Course Project for Mathematics for Engineers-II

Regression and Hypothetical Testing

A project submitted
in partial fulfilment of the requirements for the degree
of
Bachelor of Technology
in
Computer Science

SUBMITTED BY:

Yashvardhan Singh Latwal (230C2030409)
Priyanshu Kumar (230C2030358)

BML MUNJAL UNIVERSITY™

SCHOOL OF ENGINEERING AND TECHNOLOGY
BML MUNJAL UNIVERSITY, GURUGRAM

Contents

Problem Statement

Regression analysis and hypothesis testing will be used to examine the effects of the population growth rate on the Indian economy during the previous 20 years.

Expected outcome: Finding a relationship between the population growth rate and different economic indicators like GDP growth. This study will help with finding whether there is a significant correlation between changes in population growth and the GDP.

Data source:

Population Growth Rate: Official government records, such as the India Census, which is carried out every ten years, are a good source of information. Furthermore, yearly or periodic estimates can be obtained from demographic surveys and reports issued by agencies such as Macro Trends and Indian census.

Economic Indicators: The Reserve Bank of India (RBI), the Ministry of Statistics and Programme Implementation (MOSPI), the World Bank, the International Monetary Fund (IMF), and other respectable economic research organizations are good places to find information on a variety of economic indicators, including the GDP growth rate. Cumulated GDP rate was found from Macrotrends.

Time Series Data: To enable regression analysis and hypothesis testing, time-series data covering the last 20 years must be gathered for the population growth rate and economic indicators. For this, it might be necessary to gather information from several sources and make sure the time series is accurate and consistent.

Through a combination of multiple sources of data and the use of statistical techniques such as regression analysis and hypothesis testing, a thorough examination of the impact of the population growth rate on the Indian economy can be performed.

Introduction

A key factor that impacts economic development is population growth, which has a significant impact on a number of social and economic factors including GDP growth. Understanding the link between population growth and economic performance is necessary to successful decision-making and economic growth methods in the context of India, a nation set for important changes in its population. In order to shed light on the complex interactions that have developed over the past 20 years between the population growth rate and the Indian economy, regression analysis and hypothesis testing will be used in this study.

Literature Survey: In order to calculate the relationship between population growth rate and economic variables, control for confounding variables, and determine the statistical significance of observed associations, regression analysis and hypothesis testing are essential. Researchers are able to determine patterns, trends, and causal pathways that clarify the influence of population dynamics on economic performance through the estimation of regression coefficients, hypothesis testing, and the computation of measures of fit and predictive power. But it's important to recognize the methodological difficulties and constraints that come with regression analysis, like endogeneity, multicollinearity, and errors in model specification, and to use the right strategies to address these problems.

Population Growth and Economic Development: Over time, different theories have been proposed to explain the connection between population growth and economic development. Numerous frameworks have been put forth to explain the intricate relationships between population dynamics and economic performance, ranging from Thomas Malthus' pessimistic view that population would exceed food production to contemporary theories like the demographic transition theory and endogenous growth theory. The results of empirical research examining this relationship have been conflicting; while some have pointed to possible negative effects like resource depletion and environmental degradation, others have suggested a positive correlation between population growth and economic growth.

Materials and Method

Describe the statistical tools, mathematical concept used and assumptions made, parameters to be inferred, a glimpse of the data set.

Statistical Tools:
We will be using R studio to code the following problem.
Regression Analysis: By estimating the coefficients of a regression equation, this technique enables us to model the relationship between a dependent factor (such as economic indicators) and one or more independent variables (such as population growth rate). Depending on the type of data and research questions, we will use a variety of regression models, including multiple linear regression, simple linear regression, and maybe time-series data regression.
Testing Hypotheses: The statistical significance of the correlation between the population growth rate and economic indicators will be evaluated through hypothesis testing. To draw conclusions about the population parameters, this entails developing null and alternative hypotheses, choosing a suitable test statistic (such as the t-test or F-test), and calculating the p-value.

Mathematical Concepts:
Time series regression: The relationship between the independent and dependent variables varies over time, and the data in this case is dependent on time. The formula for time series regression takes into consideration the time series' structure and might incorporate certain variables linked to time.

Simple linear regression: When there is a linear relationship between the independent and dependent variables, the situation is known as simple linear regression. $Y = a + b X$ is the formula, where X is independent, Y is dependent, b represents the slope, and a represents the intercept.

Hypothesis testing: Creating a null hypothesis (H0) and an alternative hypothesis (H1) is the first step in the hypothesis testing process. Usually, the alternative hypothesis contends that there is a significant relationship between the variables, whereas the null hypothesis claims there isn't. In order to draw conclusions about the population parameters, we employ statistical tests to assess the probability of observing the sample data in the event that the null hypothesis was true.

Dataset:

Data on Population Growth Rate: For the previous 20 years, India's population growth rate has been estimated either annually or on a periodic basis. It could come from official government documents like the Indian Census or demographic studies (Macro trends).

Data on Economic Indicators: During the same time period, this dataset contains corresponding economic indicators such as GDP growth rate. Sources of data include the Reserve Bank of India (RBI), the World Bank.

