# IMPACT OF URBANISATION AND INCOME ON FERTILITY RATES IN INDIA

NANDANA P V (21/871)

LADY SHRI RAM COLLEGE FOR WOMEN DELHI

## Abstract

The study examines the impact of urbanization and income on the fertility rate from the period 1996 to 2020 in India. The study will employ a quantitative research method, using statistical analysis to examine the relationship between urban population, income, and fertility rate. Using data from World Bank open data, the study employs multiple regression analysis to examine the impact of these factors on the fertility rate. The results show that there is a negative correlation between urban population and fertility rate, suggesting that as urbanization increases, fertility rates tend to decline. Furthermore, the study finds that income also plays a significant role, with higher incomes associated with lower fertility rates. These findings have important implications for policymakers and researchers interested in understanding the demographic and economic changes that occur with urbanization and development. The findings of this study will contribute to a better understanding of the factors that influence fertility rates and can be used to inform policies and programs aimed at promoting sustainable population growth in urban areas.

Keywords: Fertility rate, urbanization, income, multiple regression analysis, India

## Introduction

Western Europe and North America experienced unprecedented economic growth during the industrial revolution, considered a new era in population dynamics. During the demographic transition, countries went from a high mortality and high fertility regime to a low mortality and low fertility regime. In the past century, the world has witnessed unprecedented demographic and economic changes that have changed the way people live, work, and reproduce. The paper explores the relationship between fertility rate, urban population, and income in India. In view that the economic liberalization policy of 1991, India witnessed extensive economic growth and development. No matter its rapid boom, India nonetheless faces many demanding situations, together with poverty, inequality, and a range of social, political, and environmental issues. But the country has made considerable progress in recent years, with improvements in regions consisting of schooling, healthcare, and infrastructure. In recent decennary, India has experienced significant changes in these demographic factors. In 2023, India stands with a 142.86 crore populous nation, according to the recent United Nation Population Fund Data. The nation has surpassed China to win the title of the most populated nation in the world. For Indias growth, this huge and expanding population brings both possibilities and difficulties.

One of the most crucial factors in population growth is the fertility rate. The number of children each woman would have if she were to survive to the end of her childbearing years and have children by the current age-specific fertility rates is the total number of children that would be born in that year. It is determined by adding the age-specific fertility rates across intervals of five years. The prevalence of fertility in the world varies significantly depending on the nation, culture, social status, and age of the individual. However, as far as human history, urban environments are quite a recent phenomenon. The way we work, travel, live, develop and establish networks have all changed as a result of this change. Under UN estimates, in 2007 there were more people residing in urban areas compared with rural regions for the first time. India is transitioning from an agriculturally based rural economy to an industry and service-focused urban one. India saw a rise in urbanization from 17.924 in 1960 to 35.3983 in 2021. Hypothesize that there is a negative correlation between fertility rate and urban population and income since urbanization and economic growth frequently lead to changes in gender roles, lifestyles, and access to education that lessen the desire and capacity for having large families.

## Literature Review

Some empirical studies have systematically studied the relationship between urbanization, and income on fertility rates. Christophe Z Guilmoto and S Irudaya Rajan have conducted research on fertility patterns at the district level in India, using data from the 2011 census. Their research provides important insights into the factors that influence fertility rates in different parts of the country. At the same time, Guilmoto and Irudaya Rajan also identified several factors that are associated with lower fertility rates, including higher levels of urbanization, greater access to family planning services, and higher levels of female education. They argue that these factors can be leveraged to promote lower fertility rates and improve reproductive health outcomes in India.

