

# **IT302 Assignment 2**

NAME: SUYASH CHINTAWAR

ROLL NO.: 191IT109

TOPIC: RANDOM VARIABLES AND  
CORRELATION

**Note:**

- 1) The colab link has been attached below. Please use nitk id to access the link. After opening the link, if it opens in drive, click on “Open with Google Colaboratory” to view the complete code.
- 2) Only output screenshots have been attached. Code for the same can be found in the colab notebook.

**Colab notebook link:**

[https://colab.research.google.com/drive/1ISf514HuXUbjP\\_EHDvCh4goeH4gabBHV?usp=sharing](https://colab.research.google.com/drive/1ISf514HuXUbjP_EHDvCh4goeH4gabBHV?usp=sharing)

**Q1.** Assume the students in a class of 100 were classified according to Gender (G) and Covid (C) as follows. Where Cp, Cr, and N denote the covid status: “now Covid Positive”, “Covid Positive but Recovered” and “Never Affected by Covid”, respectively.

	Cp	Cr	N	Total
Boys	20	32	8	60
Girls	10	5	25	40
Total	30	37	33	100

Table 1. Distribution 1

Write a program to find...

1. Probability that a randomly selected student is a boy.

Program Output:

Probability that a randomly selected student is a boy: 0.6

Explanation: As the number of boys is 60,  $n(E)$  where ‘E’ is our event is 60. Sample space,  $n(S)$ , here is 100, i.e. the total number of students in the class.

2. Probability that a randomly selected student is a boy affected by Covid Positive.

Program Output:

Probability of a boy and covid positive: 0.2

Explanation:  $n(E) = 20$ ,  $n(S) = 100$ .

3. Probability that a randomly selected student is either a Covid Positive or Recovered.

Program Output:

Probability of either covid positive or recovered: 0.67

Explanation:  $n(E) = 20+32+10+5$ ,  $n(S) = 100$ .

4. Probability that a randomly selected student is a girl who is a Covid Positive or Recovered.

Program Output:

Probability of girl either covid positive or recovered: 0.15

Explanation:  $n(E) = 10+5$ ,  $n(S) = 100$ .

**Q2.** Consider these two more distributions also and calculate the expectation, variance for a boy/girl 'Never affected by covid'.

	Cp	Cr	N	Total
Boys	16	24	10	50
Girls	15	15	20	50
Total	31	39	30	100

Table 2. Distribution 2

	Cp	Cr	N	Total
Boys	16	10	9	35
Girls	15	15	35	65
Total	31	25	44	100

Table 3. Distribution 3

## SOLUTION:

### Program Output:

Expectation of the number of students affected by covid: 35.666666666666664  
Expectation that a boy/girl is affected by covid: 0.3566666666666667

Variance: 0.0054333333333333333

Explanation: Expectation is found out using the number of students never affected by covid(both boys and girls) and multiplying it with the probability of the event occurring, i.e.  $\frac{1}{3}$  as in each distribution. Variance of the probability that a boy/girl is never affected by covid is taken, which are, [0.33,0.3,0.44].

