STATISTICAL STUDY ON VITAL REPORTS

A Report submitted to the Department of Statistics

PSG College of Arts & Science, (Autonomous)

In partial fulfilment of the requirement for the award of the Degree of

MASTER OF SCIENCE

IN

STATISTICS

Of the Bharathiar University

Submitted by

1. DEEPANEESH R V (23MST005)
2. DHANAPRIYA B (23MST006)
3. JASMIN S (23MST011)
4. MOHAMED VAZIL G (23MST018)
5. MYTHRAYAN M (23MST019)
6. SRI AISHWARYA A M (23MST032)

UNDER THE GUIDANCE OF

Dr. E. SAKTHIVEL M.Sc., M.Phil., PGDCA., Ph.D.,

Assistant Professor,

Department of Statistics,

PSG College of Arts & Science,

Coimbatore.

DEPARTMENT OF STATISTICS

PSG COLLEGE OF ARTS & SCIENCE

An Autonomous College Affiliated to Bharathiar University

Accredited with ‘A++’ Grade by NAAC (3rd cycle)

College with Potential for Excellence (Status Awarded by the UGC)

Star College Status Awarded by DBT-MST

An ISO 9001:2015 certified Institution.

Civil Aerodrome Post, Coimbatore-641014

JANUARY 2023

DECLARATION

We, **DEEPANEESH R V , DHANAPRIYA B , JASMIN S , MOHAMED VAZIL G , MYTHRAYAN M , SRI AISHWARYA A M**. A here declare that the project entitled **"STATISTICAL STUDY ON VITAL REPORTS"** submitted to the **PSG COLLEGE OF ARTS & SCIENCE** in partial fulfilment of the requirement for the award of degree of **MASTER OF SCIENCE IN STATISTICS** is a record of original research work done by us under the supervision and guidance of **Dr. E. SAKTHIVEL M.Sc., M.Phil., PGDCA., Ph.D., Assistant Professor, DEPARTMENT OF STATISTICS.**

Name of Candidates Signature of Candidates

1. Deepaneesh. R.V (23MST005)
2. Dhanapriya. B (23MST006)
3. Jasmin. S (23MST011)
4. Mohamed Vazil. G (23MST018)
5. Mythrayan. M (23MST019)
6. Sri Aishwarya. A.M (23MST032)

ACKNOWLEDGEMENT

We express our sincere and deep filling thanks to our beloved principal who had been a great moral support to us and for providing an excellent environment at college. We greatly indented to record our respectful valuable suggestion encouragement during the period of this work. The meticulous guidance of our guide **Dr. E. SAKTHIVEL**, for his constant encouragement, inspiration and support at each and every stage of my project work which made the work get completed in time and for providing necessary guidance and encouragement for doing this project.

Our unforgettable thanks and deep sense of gratitude to my parents and friends for their encouragement, help, mental and physical support for the successful completion of my work.

CONTENT

CHAPTER I

1.1 INTRODUCTION:

Biostatistics (also known as biometry) is a branch of statistics that applies statistical methods to a wide range of topics in biology. It encompasses the design of biological experiments, the collection and analysis of data from those experiments, and the interpretation of the results.

1.1.1VITAL STATISTICS:

The vital statistics themselves are a critical national information resource for understanding public health and examining such key indicators as fertility, mortality, and causes of death, and the factors associated with them. In statistics, the two statistical values used to measure the growth or decline of a population are crude birth rate (CBR) and crude death rate (CBR). The growth or decline of a population is based on literacy and economic growth of any country.

1.1.2 BIRTH RATE:

Birth rate is the number of individuals born in a population in a given amount of time. Birth rate, also known as natality, is the total number of live human births per 1,000 population for a given period divided by the length of the period in years. The number of live births is normally taken from a universal registration system for births; population counts from a census, and estimation through specialized demographic techniques. The birth rate (along with mortality and migration rates) is used to calculate population growth. The estimated average population may be taken as the mid-year population.

1.1.3 DEATH RATE:

According to the mortality rate definition, an index of the number of deaths in a given year or a specified time interval in a defined population is called the mortality rate. It is typically expressed in population units, such as deaths per 1,000 people for a recall period (or specified time frame). Mortality rate, or death rate, is a measure of the number of deaths (in general, or due to a specific cause) in a particular population, scaled to the size of that population, per unit of time. An important specific mortality rate measure is the crude death rate, which looks at mortality from all causes in a given time interval for a given population. When the crude death rate is subtracted from the crude birth rate (CBR), the result is the rate of natural increase (RNI). This is equal to the rate of population change (excluding migration).

1.1.4 BIRTH AND DEATH REGISTRATION:

Birth registration is the process of recording a child's birth. It is a permanent and official record of a child's existence and provides legal recognition of that child's identity. The death registration record may be required for settlement of inheritance, insurance claims, claiming family allowances, and other social security benefits

1.1.5 SURVIVAL RATE:

Survival rate is the proportion of people in a study or treatment group still alive at a given period of time after diagnosis. It is a method of describing prognosis in certain disease conditions and can be used for the assessment of standards of therapy.

1.2 REVIEW OF LITERATURE

There has been consistent decrease in the Birth Rate, Death Rate and Natural Growth Rate in India since 1991 to 2016. As on 2016 India has registered Birth Rate of 20.4 per 1000 populations and Death Rate of 6.4 per 1000 populations while the Natural Growth Rate was 14.0 per 1000 population in India.

Government of India is a signatory to the United Nations (UN) Sustainable Development Goals (SDGs), which adopted a global maternal mortality ratio (MMR) target of fewer than 70 deaths per 100 000 live births by 2030.

[WHO] This requires the reliable quantification of maternal deaths and trends and an understanding of the major causes of these deaths at the subnational level. India, similar to many countries with high maternal mortality, officially registers only a fraction of births, deaths and vital events.

Jha P, Gomes M, Begum R, Sati P, Dikshit R, Gupta PC, Kumar R,Maternal deaths are concentrated in remote rural areas and are among the least likely to be recorded.

Registrar General India, Centre for Global Health Research India, however, has had a functioning Sample Registration System (SRS) to monitor fertility and mortality covering over 1 million nationally representative homes for more than five decades.

UN estimates that about 24 million children were born in 2017 in India, and about 35 000 mothers died during childbirth or shortly thereafter, giving an MMR of 145 per 100 000 live births.

WHO, This rate represented 12% of global maternal deaths. According to the World Health Organization (WHO), the worldwide MMR has fallen substantially from 342 in the year 2000 to 211 in 2017, reducing global maternal deaths from

451 000 to 295 000 during this period. 5 About 40% of this absolute decline was derived from fewer maternal deaths in India.

Registrar General India, Centre for Global Health Research India, here, we evaluate national and subnational levels and trends of maternal mortality over a 20-year period in India by examining nearly 10 000 maternal deaths among 4.3 million live births identified in the SRS, and over 1500 maternal deaths assigned causes in the Million Death Study (MDS), a nationally representative, population-based mortality survey.

Dhingra N, Jha P, Sharma VP, Cohen AA, Jotkar RM, Rodriguez PS,Menon GR, Singh L, Sharma P, Yadav P, Sharma S, Kalaskar S,We focus particularly on the rates of progress in the poorer states of India, which currently lag behind the national trends. A 2006 report, 4 based on maternal deaths captured in the SRS from 1997 to 2003, suggested that progress at that time was insufficient for India to meet the 2015 UN goals. We reassess that projection and point to how India can achieve the 2030.

We applied the WHO definitions of maternal death, namely the death of a woman while pregnant or within 42 days of terminating a pregnancy, irrespective of the site and duration of the pregnancy, from any cause related to or aggravated by the pregnancy or its management, but not from accidental or incidental causes

WHO, nationally representative data sources to assess MMRs, and estimate maternal deaths and cause distribution. We extracted ten periodic MMR estimates (representing 2- or 3-year periods from 1997 to 2018) for all of India, for major regions and for individual states from the SRS Special Bulletins on Maternal Mortality,Office of the Registrar General & Census Commissioner India, which provide subnational MMR statistics.

1.3 OBJECTIVES OF THE STUDY

Statistical methods involved in carrying out a study include planning , designing, collecting data analysis gives meaning to the meaningless numbers, there by breathing life into a lifeless data. The results and inference are precise only if proper statistical test are used, understanding the quantitative and qualitative with parametric and non parametric test used for data analysis.

- To Compare the relationship between Gender , region with age of death
- To estimate the probability of mortality rate
- To estimate birth rate by region-wise
- To predict & study about the rapid increase in no. of . live birth

CHAPTER II

2. RESEARCH METHODOLOGY:

Research is systematic process of collecting and analysis information in order to increase our understanding of the phenomena about which are concerned or interested. It is a careful research or inquiry into any branch of knowledge, research in these days treated as advancement of knowledge acquire through scientific method. Research is a long process to achieve expected results keeping in view developing in view developing any branch of modern activities.

