ECONOMETRIC ANALYSIS ON TOTAL FERTILITY RATE

BY- ABANTIKA BASU , MA ECO-2024-26 , ROLL NO- 2

■ INTRODUCTION

The total fertility rate (TFR) is a critical demographic measure that estimates the average number of children a woman is expected to have over her reproductive lifetime (ages 15-49), assuming current age-specific fertility rates remain constant. As a key indicator of population dynamics, TFR plays a pivotal role in understanding the socio-economic and cultural trends influencing population growth and decline.

A TFR of 2.1 children per woman ensures a broadly stable population.

The global average TFR has dropped steadily to less than half that number, 2.3 births per woman in 2023.

In this project, we have taken the data of rural and urban areas of different states across our country and have tried to analyze the effects of different factors on TFR. The outcomes of this project will provide insights into global and local fertility patterns, highlight disparities, and inform decision-making processes in areas such as public health, education, and economic policy.

■ PROBLEM STATEMENT

This analysis project aims to investigate the factors that shape TFR, its historical and regional variations.By leveraging statistical tools and demographic data, the study seeks to answer essential questions such as:

1. How do socio-economic factors like literacy rate,age of marriage,social awareness about modern techniques, healthcare access,infant mortality rate impact fertility rates?

2. How does TFR influence population structures and long-term economic development?

■ DATA SET

1. SOURCE OF DATA: As the area of research work concentrates in health ,so the data has been collected from https://www.data.gov.in. Under the branch of Health ,I took the nfhs_5_factsheets data (survey period- June, 2019).
C:\Users\ABANTIKA\OneDrive\NFHS_5_Factsheets_Data.xls

2. DESCRIPTION OF VARIABLES:

The variables which have been considered for this analysis are:

- **Total Fertility Rate (tfr):** It is the **dependent variable** .It estimates the average number of children a woman is expected to have over her reproductive lifetime (ages 15-49).
- **Women of age 15-49 who are literate:** It is an **independent variable**. Literacy is a major component for measuring fertility. If a woman is not literate then she won't have any basic knowledge about reproduction and health isssues.

- **Women who are married before age 18:** It is an **independent variable**. Age of marriage plays an important role since in our country women are still being forced to get married early.
- **Family planning method:** It is another **independent variable**. For the sake of a woman's health and a child's proper upbringing in this century, proper family planning method is necessary.
- **Mother's nutritrion:** It is an **independent variable** used here. In our coutry, at many abundant areas proper medical and nutritional facilities are still not provided.Women do not get prenatal and postnatal care properly and that adversely affects their health.
- **Women's decision making power:** It is also an **independent variable** used in our analysis. Women's decision-making power is a crucial determinant of Total Fertility Rate (TFR), influencing reproductive choices, family planning, and overall fertility behavior. It encompasses their ability to make autonomous decisions regarding education, employment, healthcare, and family size.
- **Infant Mortality Rate:** It is an important independent variable used in our model. Infant mortality rate is the probability of a child born in a specific year or period dying before reaching the age of one, if subject to age-specific mortality rates of that period which highly affects tfr.

3. <u>SAMPLE SIZE:</u> There are all over **111 observations** which includes observations of rural and urban areas separately in each states of india and also the sum total observation of rural and urban areas of them.

4. <u>MISSING VALUES:</u> We cannot find any data of the undertaken variables for the rural area of Chandigarh. Also we get some missing values of IMR for Andaman,Sikkim,Goa.

▪ DATA EXPLORATION:

• DESCRIPTIVE STATISTICS:

Descriptive statistics are typically applied to each variable to understand its fundamental properties before conducting further analysis.

Variable	Obs	Mean	Std. Dev.	Min	Max
Womenage15~a	111	79.8582	17.46262	-69.23	99.14
States_UTs	111	19	10.7255	1	37
area	111	2	.8201995	1	3
tfr	111	39.72973	20.65383	1	78
womenmarr~18	111	55.95495	32.1127	1	110
fam_plan_m~d	111	55.28829	31.8312	1	110
mother_nut~n	111	54.82883	31.35018	1	109
women_deci~n	111	54.12613	31.14398	1	107
imr	111	42.81982	27.88587	1	94

In the above chart, we can see the mean,standard deviation, minimum value and maximum value of all variables under consideration.
OBSERVATION:

· **States_UTs** (State/UT ID)

- **Observations:** 111
- **Mean:** 19
- **Standard Deviation:** 10.73 (moderate spread).
- **Range:** 1 to 37 (likely categorical or geographical identifiers).

· **area** (for rural/urban or zones)

- **Observations:** 111
- **Mean:** 2
- **Standard Deviation:** 0.82 (low variability).
- **Range:** 1 to 3.

· **tfr** (Total Fertility Rate)

- **Observations:** 111
- **Mean:** 39.73
- **Standard Deviation:** 20.65 (high variability).
- **Range:** 1 to 78 (wide range).

· **Womenage15~a** (Percentage of women aged 15-49)

- **Observations:** 111
- **Mean:** 79.86
- **Standard Deviation:** 17.46 (moderate variability)
- **Range:** -69.23 to 99.14 (negative minimum is unusual; may indicate data errors).

- **womenmarr~18** (Percentage of women married before age 18)

  1. **Observations:** 111
  2. **Mean:** 55.95
  3. **Standard Deviation:** 32.11 (high variability).
  4. **Range:** 1 to 110.

- **fam_plan_m~d** (Family planning met demand)

  1. **Observations:** 111
  2. **Mean:** 55.29
  3. **Standard Deviation:** 31.83.
  4. **Range:** 1 to 110.

- **mother_nut~n** (Mothers' nutritional status)

  1. **Observations:** 111
  2. **Mean:** 54.83
  3. **Standard Deviation:** 31.35.
  4. **Range:** 1 to 109.

- **women_deci~n** (Women's decision-making power)

  1. **Observations:** 111
  2. **Mean:** 54.13
  3. **Standard Deviation:** 31.14.
  4. **Range:** 1 to 107.

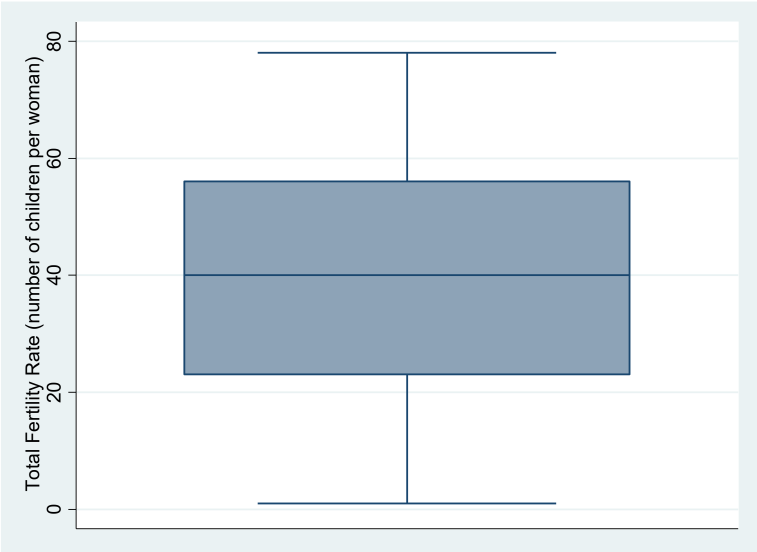
- **imr (Infant Mortality Rate):**

- **Mean**: 42.82 suggests an average infant mortality rate per 1,000 live births.
- **Std. Dev.**: 27.89 shows wide disparities.
- **Max (94)**: Regions with very high mortality rates are present.