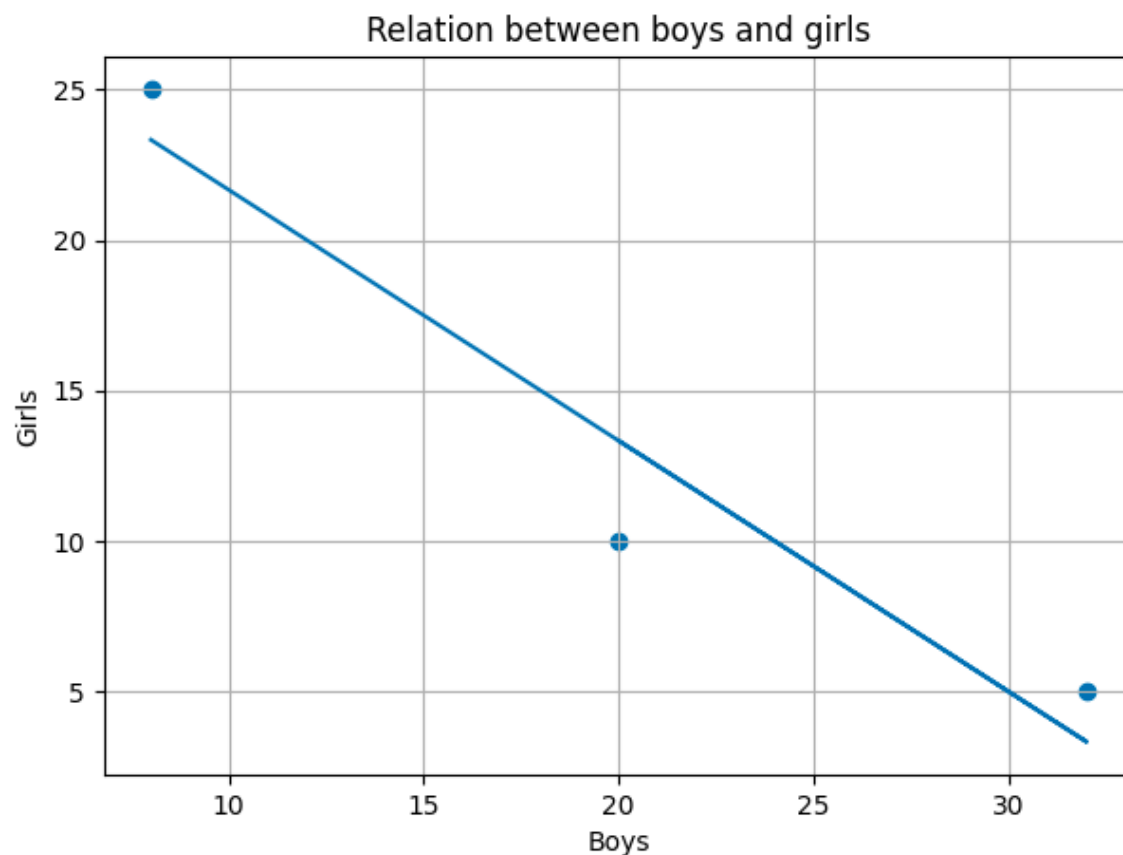
**Q3.** Plot the values in a graph and visualize the correlation between Boys and Girls.

## SOLUTION:

### Program Output:

Correlation: -0.9607689228305228

### Graph/Plot:



Explanation: To find correlation, distribution 1 has been used, Boys data as one array and girls data as another. Each containing three values as given in Table 1. Correlation coefficient has been calculated which is -0.96. Graph has been plotted to visualize the relationship between boys and girls. Polyfit in matplotlib has been used to find the best linear curve fitting on the given data.

**Q4.** Use this real-time dataset from [Kaggle](#) to find the correlation of covid cases and vaccination status and interpret/plot/visualize the results.

**SOLUTION:**

Dataset Description: Out of the three csv files in the dataset, 2 of them have been used namely, 'covid\_vaccine\_statewise.csv' and 'StateswiseTestingDetails.csv'. The first csv has 24 columns, dimensions are (7845,24), of which two have been used to extract the vaccinated statuses of each state, namely, ['State','Total Doses Administered']. The latter csv file has 5 columns, dimensions are (16336,5), two of which have been used to extract the affected or positive status of covid in each state, namely, ['State','Positive']

Final Output:

**Correlation: 0.5727398627776005**

Graphs/Plots:

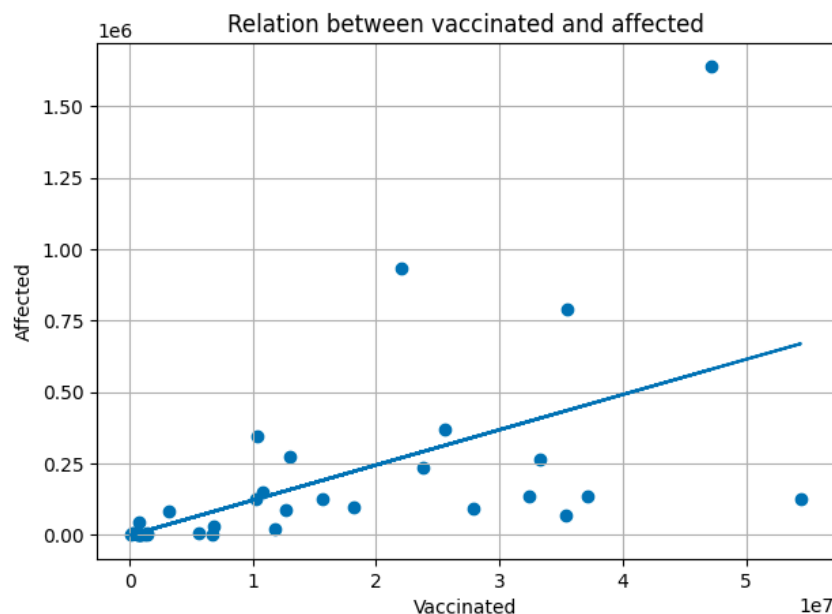


Fig. Graph 1

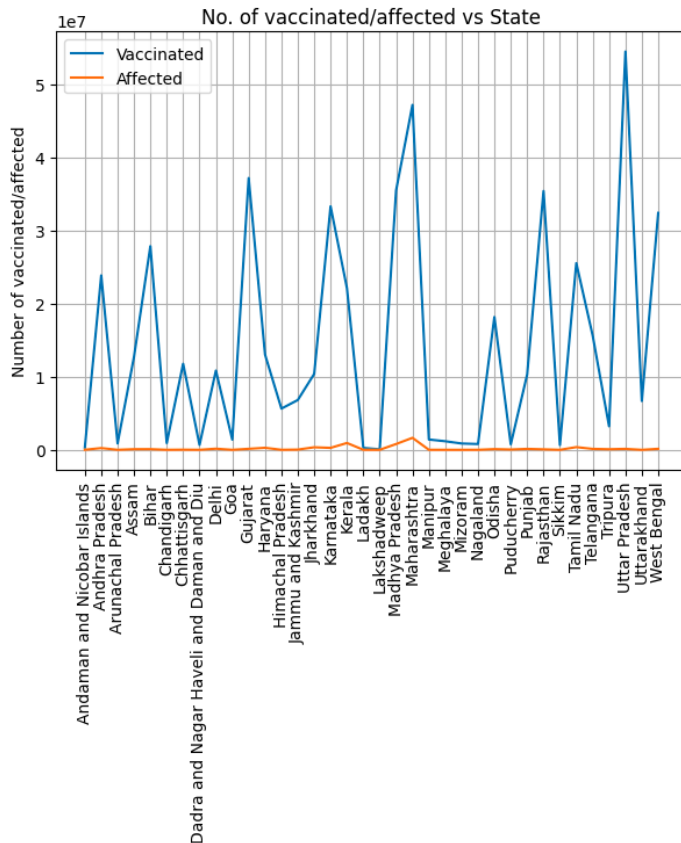


Fig. Graph 2

Explanation and Conclusion: Statewise vaccinated and affected data has been taken for the latest possible date, which gives us the cumulative data till that date for a particular state. Two arrays have been formed of the size of the number of states and union territories, which is 36. The corresponding values in the two arrays 'vaccinated' and 'affected' show the values for a particular state. Correlation coefficient obtained of these arrays is 0.572. It can be concluded that there is some relation between vaccinated and affected values as correlation coefficient of 1 tells us that the two arrays are completely related (i.e they are same). To visualize this relation, graph 1 has been used. The polyfit function in matplotlib shows us the best linear curve fitting the two arrays. Graph 2 shows us the number of vaccinated and affected people vs State.