2.1 SOURCE OF DATA AND STUDY AREA:

The data used in this study is sourced from secondary sources, Provided in the website of www.indiastat.com for the study and the study area is India.

2.2 RESEARCH TOOLS:

The collected data were processed with SPSS software, R software, Excel, Jamovi . The following statistical tools were used in tune with the objective of the study.

2.2.1 CHI-SQUARE TEST

The Chi-Square test of independence is used to determine if there is a significant relationship between two nominal (categorical) variables. The frequency of each category for one nominal variable is compared across the categories of the second nominal variable. The data can be displayed in a contingency table where each row represents a category for one variable and each column represents a category for the other variable. The chi-square test of independence can be used to examine this relationship. Null hypothesis assumes that there is no association between the two variables. On other hand, an alternative hypothesis assumes that there is an association between the two variables. After calculating the expected value, we will apply the following formula to calculate the value of the Chi-Square test of Independence:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i}$$

- Oi = observed frequency
- Ei = Expected frequency

Degrees of freedom is calculated by using the following formula: df = (r-1) ×(c-1);

- r = No. of rows and
- c = No. of columns

FOR CONDUCTING A CHI-SQUARE ANALYSIS

- **Define Research Question:** Clearly formulate a research question that involves examining the association or independence between two or more categorical variables.
- **Select Population and Sample:** Define the population of interest and select a representative sample from this population.
- **Variable Identification:** Clearly identify and categorize the variables relevant to your research question. Ensure that these variables are categorical.
- **Data Collection:** Collect data by survey, observation, or other appropriate methods, ensuring that each observation falls into one of the defined categories.
- **Contingency Table Formation:** Organize the data into a contingency table, with rows and columns representing the categories of the variables. Populate the table with observed frequencies.

- **Hypothesis Formulation:** Formulate the null hypothesis (H0) and alternative hypothesis (H1) based on the expected relationship or independence between the variables.
- **Expected Frequencies Calculation:** Calculate expected frequencies for each cell in the contingency table under the assumption of no association between variables.
- **Chi-Square Statistic Computation:** Compute the chi-square statistic using the observed and expected frequencies in the contingency table.
- **Degrees of Freedom Determination:** Calculate the degrees of freedom based on the number of rows and columns in the contingency table.
- **Critical Value or P-Value Analysis:** Compare the calculated chi-square statistic to the critical value from the chi-square distribution table or obtain the p-value using statistical software.
- **Decision Making:** Make a decision to either reject or fail to reject the null hypothesis based on the comparison. Consider the significance level chosen for the test.
- **Interpretation and Conclusion:** Interpret the results in the context of the research question, stating whether there is evidence of association or independence between the variables.

2.2.2 REGRESSION ESTIMATION WITH CATEGORICAL VARIABLE

Regression analysis is a statistical method used to examine the relationship between one or more independent variables and a dependent variable. When categorical variables are involved, various techniques are applied to integrate them into the regression model. Here's a comprehensive overview of regression estimation with categorical variables:

Introduction to Regression with Categorical Variables:

Regression analysis aims to model the relationship between a dependent variable (response) and one or more independent variables (predictors). Categorical variables, which represent qualitative characteristics, can be included in regression models to enhance their explanatory power.

Handling Categorical Variables:

Dummy Coding:

- Assign numerical codes (usually 0 or 1) to represent the different categories of the variable.
- Create dummy variables for each category, using one category as the reference level.
- Interpret the coefficients of dummy variables as the change in the dependent variable relative to the reference category.

One-Hot Encoding:

- Create binary columns for each category, with 1 indicating the presence of that category and 0 otherwise.
- Interpret coefficients of these binary columns similarly to dummy variables.

Effect Coding:

- Assign one category as -1, and the other categories as 0 or 1.
- Facilitates interpretation of coefficients as deviations from the overall mean.

Reference Level:

- Choose a reference level for categorical variables; it serves as the baseline for comparisons.
- Coefficients for other levels are interpreted as changes relative to the reference level.

Interaction Terms:

• Introduce interaction terms between categorical variables and other predictors to capture joint effects.

• Useful when the effect of one variable depends on the level of another.

Interpretation of Coefficients:

• Coefficients for dummy variables indicate the change in the dependent variable relative to the reference category.

• Interpret coefficients in the context of the chosen coding scheme.

Model Assessment:

• Check assumptions of regression (linearity, independence of errors, homoscedasticity, etc.).

• Evaluate model fit using metrics like R-squared for linear regression or others suitable for the model type.

Example Scenario:

Consider predicting employee performance (dependent variable) based on job roles (categorical variable). Dummy code or one-hot encode job roles, choosing one role as the reference level. Interpret coefficients to understand how performance varies across different roles.

Conclusion:

Regression estimation with categorical variables is a powerful tool for understanding relationships in data. Properly handling categorical variables and interpreting coefficients in the context of the chosen reference level are crucial for deriving meaningful insights from regression models.

REFERENCE LEVEL

- The reference level provides a baseline for comparison.
- Coefficients for other categories are interpreted as changes from this baseline.

Dummy Coding Example:

Consider predicting customer satisfaction (dependent variable) based on product types (categorical variable: A, B, C). Assume product type A is the reference level.

Coefficients:

- β_1 (Product B) represents the change in satisfaction compared to Product A.
- β_2 (Product C) represents the change in satisfaction compared to Product A.

Interpretation guidelines

Intercept:

Represents the expected value of the dependent variable when all predictors are zero (or in the reference category).

Dummy Variable Coefficients:

• **Positive coefficient:** The category is associated with a higher outcome.

• **Negative coefficient**: The category is associated with a lower outcome compared to the reference.

Interaction with Other Predictors:

Assess how the effect of the categorical variable changes in the presence of interaction terms.

Model Assessment:

• Check assumptions like linearity, independence of errors, and homoscedasticity.

• Evaluate goodness of fit using relevant metrics.

Thus Understanding the reference level is critical for meaningful interpretation of regression coefficients when dealing with categorical variables. Careful coding and interpretation enhance the insights gained from regression analysis, enabling informed decision-making in various fields.

2.2.3 ARIMA Model

ARIMA stands for Autoregressive Integrated Moving Average model. It is a popular time series forecasting model that is used to model and forecast data that exhibit patterns of non-stationarity.

$\boldsymbol{yt}$= $\boldsymbol{\varphi 0}$ + $\boldsymbol{\varphi 1 yt}$ -1 + $\boldsymbol{\varphi 2 yt}$ -2 +… + $\boldsymbol{\varphi p yt}$ -p + $\boldsymbol{\varepsilon t}$ - $\boldsymbol{\theta 1 \varepsilon t}$ -1- $\boldsymbol{\theta 2 \varepsilon t}$ -2-…- $\boldsymbol{\theta q \varepsilon t}$ -1 Where:

εt - represents the past error value,

yt - is the actual value,

ϕ - & Θ are the coefficients and

p and q are referred to as the autoregressive and moving average.

The ARIMA model combines three components:

i**. Autoregression (AR**): In this component, the values of a time series at a given time are modelled as a linear combination of its past values. The "order" of the autoregression component is denoted by p, and it represents the number of lags of the time series that are included in the model.

$\boldsymbol{yt}$ = $\boldsymbol{\varphi 1 yt}$ -1 + … + $\boldsymbol{\varphi p yt}$ -p + $\boldsymbol{et}$ (OR) $\boldsymbol{yt}$ = Θ + $\boldsymbol{\varphi 1 yt}$ -1 + … + $\boldsymbol{\varphi p yt}$ -p + $\boldsymbol{et}$

Where:

yt - is the current value,

$yt{-}1$ - is the lagged value of variable y,

$\boldsymbol{e}$ - defines the error term,

ϕ - is constant or drift and

p - determines the number of period lag.

ii. Integration (I): This component is used to transform a non-stationary time series into a stationary one by differencing the time series. The "order" of the integration component

is denoted by d, and it represents the number of times the differencing operation is performed.

iii. Moving Average (MA): In this component, the errors of the time series are modelled

as a linear combination of past error terms. The "order" of the moving average component is denoted by q, and it represents the number of lagged error terms that are included in the model.

$\boldsymbol{yt}$ = $\boldsymbol{et}$ + $\boldsymbol{\theta 1et}$ -1 + … + $\boldsymbol{\theta qet}$ -q (OR) $\boldsymbol{yt}$ = α + $\boldsymbol{et}$ + $\boldsymbol{\theta 1et}$ -1 + … + $\boldsymbol{\varphi pyt}$ -q Where:

yt - is the current value,

e - is the residual term,

q - is the number of moving average and is the constant term.

The ARIMA model is specified by three parameters: (p, d, q). The parameters p and q are

usually determined by looking at the autocorrelation and partial autocorrelation plots of the time series. The parameter d is determined by checking if differencing is needed to make the time series stationary.

ARIMA models can be used to forecast time series data for a specified period into the future. The accuracy of the forecast depends on how well the model captures the underlying patterns in the data.

