Introduction

The problem that we will work on during this project is about Female Education. To place it specifically, we will create a model which will predict Female Education growth in Asia, and that we shall include the four major countries of Asia during this small research.

The model technique we wish to use is simple regression using the least-squares method. Now one may need the question of what exactly simple regression is. In statistics, rectilinear regression could be a linear approach to modeling the connection between a variable and one or more independent variables. Within the case of 1 variable quantity, it's called simple statistical regression. For quite one variable quantity, the method is termed multiple simple regression.

And to see the worth of the associated variable, that might provide a minimum error within the dataset. During this process, we use the least-squares method.

Yet, to discover the mistake first, we need to utilize the quadratic misfortune work that we have learned in our specialty course known as Intelligent System. This misfortune capacity would decide how the anticipated worth is not quite the same as the genuine worth. The justification of squaring it is to stay away from negative qualities from the absolute blunder.

As an inhabitant of the Indian subcontinent (four of us), this conversation is truly significant. Till now, a vast population share of the education from the Indian subcontinent have a place with the Asian Education. Also, to decide the future financial state of female Education, and Education sway on females of Asia it is really important to decide the Education pace of females contrasted with other female Educations around there. Given the ebb and flow circumstance made by the counter Islamist government in India, this examination is truly significant. By this examination, we will actually want to decide the populace development pace of Females and anticipate the number of inhabitants in Females in the following 50 years. There is no uncertainty about the effect the Female Educations have on the sociopolitical scene of an area.

Problem Statement

In underdeveloped countries, a substantial percentage of school-aged girls do not attend school. In South Asia, around a third of girls do not attend school, and barely a quarter of females attend basic school in some areas. Women's education and reproductive health can be improved by eliminating gender inequalities in school attendance.

Educating girls is one of the most effective ways for countries to achieve their economic potential, improve health, reduce conflict, and save lives. Education is also a basic human right. At a time when growing inequality in a number of Asian countries is profoundly affecting girls' access to education, it is critical that donors investing in girls' education include Asia in their geographic focus. In a number of Asian countries, girls cannot complete their schooling due to early marriage, poverty, gender norms, and other sociocultural factors. This has a downstream effect of limiting young women's participation in fields such as science, technology, engineering, mathematics, and applied sciences (STEM), thus narrowing the scope of job opportunities for them.

Our project centers around projecting Female Education dependent on a public dataset utilizing least squares technique and anticipating its populace for the year 2025, 2027 and 2030 to decide if there would be any exceptional change in the informed scene of the world.

Objectives

- To create a linear model that can give future projections of the population of the four major Educations in Asia for the year 2030, 2050 and 2070.

- To compare the prediction for the four Countries' Female Education Growth with one another.

- To compare our prediction of the world's largest Education after 50 years with that of Pew Research Center

- To conduct a systematic review of the available data from international organizations and regional registries to explore the association between female education and fertility choices in South Asia.

Using linear regression with the least-squares method, we will be able to predict which of these Educations will be or will remain the largest Education by population in the future.

Importance

Female education is a term which includes issues and debates surrounding all types of education including primary, secondary, tertiary, and health education for girls and women. The concept of female education includes gender equality, access to education, and poverty alleviation. Education is the key that will allow many other Sustainable Development Goals

(SDGs) to be achieved. When people are able to get quality education, they can break from the cycle of poverty. Education therefore helps to reduce inequalities and to reach gender equality. Education is related to SDGs Goal-16 (Peace, Justice, and Strong Institution). Education teaches us to be kind to others, telling the truth or praying. Based on different Educations, there may be different traditions to practice. It also occupies an important place in human society. Almost all the aspects of society including economic, political regions are guided and controlled by it. But, in the end, Education helps us to build a strong society and culture.

Related work

There is a lot of research done into this specific topic, each has their own methodology of data collection and projection. We will be looking at one of the major ones.

- Female Education Externality and Inclusive Growth

This research was received on 11 May 2019 and accepted on 13 June 2019 and published on 17 June 2019. This method can be found on page 5 of the published article. Accordingly, the data for research and prediction come from the census, census, general census, population register, and statistical reports from the World Educational Database. The information gathered is normalized, so all information has a similar degree of particularity, that is, information detailing a particular personality (like an Educational division inside an Education) will be assembled into a similar classification as the primary Education. Moreover, unlike the customary perspective that human resources ought to be genderless, the current exploration features the sex specifific gets back to instruction checked by the comprehensive development measures.One deficiency of this examination is that without work market investment rates by sexual orientation and training level for every country, we can't advise whether and how much our decision in regards to the effect of sex schooling hole gets from the sex specifific nature, or sex specifific usage rate, of human resources. Specifically, the way that a more instructed lady is almost certain than a less taught partner to be utilized may enhance the (unadulterated) impact of the sexual orientation schooling hole on both conventional and comprehensive development measures. Our utilization of crosscountry information blocks disintegrating the impact of the sex explicit nature of human resources from the other impact, since it doesn't give the work market investment rate by sexual orientation and by schooling level. We look to future examinations to cure this insufficiency utilizing miniature level board information.

The change in Education is also one of the factors considered to be important for forecasting. The researchers calculated the speed of Educational conversion based on past data. The projection itself is based on the group component method. In this method, the age structure is considered when predicting the future population. Cohort here refers to a birth cohort in which people of the same age group represent a cohort, that is, they are people born at the same time. The study used a multi-state variant of this method, in which each Education is represented. As "state" ", people can switch between these states. The model is simulated on a computer to produce the projection.

Methodology

1. **Variables and Data Sources**

For the model that we intend to create, we would require data on Education and their population for at least 10 years. This is based on the assumption that the higher the number of years used in the data used to create the model, the more accurate the model would be. Since Education is a widely discussed topic, statistics on it are readily available as public data. We decided to go with the dataset published by Correlates of World Bank as it is available in a neatly organized format at their website: https://knoema.com/WBEDS2017Jun/education-statistics The dataset contains several columns. But, for our particular project, we will be utilizing the following columns: Year, Bangladesh, India, Indonesia, Japan.