George Martine, Jose Eustaquio Alves, and Suzana Cavenaghi have conducted research on the relationship between urbanization and fertility decline, with a focus on the potential for "cashing in" on structural change to accelerate fertility decline in urban areas. At the same time, Martine, Alves, and Cavenaghi argue that urbanization itself can be a powerful force for fertility decline, as urban areas tend to have lower fertility rates than rural areas. Another study was conducted by A. Dharmalingam, Sowmya Rajan, and S. Philip Morgan on the determinants of low fertility in India, examining the various social, cultural, and economic factors that have contributed to declining fertility rates in the country. Their review found that while India has traditionally had high fertility rates, particularly in rural areas, there has been a significant decline in fertility over the past few decades. They identified several potential explanations for these regional differences, including variations in economic development, urbanization, and access to family planning services. T. Paul Schultzs article, "Fertility and Income," published in the Journal of Economic Literature in 1997, explores the relationship between income and fertility rates. Schultz argues that there is a negative relationship between income and fertility rates, meaning that as income increases, fertility rates tend to decrease. He suggests that this relationship can be explained by a range of factors, including increased access to family planning services, increased education for women, and changes in cultural attitudes toward childbearing. Schultz also notes that the relationship between income and fertility can vary depending on the specific context, such as the country, region, or socioeconomic group being studied.

## A Priori Expectation

In this analysis, the dependent variable is the fertility rate and the explanatory variables are Urban population and Income. First, the fertility rate (Y) is expected to have a negative relation with the urban population(X1). Next, it is also expected to have a negative relationship with income(X2) after controlling the effects of other variables.

## Data and Methodology

This paper figures out the how fertility rate is related to the Urban population and income in India. This is a time series study carried out from the year 1996 to 2020. The sample size was estimated to be 25. Data were collected using World Bank open data, which included information on all countries from the year 1960 to 2021. Internationally recognized standards and norms are provided by bank data, creating a consistent, trustworthy source of information. This information supports or refutes the research question. because it gives the data needed to analyze the outcome. Making inferences based on solid facts and relevant insights is made possible by this data. From the open data of the world bank, the required data for the research for the past 25 years that is from 1996-2020 was curated in Excel. The collected information was analyzed in R (R-4.2.2) and the following tests were used to obtain the research goals: ANOVA, hypothesis testing, and jarque bera test. Simple Linear Regression takes up the form of, Yi=b1+b2+e, where Y is the dependent variable, X is the independent variable, b1 is the intercept, b2 is the slope and e is the error term. In the simple linear model, the effect of Urban population and income was separately tested against the fertility rate. Multiple Linear Regression of the form, Yt=B1+B2X2t+B3X3t+Ut, where Yt is the dependent variable, and X1 and X2 are the explanatory variables. B1 is the intercept coefficient and B2 and B3 are partial slope coefficients. In all tests, the significance level of 5% was considered.

## OLS Estimated Equation

## Simple Linear Regression

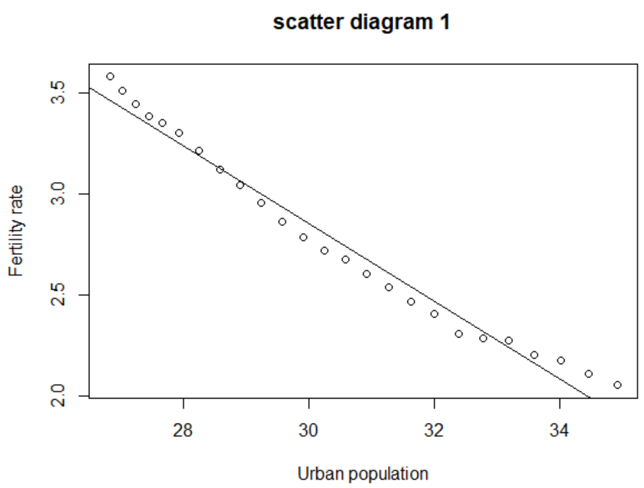
A simple linear regression is the linear relationship between two variables and the regression takes the equation as,

$$Y=b1+b2X+e \qquad (1)$$

Here, Y is the dependent variable, X is the independent variable, b1 is the intercept, b2 is the slope and e is the error term. According to the analysis, fertility rate(Y) is the dependent variable and the independent variables are urban population and income.

## Case 1-Regressing Fertility Rate in Urban Population

The scatter plot is used to show how the urban population and fertility rate are linearly related.



It is obvious from the scatter plot above that there is a linear link between urban population and fertility rate. Over time, the fertility rate declines as the urban population increases.