CHAPTER III

3 .ANALYSIS AND INTERPRETATIONS:

3.1 CHI – SQUARE TEST:

The Chi-Square test of independence is used to determine if there is a significant relationship between two nominal (categorical) variables.

3.1.1. REGION * AGE OF DEATH

AIM: To test wheather there is relationship between region and age of death

HYPOTHESIS:

H0 : there is no relationship between region and age of death

H1: there is a relationship between region and age of death

CHI- SQUARE TABLE:

year	Chi - value	P – value		Year	Chi – value	P – value
2000	4.31	0.505		**2010**	3.48	0.627
2001	4.91	0.427		**2011**	3.29	0.655
2002	4.04	0.543		**2012**	2.86	0.721
2003	3.80	0.578		**2013**	2.64	0.756
2004	3.46	0.629		**2014**	2.45	0.784
2006	4.58	0.469		**2015**	2.33	0.802
2007	3.92	0.561		**2016**	2.17	0.825
2008	3.71	0.592		**2017**	2.13	0.830
2009	3.27	0.658		**N=200,df=5**		

INFERENCES:

The significant value is greater(**p -value**) than the alpha value **0.05**.. So,there is a relationship between region and age of death of person. Year by year , the p-value for the test is constantly increasing. This indicates that the relationship between the region and age of death is strengthening strongly.

3.1.2. GENDER * AGE OF DEATH

AIM: To test wheather there is relationship between gender and age of death

HYPOTHESIS:

H0 : there is no relationship between region and age of death

H1: there is a relationship between region and age of death

year	Chi - value	P – value		Year	Chi – value	P – value
2000	1.37	0.928		**2010**	1.63	0.897
2001	1.46	0.917		**2011**	2.46	0.782
2002	1.46	0.918		**2012**	2.77	0.735
2003	2.03	0.846		**2013**	2.05	0.842
2004	1.28	0.937		**2014**	1.01	0.962
2006	1.92	0.860		**2015**	0.82	0.976
2007	2.29	0.808		**2016**	1.42	0.922
2008	2.27	0.811		**2017**	1.52	0.911
2009	2.43	0.787		**N=200,df=5**		

INFERENCES:

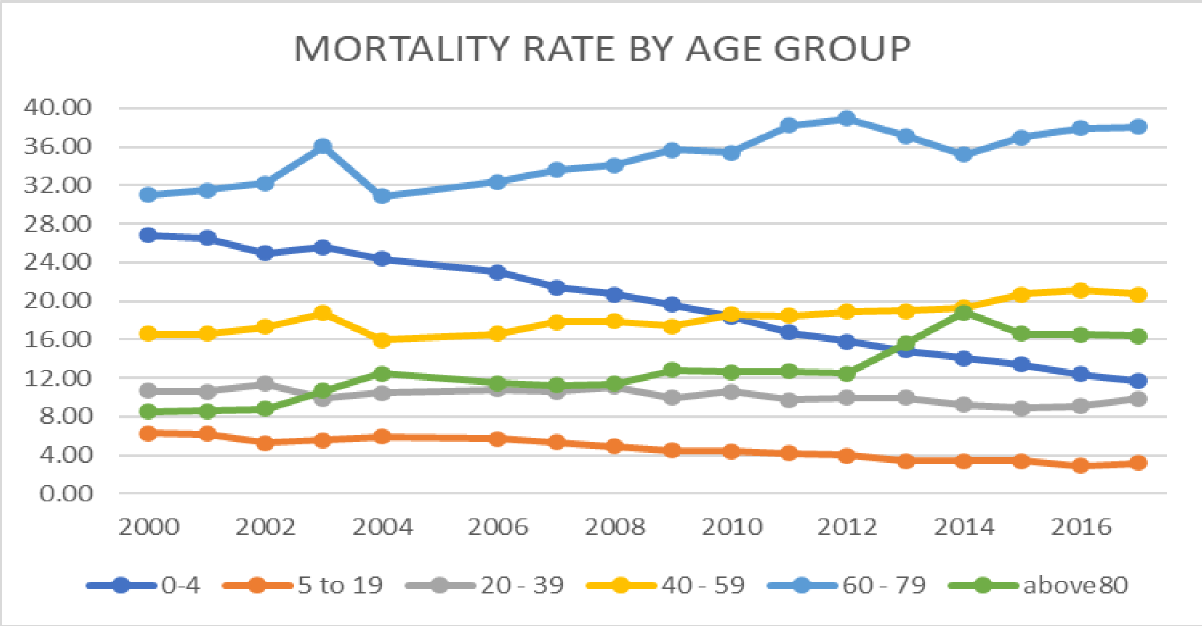
The significant value is greater(**p -value**) than the alpha value **0.05**.. So, There is a relationship between gender and age of death of person. Year by year , the p-value for the test is non-constantly increasing. This indicates that the relationship between the gender and age of death is strengthening strongly.

3.2 REGRESSION ESTIMATION WITH CATEGORICAL VARIABLE

Regression analysis aims to model the relationship between a dependent variable (response) and one or more independent variables (predictors). Categorical variables, which represent qualitative characteristics, can be included in regression models to enhance their explanatory power.

3.2.1 REGRESSION ESTIMATION FOR MORTALITY RATE

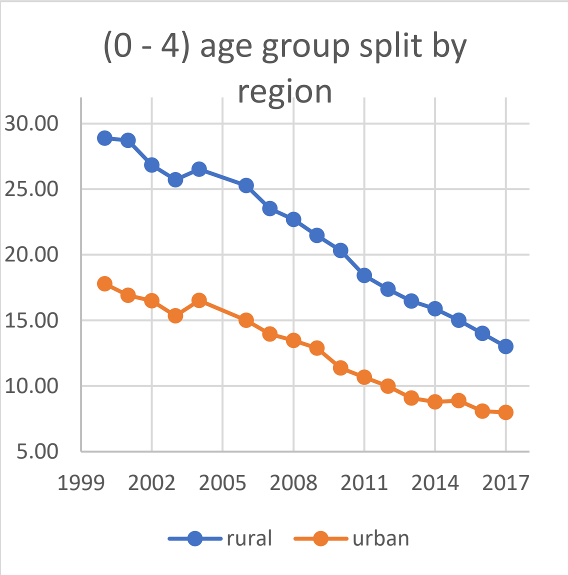
In order to estimate the future mortality rate , we need to split the mortality rate by age group.

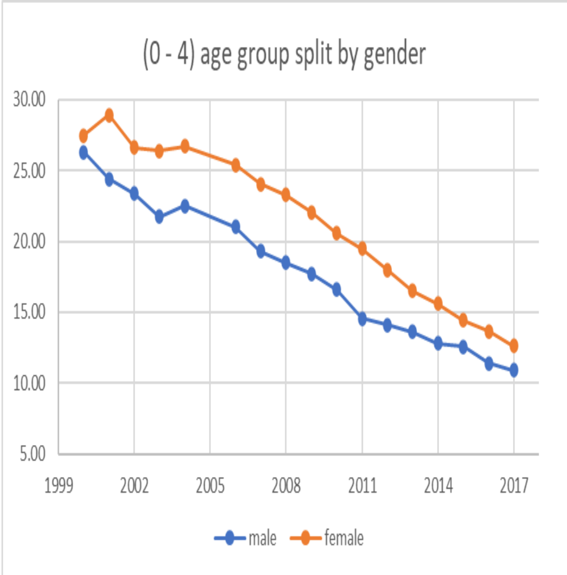


YEAR	0-4	5-19	20-39	40-59	60-79	80&ABOVE
2000	26.83	6.31	10.7	16.62	31	8.51
2017	11.7	3.2	9.9	20.7	38	16.4
CHANGING(%)	-56.39	-49.29	-7.48	+24.55	+22.58	+92.72

To project future-year mortality rates across different age groups, it is essential to estimate each age group separately. Finally, the cumulative sum of the age groups should be converted to a total of 100

MORTALITY RATE FOR (0 – 4) AGE GROUP





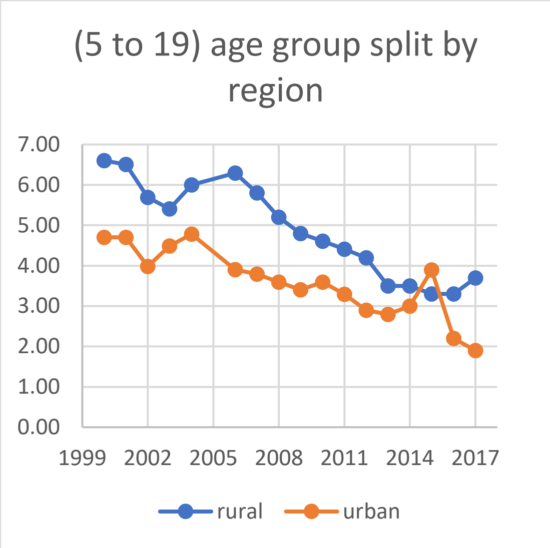
With help of above diagrams , we can conclude that urban peoples have less mortality rate in the basis of region & male have a less mortality rate in the basis of gender in (0 – 4) age group mortality rate

