

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

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	Ratio
Height	Ratio
Type of living accommodation	Ordinal
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	Nominal
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: Total Possible Events: 8, No of desired events: 3

So the probability is  $P = 3/8$

**P = 0.375**

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

- a) Equal to 1
- b) Less than or equal to 4
- c) Sum is divisible by 2 and 3

ANS: **a) 0** because if two dices are rolled minimum result will be two

b) Total Possible Events: 36,

No of desired events: 6

So probability =  $6/36 = \mathbf{0.1666}$

C) Total Possible Events: 36,

No of desired events: 24

So probability  $P = 24/36 = \mathbf{0.66}$

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?

ANS:

Total number of balls =  $(2 + 3 + 2) = 7$

Let  $S$  be the sample space.

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

$$= {}^7C_2 = 21$$

Let  $E$  = Event of drawing 2 balls, none of which is blue.

$\therefore n(E)$  = Number of ways of drawing 2 balls out of  $(2 + 3)$  balls.

$$= {}^5C_2 = 10$$

$$\therefore P(E) = n(S) / n(E)$$

**So the answer is = 21/10**

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

$$\text{ANS:} = 1 \times 0.015 + 4 \times 0.20 + 3 \times 0.65 + 5 \times 0.005 + 6 \times 0.01 + 2 \times 0.120 = 3.09$$

**Expected number of candies for a randomly selected child  $\approx 3.09$**

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:**

Ans is in Jupiter notebook “Q. which contain Data sets” named file

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 of the Weight of that patient =  $\sum x/n$

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

= 145.33

So the Expected Value of the Weight of that patient is 145.33

**Q9) Calculate Skewness, Kurtosis & draw inferences on the following data Cars speed and distance**

**Use Q9\_a.csv**

**Ans is in Jupiter notebook “Q. which contain Data sets” named file**

**Use Q9\_b.csv**

Ans is in Jupiter

notebook “Q.

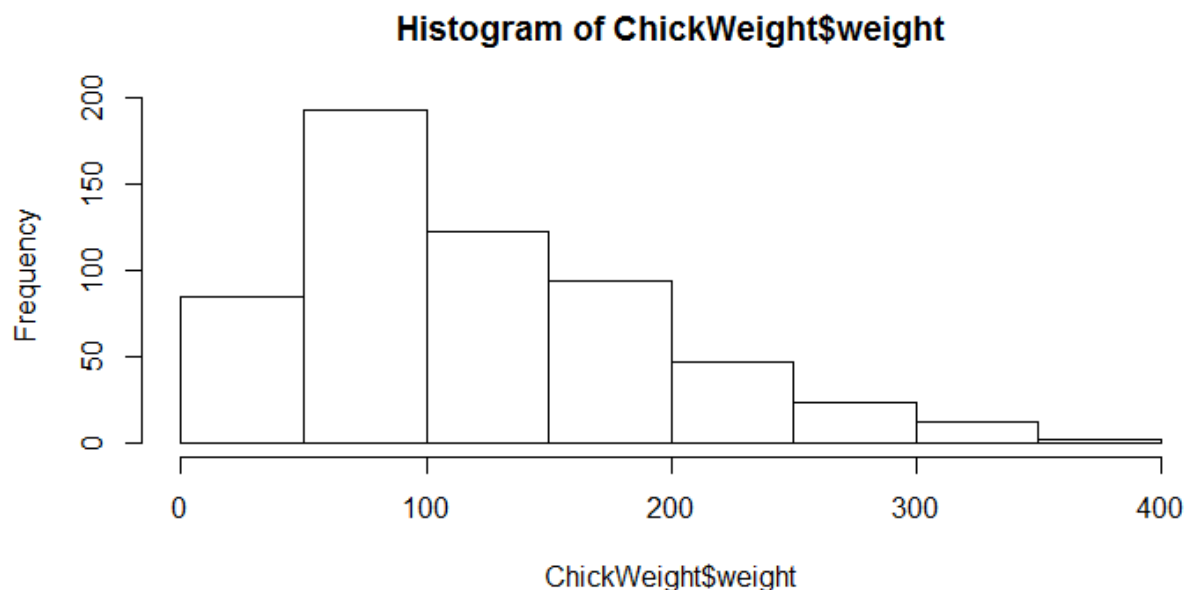
which contain

Data sets”

named file

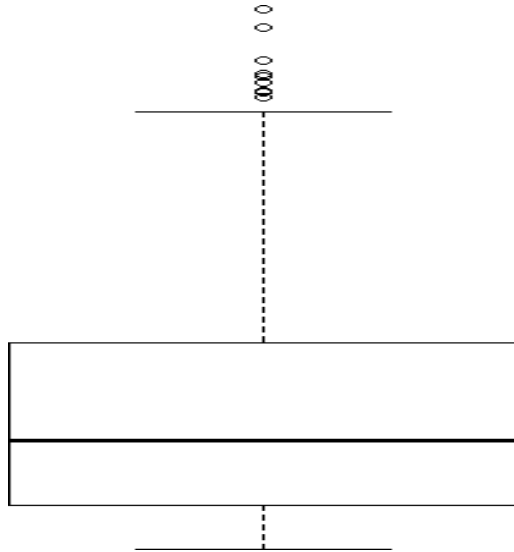
**Q10) Draw inferences about the following boxplot & histogram**

---



ANS :

- a) Majority of the Chicks has weight in range 50 – 150
- b) The data is positively Skewed



ANS:

- a) Data has outliers
- b) Data 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:**

	94%	98%	96%
Upper	201.04	201.38	201.17
Lower	198.96	198.62	198.83

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.

