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

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) | 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 | INTERVAL |
| Years of Education | RATIO |

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

Ans. Sample space(s)={(HHH),(HTT),(HTH),(HHT),(TTT),(THH),(THT),(TTH)}

Total no. of outcomes=8

Favourable outcomes for two heads(E1) =3 P(E1)=3/8

Favourable outcomes for one tail(E2)=3 p(E2)=3/8

Q4) Two Dice are rolled, find the probability that sum is

* Equal to 1
* Less than or equal to 4
* Sum is divisible by 2 and 3

Ans. S={(1,1)(1,2)(1,3)(1,4)(1,5)(1,6)(2,1)(2,2)(2,3)(2,4)(2,5)(2,6)(3,1)(3,2)(3,3)(3,4)(3,5)(3,6)(4,1)(4,2)(4,3)(4,4)(4,5)(4,6)(5,1)(5,2)(5,3)(5,4)(5,5)(5,6)(6,1)(6,2)(6,3)(6,4)(6,5)(6,6)}

a.0

b.1/6

c.24/36=2/3

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, Sample Space=S

Then, n(S) = Number of ways of drawing 2 balls out of 7 =7C2=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=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**

Ans.

**cars\_data\_mean**

Points 3.596563

Score 3.217250

Weigh 17.848750

cars\_data\_median

Points 3.695

Score 3.325

Weigh 17.710

#MODE

cars\_data\_Points\_mode

0 3.07

1 3.92

cars\_data\_Score\_mode

0 3.44

cars\_data\_Weigh\_mode

0 17.02

* 18.90

#VARIANCE

cars\_data.var()

Points 0.285881

Score 0.957379

Weigh 3.193166

#STD. DEV.

cars\_data.std()

Points 0.534679

Score 0.978457

Weigh 1.786943

#RANGE

cars\_data.describe()

|  |  |  |
| --- | --- | --- |
| **Points** | **Score** | **Weigh** |
| **count** | 32.000000 | 32.000000 | 32.000000 |
| **mean** | 3.596563 | 3.217250 | 17.848750 |
| **std** | 0.534679 | 0.978457 | 1.786943 |
| **min** | 2.760000 | 1.513000 | 14.500000 |
| **25%** | 3.080000 | 2.581250 | 16.892500 |
| **50%** | 3.695000 | 3.325000 | 17.710000 |
| **75%** | 3.920000 | 3.610000 | 18.900000 |
| **max** | 4.930000 | 5.424000 | 22.900000 |

cars\_data\_Points\_Range=cars\_data.Points.max()-cars\_data.Points.min()

cars\_data\_Points\_Range

2.17

cars\_data\_Score\_Range=cars\_data.Score.max()-cars\_data.Score.min()

cars\_data\_Score\_Range

3.9110000000000005

cars\_data\_Weigh\_Range=cars\_data.Weigh.max()-cars\_data.Weigh.min()

cars\_data\_Weigh\_Range

8.39999999999999



9

Q8) Calculate Expected Value for the problem below

* 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

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.**

|  |  |  |
| --- | --- | --- |
|  | speed | dist |
| skewness | -0.11750986144663393 | 0.8068949601674215 |
| kurtosis | -0.5089944204057617 | 0.4050525816795765 |

**We calculated the skewness and kurtosis of given data, we found that skewness of speed is -0.117 so it is -ve skewness and the data is fairly symmetrical. The skewness of dist is 0.806 so it is positively skewness and the data is moderately skewed. We found that kurtosis of speed is -0.508 so it is platykurtic that it is lack of outliers. The kurtosis of dist is 0.405 so it is platykurtic it has profusion outliers.**

**SP and Weight(WT)**

**Use Q9\_b.csv**

**Ans.**

|  |  |  |
| --- | --- | --- |
|  | SP | WT |
| skewness | 1.61145 | -0.61475 |
| kurtosis | 2.977329 | 0.950291 |

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



Ans. Right side skewed or positively skewed



Ans. : the interface for this box plot is positively 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.

Sample mean = 200

Sample SD = 30

n = 2000

➢ Avg. weight of Adult in Mexico with 94% CI

stats.norm.interval(0.94,200,30/(2000\*\*0.5)

(198.738325292158, 201.261674707842)

➢ Avg. weight of Adult in Mexico with 98% CI

stats.norm.interval(0.98,200,30/(2000\*\*0.5)) (198.43943840429978, 201.56056159570022)

➢ Avg. weight of Adult in Mexico with 96% CI

stats.norm.interval(0.96,200,30/(2000\*\*0.5))

(198.62230334813333, 201.37769665186667)

**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**

* Find mean, median, variance, standard deviation.
* What can we say about the student marks?

Ans. : 1) mean=sum of the terms/number of terms => 738/18=41

Median = 1

Variance = 22.705

Standard deviation = 4.764

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

Ans:Normalized skewness

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

Ans:Right skewness

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

Ans. Left skewness

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

Ans. Sharp peak in the plot less gap between tails and to X-axias

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

Ans. : Border peak under the curve and more gap between tails and x-axis

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



What can we say about the distribution of the data?

Ans. : The data is distributed in De-assigned format

What is nature of skewness of the data?

Ans. Left side skewed

What will be the IQR of the data (approximately)?   
Ans. Q3-Q1

= 18-10

= 8 is IQR  
  
  
Q19) Comment on the below Boxplot visualizations?



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

Ans. The box plot 1 designed with range = 3 , The second one range is = 1.5

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

* P(MPG>38)

Ans: 1**-**stats**.**norm**.**cdf(38,cars**.**MPG**.**mean(),cars**.**MPG**.**std())

0.3475939251582705

* P(MPG<40)

Ans. stats**.**norm**.**cdf(40,cars**.**MPG**.**mean(),cars**.**MPG**.**std())

0.7293498762151616

* P (20<MPG<50)

Ans. stats**.**norm**.**cdf(0.50,cars**.**MPG**.**mean(),cars**.**MPG**.**std())**-**stats**.**norm**.**cdf(0.20,cars**.**MPG**.**mean(),cars**.**MPG**.**std())

1.2430968797327613e-05

Q 21) Check whether the data follows normal distribution

* Check whether the MPG of Cars follows Normal Distribution

# plotting distribution for cars\_data

sns.distplot(cars\_data.MPG)

plt.ylabel('HP') Dataset: Cars.csv



-Normally distributed

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

Dataset: wc-at.csv

Ans.

*# plotting distribution for Waist Circumference (Waist)*

sns**.**distplot(wcat**.**Waist)

plt**.**ylabel('density')

-normally distributed



# plotting distribution for Adipose Tissue (AT)

sns.distplot(wcat.AT)

plt.ylabel('density')

-normally distributed



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

Ans. 90%= 1.645

94%= 1.880

60% = 0.253

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

Ans. To compute the 95% confidence interval, start by computing the mean and standard error: M = (2 + 3 + 5 + 6 + 9)/5 = 5. σM = = 1.118. Z.95 can be found using the normal distribution calculator and specifying that the shaded area is 0.95 and indicating that you want the area to be between the cutoff points

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. t - statistics for the data is given as follows:

t = (x-μ ) /  (s/**√** n)

x = mean of the sample of bulbs = 260

mu= population mean = 270

s = standard deviation of the sample = 90

n = number of items in the sample = 18

t=(260-270) / (90/**√**18)

t=-0.471

Submitted by:

MUHAMMED ADILSHA TP