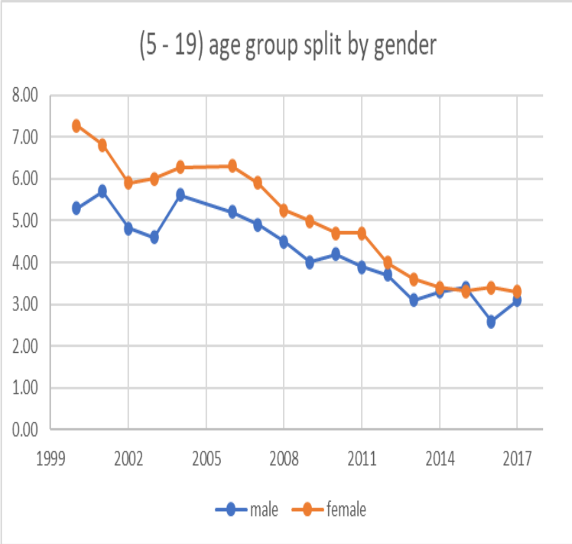
In order to estimate , we use regression estimation with categorical variable .In gender we taking male as reference level & in region we taking urban as reference level

Model Fit Measures		
Model	**R**	**R^2**
1	0.879	0.773

Model Co-efficients (0 – 4) age group				
Predictor	**Estimate**	**SE**	**t**	**P**
Intercept[a]	1464.892	146.2672	10.02	<0.001
Year	-0.724	0.0728	-9.94	<0.001
Gender Female-male	2.656	0.7670	3.46	<0.001
Region rural-urban	7.956	0.7670	10.37	<0.001

MORTALITY RATE FOR (5 – 19) AGE GROUP





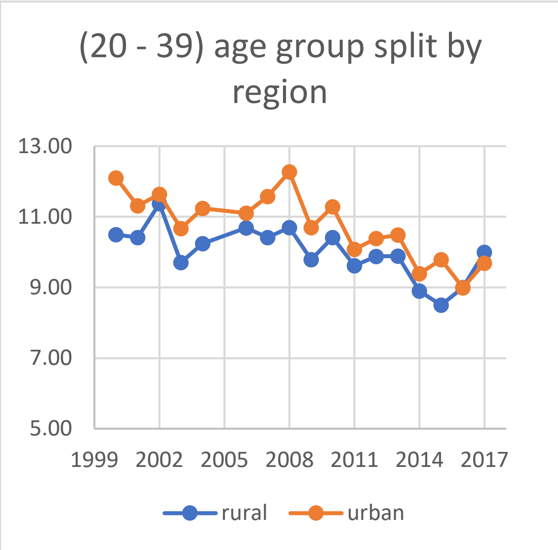
With help of above diagrams , we can conclude that urban peoples have less mortality rate in the basis of region & male have a less mortality rate in the basis of gender in (5 - 19) age group mortality rate

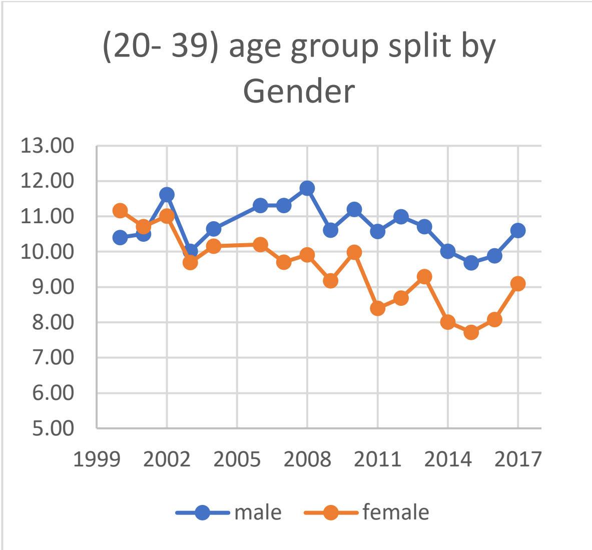
In order to estimate , we use regression estimation with categorical variable .In gender we taking male as reference level & in region we taking urban as reference level

Model Fit Measures		
Model	**R**	**R^2**
1	0.907	0.823

Model Co-efficients (5 - 19) age group				
Predictor	**Estimate**	**SE**	**t**	**P**
Intercept[a]	366.454	27.2665	13.44	<0.001
Year	-0.181	0.0136	-13.32	<0.001
Gender Female-male	0.815	0.1430	5.70	<0.001
Region rural-urban	1.338	0.1430	9.36	<0.001

MORTALITY RATE FOR (20-39) AGE GROUP





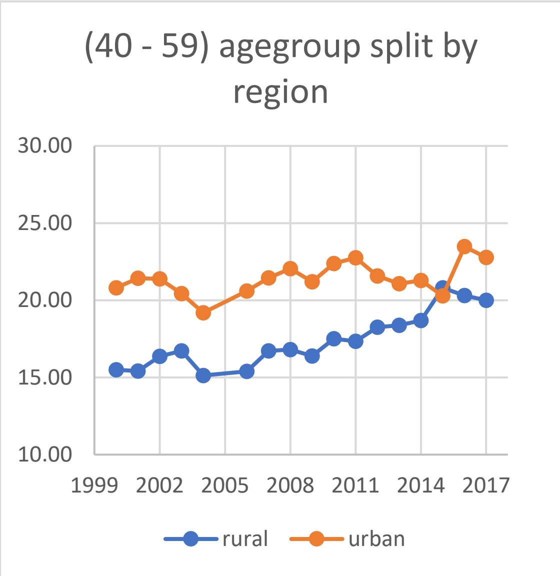
With help of above diagrams , we can conclude that rural peoples have less mortality rate in the basis of region & female have a less mortality rate in the basis of gender in (20 - 39) age group mortality rate

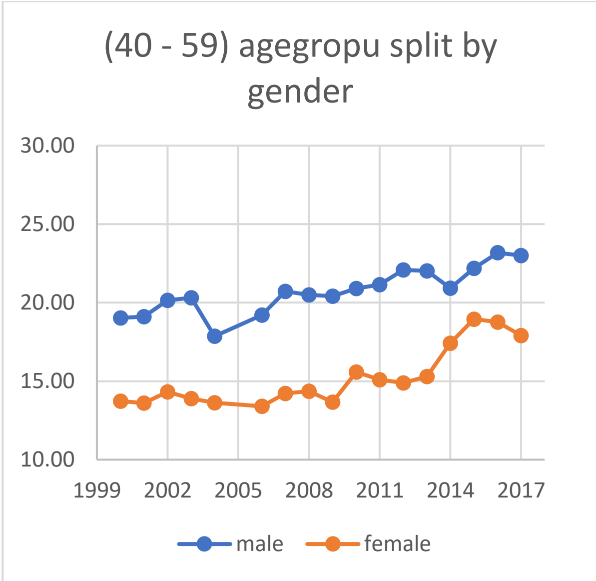
In order to estimate , we use regression estimation with categorical variable .In gender we taking female as reference level & in region we taking rural as reference level.

Model Fit Measures		
Model	**R**	**R^2**
1	0.773	0.597

Model Co-efficients (20 - 39) age group				
Predictor	**Estimate**	**SE**	**t**	**P**
Intercept[a]	265.761	44.6088	5.96	<0.001
Year	-0.128	0.0222	-5.75	<0.001
Gender Male-female	1.762	0.2339	7.53	<0.001
Region Urban-rural	0.519	0.2339	2.22	0.030

MORTALITY RATE FOR (40 - 59) AGE GROUP





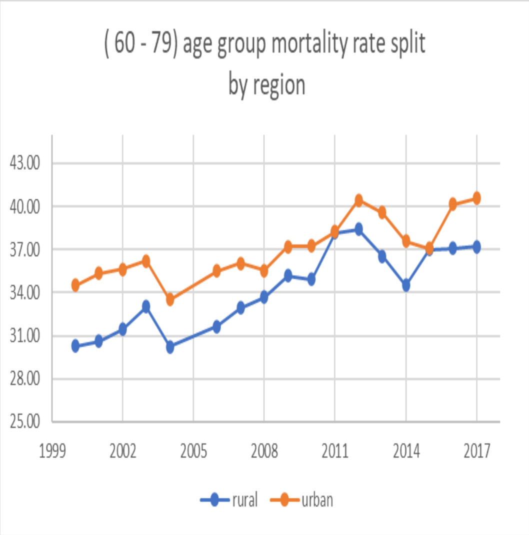
With help of above diagrams , we can conclude that rural peoples have less mortality rate in the basis of region & female have a less mortality rate in the basis of gender in (40 - 59) age group mortality rate