## Development of linear model using R

```
Call:
lm(formula = d2$St ~ d2$Q)

Residuals:
      Min        1Q    Median        3Q       Max 
-0.092127 -0.074700 -0.009091  0.046159  0.141643 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept)  8.624858   0.189216   45.58   <2e-16 ***
d2$Q        -0.192278   0.006199  -31.02   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07643 on 23 degrees of freedom
Multiple R-squared:  0.9767,	Adjusted R-squared:  0.9756 
F-statistic: 962.1 on 1 and 23 DF,  p-value: < 2.2e-16
```

From the output obtained from R, OLS estimated the equation as follows,

Fertility rate =8.624858-0.192278*Urban Population+ Error (2)

This means that the increase in the Urban population, an estimated mean value of the fertility rate in India decreases by 0.192278. The intercept value of 8.624858 tells that when Xi is zero fertility rate will be equal to 8.624858.

## Hypothesis test for Linear Regression

Ho:b2=0

Ha:b2 ≠0

The null hypothesis states that coefficient b2 is equal to zero and the alternative hypothesis states that the coefficient b2 is not equal to zero.

The p-value of the Urban Population (p-value: <2.2e-16) is significantly less than 0.05, which implies that there is a statistically significant relation between fertility rate and urban population. So, we reject the null hypothesis, b2=0. Hence, there is a significant relation between the variable in the linear regression model.

## Residual standard error

The average distance that the observed values fall from the regression line. A regression line can more closely reflect the observed data, the lower the number is. Here, 0.07643 points away from the fertility rate predicted by the regression line.

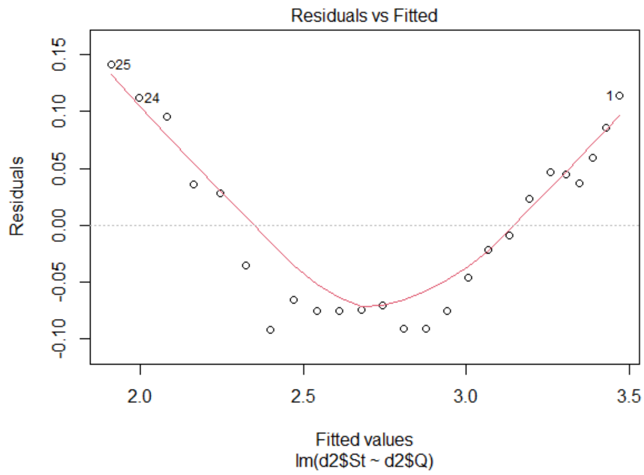
## Multiple R-squared

In general, this number tells us the percentage of the variation in the fertility rate can be explained by the urban population. The larger the R-squared value of a regression model the better the predictor variables can predict the values of the variable. In this case, 97.67% of the variation in the fertility rate can be explained by the urban population.

## F-statistic and p-value

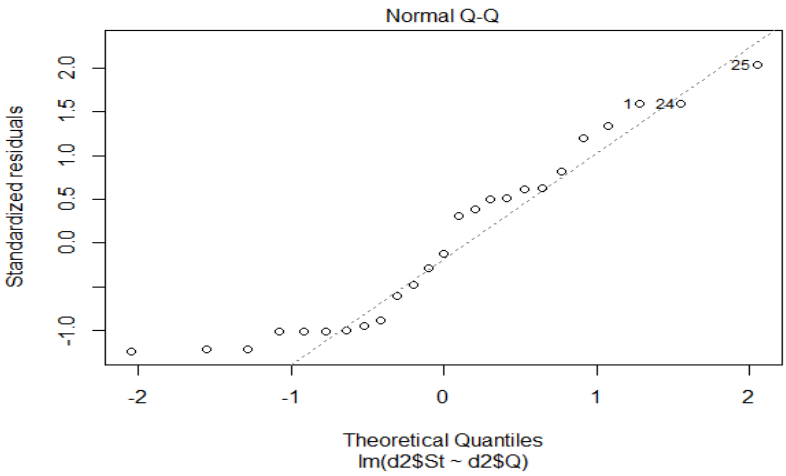
The f-statistic is 962.1 and the corresponding p-value <2.2e-16, which implies the overall significance of the regression model.

