|  |  |
| --- | --- |
| 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 |

**Nida Mulla. - 8999546939**

**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) | Ratio |
| 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 | Ratio |
| Years of Education | Ratio |

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

Ans: {HHH, HTH, THH, HHT, TTH, THT, HTT, TTT}

P= 3/8 = 0.375

**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: If two dice are rolled, there are 36 possibilities.

1. Sum is Equal to 1 = 0/36
2. Sum is Less than or equal to 4= {(1,3), (2,2), (3,1), (1,1), (1,2), (2,1)}

P= 6/36= 1/6

1. Sum is divisible by 2 and 3 = {(1,5) (3,3) (4,2) (5,1) (6,6)}

P= 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: p= 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\*0.120)

= 3.09

**Q7)** Calculate Mean, Median, Mode, Variance, Standard Deviation, Range & comment about the values / draw inferences, for the given dataset

* For Points, Score, Weight

Find Mean, Median, Mode, Variance, Standard Deviation, and Range and also Comment about the values/ Draw some inferences.

**Use Q7.csv file**

Ans:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Mean** | **Mode** | **Median** | **Variance** | **Standard Deviation** | **Range** |
| **Points** | 3.596563 | 3.92 | 3.695 | 0.28588 | 0.534678736 | 2.17 |
| **Score** | 3.21725 | 3.44 | 3.325 | 0.95738 | 0.978457443 | 3.911 |
| **Weight** | 17.84875 | 17.02 | 17.82 | 3.1931 | 1.786943236 | 8.4 |

**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 of the Weight of that patient = Exp value= Probability\*Value

:- Probability for selecting each patient= 1/9

:- Exp 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

= 145.33

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

Cars speed and distance

**Use Q9\_a.csv**

Ans:

* Skewness for Speed = (-0.11395477012828319)
* Skewness for Distance = (0.7824835173114966)
* Kurtosis for Speed = (-0.5771474239437371)
* Kurtosis for Distance= (0.24801865717051808)

**SP and Weight(WT)**

**Use Q9\_b.csv**

* Skewness for SP = (1.5814536794423764)
* Skewness for WT = (-0.6033099322115126)
* Kurtosis for SP = (2.7235214865269244)
* Kurtosis for WT = (0.8194658792266849)

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



Ans:

1.Histogram: The most of the data points are concerted in the range of 50-100 with frequency of 200. Least range of weight is 400 with the frequency somewhere around 0-10… So the expected value of the distribution is 75.

Skewness: It is clearly shown that long tail towords right so that it is heavily right skewed.

2.Boxplot: As shown we have outliers on the upper side of the boxplot. Also there are less data points between Q1 and bottom points.

Skewness: Median is less than mean i.e this distribution is right skewed.

**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: Using the t-distribution,

Mean-200

Sample size- 2000

Std deviation- 30

**t-value for -**

The 94% confidence interval is 1.880

The 96% confidence interval is 2.054

The 98% confidence interval is 2.326

**The interval is () =**

The 94% confidence interval is (198.73, 201.26)

The 96% confidence interval is (198.62, 201.38)

The 98% confidence interval is (198.44, 201.56)

**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:

1. Mean: sum of n/total count= 738/18= 41
2. Median: 40.5
3. Variance: 25.529411764705884
4. Standard Deviation: 5.05266382858645

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

Ans: When mean and median of data are equal, then the skewness of that data is Symmetrical in nature.

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

Ans: When mean > median, then the skewness of that data is Right skewed in nature.

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

Ans: When median > mean, then the skewness of that data is Left skewed in nature.

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

Ans: It shows the normally distributed data, and kurtosis value 0.

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

Ans: It shows that the data is not normally distributed as it has lighter tails and flatter peaks.

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



**a)**What can we say about the distribution of the data?

Ans: As shown above, the 50% of the data is covered above 10th place. Let’s example, the given data is about the age of the students in a school. It means approximately 50% students age is above 10yrs.

**b)**What is nature of skewness of the data?

Ans: Left skewed, median > mean.

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

Ans: Approx. 8

**Q19)** Comment on the below Boxplot visualizations?



Draw an Inference from the distribution of data for Boxplot 1 w.r.t Boxplot 2.

Ans: By observing both the plots whisker’s level is high in boxplot 2.

Also Mean and Median are nearly same or equal hence the distribution is symmetrical in nature.

**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. subset(MPG,MPG>38) = 33/81

b. subset(MPG,MPG<40) = 67/81

c. subset(MPG,MPG>20 & MPG<50) = 69/81

**Q 21)** Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: Yes

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

Dataset: wc-at.csv

Ans: Waist circumference(Waist)=(90.9, 90.8)

Adipose Tissue= (101.89, 96.54)

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

Ans:

from scipy. import stats

From scipy. Stats import norm

# Z-score of 90% confidence interval stats.norm.(0.95)

# Z-score of 94% confidence interval stats.norm. (.97)

#Z- score of 60% confidence interval stats.norm .(60)

A 94% confidence interval has two tails of 6/2=3%

So it goes from 3% to 97% which leaves 94% in the middle

So look up the Z for

P(z<Z) = 0.97

Two closest value in the z-table

P(z<1.88)= 0.96995

P(z<.89)=0.97065

Interploting

1.88 +(0.97-0.96995)\*(0.01)/(0.97062-0.96995)

Z critical = appox.1.880746

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

Ans:

The formula for a 100 (1-a) % confidence interval for µ is X+ (tα/2),

(n-1) is the point on the t- distribution

100 (α/2) % of the distribution area to its right

For a 95% confidence interval,α=0.05, because

100(1-0.05) %

100(0.95)%

95%

=25.9 + (1.987× 0.2846)

=25.9 +(0.5655)

=(25,33,26,47)

**Q24)** 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:

Assume H0= An average life of bulb>=260 days

X=260

Using python code

T=(260-270)/(90/18\*\*0.5)

t-0.4714045207910317

P value=0.32167411684460556

P0.005 so reject h0

So,an average life of bulb<260 days