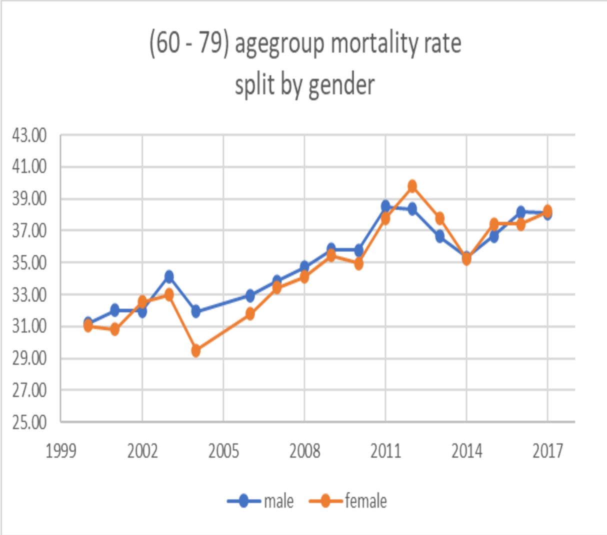
In order to estimate , we use regression estimation with categorical variable .In gender we taking female as reference level & in region we taking rural as reference level

Model Fit Measures		
Model	**R**	**R^2**
1	0.923	0.851

Model Co-efficients (40 - 59) age group				
Predictor	**Estimate**	**SE**	**t**	**P**
Intercept[a]	-344.433	73.5678	-4.68	<0.001
Year	0.179	0.0366	4.88	<0.001
Gender Male-female	6.143	0.3858	15.93	<0.001
Region Urban-rural	3.649	0.3858	9.46	<0.001

MORTALITY RATE FOR (60 - 79) AGE GROUP





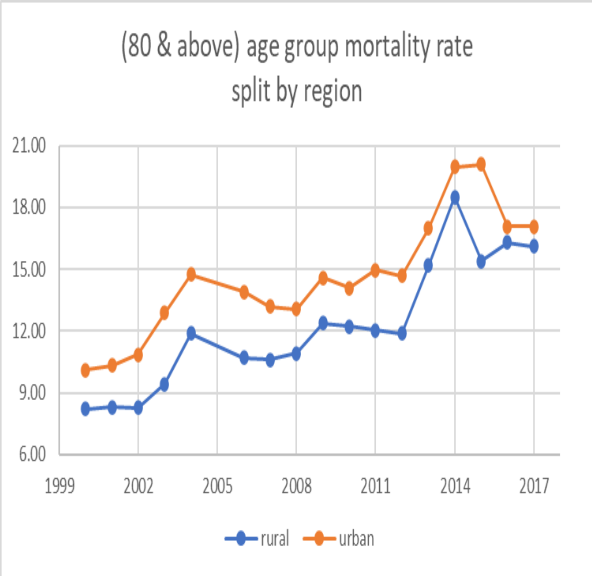
With help of above diagrams , we can conclude that rural peoples have less mortality rate in the basis of region & male have a less mortality rate in the basis of gender in (60 - 79) age group mortality rate

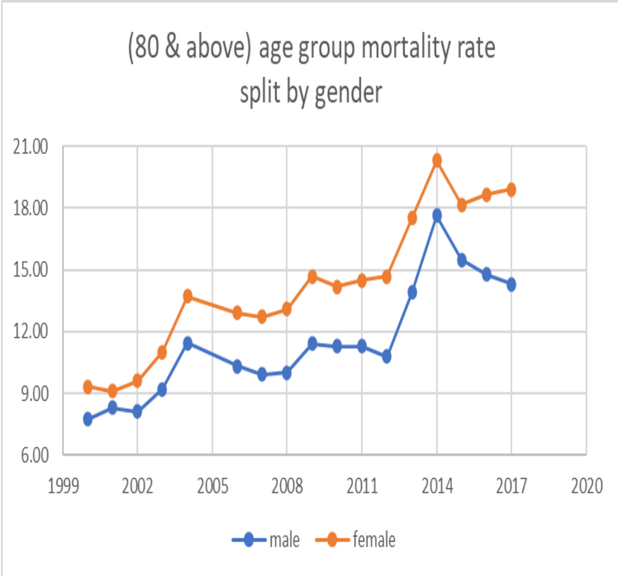
In order to estimate , we use regression estimation with categorical variable .In gender we taking male as reference level & in region we taking rural as reference level

Model Fit Measures		
Model	**R**	**R^2**
1	0.733	0.538

Model Co-efficients (60 - 79) age group				
Predictor	**Estimate**	**SE**	**t**	**P**
Intercept[a]	-682.510	102.5207	-6.66	<0.001
Year	0.357	0.0510	6.99	<0.001
Gender Female-Male	2.643	0.5376	4.92	<0.001
Region Urban-rural	0.626	0.5376	1.16	0.249

MORTALITY RATE FOR (80 & ABOVE) AGE GROUP





With help of above diagrams , we can conclude that rural peoples have less mortality rate in the basis of region & male have a less mortality rate in the basis of gender in (80 & above) age group mortality rate

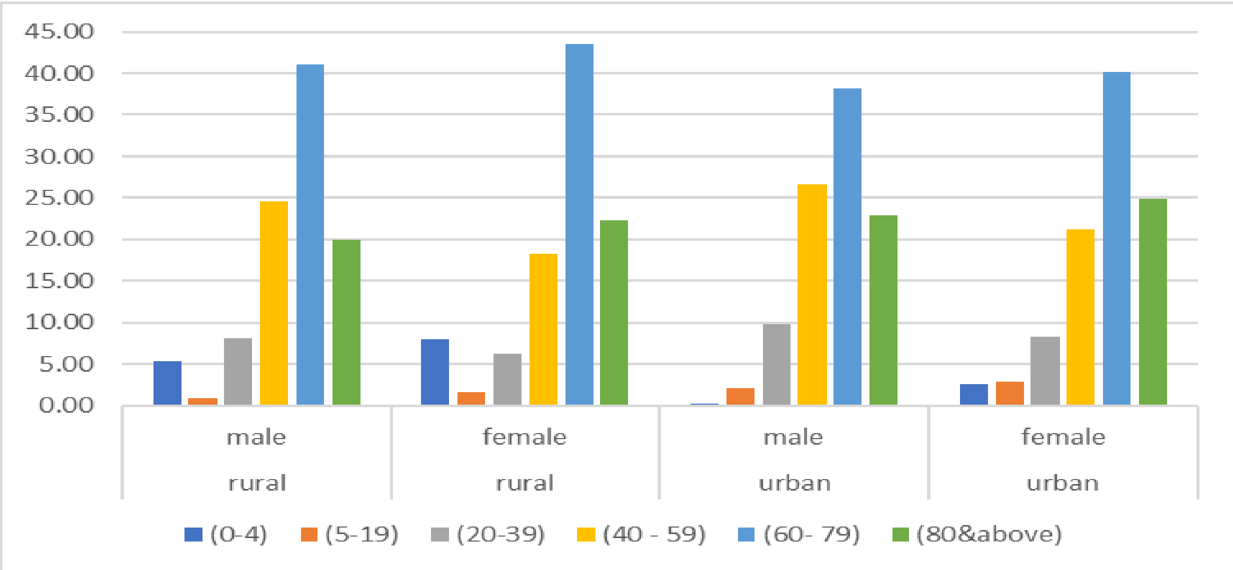
In order to estimate , we use regression estimation with categorical variable .In gender we taking male as reference level & in region we taking rural as reference level

Model Fit Measures		
Model	**R**	**R^2**
1	0.903	0.815

Model Co-efficients (80&above) age group				
Predictor	**Estimate**	**SE**	**t**	**P**
Intercept[a]	-987.396	78.7707	-12.54	<0.001
Year	0.497	0.0392	12.67	<0.001
Gender Female-Male	2.482	0.4131	6.01	<0.001
Region Urban-rural	3.808	0.4131	9.22	<0.001

With the help of above estimated equations , we have predicted the mortality rate for 2027 which is also split by both gender and region

region	gender	(0-4)	(5-19)	(20-39)	(40 -59)	(60-79)	(80& above)
rural	male	5.30	0.91	8.07	24.55	41.14	20.03
rural	female	7.90	1.71	6.26	18.28	43.49	22.36
rural	person	6.61	1.31	7.16	21.40	42.32	21.20
urban	male	0.29	2.19	9.88	26.60	38.16	22.88
urban	female	2.54	2.87	8.33	21.23	40.18	24.85
urban	person	1.42	2.53	9.10	23.91	39.17	23.87
total	male	2.60	1.60	9.05	25.66	39.53	21.57
total	female	5.01	2.33	7.38	19.87	41.70	23.70
total	person	3.81	1.97	8.21	22.76	40.62	22.64