A residual versus fitted plot is made in order to validate the homoscedasticity presumption.



Fitted values are shown on the x-axis, while residual values are shown on the y-axis. The homoscedasticity assumption is upheld as long as the residual seems to be randomly distributed over the plot surrounding zero. The assumption is supported by the preceding graph, where residuals seem to be dispersed at random about zero.

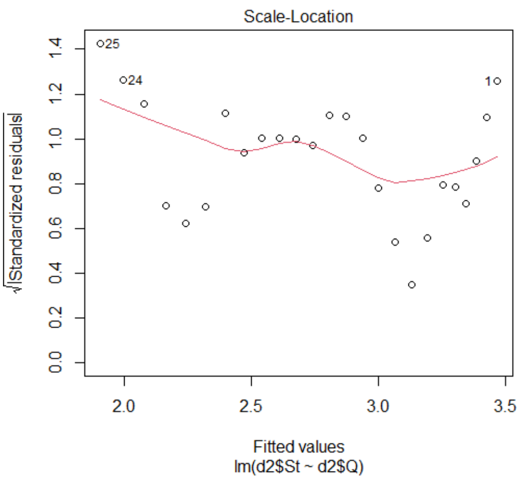
## Normal Q-Q




The data is normally distributed if the points on the plot fall along a relatively straight line at a 45-degree angle. The residuals are distributed at a 45° angle. It is true that normality is assumed. The OLS regressions assumptions are fulfilled since the residuals are hom oscedastic and normally distributed.

## Scale-Location

The fitted values of a regression model are shown along the x-axis of a scale-location plot, and the square root of the standardized residuals is shown along the y-axis.




The red line can be seen to be about horizontal across the area in the Scale-location plot. The regression model satisfies the homoscedasticity assumption. In other words, at all fitted values, the dispersion of the residuals is nearly equal. The three observations with the greatest standardized residuals, denoted as 1, 24, and 25, are shown in the plot.

```
ANOVA(Reg1)
Analysis of Variance Table

Response: d2$St
          Df Sum Sq Mean Sq F value    Pr(>F)    
d2$Q       1 5.6205  5.6205   962.1 < 2.2e-16 ***
Residuals 23 0.1344  0.0058                      
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The degrees of freedom for income is 1(2-1) and the degrees of freedom for residuals is 23(25 -2). The sum square associated with urban population is 5.6205 and the sum square associated with residuals or error is 0.1344. The mean Sq. of the urban population is 5.6205(5.6205/1) an d the mean Sq of residuals is 0.0058(0.0058/23).

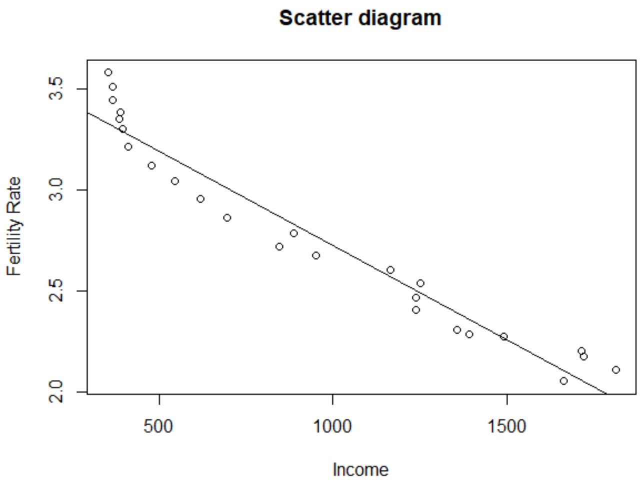
The overall F-statistic of the ANOVA model is 962.1 (5.6205/0.0058). The p-value associated with the F-statistic with numerator df=1 and denominator df=23 is less than 2.2e-16. The p-va lue is less than 0.05, so the model is statistically significant.