Ans is in Jupiter notebook "Q. which contain Data sets" named file

2) What can we say about the student marks? **ANS:**

- a) Mean is bigger than median that mean data is slightly skewed Towards right.
- b) Data don't have outlier
- c) Majority of the students scored between 35 – 45 Marks

Q13) What is the nature of skewness when mean, median of data are equal? **ANS:** If the distribution is symmetric, then the mean is equal to the median, and the distribution has **zero** skewness. If the distribution is both symmetric and unimodal, then the mean = median = mode.

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

a) **ANS:** If the mean is greater than the median, the distribution is positively skewed.

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

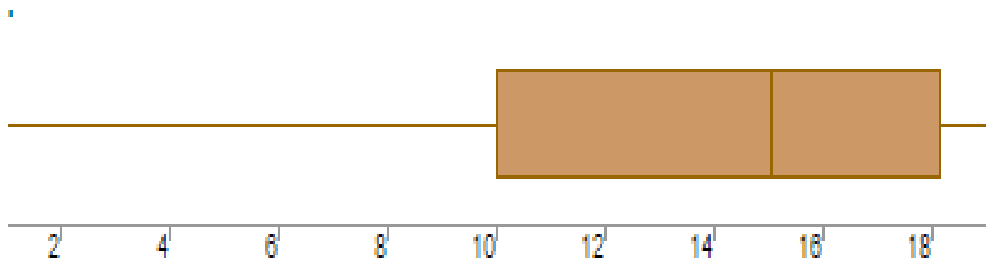
**Ans:** If the median is greater than the mean on a set of test scores, so the data are "skewed to the left.



Q16) What does positive kurtosis value indicates for a data ? **ANS: Positive values of kurtosis indicate that distribution is peaked and possesses thick tails**

Q17) What does negative kurtosis value indicates for a data? **ANS: Negative excess values of kurtosis ( $<3$ ) indicate that a distribution is flat and has thin tails.**

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



What can we say about the distribution of the data?

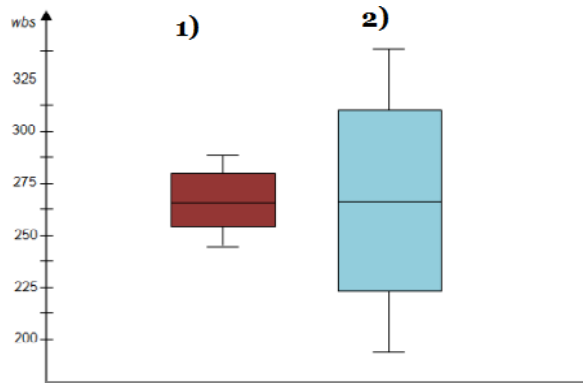
**ANS : The data is not symmetric. Data is more concentrated towards right side**

What is nature of skewness of the data? **ANS : left skewed**

What will be the IQR of the data

(approximately)? **ANS: IQR data is  $(18-10 = 8)$**

Q19) Comment on the below Boxplot visualizations?



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

ANS:

- 1) Median of the two boxplots are same = 260
- 2) The boxplots are not skewed in +ve or -ve directions
- 3) No outliers in both boxplots

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

- a.  $P(\text{MPG} > 38)$
- b.  $P(\text{MPG} < 40)$
- c.  $P(20 < \text{MPG} < 50)$

Ans is in Jupiter notebook "Q. which contain Data sets" named file

Q 21) Check whether the data follows normal distribution

- a) Check whether the MPG of Cars follows Normal Distribution Dataset: Cars.csv

Ans is in Jupiter notebook “Q. which contain Data sets” named file

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

Dataset: wc-at.csv

ANS: Both AT and Waist doesn't follow Normal Distribution

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

ANS: stats.norm.ppf(0.90)

90%	$\pm 1.711$
94%	$\pm 1.828$
60%	$\pm 2.492$

Formula:

`np.round(stats.norm.ppf(.95),2)`

`np.round(stats.norm.ppf(.97), 2)`

`np.round(stats.norm.ppf(.80), 2)`

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

ANS:

95%	$\pm 2.060$
96%	$\pm 2.167$
99%	$\pm 2.787$

Formula:

`np.round(stats.t.ppf(0.975, df = 24), 2)`

```
np.round(stats.t.ppf(0.98, df = 24), 2)
```

```
np.round(stats.t.ppf(0.995, df = 24), 2)
```

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:

critical value = -0.4714,

Degree of freedom = 17

probabilty for average life of no more than 260 days is 0.32 = 0.3216725 Formulas:

```
t_value = (260 - 270)/(90/np.sqrt(18)) print(np.round(t_value, 2))
```

```
print(np.round(stats.t.cdf(t_value, df=17), 2))
```