In 2027 ,

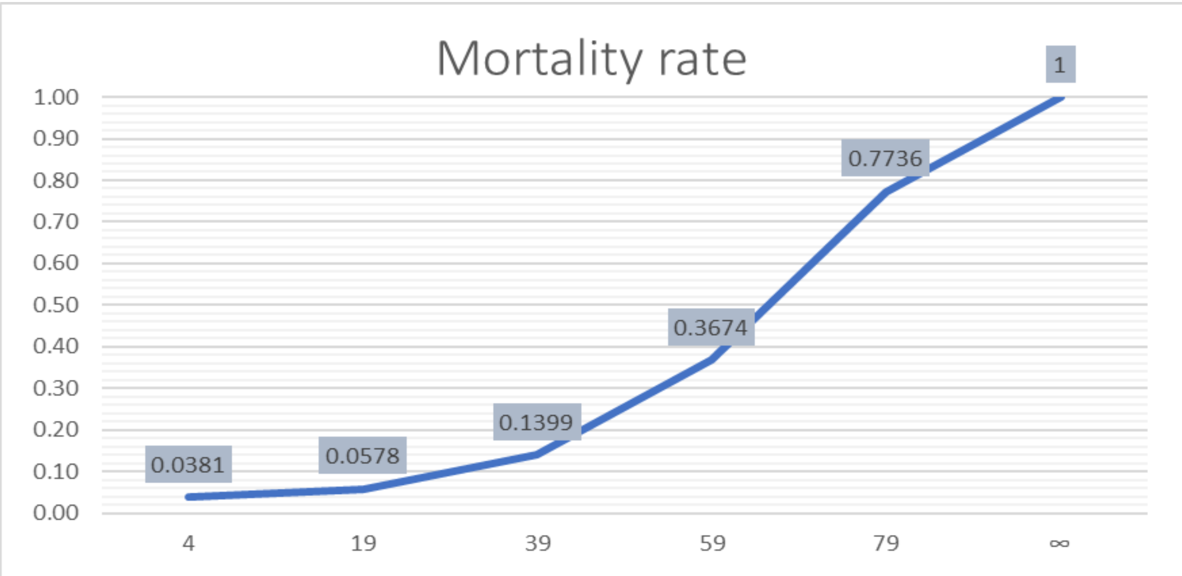
Urban -male will have less level of morality rate while compare to ruban – female in the age group of (0 -4). Rural – male will have less level of mortality rate while compare to urban – female in the age group of (5 – 19). Rural – female will have less level of mortality rate while compare to urban – male in the age group of (20 – 39) .

rural – female will have less level of mortality rate while compare to urban – male in the age group of (40 – 59). Urban – male will have less level of mortality rate while compare to rural -female in the age group of (60 – 79) . rural – male will have less level of mortality rate while compare to urban – female in (80&above) age group.

PROBABILTY FOR MORTALITY RATE IN 2027

With the help of above predicted mortality rate , we can able to constract the probability of mortality rate in age specific by using cummulative relative frequency method .

Survival rate + Mortality rate =1 , with the help of this methamatical equcation .

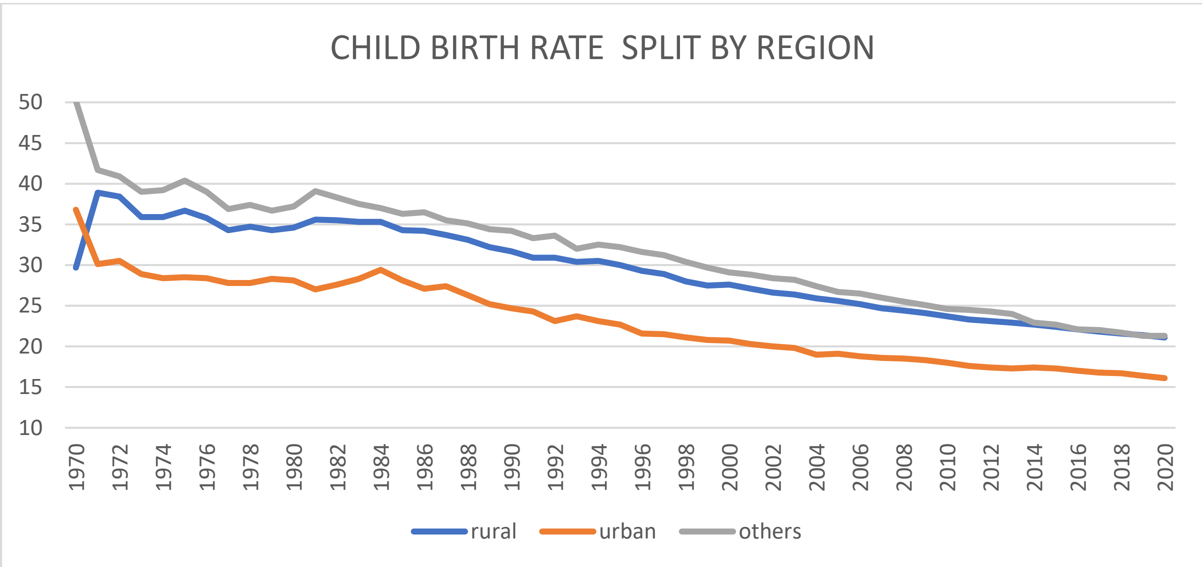


INFERENCE:

In 2027, there will be a Probability for a person will die under the age of 4 is 0.0381, will die under the age of 19 is 0.0578, will die under the age of 39 is 0.1399, will die under the age of 59 is 0.3674, will die under the age of 79 is 0.7736.

3.2.2 REGRESSION ESTIMATION FOR CHILD BIRTH RATE

In order to estimate the future child birth rate , we need to split the child birth rate by region .



With help of above diagrams , we can conclude that rural have less child birth rate in the basis of region

In order to estimate , we use regression estimation with categorical variable .In region we taking rural as reference level

Model Fit Measures		
Model	**R**	**R^2**
1	0.974	0.949

Model Co-efficients for birth rate				
Predictor	**Estimate**	**SE**	**t**	**P**
Intercept[a]	761.062	16.84172	45.2	<0.001
Year	-0.370	0.00844	-43.8	<0.001
Region (rural)	6.324	0.30435	20.8	<0.001
Region (others)	8.518	0.30435	28.0	<0.001

Every year the birth rate is reduced in the percentage of 0.37 of india`s population.

For 2027

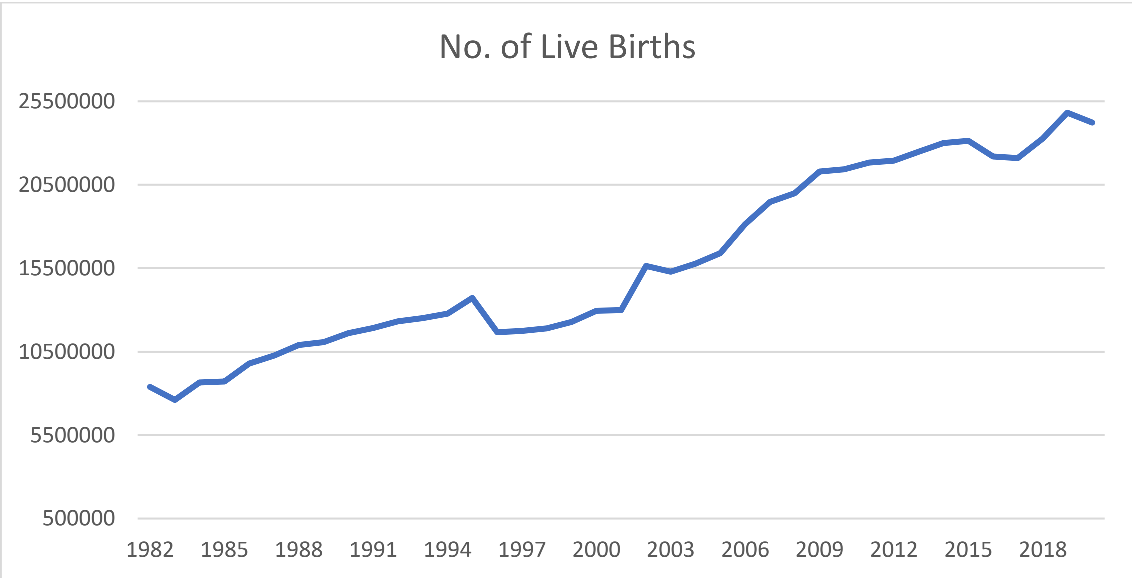
Region	Urban	Rural	Others	Overall
Birth rate(%)*	11.07	17.40	19.59	16.02

INFERENCE:

- In 2027, the overall birth rate will be 16.02%
- In 2027, the birth rate in urban will be 11.07% , rural will be 17.40% , and in other region will be 19.59%

