**Q1)** Identify the Data type for the Following:

| **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 | Nominal |
| Number of kids | Discrete |
| Number of tickets in Indian railways | Discrete |
| Number of times married | Discrete |
| Gender (Male or Female) | Ordinal |

**Q2)** Identify the Data types, which were among the following

Nominal, Ordinal, Interval, Ratio.

| **Data** | **Data Type** |
| --- | --- |
| Gender | Ordinal |
| 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) | Ordinal |
| Sales Figures | Ratio |
| Blood Group | Nominal |
| Time Of Day | Ordinal |
| Time on a Clock with Hands | Ordinal |
| Number of Children | Ratio |
| Religious Preference | Nominal |
| Barometer Pressure | Ratio |
| SAT Scores | Interval |
| Years of Education | Ratio |

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

**Answer :** Let “A” is the event in which 3 coins are tossed.

Then, **A = {HHH, HHT, HTT, HTH, THT, TTH, THH, TTT}**

n(A) = 8

Now, “B” is the set of getting 2 heads and 1 tail

**B = {HHT, HTH,THH}**

n(B) = 3

P = probability of getting 2 heads and 1 tail from event A

= **P(B) = n(B) / n(A)**

= **3/8 = 0.375 = 37%**

**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

**Answer :** Let event A is rolling of 2 dice.

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)}

n(A) = 36

Now “a” is event of getting sum is equal to 1

a = {}

n(a) = 0

Hence P(a) = 0/36 = 0%

the probability of getting only one output is always zero because it is a random variable.

Now, “b” is event of getting sum is Less than or equal to 4

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

n(b) = 6

Hence, P(b) = 6/36 = 0.166 = 16%

Now, “c” is the event of getting Sum which is divisible by 2 and 3

c = {(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)}

n(c) = 24

Hence, P(c) = 24/36 = 0.666 = 66%

**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?

**Answer :** Here, S={RR,RG,RB,GR,GG,GB,BR,BG,BB}

n(S) = 9

Now let ‘T’ be the event of getting none of the ball drawn is blue

T = {RR, RG GR GG}

n(T) = 4

P(T) = 4/9 = 0.444 = 44%

**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

**Answer :**

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**

**Answer :**

Mean(Points) = 3.596563

Mean(Score) = 3.217250

Mean(Weigh) = 17.848750

Median(Points) = 3.695

Median(Score) = 3.325

Median(Weigh) = 17.710

Mode(Points) = 3.92

Mode(Score) = this column is having uniform data there is no mode

Mode(Weigh) = 18.90

Var(Points) = 0.285881

Var(Score) = 0.957379

Var(Weigh) = 3.193166

std(Points) = 0.534679

std(Score) = 0.978457

std(Weigh) = 1.786943

Range(Points) = 2.17

Range(Score) = 3.9110000

Range(Weigh) = 8.399999

Inferences are plotted using boxplots in python in the following notebook along with code for calculating mean, median, mode, variance, standard deviation, range.

link for notebook = <https://colab.research.google.com/drive/1SNmFTUrJV9rq1mgfG2Po9KNVS3ocnXG9?usp=sharing>

**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?

**Answer :** The expected weight of randomly selected person can be calculated by calculating the mean(average weight) of all the patients.

Here, X = [108, 110, 123, 134, 135, 145, 167, 187, 199]

n(X) = 9

mean(X) = sum of weights of all the patients/ total number of patients

= 108 + 110 + 123 + 134 + 135 + 145 + 167 + 187 + 199 / 9

= 1308/9

= 145.33

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

Cars speed and distance

**Use Q9\_a.csv**

**Answer:**

skewness(speed) = -0.11395477012828319

kurtosis(speed) = -0.5771474239437371

skewness(distance) = 0.7824835173114966

kurtosis(distance) = 0.24801865717051808

link for notebook = <https://colab.research.google.com/drive/1SNmFTUrJV9rq1mgfG2Po9KNVS3ocnXG9?usp=sharing>

**SP and Weight(WT)**

**Use Q9\_b.csv**

skewness(SP) = 1.5814536794423764

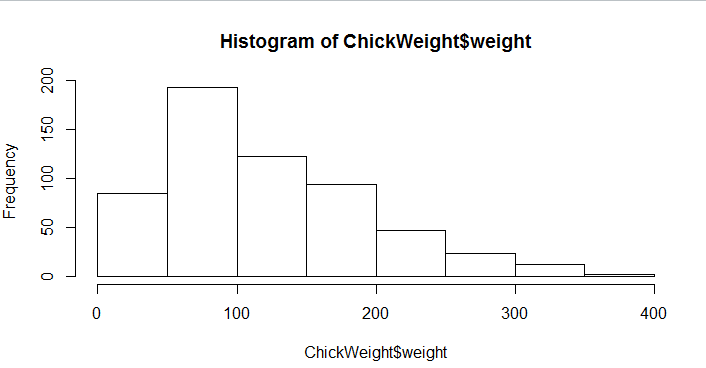
kurtosis(SP) = 2.7235214865269244

skewness(WT) = -0.6033099322115126

kurtosis(WT) = 0.8194658792266849

link for notebook = <https://colab.research.google.com/drive/1SNmFTUrJV9rq1mgfG2Po9KNVS3ocnXG9?usp=sharing>

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





**Answer :** by watching the above histogram we can see that the majority of chick weight$weight data lies between 0-200. we can observe that there is a right skewness present in our dataset. It shows the presence of outliers in the dataset. And by watching the boxplot we conclude that the mean is affected by presence of outliers in the dataset and we can clearly see there are dots which represent outliers in the dataset.