```
jarque.bera.test(residuals(Reg1))

        Jarque Bera Test

data:  residuals(Reg1)
X-squared = 1.9672, df = 2, p-value = 0.374
```

Ho= u is normally distributed
Ha= u is not normally distributed

From the above output, the test statistic value is 1.9672 and the p-value of the test is 0.374, which implies that fails to reject the null hypothesis that the data is normally distributed.

## Case 2- Fertility rate regressed on income




From the above scatter plot, it is clearly visible that there is a linear relationship between fertility rate and income. As income increases fertility rate decreases over the years.

## Development of linear model using R

```
Call:
lm(formula = d2$St ~ d2$L)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.15308 -0.09349 -0.03655  0.08493  0.24904 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  3.6613830  0.0496488   73.75  < 2e-16 ***
d2$L        -0.0009343  0.0000463  -20.18 3.99e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.1157 on 23 degrees of freedom
Multiple R-squared:  0.9465,   Adjusted R-squared:  0.9442 
F-statistic: 407.1 on 1 and 23 DF,  p-value: 3.989e-16
```

From the output obtained from R, OLS estimated the equation as follows:

Fertility rate = 3.6613830 - 0.0009343 * Income + Error   (3)

This means that with the increase in income, an estimated mean value of the fertility rate in India decreases by 0.0009343. When $X_i$ is equal to zero, the fertility rate will be equal to the intercept value which is equal to 3.6613830.

## Hypothesis testing

$H_o: b_2 = 0$

$H_a: b_2 \neq 0$

The null hypothesis states that coefficient b2 is equal to zero and the alternative hypothesis states that the coefficient b2 is not equal to zero.

The p-value of the income (p-value: 3.989e-16) is significantly less than 0.05, which implies that there is a statistically significant relation between fertility rate and income. So, we reject the null hypothesis, $b_2 = 0$. Hence, there is a significant relation between the variable in the linear regression model.

## Residual standard error

The average distance that the observed values fall from the regression line. The lower this value, the more closely a regression line can match the observed data. Here, 0.1157 points away from the fertility rate predicted by the regression line.

## Multiple R-squared

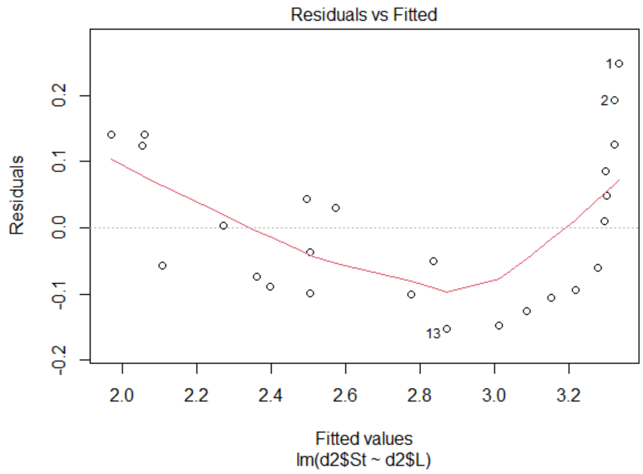
In general, this number tells us the percentage of the variation in the fertility rate can be explained by income. The larger the R-squared value of a regression model the better the predictor variables can predict the values of the variable. In this case, 94.65% of the variation in the fertility rate can be explained by income.

## F-statistic and p-value

F-statistic is 407.1 and the corresponding p-value is 3.989e-16, which implies the overall significance of the regression model.

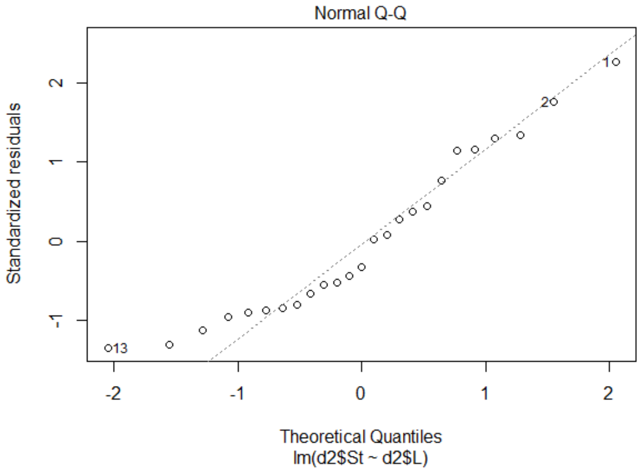
To verify the assumption of homoscedasticity, a residual vs fitted plot is created.

The x-axis displays fitted values and the y-axis displays residual values. As long as the residual appears to be randomly distributed throughout the plot around the value zero, the assumption of homoscedasticity is not violated. In the above graph, residuals appear to be randomly scattered around zero, so the assumption is met.



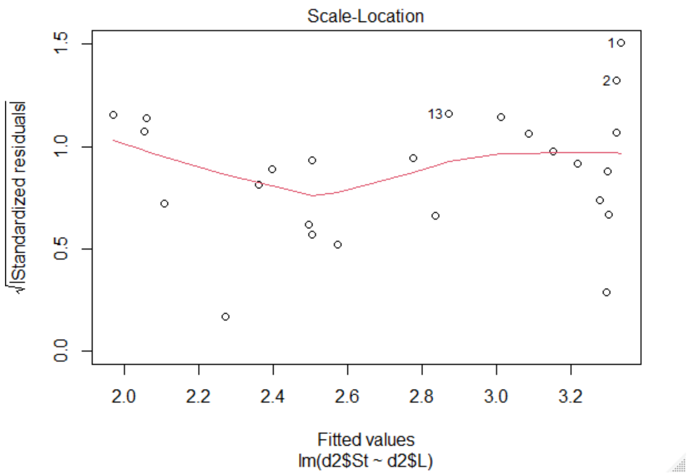

## Normal Q-Q

If the points in the plot fall along a roughly straight line at a 45-degree angle, then the data is normally distributed. The residuals fall along a 45-degree angle. The assumption of normality is met. Since, the residuals are normally distributed and homoscedastic, assumptions of OLS regression are met.




## Scale-Location

A scale-location plot is a type of plot that displays the fitted values of a regression model along the x-axis and the square root of the standardized residuals along the y-axis.




From the Scale-location plot, it is seen that the red line is roughly horizontal across the plot. The assumption of homoscedasticity is satisfied for the regression model. That is, the spread of the residuals is roughly equal at all fitted values. The three observations in the plot have the highest standardized residuals labeled as 1,2,13.