3.3 ARIMA MODEL FOR NUMBER OF LIVE BIRTH

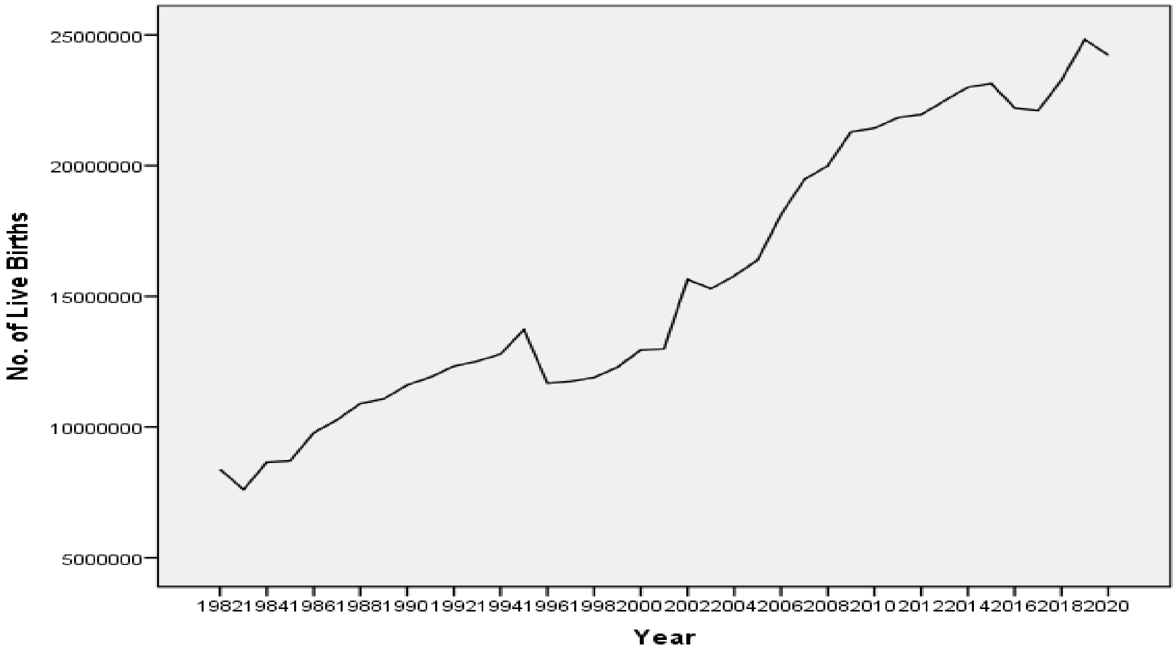
In this chapter, the birth registered in India is examined using a time series projection. The use of Box and Jenkins' (ARIMA models) methodology in this study's examination of the data from this yearly series gave it significant value. In more detail, we will look at the time series, find the appropriate ARIMA model for these data, and estimate the model's parameters. The residuals will then be checked for autocorrelation as part of the diagnostic process. Finally, we will employ the proper ARIMA model for forecasting, relying on it to guide future planning by decision-makers



This figure shows that there is a increase in the live birth in India between the years 1982 to2020

There is a increase of 189.04% in the no of live birth by 2020 when compare to 1982

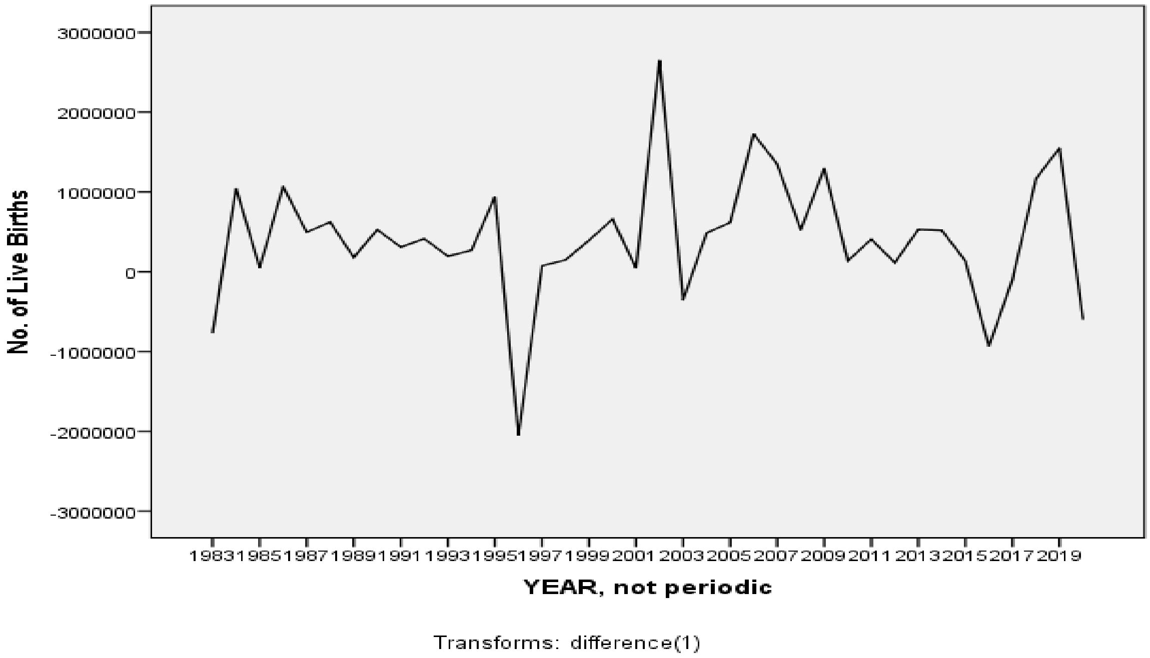
Checking stationarity , partial auto correlation , auto correlation for no of live birth data set .

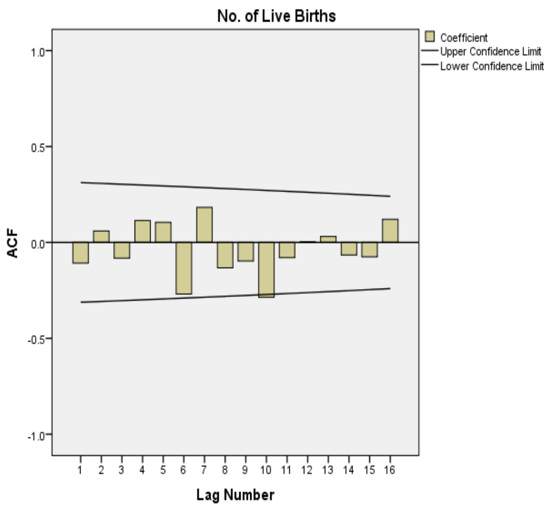


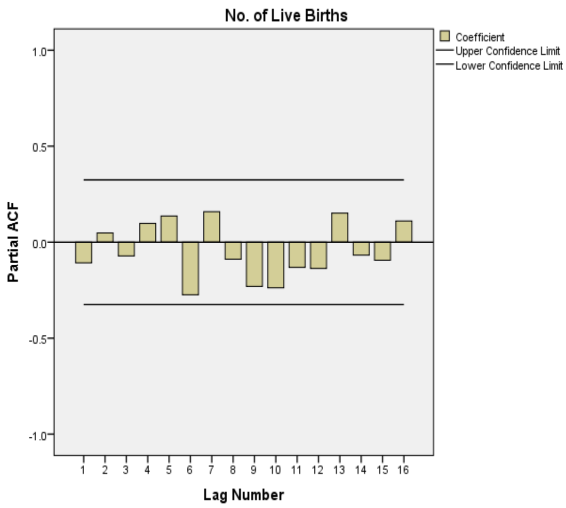
No of live birth data set is in the form of non – stationary data, ARIMA is calculated under stationary data set , in order to convert the data into stationary , we using differencing level.

Differencing level :1

With the help of differencing , we can calucualte the p value(partial auto correlation value)&q value(auto correlation value)







In differencing level :1 , we got our data into stationary set .no of outliers in **ACF** is zero ,so we taking **q** value as 0 . no of outliers in **PACF** is zero , so we taking **p** value as 0.

