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

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 | Ordinal |
| 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?

N={HHH,THH,TTH,TTT,HTT,HHT,THT,HTH}

P(n)=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
4. n(S)=36

the sum is equal to1=0

p=n(c)/n(S) =0/36=0

b)

A={(1,3),(3,1),(2,2),(1,1),(1,2),(2,1)}

P(A)=n(A)/n(S)

=6/36

=0.16

C) n(B)={(1,1),(1,2),(1,3),(1,5)

(2,1),(2,2),(2,4),(2,6)

(3,1),(3,3)(3,5),(3,6)

(4,2),(4,4),(4,5),(4,6)

(5,1),(5,3),(5,4),(5,5)

(6,2),(6,3),(6,4),(6,6)}

P(A)=n(A)/n(s)

P(A)=24/36

P(A)=0.6666

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?

T(B)=(2+3+2)

T(B)=7

n(S)=number of drawing 2 balls out of 7

n(S)=(7\*6\*5\*4\*3\*2\*1)/(2\*1)(5\*4\*3\*2\*1)

n(S)=(7\*6)/(2\*1)

n(S)=42/2

n(S)=21

n(E)=number of ways of drawing 2 balls out of(2+3) balls

=(5\*4)/(2\*1)

=20/2

=10

P(E)=n(E)/n(S)

P(E) =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

Sol: Expected number of candies for a randomly selected child =n(S)

n(S)=1\*.015+4\*0.20+3\*0.65+5\*0.005+6\*0.01+2\*0.120

n(S)=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.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Points | Score | Weigh |
| Mazda RX4 | 3.9 | 2.62 | 16.46 |
| Mazda RX4 Wag | 3.9 | 2.875 | 17.02 |
| Datsun 710 | 3.85 | 2.32 | 18.61 |
| Hornet 4 Drive | 3.08 | 3.215 | 19.44 |
| Hornet Sportabout | 3.15 | 3.44 | 17.02 |
| Valiant | 2.76 | 3.46 | 20.22 |
| Duster 360 | 3.21 | 3.57 | 15.84 |
| Merc 240D | 3.69 | 3.19 | 20 |
| Merc 230 | 3.92 | 3.15 | 22.9 |
| Merc 280 | 3.92 | 3.44 | 18.3 |
| Merc 280C | 3.92 | 3.44 | 18.9 |
| Merc 450SE | 3.07 | 4.07 | 17.4 |
| Merc 450SL | 3.07 | 3.73 | 17.6 |
| Merc 450SLC | 3.07 | 3.78 | 18 |
| Cadillac Fleetwood | 2.93 | 5.25 | 17.98 |
| Lincoln Continental | 3 | 5.424 | 17.82 |
| Chrysler Imperial | 3.23 | 5.345 | 17.42 |
| Fiat 128 | 4.08 | 2.2 | 19.47 |
| Honda Civic | 4.93 | 1.615 | 18.52 |
| Toyota Corolla | 4.22 | 1.835 | 19.9 |
| Toyota Corona | 3.7 | 2.465 | 20.01 |
| Dodge Challenger | 2.76 | 3.52 | 16.87 |
| AMC Javelin | 3.15 | 3.435 | 17.3 |
| Camaro Z28 | 3.73 | 3.84 | 15.41 |
| Pontiac Firebird | 3.08 | 3.845 | 17.05 |
| Fiat X1-9 | 4.08 | 1.935 | 18.9 |
| Porsche 914-2 | 4.43 | 2.14 | 16.7 |
| Lotus Europa | 3.77 | 1.513 | 16.9 |
| Ford Pantera L | 4.22 | 3.17 | 14.5 |
| Ferrari Dino | 3.62 | 2.77 | 15.5 |
| Maserati Bora | 3.54 | 3.57 | 14.6 |
| Volvo 142E | 4.11 | 2.78 | 18.6 |

Sol: For points column

Mean=115.09/32 , Median=3.62 , Mode=’Numeric’ , variance= 0.285

=3.5965

Standerd deviation = 0.5338

|  |
| --- |
|  |

For Score:Mean=102.9/32 , Median=3.32, Mode=’Numeric’ , variance= 0.957

Mean=3.217

Standard deviation=0.9784

For Score:Mean=571/32 , Median=17.71, Mode=’Numeric’ , variance= 3.193

Mean=17.84

Standard deviation=1.786

Note:’Point’ &’Score’s mean are almost near to each other

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?