```
> ANOVA(Reg3)
Analysis of Variance Table
```

```
Response: d2$St
          Df Sum Sq Mean Sq F value    Pr(>F)    
d2$L       1 5.4471  5.4471  407.12 3.989e-16 ***
Residuals 23 0.3077  0.0134                      
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

The degrees of freedom for income is 1(2-1) and the degrees of freedom for residuals is 23(25 -2). The sum square associated with income is 5.4471 and the sum square associated with resi duals or error is 0.0134. The mean Sq. of income is 5.4471(5.4471/1) and the mean Sq of resi duals is 0.0134(0.3077/23).

The overall F-statistic of the ANOVA model is 407.12 (5.4471/0.0133). The p-value associate d with the F-statistic with numerator df=1 and denominator df=23 is 3.989e-16. The p-value i s less than 0.05, so the model is statistically significant.

```
>jarque.bera.test(residuals(Reg3))

        Jarque Bera Test

data:  residuals(Reg3)
X-squared = 1.8675, df = 2, p-value = 0.3931
```

Ho= u is normally distributed

Ha= u is not normally distributed

From the above output, the test statistic value is 1.8675and the p-value of the test is 0.3931, which implies that fails to reject the null hypothesis that the data is normally distributed.

## Multiple Linear Regression

Multiple linear regression establishes the relationship between a dependent variable and two or more independent variables. The model follows the regression equation as,

$$Yt=B1+B2X2t+B3X3t+Ut \quad (4)$$

Yt is the dependent variable, and X1 and X2 are the explanatory variables. B1 is the intercept coefficient and B2 and B3 are partial slope coefficients.

Here, the Fertility rate is the dependent variable, Y and the independent variables are Urban population, X1, and Income, X2.