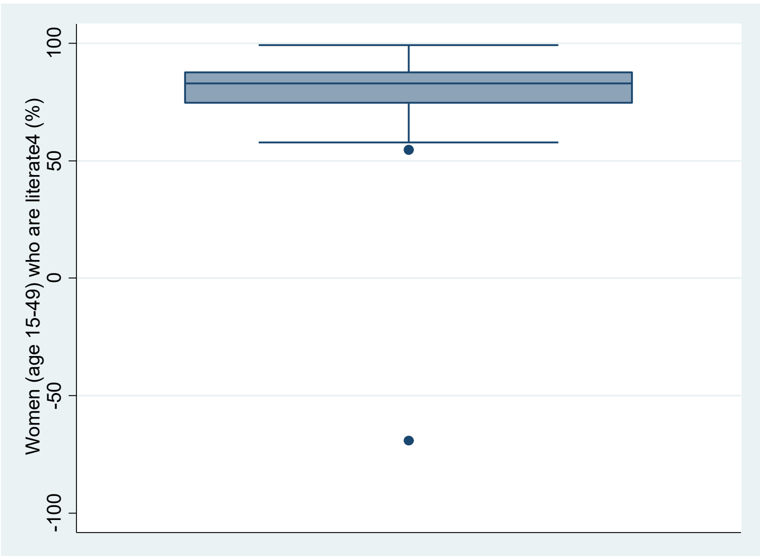
BOX PLOTS:

Box Plot is a graphical method to visualize data distribution for gaining insights and making informed decisions.

1.Total Fertility Rate

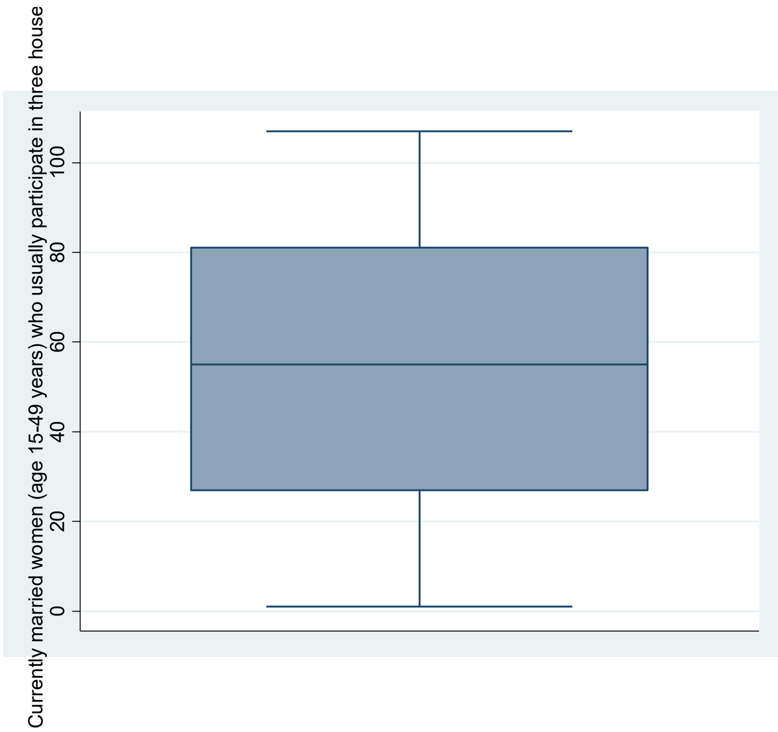


1. Women Literay Rate

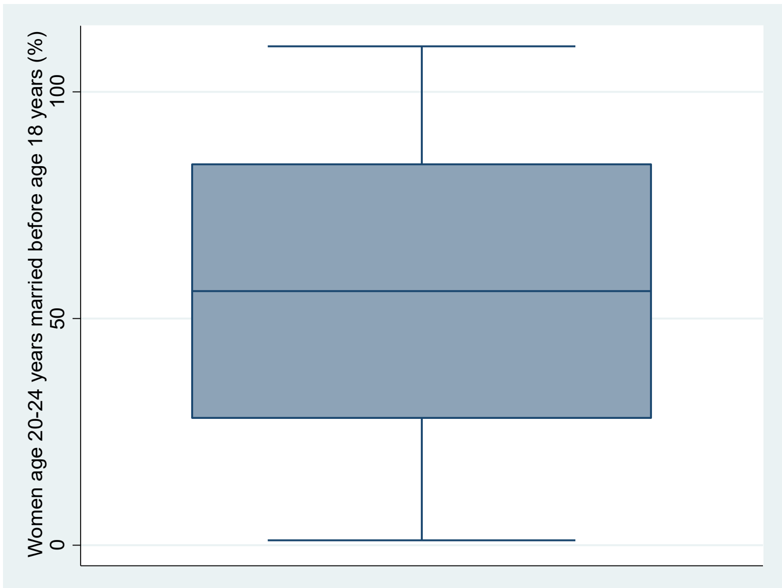


Only one outlier is observed.

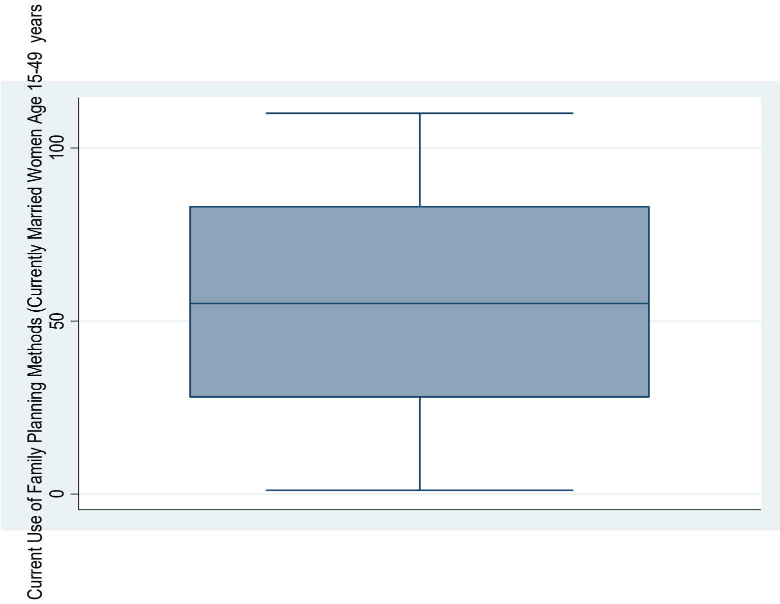
2. Women who have the right to take decision