So our **ARIMA MODEL IS (0,1,0)**

MODEL DESCRIPTION

	Model type
Model ID : no of live births Model-1	**ARIMA(0,1,0)**

MODEL SUMMARY

Fit statistic	VALUE
R-SQUARED	0.977
RMSE	798515.751
MAPE	3.723
MAE	540925.593
MaxAE	2471744.759
Normalized BIC	27.277

MODEL STATISTICS

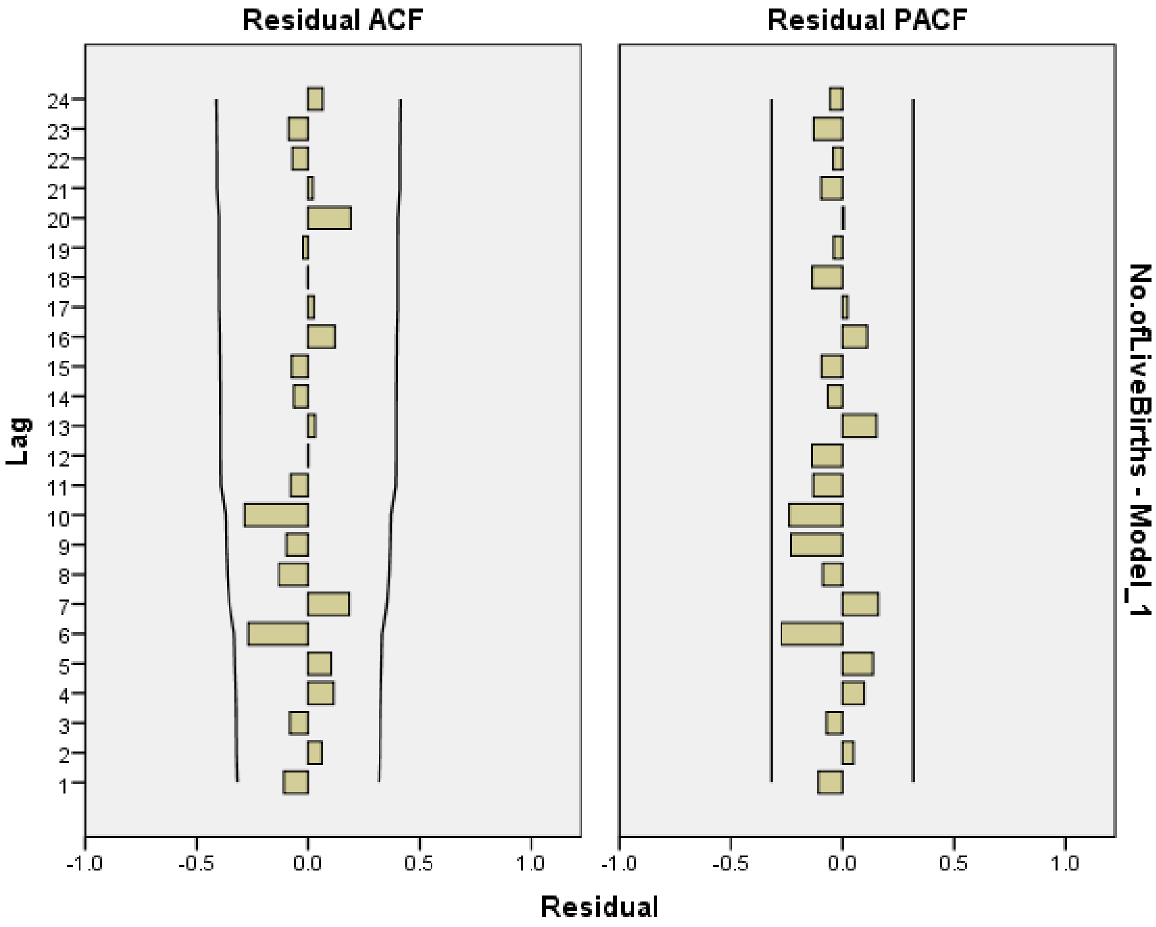
MODEL	**VALUE**
R^2	0.977
MAPE	3.723
MAE	540925.593
BIC	27.277

LJUNG-BOX

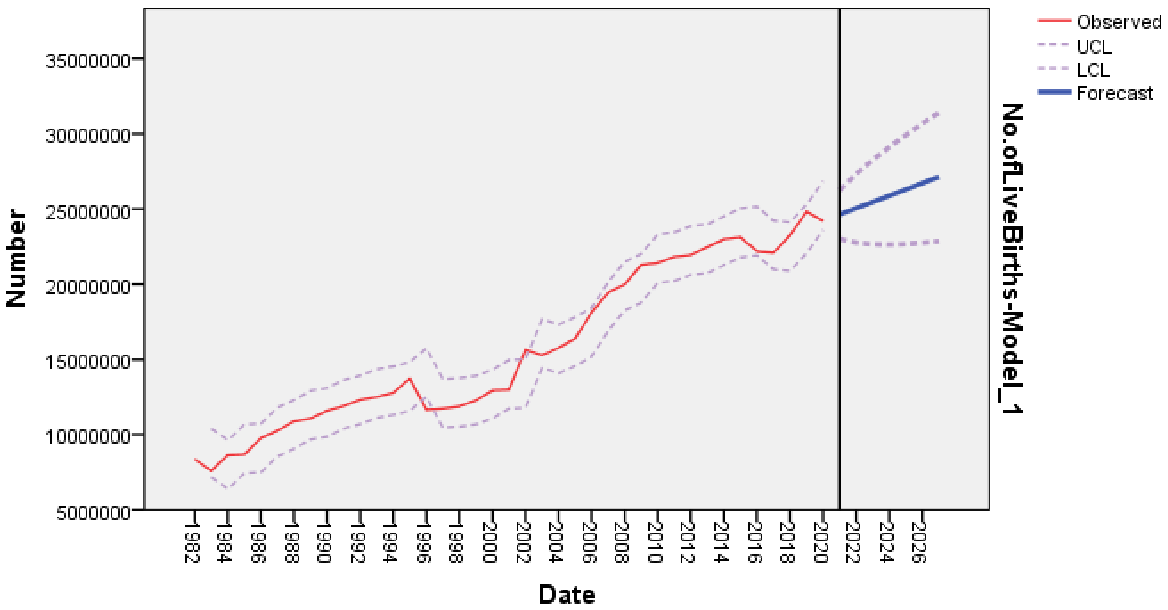
Statistics	15.018
DF	18
Sig.	0.461
Number of outliers	0

FORCAST

Model		2021	2022	2023	2024	2025	2026	2027
No. of Live Births-Model_1	Forecast	24639339	25056234	25473128	25890023	26306918	26723813	27140707
	UCL	26257285	27344356	28275494	29125916	29924756	30686956	31421392
	LCL	23021392	22768111	22670763	22654130	22689079	22760669	22860023



Residual ACF
Residual PACF
Lag
No.ofLiveBirths - Model_1
Residual



Observed
UCL
LCL
Forecast
Number
Date
No.ofLiveBirths-Model_1

INTERPRETATION:

The data are analysed, and an ARIMA model is constructed with the intention of transforming non-stationary data into stationary data. ACF and PACF are used to determine the data's stationarity and to select the optimal model by minimising BIC (Bayesian Information Criterion). The forecasting of the future value is finally depicted . We draw the conclusion that No of live birth would rise, with future estimates estimated to range from 2021 to 2027.

CHAPTER IV

4. CONCLUSION:

The relationship between region and age of death of a person is evident in various studies and demographics. It has been observed that certain regions have higher average ages of death compared to others. Similarly, the gender of a person also plays a role in determining the age of death. Statistics and research suggest that there are differences in life expectancy between males and females.

Looking specifically at the year 2027, the probability of a person dying under the age of 59 has been estimated to be 0.3674. This statistics indicates the risk associated with premature deaths before reaching this age threshold.

Furthermore, when examining the birth rates in different areas, it is projected that urban areas will experience a birth rate of 11.07% in 2027. In rural areas, the birth rate is anticipated to be 17.04% while other regions will have a birth rate of 19.59%. Combining these figures, the overall birth rate for 2027 is projected to be 16.02%

Interestingly, there is an anticipated increase of 12.05% in the number of live births by 2027 when compared to 2020. This growth signifies a potential rise in population in the coming years, indicating various factors such as improvements in healthcare, changes in socio-economic conditions, and advancements in maternal and child health.

In summary, analysing the relationship between factors such as region, gender ,age of death and birth rates can provide valuable insights into population dynamics and trends. By examining these statistics, researchers and policymakers can better understand the demographic changes that may occur in the future.

CHAPTER V

5.REFERENCES:

Dhingra N, Jha P, Sharma VP, Cohen AA, Jotkar RM, Rodriguez PS, et al(2010). Adult and child malaria mortality in India: a nationally representative mortality survey. Lancet;376:1768–74. [PMC free article] [PubMed] [Google Scholar]

Gokhale MK, Rao SS, Garole VR.(2002) Infant mortality in India: Use of maternal and child health services in relation to literacy status. J Health Popul Nutr;20:138–47. [PubMed] [Google Scholar]

Gomes M, Begum R, Sati P, Dikshit R, Gupta PC, Kumar R, et al.(2017). Nationwide mortality studies to quantify causes of death: relevant lessons from India's million death study. Health Aff;36:1887–95.

Graff M, Yount KM, Ramakrishnan U, Martorell R, Stein AD. Childhood nutrition and later fertility: Pathways through education and pre-pregnant nutritional status. Demography. 2010;47:125–44. [PMC free article] [PubMed]

IIPS(2000) . National Family Health Survey (NFHS-2), 1998-99. Mumbai, India: International Institute for Population Sciences (IIPS) and ORC Macro; 2000. [Google Scholar]

Jha P. (2014) Reliable direct measurement of causes of death in low- and middle-income countries. BMC Med;12:1–10. [PMC free article] [PubMed]

Kravdal Ø. Education and fertility in sub-Saharan Africa: Individual and community effects. Demography. 2002;39:233–50. [PubMed] [Google Scholar]

WHO(2012) . The WHO Application of ICD-10 to Deaths during Pregnancy, Childbirth and the Puerperium: ICD-MM. Geneva, Switzerland: WHO.