- Country

This column represents the country in a row for which the aggregated value was calculated. This particular column is useful to find the educated trends in a particular region, should we wish to do so. At this point in time, it is a secondary objective where we wish to give a few insights on the educated landscape in Asia

Year

This represents the year to which the data corresponds to. This is important to our modelling process, as the trends would be used to create the matrix required for the model. The year in the dataset ranges from 2001 to 2013, recorded without any intervals.

- Columns

The dataset contains columns where each corresponds to a country of female Education growth. The columns here are labelled in shorthand. But the columns of interest to us are:

  - Bangladesh—Total No. of Adherents
  - Japan—Total No. of Adherents
  - India —Total No. of Adherents
  - Indonesia —Total No. of Adherents

The above description of columns was obtained from the codebook provided alongside the dataset. Each record under these columns gives us the total number of females belonging to that particular area for the year described by the year attribute of the record. This would be a major part of the modelling process as it would become the target variable which we will be predicting through our model.

2. **Model Type**

The model would be created using linear regression by least squares method. This particular technique is very reliable for population growth modelling and is used by census organizations themselves due to its simplicity. Variants of it such as the Partial Least Squares method have also been shown to possess high predictive capability on highly dimensional data (Yeniay & Göktaş, 2001).

For our problem, we have four countries under consideration for which we need to create four separate models. Each of these models will allow us to predict the female educated population for that country. Since we are using linear regression using the least squares method, the model is created by obtaining a best fit line to fit the data points. The data points here represent the total population of the religious group plotted against the year. Once the model is created, it allows us to predict the Female population prediction in the future by getting the intersection of the line that was plotted with the year for which we wish to predict the population. The point of intersection gives us the predicted population.

A methodology widely used in Regression Analysis is the least-squares regression method. It is a mathematical technique used to find the best fit line that expresses the relationship between a variable that is independent and dependent.The line is a mathematical model used for a given x to predict the value of y.

The concept of a regression model is to analyze the effect on a dependent variable of interest of one or more independent variables. Linear regression analyses such as these are based on a basic equation:

$$y = mx + c$$

- y = dependent variable.
- m = slope of the line.
- x = independent variable.
- c = intercept.

Thus, to extract the value of the dependent variable, the goal is to calculate the slope, y-intercept values and replace the corresponding 'x' values in the equation.

We calculate two metrics for evaluation of the models: root mean square error and the r-squared (R2) score. The root mean square error is the average of squared errors of the model. It is the standard deviation of the actual data from the regression line. It tells us how far the predicted point on the line is from the actual point. The smaller it is, the better the model is. R2 score on the other hand, is a measure of fit. It tells us how well the best fit line fits our model. The closer it is to 1 the better is the model. Unlike root mean square, R2 score is relative and gives us a more interpretable value.

Results

1. **Descriptive Analysis of the Dataset**

- **Question 1:** What are the percentage of female educated groups in Asian countries?

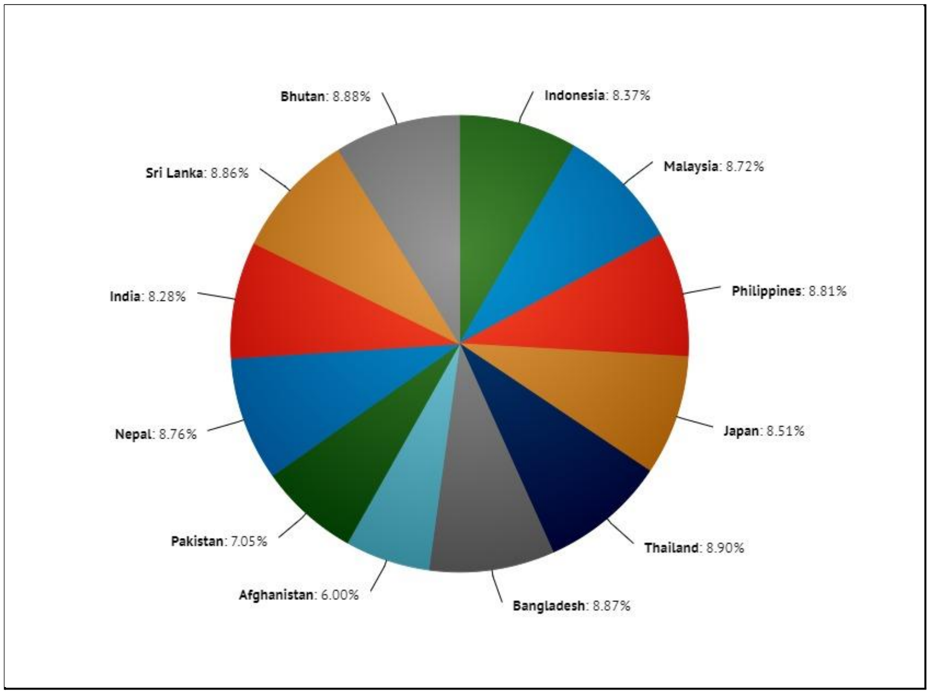

Figure 1: Largest educated female groups

- **Question 2:** What is the largest female educated in the world by population?