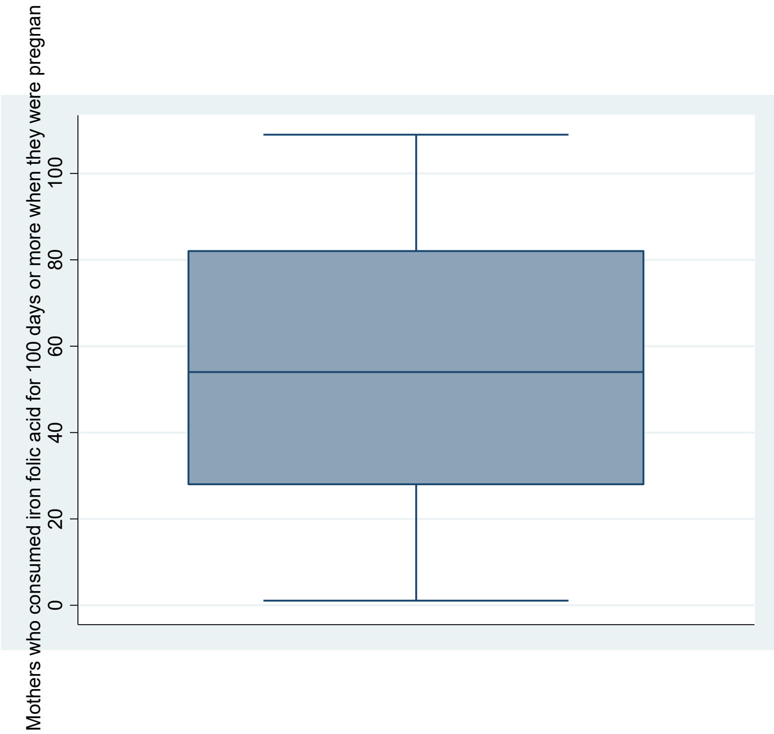
3. Women who got married before age 18



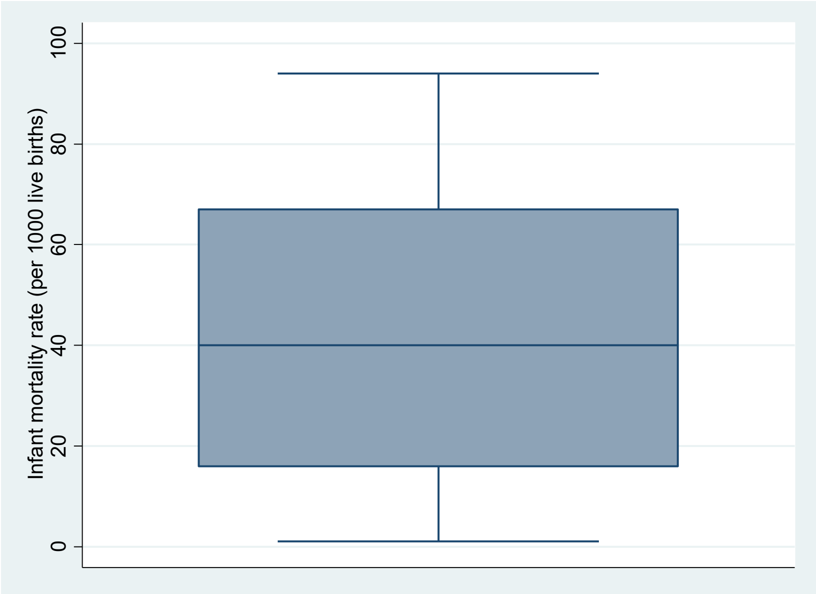
4. Family planning method



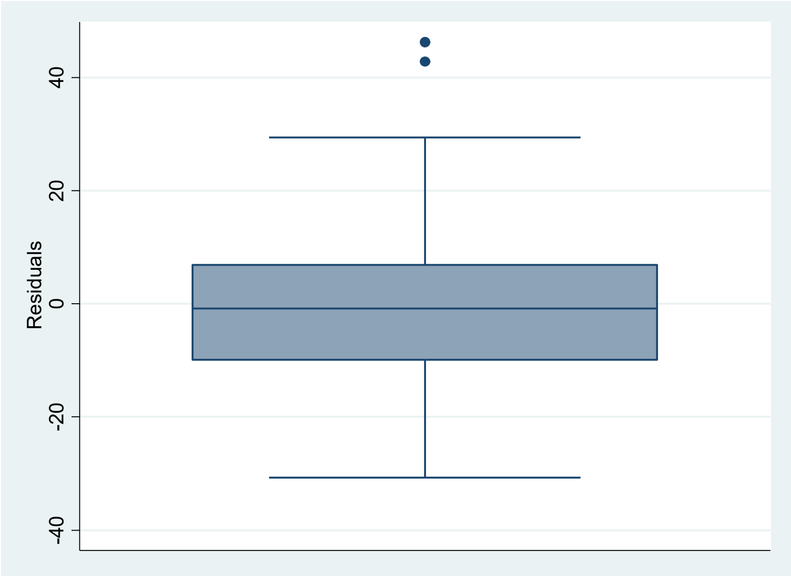
5. Mother’s nutrition



6.Infant Mortality Rate



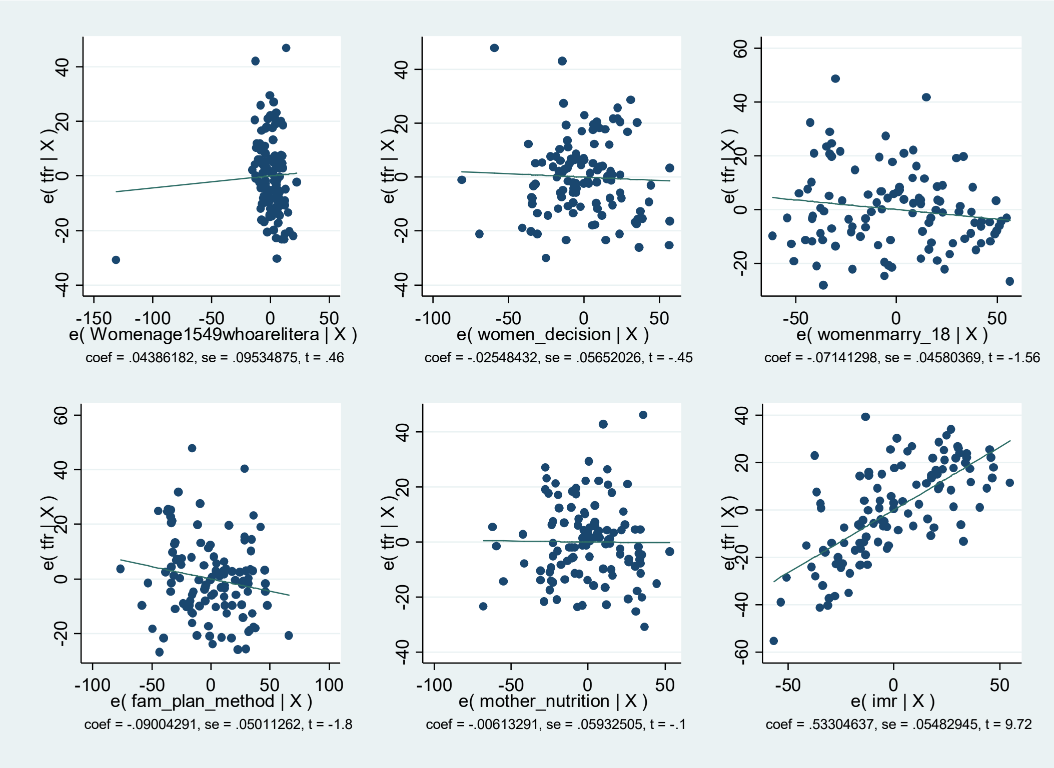
7.Residual



Two outliers are observed.

■ LINEARITY:

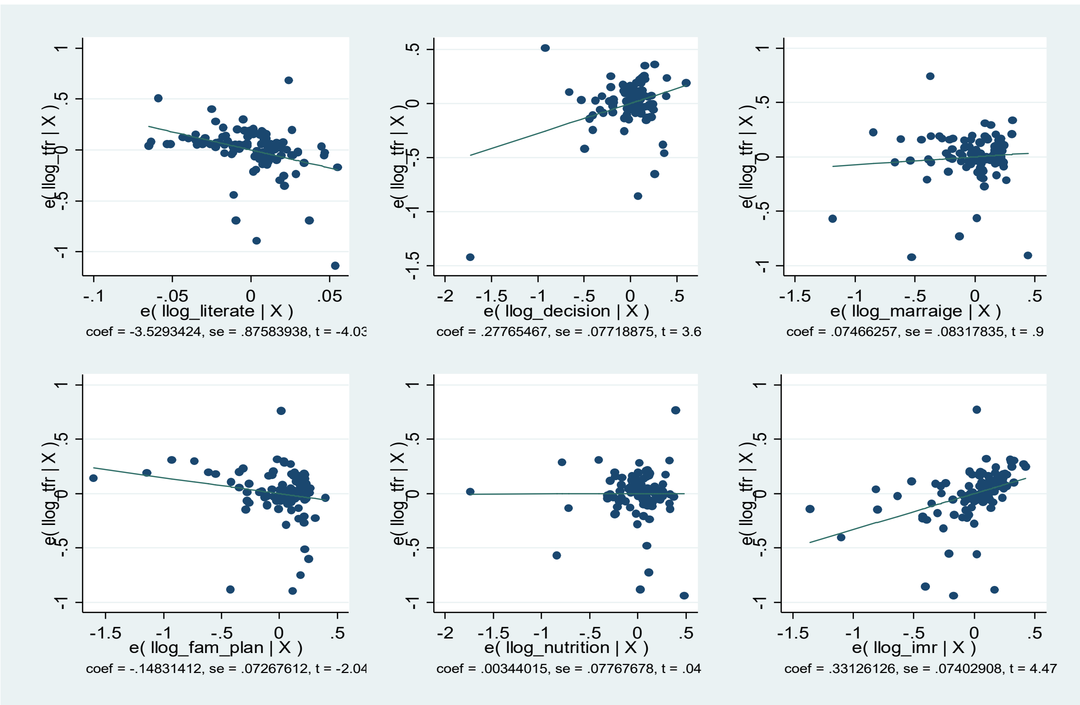
To check linearity,we first run the regression model. And then check the partial regression plots.



We can see that the avplots doesn't show linearity between the dependent and each independent variable.The plots are scattered away from the linear line. So we will consider Box-Tidwell transformation I.e the logarithmic transformation of each variable to satisfy linearity.

- **<u>REMEDIAL MEASURE:</u>**

Taking double log transformation of all the variables we can finally eliminate the non linearity issue.