**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?

**Answer :** Confidence interval is calculated in Python

Mean at 94% confidence interval is (14.5761, 256.4238)

Mean at 98% confidence interval is (130.2095, 269.7904)

Mean at 96% confidence interval is (138.3875, 261.6124)

Notebook link = <https://colab.research.google.com/drive/1SNmFTUrJV9rq1mgfG2Po9KNVS3ocnXG9?usp=sharing>

**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?

**Answer :**

1. Mean = 41 , Median = 40.5, Mode = 41, Variance = 25.529412, Standard Deviation = 5.052664
2. Majority of students scored around 41 marks out of which most of the students scored exactly 41 marks. Looking towards the marks data the spread of data is 25.529412 with standard deviation as 5.052664

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

**Answer :** When Mean = Median there is no skewness present. There is a perfect symmetrical distribution.

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

**Answer :** skewness with right sided tail

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

**Answer** : Skewness with left sided tail

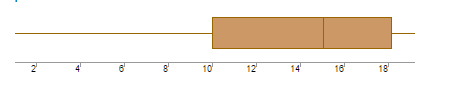
**Q16)** What does positive kurtosis value indicate for a data ?

**Answer :** It represents the Leptokurtic Kurtosis with a thick tail where the majority of datapoints are at tails instead around the mean.

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

**Answer :** It means the curve will be flatter and spreaded.

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



**A .** What can we say about the distribution of the data?

**Answer :** The data is not normally distributed. B. What is the nature of skewness of the data?

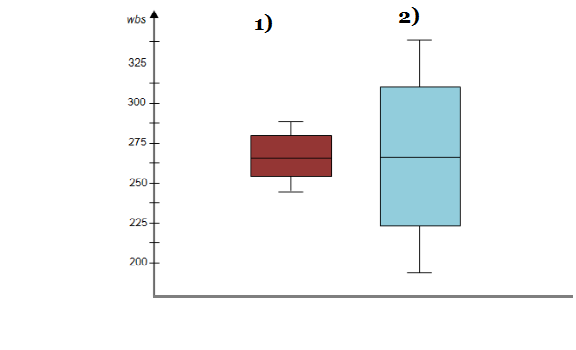
**C.** What is the nature of skewness of data

**Answer :** There is skewness present on the left side of distribution (Negative Skewness).

**C.** What will be the IQR -of the data (approximately)?

**Answer :** IQR = Q3-Q1 = 18-10 = 8

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



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

**Answer** : Boxplot 1 represents column 1 and Boxplot 2 represents column 2 of the dataset. The data of both the columns is normally distributed. There are no outliers present in both the columns. The median of both the columns is the same which lies between 275 to 250.

**Q 20) A .** Calculate probability from the given dataset for the below cases

Data \_set: Cars.csv

**B.** Calculate the probability of MPG of Cars for the below cases.

MPG <- Cars$MPG

* 1. P(MPG>38)

**Answer :** P(MPG>38) = 1-stats.norm.cdf(38,34.422076,9.131445) = 0.35 = 35%

* 1. P(MPG<40)

**Answer :** P(MPG<40) = stats.norm.cdf(40,34.422076,9.131445) = 0.73 = 73%

**c.**  P (20<MPG<50)

**Answer :** P(20<MPG<50) = stats.norm.cdf(50,34.422076,9.131445) - stats.norm.cdf(20,34.422076,9.131445) = 0.90 = 90%

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

1. Check whether the MPG of Cars follows Normal Distribution

Dataset: Cars.csv

**Answer :** MPG of cars is not following Normal Distribution. I have plotted Histogram, Boxplot and density plot

Notebook link = <https://colab.research.google.com/drive/1SNmFTUrJV9rq1mgfG2Po9KNVS3ocnXG9?usp=sharing>

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

Dataset: wc-at.csv

**Answer :** By plotting the density plot and histogram we can conclude that the Adipose Tissue (AT) and Waist Circumference(Waist) from the wc-at dataset does not follow Normal Distribution.

Notebook link = <https://colab.research.google.com/drive/1SNmFTUrJV9rq1mgfG2Po9KNVS3ocnXG9?usp=sharing>

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

**Answer :** Notebook link = <https://colab.research.google.com/drive/1SNmFTUrJV9rq1mgfG2Po9KNVS3ocnXG9?usp=sharing>

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

**Answer :** Notebook link = <https://colab.research.google.com/drive/1SNmFTUrJV9rq1mgfG2Po9KNVS3ocnXG9?usp=sharing>

**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

**Answer :** µ = 270 , n = 18, s = 90, df = 17, α = 0.05, sample mean = 260

As we have not provided the population data we have only sample data, here we have to do the t-test.

Probability that 18 randomly selected bulbs have an average life of no more than 260 days: 32.16725356709836 %

Notebook link = <https://colab.research.google.com/drive/1SNmFTUrJV9rq1mgfG2Po9KNVS3ocnXG9?usp=sharing>