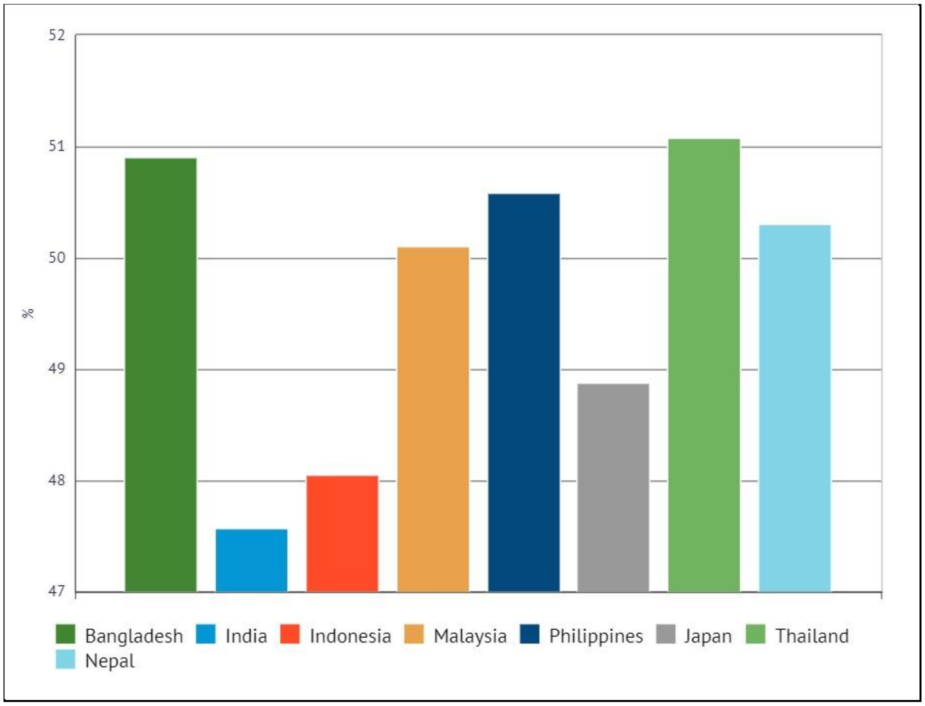

Figure 2: Largest educated females

- **Question 3:** Which country is growing fastest in female education?

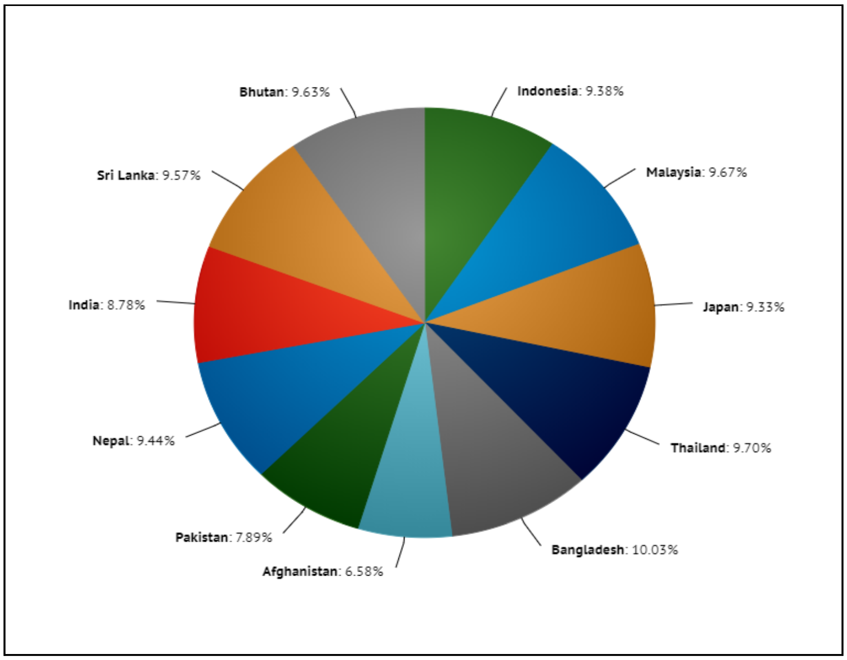

Figure 3: Bangladesh in 2008

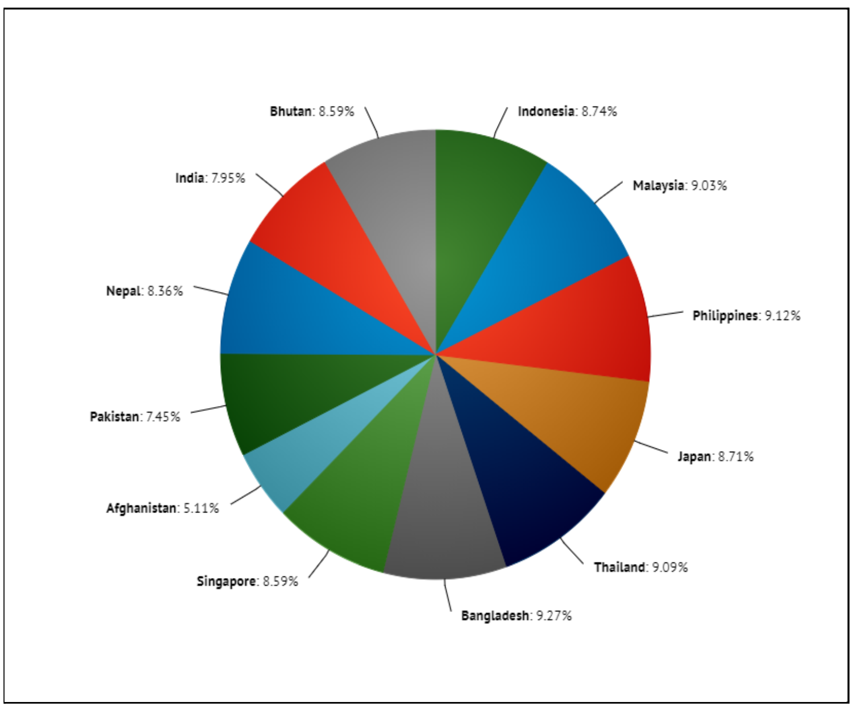

Figure 4: Bangladesh in 2011

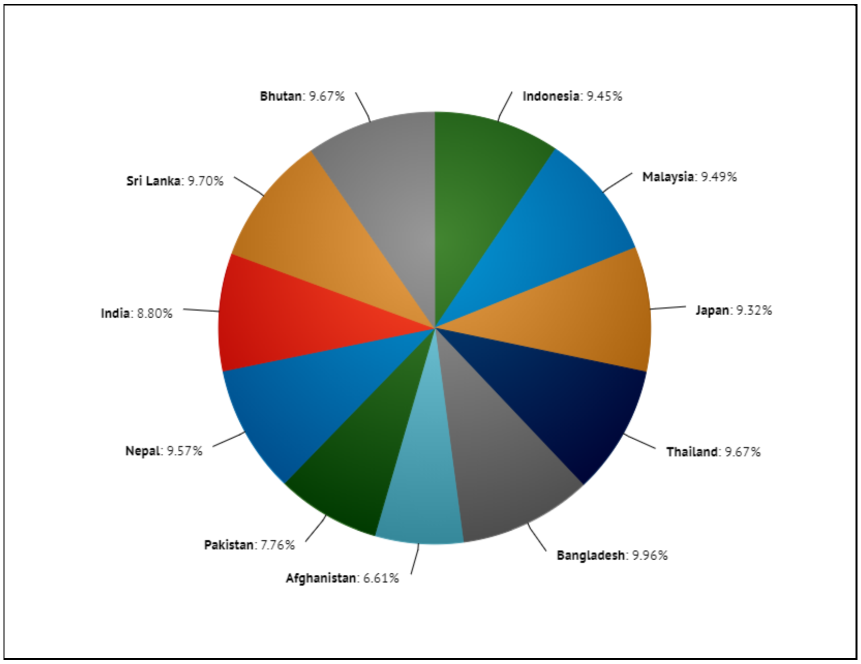

Figure 5: Bangladesh in 2012

- **Question 4: Bangladesh** vs Japan female educated Growth rate in Asia?