OBSERVATION

- **Direction and strength of the Relationship:**

1. The first and fourth graph shows the relationship between tfr and literacy rate ,tfr and family planning method is negative since they are downward sloping.
2. The second and sixth graph shows positive relationship between tfr and women's decision making power,tfr and imr since it is upward sloping.
3. The third and fifth graph shows a linear/horizontal line i.e positive relationship between tfr and age of marriage ,tfr and mother's nutrition.

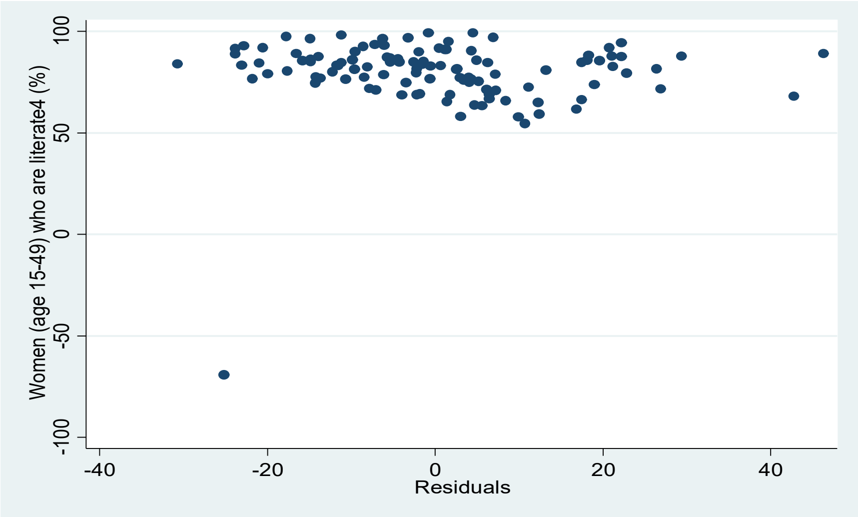
- **Spread of Data Points:**

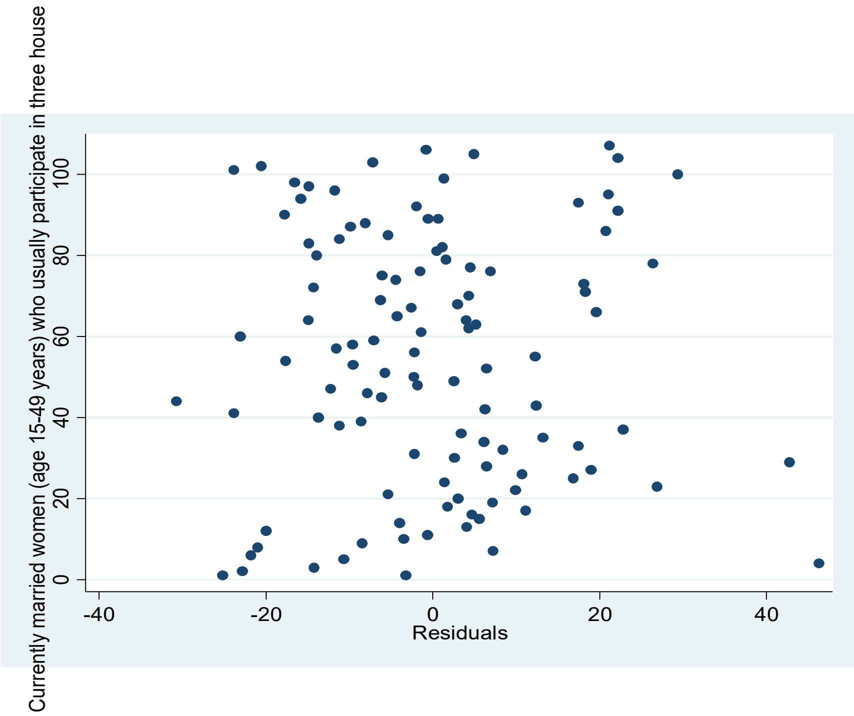
  A tighter cluster of points around the fitted line indicates a stronger relationship between the dependent variable ,tfr and each independent variable This implies the regressors explain the variation in total fertility rate in a good way.

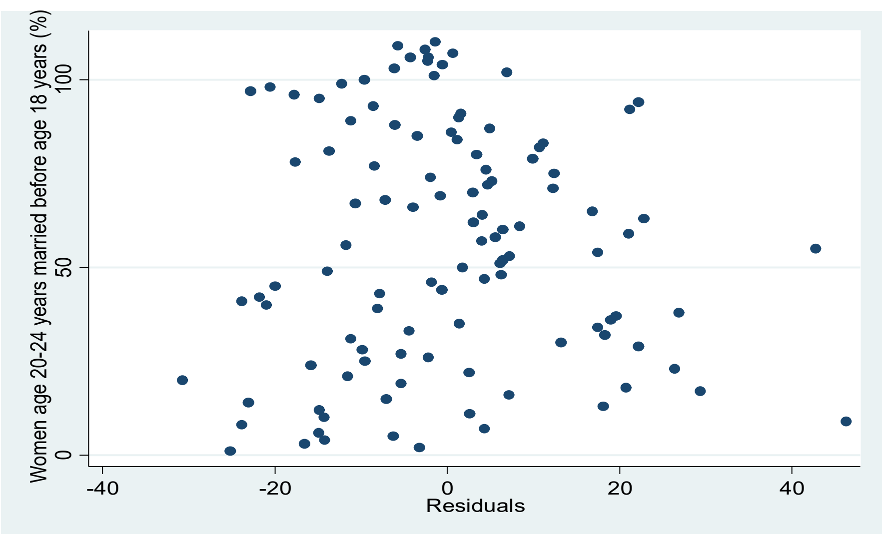
■ HETEROSCEDASTICITY

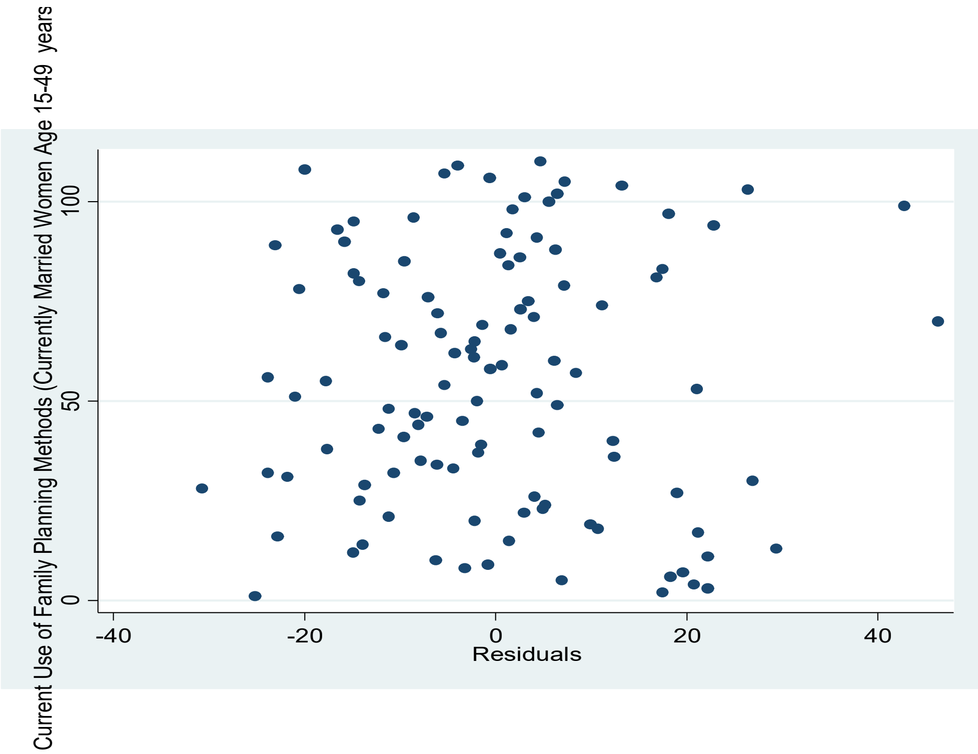
Heteroskedasticity refers to situations where the variance of the residuals is unequal over a range of measured values. When running a regression analysis, heteroskedasticity results in an unequal scatter of the residualsWhen heteroscedasticity occurs, it can make a regression model less robust and suggest that the model is misspecified.

