T1-α = t.95

α=0.05,α/2=.05/2=.025

Df = N-1 = 24

Stats.t.ppf(.975,24) = 2.06

For 96% = 2.17

For 99% = 2.79

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 – 1)Ho(Null Hypothesis – average life of bulbs >=260 days

Alternate Hypothesis – average life of bulbs is < 260 days

2)t test

3)one tail test

4)α=.05

5)Probability Calculation

Tscore =X -µ/s/√n=(260-270)/90/√18= -.471

Df = 17

p = stats.t.cdf(t\_score,17)

p = .32

6) a) if p < α, Reject Ho

b) if p > α, Fail to reject Ho

Hence, Fail to reject Null Hypothesis. Average life of Bulb is greater than 260 Days