Time	Bangladesh	Japan
Units	%	%
2013	50.91	48.88
2012	52.19	48.88
2011	52.53	48.87
2009	50.90	48.87
2008	51.98	48.83
2010	51.88	48.85
2007	50.32	48.83
2006	50.38	48.82
2005	50.55	48.83
2004	49.58	48.87
2003	51.31	48.93
2002	51.34	49.00
2001	50.92	49.05

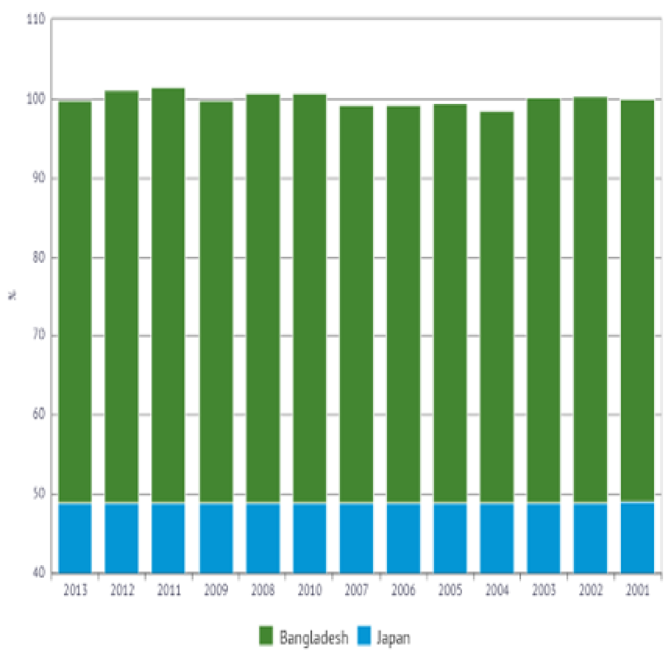

Figure 6: Bangladesh vs Japan female education Growth in Europe 2001 - 2013

- **Question 5:** Which county and region has the highest female educated rate?

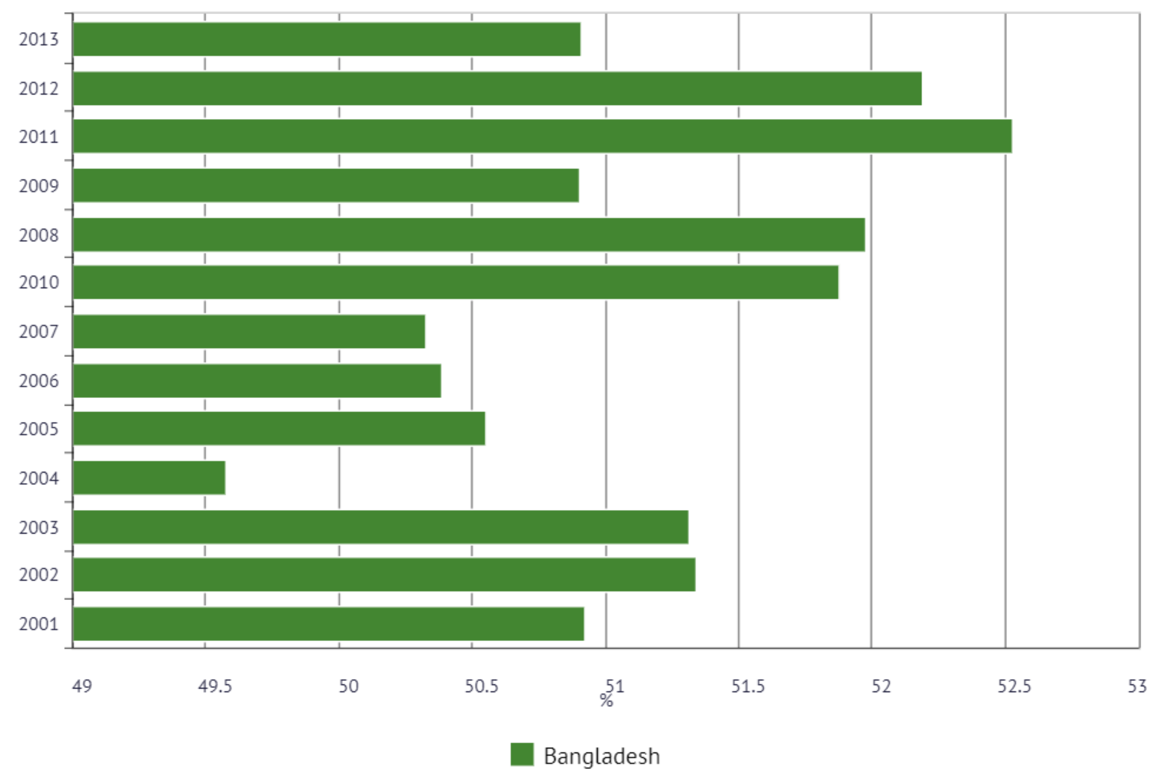

Figure 7: Country with highest female educated rate

1. **Model Analysis**

a. **Bangladesh** female education rate model

The first one among our models is the Bangladesh female educated rate model. This model was created based on two columns of our dataset: Year and percentage of females. The following figure shows the best fit regression line obtained obtained by linear regression using least squares method alongside the data points:

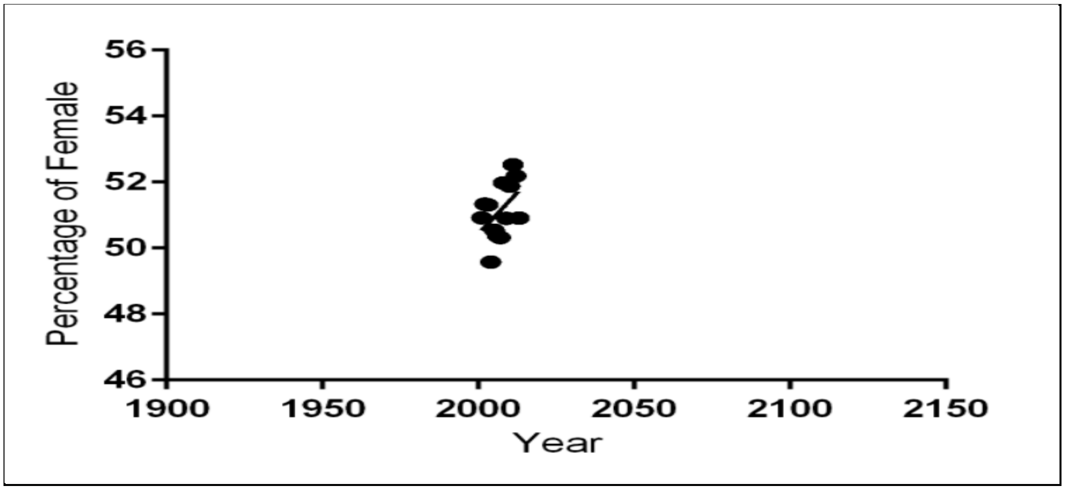

Figure 9: Linear regression plot for Bangladesh female educated rate

The following table shows the calculation done to obtain the coefficients of our regression line:

	X	Y(%)	X - mean_x	Y - mean_y	(X - mean_x) * (Y - mean_y)	(X - mean_x) ^ 2
0	2001	50.92	-6	-0.21	1.26	36
1	2002	51.34	-5	0.21	-1.05	25
2	2003	51.31	-4	0.18	-0.72	16
3	2004	49.58	-3	-1.55	4.65	9
4	2005	50.55	-2	-0.58	-1.16	4

5	2006	50.38	-1	-0.75	0.75	1
6	2007	50.32	0	-0.81	0	0
7	2008	51.98	1	0.85	0.85	1
8	2009	50.90	2	-0.23	-0.46	4
9	2010	51.88	3	0.75	2.25	9
10	2011	52.53	4	1.4	5.6	16
11	2012	52.19	5	1.06	5.3	25
12	2013	50.91	6	-0.22	-1.32	36

Bangladesh:

Mean_x=2007

Mean_y=51.13

Σ((X[i] - mean_x) * (Y[i] - mean_y)) =15.95

Σ((X[i] - mean_x) ^ 2)=182

m = Σ((X[i] - mean_x) * (Y[i] - mean_y)) / Σ((X[i] - mean_x) ^ 2)

m = (15.95 / 182)

m = 0.0876374

c = mean_y - (m * mean_x)

c = 51.13 - (0.0876374 * 2007)

c = -124.5882618

Similarly, we could also calculate the Root Mean Squared Error and the R2 score by using the following table:

	Y	Y-y_pred=c+ m*X[i]	Y - mean_y
0	50.92	0.145824	-0.21
1	51.34	0.478187	0.21
2	51.31	0.36055	0.18
3	49.58	-1.45709	-1.55
4	50.55	-0.57473	-0.58
5	50.38	-0.83236	-0.75

6	50.32	-0.98	-0.81
7	51.98	0.592363	0.85
8	50.90	-0.57527	-0.23
9	51.88	0.317088	0.75
10	52.53	0.87945	1.4
11	52.19	0.451813	1.06
12	50.91	-0.91582	-0.22

From the table, the calculation is as follows:

Total no. of records (n) = 13

Σ((Y[i] - y_pred) ^ 2) = 34004.37

Root Mean Square Error = √(Σ((Y[i] - y_pred) ^ 2)/n)

Root Mean Square Error = √(34004.37/ 13)

Root Mean Square Error = 2615.72

Σ((Y[i] - mean_y) ^ 2) = 8.548

R2 Score = 1 - Σ((Y[i] - mean_y) ^ 2)/Σ((Y[i] - y_pred) ^ 2)

R2 Score = 1 - (8.548) / (34004.37)

R2 Score = 0.9997

Therefore, the model obtained has the following associated metrics:

- **Coefficients:** [8.548] [34004.37]
- **Root Mean Square Error:** 2615.72
- **R2 Score:** 0.9997

From the above, we can see the model is quite accurate as given by the R2 Score, as it is close to 1. So, it can give predictions that would be pretty close to the actual value. The female educated rate prediction obtained from the above model is listed below:

- **2025: 50.61**
- **2027: 52.02**
- **2030: 52.34**

As we can see from the predictions, the female education rate will reach 52.34 by the year 2030. The growth here is the largest among our models.

b. India female educated rate model

The next model we created was for the India female educated rate. Like the last model, this is the plot of the regression line and data points we obtained from our model:

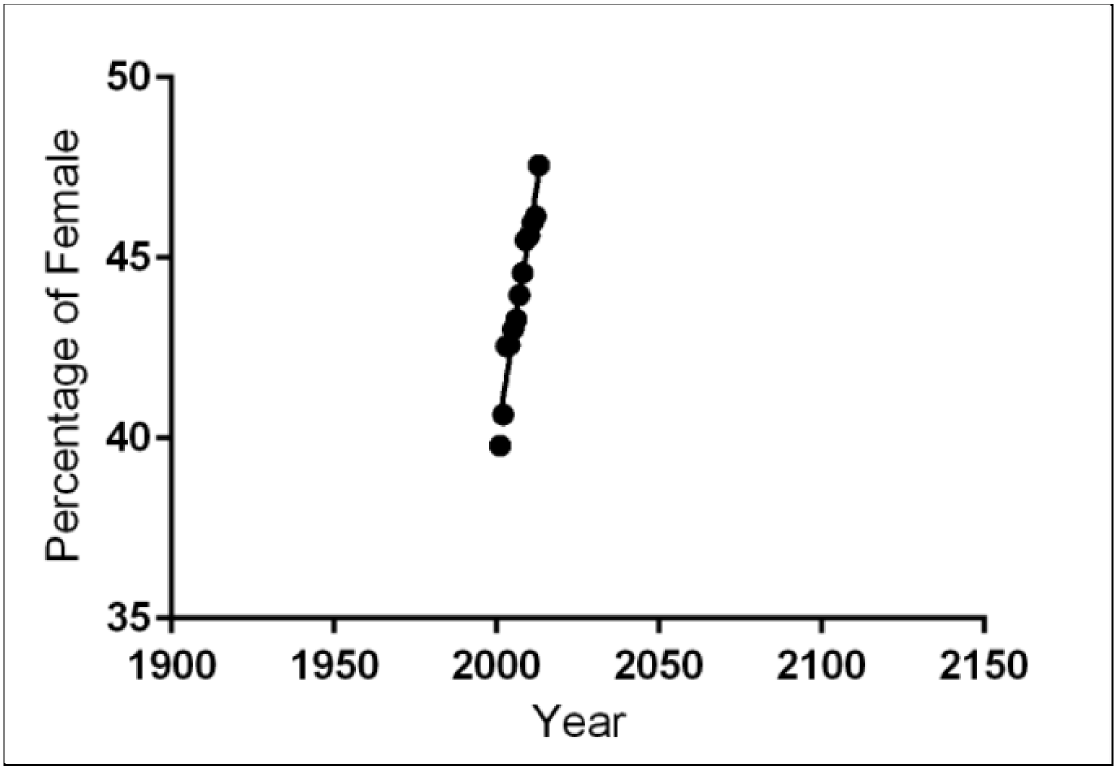

Figure 10: Linear regression plot for India female educated rate

In this case, we had the following table for calculating the coefficients for the line:

	X	Y(%)	X - mean _x	Y - mean _y	(X - mean _x) * (Y - mean _y)	(X - mean _x) ^ 2
0	200 1	39.7 9	-6	-9.246	55.47 6	36
1	200 2	40. 65	-5	-8.385	41.92 5	25

2	2003	42.56	-4	-6.475	25.9	16
3	2004	42.58	-3	-6.455	19.365	9
4	2005	43.02	-2	-6.015	12.03	4
5	2006	43.3	-1	-5.735	5.735	1
6	2007	43.97	0	-5.065	0	0
7	2008	44.58	1	-4.455	-4.455	1
8	2009	45.49	2	-3.545	-7.09	4
9	2010	45.62	3	-3.415	-10.245	9
10	2011	45.98	4	-3.055	-12.22	16
11	2012	46.15	5	-2.885	-14.425	25
12	2013	47.57	6	-1.465	-8.79	36

X mean=2007

Y mean= 43.94308

Σ((X[i] - mean_x) * (Y[i] - mean_y)) = 103.206

Σ((X[i] - mean_x) ^ 2)=182

m = Σ((X[i] - mean_x) * (Y[i] - mean_y)) / Σ((X[i] - mean_x) ^ 2) = .567

c = mean_y - (m * mean_x) = 1181.91

	Y(%)	Y-y-pred	Y - mean_y
0	39.79	-2276.69	-9.246
1	40.65	-2276.39	-8.385
2	42.56	-2275.05	-6.475
3	42.58	-2275.6	-6.455
4	43.02	-2275.73	-6.015
5	43.3	-2276.01	-5.735
6	43.97	-2275.91	-5.065
7	44.58	-2275.87	-4.455
8	45.49	-2275.52	-3.545

9	45. 62	-2275.96	-3.415
10	45. 98	-2276.17	-3.055
11	46.1 5	-2276.56	-2.885
12	47.5 7	-2275.71	-1.465

From the table, the calculation is as follows:

Total no. of records (n) = 13

Σ((Y[i] - y_pred) ^ 2) = 62159638

Root Mean Square Error = √(Σ((Y[i] - y_pred) ^ 2)/n)

Root Mean Square Error = √(62159638 / 13)

Root Mean Square Error = 2186.67

Σ((Y[i] - mean_y) ^ 2) = 395.8466

R2 Score = 1 - Σ((Y[i] - mean_y) ^ 2)/Σ((Y[i] - y_pred) ^ 2)

R2 Score = 1 - (395.8466) / (62159638)

R2 Score = 0.999

Therefore, the model obtained has the following associated metrics:

- **Coefficients:** [62159638] [395.8466]
- **Root Mean Square Error:** 2186.67
- **R2 Score:**0.999

From the above, we can see the model is quite accurate as given by the R2 Score, as it is close to 1. So, it can give predictions that would be pretty close to the actual value. The female educated rate prediction obtained from the above model is listed below:

- **2025: 46.01**
- **2027: 46.12**
- **2030: 47.13**

As we can see from the predictions, the female education rate will reach 47.13 by the year 2030. The growth here is the second largest among our models.

c. Indonesia female educated rate model

The third linear regression model was based on the Indonesia female educated rate. In this case, we have the following plot