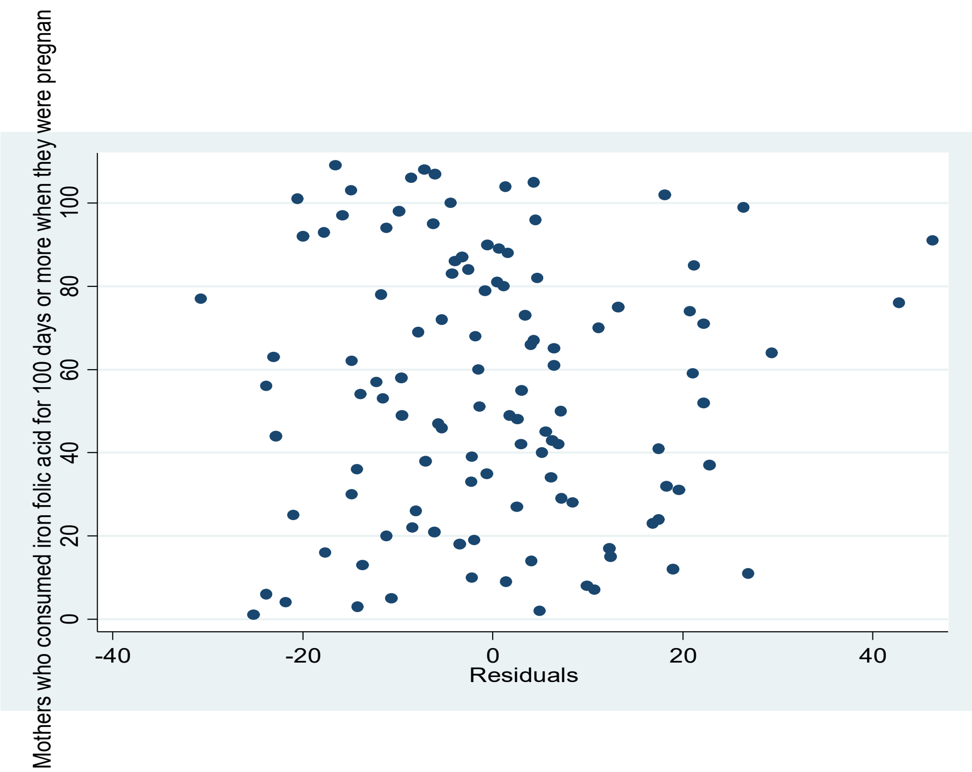
- **Residual vs Regressors plot**

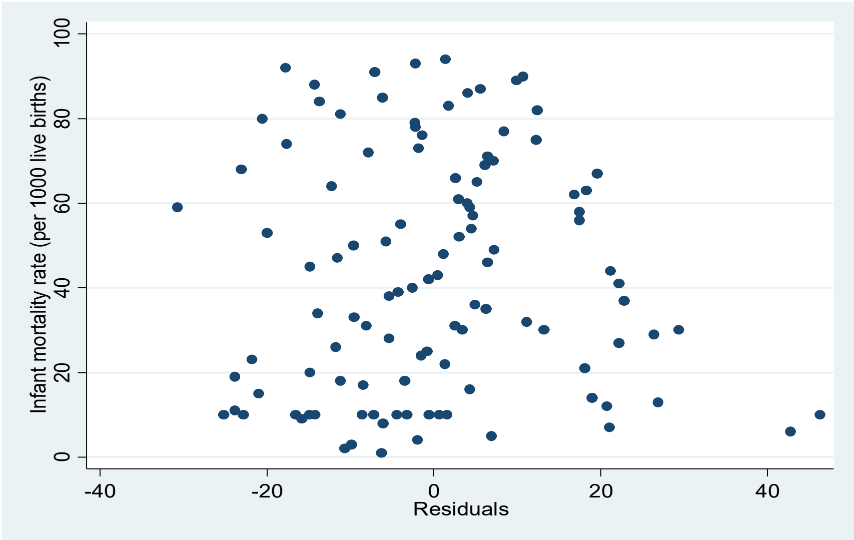

Women (age 15-49) who are literate4 (%)
Residuals


Currently married women (age 15-49 years) who usually participate in three house
Residuals


Women age 20-24 years married before age 18 years (%)
Residuals


Current Use of Family Planning Methods (Currently Married Women Age 15-49 years
0
50
100
-40
-20
0
20
40
Residuals


Mothers who consumed iron folic acid for 100 days or more when they were pregnan
0
20
40
60
80
100
-40
-20
0
20
40
Residuals



We can see that the plots are scattered showing heteroscedasticity.

reg tfr Womenage1549whoarelitera women_decision womenmarry_18 fam_plan_method mother_nutrition imr

Source	SS	df	MS		
				Number of obs =	111
				F(6, 104) =	18.52
Model	24235.7828	6	4039.29714	Prob > F =	0.0000
Residual	22688.109	104	218.154895	R-squared =	0.5165
				Adj R-squared =	0.4886
Total	46923.8919	110	426.580835	Root MSE =	14.77

tfr	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Womenage1549whoarelitera	.0438618	.0953487	0.46	0.646	-.1452183	.232942
women_decision	-.0254843	.0565203	-0.45	0.653	-.1375661	.0865975
womenmarry_18	-.071413	.0458037	-1.56	0.122	-.1622434	.0194175

fam_plan_method |	-.0900429	.0501126	-1.80	0.075	-.1894181	.0093323
mother_nutrition |	-.0061329	.0593251	-0.10	0.918	-.1237767	.1115109
imr |	.5330464	.0548294	9.72	0.000	.4243175	.6417752
_cons |	24.09191	7.814412	3.08	0.003	8.595637	39.58818

Here we can see, the standard errors of the coefficients are underestimated, leading to incorrect significance tests.So we conduct the formal Bruesh Pagan test. The prob>chi2 value is coming 0.02 impling presence of heteroscedasticity.

- **REMEDIAL MEASURE**: We take robust regression with the log transformed independent variables here.

reg tfr llog_tfr llog_literate llog_decision llog_marraige llog_fam_plan llog_nutrition llog_imr,robust

Linear regression   Number of obs = 98

F(7, 90) = 68.86

Prob > F = 0.0000

R-squared = 0.8259

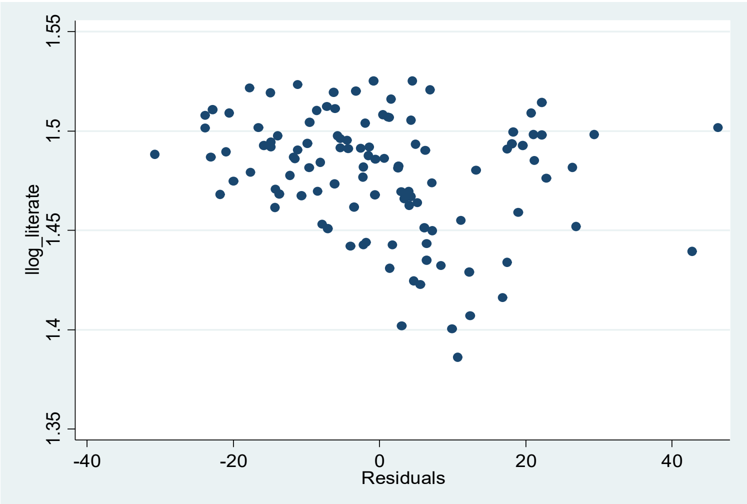
Root MSE = 8.3768

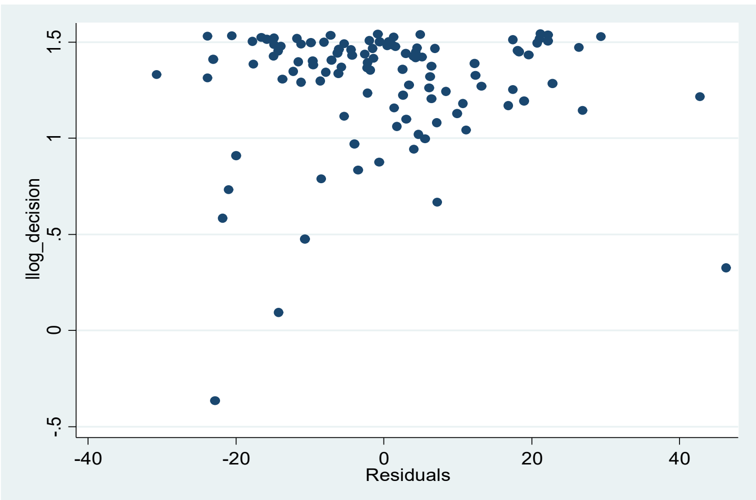
tfr |	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
llog_tfr |	52.63513	10.05773	5.23	0.000	32.6537	72.61657
llog_literate |	-224.9966	45.28954	-4.97	0.000	-314.9722	-135.021
llog_decision |	13.53897	4.676266	2.90	0.005	4.248751	22.82919