Sol: Mean=1308,Mue=n(x)/n-1=1308/8=163.5-

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

**Cars speed and distance**

**Use Q9\_a.csv**

**Sol:**[**https://colab.research.google.com/drive/1J4W7-WUv71FHtr5uF29XwAfINocoK6pU**](https://colab.research.google.com/drive/1J4W7-WUv71FHtr5uF29XwAfINocoK6pU)

**The speed skewnees for speed is negative,so it is left skewed**

**So that the magnitude is greater then 0 it is slightly left skewed.**

**And the distance value is positive and it greater than 0**

**So it is right skewed**

**SP and Weight(WT)**

**Use Q9\_b.csv**

[**https://colab.research.google.com/drive/1ineVMtAD09qdiSFWvdk8I4v2zE5IeDhv**](https://colab.research.google.com/drive/1ineVMtAD09qdiSFWvdk8I4v2zE5IeDhv)

**All the value of the SP and WT are right skewed and all are positive**

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



The most of the data are connected in the range b/w 50-100 with the frequency 200

The least range of weight is 400somewere around 0-10

So the expected value is 75 for the above distribution

The above histogram shows it is right heavily skewed

here the meadian is less than the mean so its right skewed and here the outliers are upperside of the box plot and the data points between q1 and bottom point are very less.



**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?

<https://colab.research.google.com/drive/1IX5CBHAnTPzzfTmGGaDiDCOISNLYEPSz>

**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.

Sol: <https://colab.research.google.com/drive/1hBgiOkRn_xZOgmt3ZpVYEzf4V6KuDkVM>

1. What can we say about the student marks?

Sol: The marks of the student are likely in passing range

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

Sol: Symetrical

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

Sol: Right skewed

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

Sol: Left skewed

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

Sol: The data is normally distributed and the kurtosis value is 0

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

Sol:The data distribution has lighter tails and flatter peaks then normal

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



What can we say about the distribution of the data?

Sol: assume that box plot is about ages of the student in a school.

50% of the people are above 10 years old and remaining are less and students whose age is above 15 are approx 40%

What is nature of skewness of the data?

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

IRQ = 3rd Qurtaile – 1st Qurtaile

IQR=18-10

Approximately = 8

Q19) Comment on the below Boxplot visualizations?



Sol:By observing the above whisker’s plot we came to conclude that 2)boxplot is higher than 1)boxplot, Hence mean and median are equally distributed and it is symetrical

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

sol:for a&b ans

<https://colab.research.google.com/drive/1nPELVwExwlLubIErDTJGIeJjF_SZ-hMY>

* 1. P(20>MPG>50)=P(MPG>20)-P(MPG>50)

=0.057-0.044

=0.013

Q 21) Check whether the data follows normal distribution

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

Sol: <https://colab.research.google.com/drive/1bwn8-ryJ2zr0nDEmuX-8ZVPOgkKUR3v9>

The distribution is not normal it is left skewed

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

Dataset: wc-at.csv

Sol: <https://colab.research.google.com/drive/1wYGjUqrKyubtqk13uok3-Pg3u2zIIayx>

* In Waist lightly right skewed in the Density graph
* In AT Highly right skewed in the Density graph

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

Sol:

<https://colab.research.google.com/drive/1cDCYVPvoDWM-yV7WXYYmdJM6kEGtUZpv>

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

Sol:<https://colab.research.google.com/drive/1VBcNIRSHWVt7XLH9i0q7hXFvGc2YffCu>

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

<https://colab.research.google.com/drive/17TJpDVHciyB8Aez5AtCb6xBWZbqena5D>

t\_score=0.48=48%