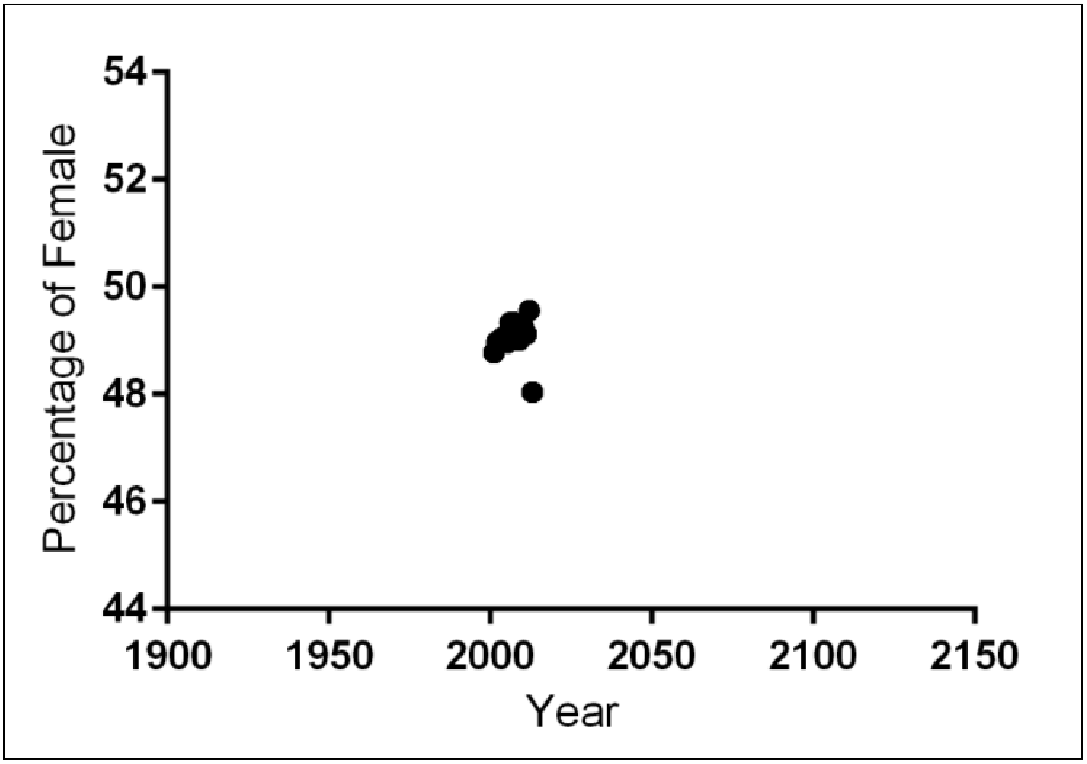

Figure 11: Linear regression plot for Indonesia female educated rate.

We obtained the following table for calculating the coefficients for the regression line for it:

	X	Y(%)	X - mean_x	Y - mean_y	(X - mean_x) * (Y - mean_y)	(X - mean_x) ^ 2
0	2001	48.78	-6	-0.25	1.5	36
1	2002	48.98	-5	-0.05	0.25	25
2	2003	48.97	-4	-0.06	0.24	16

3	2004	49.07	-3	0.04	-0.12	9
4	2005	48.96	-2	-0.07	0.14	4
5	2006	49.34	-1	0.31	-0.31	1
6	2007	49.35	0	0.32	0	0
7	2008	49.01	1	-0.02	0.02	1
8	2009	49.01	2	-0.02	-0.04	4
9	2010	49.26	3	0.23	0.69	9
10	2011	49.13	4	0.1	0.4	16
11	2012	49.56	5	1.53	7.65	25
12	2013	48.05	6	0.02	0.12	36

Mean_x =2007

Mean_y=49.03

Σ((X[i] - mean_x) * (Y[i] - mean_y)) = 10.54

Σ((X[i] - mean_x) ^ 2) = 182

m = Σ((X[i] - mean_x) * (Y[i] - mean_y)) / Σ((X[i] - mean_x) ^ 2)

m = (10.54 / 182)

m =0.057912087

c = mean_y - (m * mean_x)

c = 49.03 - (0.057912087 * 2007)

c = -67.19955861

Similarly, we could also calculate the Root Mean Squared Error and the R2 score by using the following table:

	X	Y-y_pred=c+ m*X[i]	Y - mean_y
0	2001	0.09747252	-0.25
1	2002	0.23956044	-0.05
2	2003	0.17164835	-0.06
3	2004	0.21373626	0.04
4	2005	0.04582417	-0.07
5	2006	0.36791209	0.31

6	2007	0.32	0.32
7	2008	-0.07791209	-0.02
8	2009	-0.13582417	-0.02
9	2010	0.05626374	0.23
10	2011	-0.13164835	0.1
11	2012	0.24043957	1.53
12	2013	-1.32747252	0.02

From the table, the calculation is as follows:

Total no. of records (n) = 13

Σ((Y[i] - y_pred) ^ 2) = 0.484722

Root Mean Square Error = √(Σ((Y[i] - y_pred) ^ 2)/n)

Root Mean Square Error = √(0.484722 / 13)

Root Mean Square Error = 0.1931

Σ((Y[i] - mean_y) ^ 2) = 2.6786

R2 Score = 1 - Σ((Y[i] - mean_y) ^ 2)/Σ((Y[i] - y_pred) ^ 2)

R2 Score = 1 - (2.6786) / (0.1931)

R2 Score = -12.87

Therefore, the model obtained has the following associated metrics:

- **Coefficients:** [0.484722] [395.8466]
- **Root Mean Square Error:** 0.1931
- **R2 Score:**-12.87

From the above, we can see the model is quite accurate as given by the R2 Score, as it is close to 1. So, it can give predictions that would be pretty close to the actual value. The female educated rate prediction obtained from the above model is listed below:

- **2025: 48.01**
- **2027: 48.12**
- **2030: 49.03**

As we can see from the predictions, the female education rate will reach 49.03 by the year 2030. The growth here is the second largest among our models.

d. **Japan** female educated rate model

The last model we created was for predicting the Japanese female education rate. The plot is as follows:

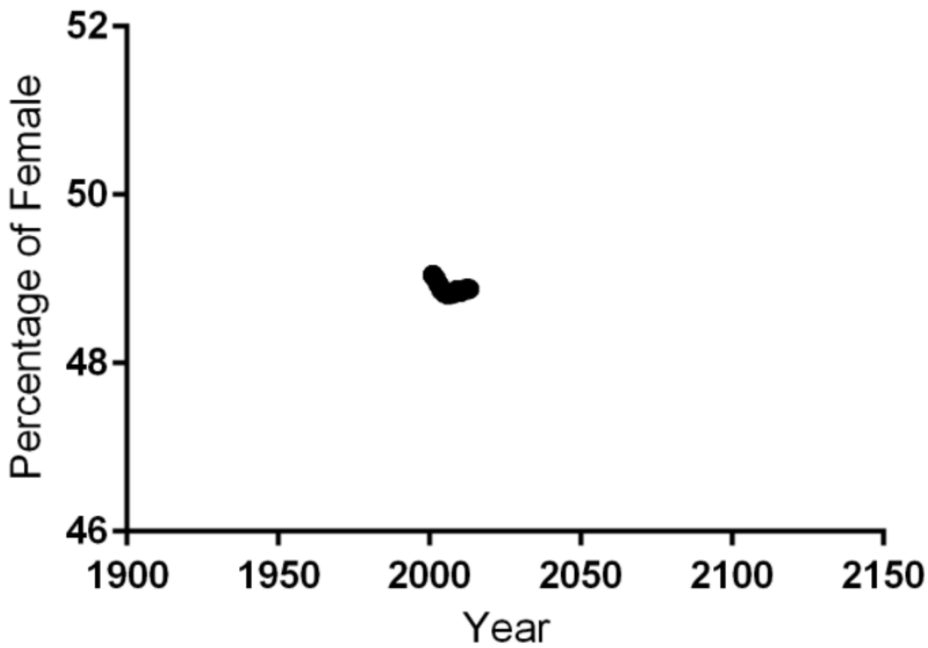

Figure 12: Linear regression plot for Japan female educated rate.