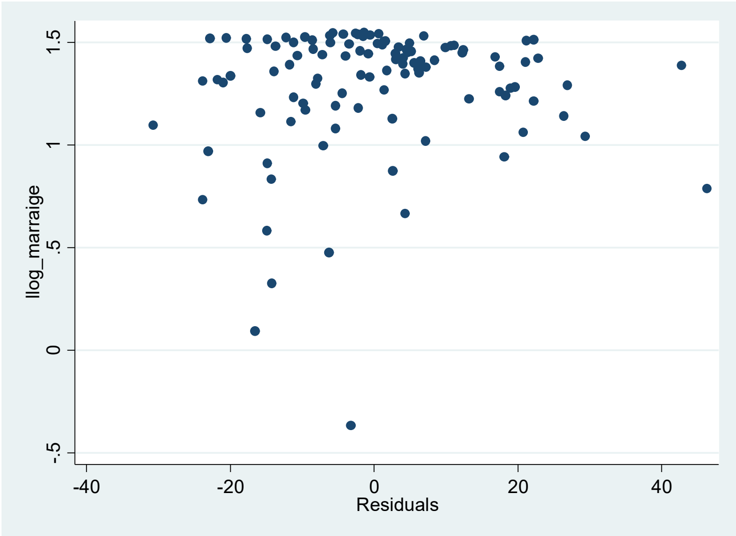
llog_marraige	-12.11102	5.061575	-2.39	0.019	-22.16672	-2.055314
llog_fam_plan	-12.56941	4.011823	-3.13	0.002	-20.5396	-4.599228
llog_nutrition	-1.562988	3.205053	-0.49	0.627	-7.930386	4.80441
llog_imr	10.67522	3.817156	2.80	0.006	3.091769	18.25866
_cons	311.6224	71.94883	4.33	0.000	168.6835	454.5613

We an see that estimates are unbiased,robust standard errors are higher than earlier which is minimizing the heteroscedasticity though results are inefficient still now. But p values are significant now whch is a positive sign. However we may not remove heteroscedasticity fully from this data plot.

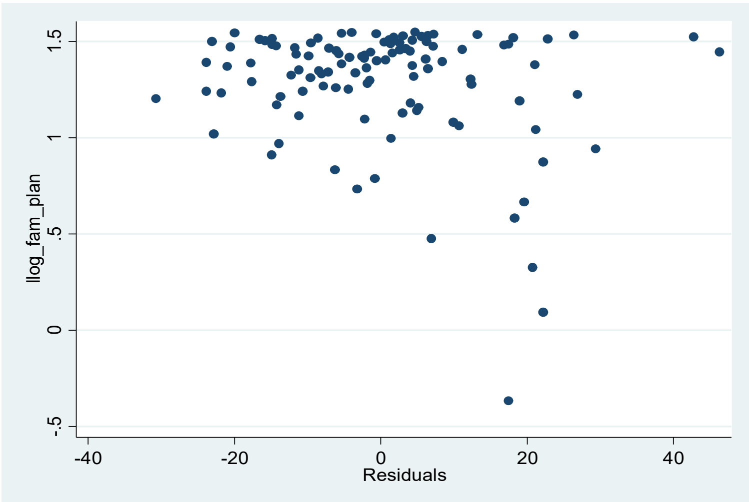
- **RESIDUAL VS REGRESSOR PLOT AFTER TRANSFORMATION**



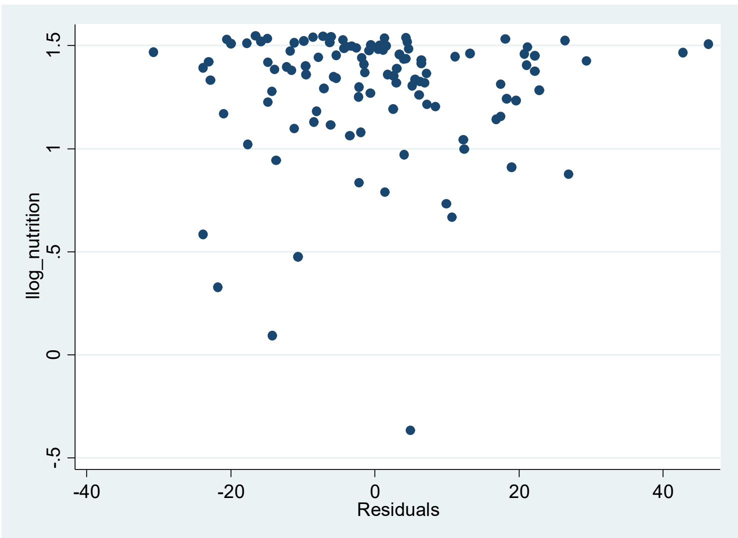




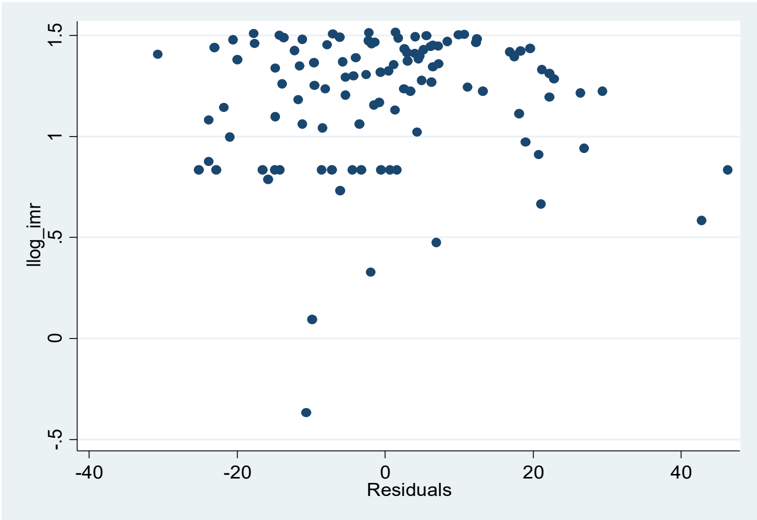
llog_marraige
Residuals



llog_fam_plan
Residuals



llog_nutrition
Residuals

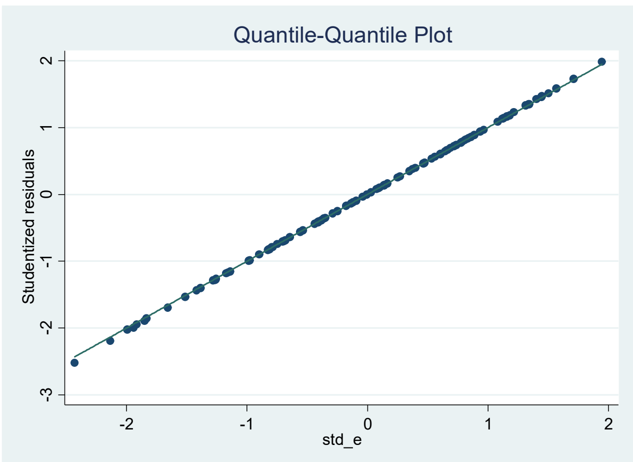


We can see dataplots are less scattered than earlier without any recognizable pattern lowering heteroscedasticity.

■ NORMALITY

Normality refers to the assumption that a dataset or a variable is approximately distributed according to a **normal distribution**, also known as a Gaussian distribution.If probability is grater than 0.05 then the normality of the model is satisfied.

OBSERVATION



A **Q-Q plot** (Quantile-Quantile plot) is a graphical tool used to assess whether a dataset follows a specific theoretical distribution, most commonly the **normal distribution**. We have taken studentized and standardized residuals and generated the q-q plot. We see that the data are plotted along the 45 degree line impliying that the model follows normal distribution.