## Development of multiple linear model using R

```
Call:
lm(formula = d2$St ~ d2$Q + d2$L)

Residuals:
     Min       1Q   Median       3Q      Max 
-0.09279 -0.07183 -0.00469  0.04852  0.14965 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  8.847e+00  9.717e-01   9.104 6.47e-09 ***
d2$Q        -2.009e-01  3.763e-02  -5.340 2.32e-05 ***
d2$L         4.331e-05  1.857e-04   0.233    0.818    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07805 on 22 degrees of freedom
Multiple R-squared:  0.9767,	Adjusted R-squared:  0.9746 
F-statistic: 461.3 on 2 and 22 DF,  p-value: < 2.2e-16
```

From the output obtained from R, OLS estimated the equation as follows,

Fertility rate = 8.847-0.2009*Urban Population+ 0.00004331*Income+ Error

The fertility rate decreased by 20.09% or 0.2009 for every 1% increase in the Urban population by holding income constant. The fertility rate increased by 0.00004331 for every 1% increase in income. The residual standard error for the multiple regression is 0.07805, Since the value is low more closely the regression line can match the data. The p-value is less than 2.2e-16, which states that the data is statistically significant.

```
Analysis of Variance Table

Response: d2$St
          Df Sum Sq Mean Sq  F value Pr(>F)    
d2$Q       1 5.6205  5.6205 922.5426 <2e-16 ***
d2$L       1 0.0003  0.0003   0.0544 0.8178    
Residuals 22 0.1340  0.0061                    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> jarque. bera.test(residuals(Reg2))

	Jarque Bera Test

data:  residuals(Reg2)
X-squared = 1.7634, df = 2, p-value = 0.4141
```

## Conclusion

The increase in urban population, as well as income over the period 1996-2020, has contributed to the decrease in the fertility rate in India. Through analysis of data sets and statistical methods, it has become clear that there is a significant negative correlation between fertility rate, urban population, and income. As the urban population increases, fertility rates decline. Additionally, there is a strong correlation between urban population and income, with urban areas generally having higher income levels than rural areas. Reducing fertility rates can contribute to economic growth and development, as it allows for greater investment in human capital and infrastructure. This study also highlights the importance of access to quality healthcare and social services which play a crucial role in reducing fertility rates. Also, highlights the need for policies and programs that promote family planning and reproductive health, as well as investments in education, healthcare, and infrastructure, to support sustainable urban growth and economic development. Other factors such as education, access to healthcare, and cultural norms also play a significant role in shaping population growth and economic development.

## Limitations of the study

The availability of data for many variables was not accessible during the study projects execution because the study has to be condensed into a two-variable analysis. Regression analysis results are only relevant within the range of the analysiss data; extrapolating them outside of that range may not be reliable.

## Reference

1. Ranganathan, S., Swain, R. B., & Sumpter, D. J. (2015). The demographic transition and economic growth: Implications for development policy. Palgrave Communications, 1(1), Article 1. https://doi.org/10.1057/palcomms.2015.33.

2. Tripathi, D. S. (n.d.). Relationship between urbanization and health outcomes in Indian states.
3. The demography of fertility and infertility. (n.d.). Retrieved April 22, 2023, from https://www.gfmer.ch/Books/Reproductive_health/The_demography_of_fertility_and_infertility.htm.
4. Demography—Fertility rates—OECD Data. (n.d.). TheOECD. Retrieved April 22, 2023, from http://data.oecd.org/pop/fertility-rates.htm.
5. Ritchie, H., & Roser, M. (2018). Urbanization. Our World in Data. https://ourworldindata.org/urbanization
6. Guilmoto, C. Z., & Rajan, I. (n.d.). Fertility at District Level in India: Lessons from the 2011 Census.
7. Martine, G., Alves, J. E., & Cavenaghi, S. (n.d.). Urbanization and fertility decline:
8. Dharmalingam, A., Rajan, S., & Morgan, S. P. (2014). The Determinants of Low Fertility in India. Demography, 51(4), 1451–1475. https://doi.org/10.1007/s13524-014-0314-9
9. World Bank Open Data. (n.d.). World Bank Open Data. Retrieved April 22, 2023, from https://data.worldbank.org
10. Schultz, T. P. (2005). Fertility and Income (SSRN Scholarly Paper No. 838227). https://doi.org/10.2139/ssrn.838227