|  |  |
| --- | --- |
| Activity | Data Type |
| Number of beatings from Wife | Discrete |
| Results of rolling a dice | ordinal |
| Weight of a person | ratio |
| Weight of Gold | ratio |
| Distance between two places | ratio |
| Length of a leaf | ratio |
| Dog's weight | ratio |
| Blue Color | nominal |
| Number of kids | discrete |
| Number of tickets in Indian railways | ordinal |
| 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 | ordinal |
| Level of Agreement | ordinal |
| IQ(Intelligence Scale) | ratio |
| Sales Figures | interval |
| Blood Group | nominal |
| Time Of Day | ratio |
| Time on a Clock with Hands | ratio |
| Number of Children | ordinal |
| Religious Preference | nomial |
| 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?

ANS: - if three coins are tossed, then 8 outcomes are possible, those are

HHH,HHT,HTH,HTT,THH,THT,TTH,TTT

So, three outcomes give two heads and one tail (HHT, HTH, THH)

Therefore, the probability is 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: Assumptions

1. The dice are “fair “, that is, not biased in any manner.
2. The dice are both six-sided dice, that is both have 6 faces, with each face on each dice, showing one of the numbers, 1 to 6, with no number repeated on the same dice.

Analysis

With two dice, there are ( 6 ) \* ( 6 ) = ( 36 ) possible combinations of numbers.

The minimum sum possible for the two dice thrown is (1, 1) = a sum of (2 )

The maximum sum possible for the two dice thrown is (6, 6) = a sum of (12).

Sum = (1).

The minimum possible sum is (1, 1) = ( 2 ).

Therefore P( 1 ) = ( 0 )/( 36 ) = 0

Sum = (4)

A sum of (4) can be achieved with number combinations ( 2, 2), (1, 3) and (3, 1), that is only with 3 combinations of numbers.

P(sum = 4) = ( 4 / 36 ) = ( 1 / 9 )

So, the answer is B) Less than or equal to 4.

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: There are 7 balls originally with 2 of them blue so the probability of the first ball not being blue is 5/7. This leaves 6 balls with 2 blue. The probability of the second ball not being blue assuming that the first wasn’t is 4/6. The probability that neither ball drawn was blue is (5/7)\*(4/6)=20/42=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

= 3.090

= 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

Ex 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**

**Ans:** By using the data set " **Q9\_a.csv**"

Skewness of Car speed=(-0.11751)

Kurtosis=(-0.50899)

Skewness of distance=(0.806895)

Kurtosis of distance=(0.405052)

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans:**

Skewness of SP=(1.61145)

Kurtosis of SP=(3.24148)

Skewness of WT=(-0.61475)

Kurtosis of WT=(0.902537)

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



Ans:By checking the bar graph we can identify that mean> median, so thes is in positively skewed data.

By checking the boxplot, we can identify that there is an 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:- Confidence interval is 9%

**Q12)** **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: from above data we can say that mean of marks of student is 41 which is slightly greater than median.   
Most of the students got marks in between 41-42, there are two outlier 49,56.

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

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

Ans: Positively skewed

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

Ans: Positively skewed

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

Ans: The distribution of data is peaked.

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

Ans: The distribution of data is flat or having thin tails.

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



What can we say about the distribution of the data?

What is nature of skewness of the data?

What will be the IQR of the data (approximately)?   
  
Ans: the nature of skewness of data is positively skewed.  
Q1 = 10  
Q2 = 14

Q3 = 18

IQR = 1-10 =8

IQR = 8.

Q19) Comment on the below Boxplot visualizations?



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

Ans: **Inferences::**

Min value for the Boxplot 1 is 240 (approx.)

Min value for the Boxplot 2 is 180 (approx.)

Median of the Boxplot1 is 262.5

Median of the Boxplot2 is 262.5

Max value of the Boxplot1 is 282.5(approx)

Max value of the Boxplot2 is 350(approx)

Boxplot2 mean varies much less than Boxplot 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)

c. P (20<MPG<50)

Ans::

Prob\_MPG\_greater\_than\_38 = np.round(1 - stats.norm.cdf(38, loc= q20.MPG.mean(), scale= q20.MPG.std()),3)

print('P(MPG>38)=',Prob\_MPG\_greater\_than\_38)

P(MPG>38)= 0.348

B) P(MPG<40)

Ans: prob\_MPG\_less\_than\_40 = np.round(stats.norm.cdf(40, loc = q20.MPG.mean(), scale = q20.MPG.std()),3)

print('P(MPG<40)=',prob\_MPG\_less\_than\_40)

P(MPG<40)= 0.729

C) P (20<MPG<50)

Ans: prob\_MPG\_greater\_than\_20 = np.round(1-stats.norm.cdf(20, loc = q20.MPG.mean(), scale = q20.MPG.std()),3)

print('p(MPG>20)=',(prob\_MPG\_greater\_than\_20))

p(MPG>20)= 0.943

prob\_MPG\_less\_than\_50 = np.round(stats.norm.cdf(50, loc = q20.MPG.mean(), scale = q20.MPG.std()),3)

print('P(MPG<50)=',(prob\_MPG\_less\_than\_50))

P(MPG<50)= 0.956

prob\_MPG\_greaterthan20\_and\_lessthan50= (prob\_MPG\_less\_than\_50) - (prob\_MPG\_greater\_than\_20)

print('P(20<MPG<50)=',(prob\_MPG\_greaterthan20\_and\_lessthan50))

P(20<MPG<50)= 0.013000000000000012

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Ans: MPG follows normal distribution

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 doesn't follow normal distribution

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

interval, 60% confidence interval

ANS: # z value for 90% confidence interval

print('Z score for 60% Confidence Interval =',np.round(stats.norm.ppf(.05),4))

Z score for 60% Confidence Interval = -1.6449

# z value for 94% confidence interval

print('Z score for 60% Coincidence Interval =',np.round(stats.norm.ppf(.03),4))

Z score for 60% Coincidence Interval = -1.8808

# z value for 60% confidence interval

print('Z score for 60% Confidence Interval =',np.round(stats.norm.ppf(.2),4))

Z score for 60% Confidence Interval = -0.8416

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

Ans: # t score for 95% confidence interval

print ('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.025,df=24),4))

T score for 95% Confidence Interval = -2.0639

# t value for 94% confidence interval

print ('T score for 94% Confidence Inteval =',np.round(stats.t.ppf(0.03,df=24),4))

T score for 94% Confidence Inteval = -1.974

# t value for 99% Confidence Interval

print ('T score for 95% Confidence Interval =',np.round(stats.t.ppf(0.005,df=24),4))

T score for 95% Confidence Interval = -2.7969

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 à pt(tscore,df)

df à degrees of freedom

Ans: mean=(x)=260

Population mean=270(mu)

S.D=90

n= no of samples=18

t=-0.471

The probability is t<-.471 assuming the population mean is True.

The probability of the bulbs lasting less than 260 days on an Avg of mean life of 300 days.