We have performed the **Saphiro-Wilk** test generating p value less than 0.05 which implies the model doesnot satisfy normality.

- ## Remedial Measure

For this we conduct **Box-Cox transformation** to make the residual follow normal distribution.

boxcox tfr Womenage1549whoarelitera womenmarry_18 fam_plan_method mother_nutrition women_decision imr

Fitting comparison model

Iteration 0: log likelihood = -493.09692

Iteration 1: log likelihood = -491.97829

Iteration 2: log likelihood = -491.96013

Iteration 3: log likelihood = -491.96013

Fitting full model

Iteration 0: log likelihood = -452.76584

Iteration 1: log likelihood = -452.10798

Iteration 2: log likelihood = -452.10319

Iteration 3: log likelihood = -452.10319

Number of obs = 111

LR chi2(6) = 79.71

Log likelihood = -452.10319 Prob > chi2 = 0.000

----

tfr | Coef. Std. Err. z P>|z| [95% Conf. Interval]

```
-------------+----------------------------------------------------------------
      /theta |   .8515845   .1255978     6.78   0.000     .6054175    1.097752
------------------------------------------------------------------------------

Estimates of scale-variant parameters
----------------------------
             |        Coef.
-------------+--------------
Notrans      |
Womenage15~a |     .0377796
womenmarr~18 |    -.0389943
fam_plan_m~d |    -.0481796
mother_nut~n |    -.0029458
women_deci~n |     -.014764
         imr |     .3174621
       _cons |     14.49407
-------------+--------------
      /sigma |     8.502173
----------------------------

---------------------------------------------------------
   Test         Restricted     LR statistic      P-value
    H0:       log likelihood       chi2        Prob > chi2
---------------------------------------------------------
theta = -1      -634.84535       365.48           0.000
```

theta = 0	-483.17643	62.15	0.000
theta = 1	-452.76584	1.33	0.250

Now the formal **Saphiro-Wilk** test is giving a significant p-value.

swilk std_e

Shapiro-Wilk W test for normal data

Variable	Obs	W	V	z	Prob>z
std_e	111	0.97890	1.901	1.433	**0.07591**

We see that p-value is now 0.07591 which is grater than 0.05 satisfying normality.

■ LEVERAGES AND OUTLIERS

Leverage refers to the influence of an individual data point on the estimation of a regression model. It measures how far an independent variable's value is from the mean of all the independent variable values. High-leverage points can have a substantial effect on the fitted regression line, even if their corresponding dependent variable values are not outliers.

· Leverage quantifies the degree to which a data point contributes to the determination of the regression line or plane.
· It is derived from the **hat matrix** H, where hii(diagonal elements) measures the leverage of the i-th observation.
Here we have predited the leverage values ad the values are low which implies the regression model is well fitted.

■ MODEL SELECTION

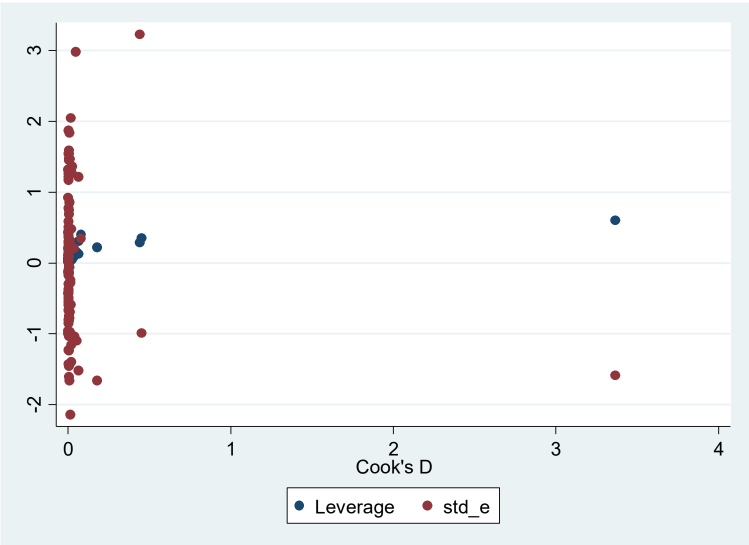
Model selection is the process of choosing the most appropriate statistical or machine learning model from a set of candidate models to best explain the underlying data or make accurate predictions. The goal is to find a model that balances complexity and performance, ensuring it generalizes well to unseen data.

We have run the regressions eliminating one regressor at a time and observed that the Rsquared values are coming more worse. So we can say that the model is better when selected with the undertaken regressors.These regressors better explain the variation in the dependent variable compared to other regressions done after omitting regressors one by one. Though this model has moderate adjusted R square value but a model with moderate Rsquare value can also produce good result if it's p values are significant. And from the above results we can see that the p values are significant and model has satisfied linearity, heteroscedasticity and normality after transformation.

- **INFLUENTIAL STATISTICS AND THEIR INTERPRETATION**

- **COOK'S DISTANCE:** Cook's distance is a statistical measurement that indicates how much a data point influences the results of a regression analysis.Cook's distance measures how much the regression coefficients would change if a particular data point were excluded from the analysis.Cook's distance is used to identify influential data points, or outliers, in a regression analysis. It can also help identify regions of the design space where more data points might be useful.

OBSERVATION



- A rule of thumb is that Cook's Distance > 1 indicates a highly influential point.
- **Points near or above the threshold line:**

    - These observations have a strong impact on the regression model. Investigate their validity and whether they represent outliers, data errors, or unique conditions.
- **Uniformly low values:**
    - Indicates no single data point excessively influences the regression.

Points with high leverage:

- Lie far from the center of the predictor variable space.
- Do not necessarily indicate an issue unless they also have large residuals.

High Leverage and Low Residual:

- The point may be valid and fits the model well.

High Leverage and High Residual:

- The point is both influential and poorly predicted, which might distort the model.

- Random scatter plot of residuals indicate the model is appropriate.

1. Very few observations with cook's distance >1 implies that they disproportionately affect tfr predictors.
2. High-leverage points might represent regions with unusual predictor combinations (e.g., extremely high women's decision-making power or high consumption of nutrition or good literacy rate).
3. The random pattern of residuals indicates that the model has no need of any interaction terms ; tfr is good predicted by the regressors.

■ TEST OF HYPOTHESIS

● MOEDL COMPARISON USING ANOVA

We have tested one way ANOVA for our model. To test the significance of the model we have considered the F-stat values from the table.

The F-statistic in a regression model tests the null hypothesis that all coefficients in the model (except the intercept) are equal to zero. For the model to be statistically significant, the F-statistic must be large enough(grater than the critical value 0.05) to reject this null hypothesis.

We consider the null hypothesis H0: means of all groups are same.

OBSERVATION

oneway llog_tfr llog_literate

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	8.96502857	109	.082247968		
Within groups	0	0	.		
Total	8.96502857	109	.082247968		

. oneway llog_tfr llog_decision

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	8.89503983	105	.084714665	3.63	0.1569
Within groups	.069970269	3	.023323423		
Total	8.9650101	108	.083009353		

Bartlett's test for equal variances: chi2(2) = 5.6725 Prob>chi2 = 0.059

note: Bartlett's test performed on cells with positive variance:

103 single-observation cells not used

. oneway llog_tfr llog_marraige

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	8.95224131	108	.082891123	6.48	0.3047
Within groups	.012787257	1	.012787257		
Total	8.96502857	109	.082247968		

. oneway llog_tfr llog_fam_plan

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	8.93356962	108	.082718237	2.63	0.4613
Within groups	.031458949	1	.031458949		
Total	8.96502857	109	.082247968		

. oneway llog_tfr log_nutrition

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	8.91990808	107	.083363627	3.70	0.2366
Within groups	.045120483	2	.022560241		
Total	8.96502857	109	.082247968		

Bartlett's test for equal variances: chi2(1) = 0.0555 Prob>chi2 = 0.814

note: Bartlett's test performed on cells with positive variance:

106 single-observation cells not used

. oneway llog_tfr llog_imr

Analysis of Variance

Source	SS	df	MS	F	Prob > F
Between groups	6.5202841	92	.070872653	0.47	0.9874
Within groups	2.42539473	16	.15158717		
Total	8.94567882	108	.082830359		

Bartlett's test for equal variances: chi2(4) = 10.1185 Prob>chi2 = 0.038

note: Bartlett's test performed on cells with positive variance:

88 single-observation cells not used

1. We cannot get any F-stat value for tfr and literacy rate just because maybe the means of all groups are same.

2. For the test using tfr and Women decision ,the value of F-stat is coming 3.63 which implies that the null hypothesis is rejected-means are not same across groups and the model is significant.