The calculations for this is Japan female educated rate **summarized in the following table:**

	X	Y(%)	X - mean_x	Y - mean_y	(X - mean_x) * (Y - mean_y)	(X - mean_x) ^ 2
0	2001	49.05	-6	0.17	-1.02	36
1	2002	49.00	-5	0.12	0.6	25
2	2003	48.93	-4	0.05	-0.2	16

3	2004	48.87	-3	-0.01	0.03	9
4	2005	48.83	-2	-0.05	0.1	4
5	2006	48.82	-1	-0.06	0.06	1
6	2007	48.83	0	-0.05	0	0
7	2008	48.83	1	-0.05	-0.05	1
8	2009	48.87	2	-0.01	-0.02	4
9	2010	48.85	3	-0.03	-0.09	9
10	2011	48.87	4	-0.01	-0.04	16
11	2012	48.88	5	0	0	25
12	2013	48.88	6	0	0	36

mean_x=2007

mean_y=48.88

Σ((X[i] - mean_x) * (Y[i] - mean_y)) = -0.63

Σ((X[i] - mean_x) ^ 2) =182

m = Σ((X[i] - mean_x) * (Y[i] - mean_y)) / Σ((X[i] - mean_x) ^ 2)

m = (-0.63 / 182)

m = -3.462*10^-03

c = mean_y - (m * mean_x)

c = 48.88 - (-3.462*10^-03 - 2007)

c = 48.88-(-2007.003462)

c=2055.88462

For calculation of the metrics, we used the following table:

	Y(%)	Y-y_pred=c+ m*X[i]	Y - mean_y
0	49.05	-1999.90808	0.17
1	49.00	-1999.95462	0.12
2	48.93	-2000.02116	0.05
3	48.87	-2000.07769	-0.01
4	48.83	-2000.11423	-0.05

5	48.82	-2000.12077	-0.06
6	48.83	-2000.10731	-0.05
7	48.83	-2000.10385	-0.05
8	48.87	-2000.06039	-0.01
9	48.85	-2000.07692	-0.03
10	48.87	-2000.05346	-0.01
11	48.88	-2000.04	0
12	48.88	-2000.03654	0

From the table, the calculation is as follows:

Total no. of records (n) = 13

Σ((Y[i] - y_pred) ^ 2) = 52002700

Root Mean Square Error = √(Σ((Y[i] - y_pred) ^ 2)/n)

Root Mean Square Error = √(52002700 / 13)

Root Mean Square Error = 2000.05

$\Sigma((Y[i] - mean\_y)^2) = 0.0581$

$R2\ Score = 1 - \Sigma((Y[i] - mean\_y)^2)/\Sigma((Y[i] - y\_pred)^2)$

$R2\ Score = 1 - (2000.05) / (52002700)$

$R2\ Score = 0.999$

Therefore, the model obtained has the following associated metrics:

- **Coefficients:** [0.0581] [52002700]
- **Root Mean Square Error:** 2000.05
- **R2 Score:** 0.999

From the above, we can see the model is quite accurate as given by the R2 Score, as it is close to 1. So, it can give predictions that would be pretty close to the actual value. The female educated rate prediction obtained from the above model is listed below:

- **2025: 48.11**
- **2027: 48.02**
- **2030: 48.34.**

As we can see from the predictions, the female education rate will reach 48.34 by the year 2030. The growth here is the third largest among our models.

Conclusion:

Education has constantly been in the middle of humanity at some point of man's existence. In this assignment, we delved deeper into the lady training panorama of the world. We analyzed diverse traits and additionally created for one-of-a-kind fashions to move past what recorded information. Our linear regressions fashions allowed us to get an approximation for what the lady training increase might be like for the 4 main asian nations: Bangladesh, India, Indonesia and Japan. The traits from the fashions we created inform us that each one of those 3 nations might preserve their ranks even in 2030.The studies paper we accompanied is a lot specific and this may be due to the fact our version is a lot easier than the only that became created through them. They covered many elements of their prediction modelling which includes start charge, demise charge, charge of conversion and others as mentioned withinside the associated works phase of this assignment. On the other hand, we used an easy least squares approach to create our version that is primarily based totally completely at the formerly recorded lady training charge of the 4 nations. That is why , for example, our pleasant suit line is incorrect in comparison to the rest. This is as it would not display a fashion of linear increase. Despite the easy nature and shortcomings of the least squares approach, it nonetheless controlled to version the 3 different religions pretty well, with an R2 rating this is very near 1, mainly withinside the case of Christianity. In phrases of improvements, we may want to doubtlessly gain a greater correct version through utilizing a polynomial least squares linear regression approach. It can possibly version the non-linear increase of the spiritual populations greater accurately. Additionally, a greater updated dataset might have additionally helped growth the accuracy of the predictions. The statistics withinside the dataset we used results in 2010, and subsequently we've over a decade of statistics unavailable to us. The United Nations Statistics Division has a greater updated dataset, however it's far very unclean and subsequently couldn't be used for growing our version, as cleansing it might have taken a long way too lengthy to be really well worth the effort. Lastly, a nearby spiritual populace version, in place of a worldwide spiritual version, might be greater insightful in information Educational traits. Overall, we controlled to gain the targets that we got down to gain withinside the starting of the assignment. This assignment has been an insightful revel in for anyone worried and we are hoping that it advantages the reader.

References

https://knoema.com/WBEDS2017Jun/education-statistics

https://www.mdpi.com/2071-1050/11/12/3344/pdf

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6054002/

https://knoema.com/WBEDS2017Jun/education-statistics?accesskey=phlxfff