3. For the test using tfr and age of marriage,the value of F-stat is 6.48 that means here also the null hypothesis is rejected and the model is significant.

4. ANOVA test using tfr and family planning method gives F-stat value 2.63 which is good because a higher F-stat value implies significance of the model.

5. The test performed using tfr and mother’s nutrition generates F-stat value 3.70 which is grater than the critical value. So we can reject the null hypothesis and the model is significant.

6. ANOVA test using tfr and imr shows F-stat value 0.47 which is obviously grater than the critical value implying the model to be significant.

■ MULTICOLLINEARITY

Multicollinearity is the occurrence of high intercorrelations among two or more independent variables in a multiple regression model. Multicollinearity can lead to skewed or misleading results when a researcher or analyst attempts to determine how well each independent variable can be used most effectively to predict or understand the dependent variable in a statistical model.

In general, multicollinearity can lead to wider confidence intervals that produce less reliable probabilities in terms of the effect of independent variables in a model.

OBSERVATION

Variable	VIF	1/VIF
llog_tfr	1.85	0.540963
llog_liter~e	1.76	0.567990
llog_imr	1.52	0.656740
llog_decis~n	1.48	0.677309
llog_nutri~n	1.32	0.756781
llog_fam_p~n	1.16	0.859638
llog_marra~e	1.06	0.946597
Mean VIF	1.45	

VIF < 5: Multicollinearity is low and generally not a concern.

A VIF of 1.85 indicates very low multicollinearity. It is well below the commonly used threshold of 5, so there’s no need to take action.

Similarly,we an see that for all variables ,multicollinearity is not exceeding 5. Though the explanatory variables are correlated but it has not so strong impact in our analysis.

In general,we can say that total fertility rate can be explained by literacy rate, women's decision making power, age of marriage, family planning method and mother's nutrition in a well manner.All the variables have significant impact on total fertility rate.

■ STATA CODES

```
*====================================================

*Ecotrix Assignment( by Abantika Basu)

*====================================================

clear all

set more off

cd"C:\Users\ABANTIKA\Desktop\stata assignment iift"

capture log close

log using "ecotrix assignment",text replace

use"C:\Users\ABANTIKA\Downloads\nfhs_5_factsheet_data_final.dta",clear

br

encode StatesUTs,gen(States_UTs)

encode Area,gen(area)

encode TotalFertilityRatenumberof,gen(tfr)

encode Womenage2024yearsmarriedbe ,gen(womenmarry_18)

encode CurrentUseofFamilyPlanningM,gen(fam_plan_method)

encode Motherswhoconsumedironfolic,gen(mother_nutrition)

encode Currentlymarriedwomenage15,gen(women_decision)

encode Infantmortalityrateper1000,gen(imr)
```

```
br

des

sum

graph box tfr

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph box plot of tfr.gph"

graph box Womenage1549whoarelitera

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph box plot of first regressor.gph"

graph box women_decision

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph box plot of women decision.gph"

graph box womenmarry_18

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph box plot of marry 18.gph"

graph box fam_plan_method

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph box plot of family plan.gph"

graph box mother_nutrition

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph box plot of mom nuitrition.gph"

graph box imr

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph box plot of imr.gph"

reg tfr Womenage1549whoarelitera women_decision womenmarry_18 fam_plan_method mother_nutrition imr

predict e,resid

graph box e

graph save Graph "C:\Users\ABANTIKA\Downloads\box plot graph.gph"
```

```
reg tfr Womenage1549whoarelitera

reg tfr women_decision

reg tfr womenmarry_18

reg tfr fam_plan_method

reg tfr mother_nutrition

reg tfr imr

reg tfr Womenage1549whoarelitera women_decision womenmarry_18
fam_plan_method mother_nutrition imr

avplots

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph avplot
prior log.gph"

gen log_tfr=log(tfr)

gen log_literate=log( Womenage1549whoarelitera )

gen log_decision=log( women_decision )

gen log_marraige=log( womenmarry_18 )

gen log_fam_plan=log( fam_plan_method )

gen log_nutrition=log( mother_nutrition )

gen log_imr=log(imr)

reg log_tfr log_literate log_decision log_marraige log_fam_plan log_nutrition
log_imr

avplots

gen llog_tfr=log( log_tfr )

gen llog_literate=log( log_literate )

gen llog_decision=log( log_decision )

gen llog_marraige=log( log_marraige )

gen llog_fam_plan=log( log_fam_plan )

gen llog_nutrition=log( log_nutrition )
```

```
gen llog_imr=log(log_imr)

reg llog_tfr llog_literate llog_decision llog_marraige llog_fam_plan llog_nutrition llog_imr

avplots

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph partial regression plot-avplots.gph"

reg tfr Womenage1549whoarelitera women_decision womenmarry_18 fam_plan_method mother_nutrition imr

twoway scatter Womenage1549whoarelitera e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph women literate e.gph"

twoway scatter women_decision e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph women decision e.gph"

twoway scatter womenmarry_18 e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph marraige age e.gph"

twoway scatter fam_plan_method e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph fam plan e.gph"

twoway scatter mother_nutrition e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph mother nutrition e.gph"

twoway scatter imr e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph imr e.gph"

estat hettest

reg tfr llog_tfr llog_literate llog_decision llog_marraige llog_fam_plan llog_nutrition llog_imr,robust

twoway scatter llog_literate e
```

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph literacy plot after log.gph"

twoway scatter llog_decision e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph decision after log.gph"

twoway scatter llog_marraige e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph marry after log.gph"

twoway scatter llog_fam_plan e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph fam plan after log.gph"

twoway scatter llog_nutrition e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph log nutrition.gph"

twoway scatter llog_imr e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph log imr r.gph"

swilk e

boxcox tfr Womenage1549whoarelitera womenmarry_18 fam_plan_method mother_nutrition women_decision imr

reg tfr Womenage1549whoarelitera women_decision womenmarry_18 fam_plan_method mother_nutrition imr

predict r,rstudent

sum

gen std_e=(e--5.10e-08)/14.3616

qqplot r std_e

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph qqplot.gph"

swilk std_e

predict leverage,hat

```
predict cooks,cooksd

scatter leverage std_e cooks

graph save Graph "C:\Users\ABANTIKA\Desktop\stata assignment iift\Graph leverage,e,cook.gph"

oneway llog_tfr llog_literate

oneway llog_tfr llog_decision

oneway llog_tfr llog_marraige

oneway llog_tfr llog_fam_plan

oneway llog_tfr log_nutrition

oneway llog_tfr llog_imr

vif

estat vif

exit,clear
```

■ CONCLUSION

The analysis of Total Fertility Rate (TFR) provided critical insights into demographic trends and their potential implications for socio-economic development.TFR has shown a consistent decline/increase (replace with actual trend) over the analyzed period, reflecting changes in socio-economic and healthcare factors.Regions with higher/lower TFR were identified, highlighting disparities that may require targeted policy interventions.Statistical models revealed that variables such as education, income, healthcare access, significantly influence TFR. Multicollinearity was low, ensuring robust model results.

The study faced limitations in data availability and scope, which could be addressed in future research by incorporating more granular or longitudinal data. As it is an in-person survey data ,so it is very much obvious that people are unwilling to response. And that is why we face very much disparities in our model but the overall impact of the model is good. The regressors taken into account have explained variations in total fertility rates and states that these are the major factors responsible for the high fertility rate across the country. This provides insightful knowledge that literacy rate should be increased, medical and nutritional facilities at hospital and household must be taken care of and awareness needed to be spread among women regarding usage of proper family planning method, health issues which may cause. Also the economic impact of this states that tfr is a major reason of high population growth. For a developing country like India, population growth needs to be controlled otherwise people will face high unemployment, shortage of food and many more complications. So, social and biological awareness should be spread among women to control the problem of high fertility rate.