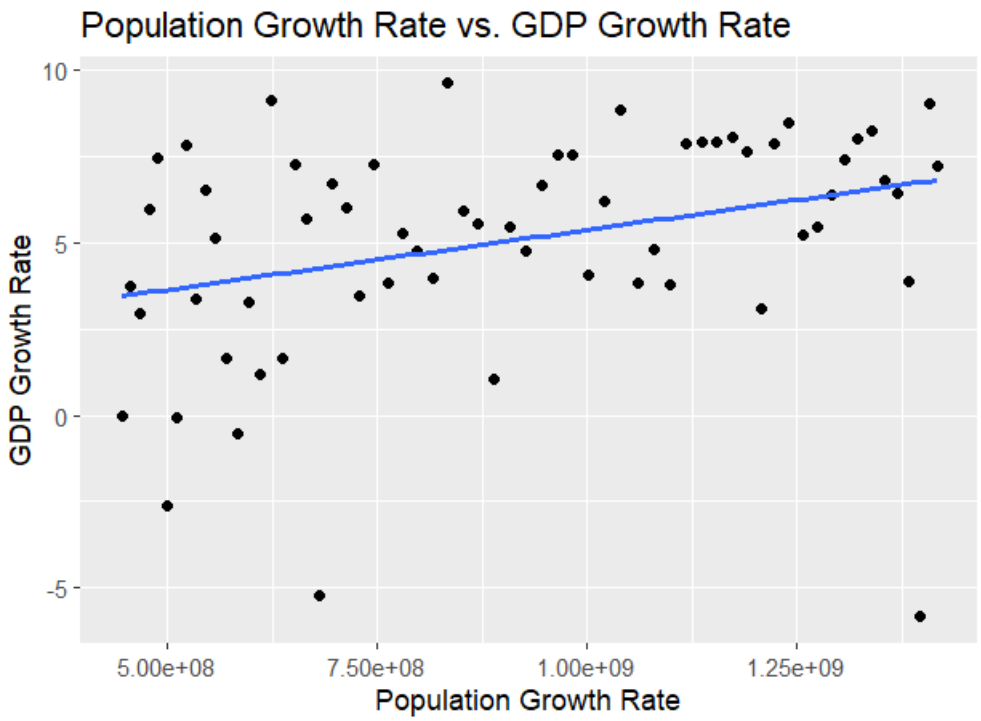
To facilitate long-term examination of population-economy factors, the dataset will be arranged in a time-series design, where every observation will correspond to a different year.

DATASET:

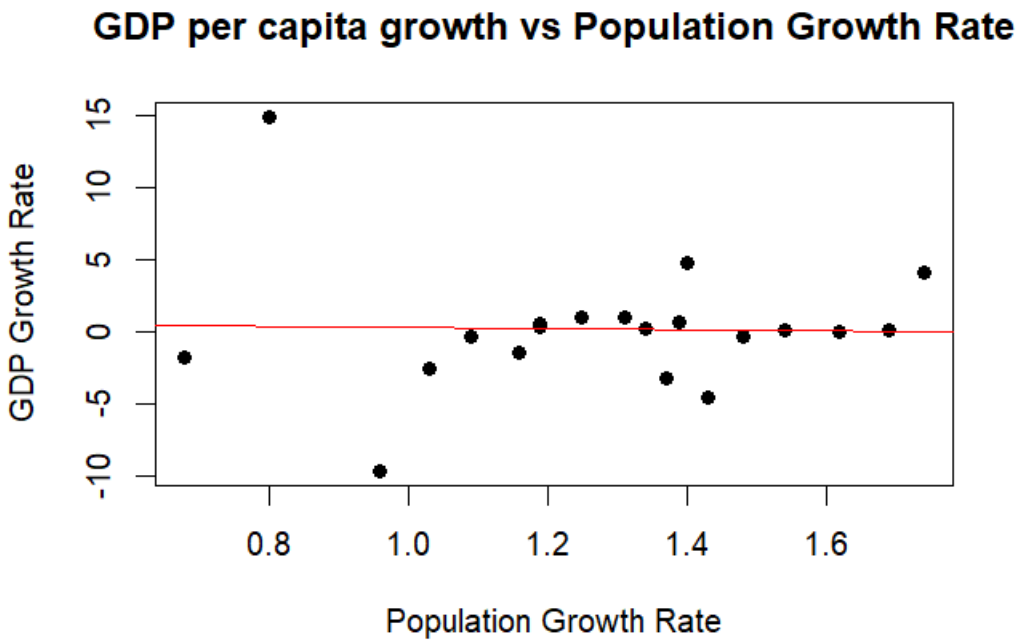
	A	B	C
1			
2	Date	Population_Growth	GDP_Growth
3	12/31/2003	1.74	4.06
4	12/31/2004	1.69	0.06
5	12/31/2005	1.62	0
6	12/31/2006	1.54	0.14
7	12/31/2007	1.48	-0.4
8	12/31/2008	1.43	-4.57
9	12/31/2009	1.4	4.78
10	12/31/2010	1.39	0.64
11	12/31/2011	1.37	-3.26
12	12/31/2012	1.34	0.22
13	12/31/2013	1.31	0.93
14	12/31/2014	1.25	1.02
15	12/31/2015	1.19	0.59
16	12/31/2016	1.19	0.26
17	12/31/2017	1.16	-1.46
18	12/31/2018	1.09	-0.34
19	12/31/2019	1.03	-2.58
20	12/31/2020	0.96	-9.7
21	12/31/2021	0.8	14.88
22	12/31/2022	0.68	-1.81

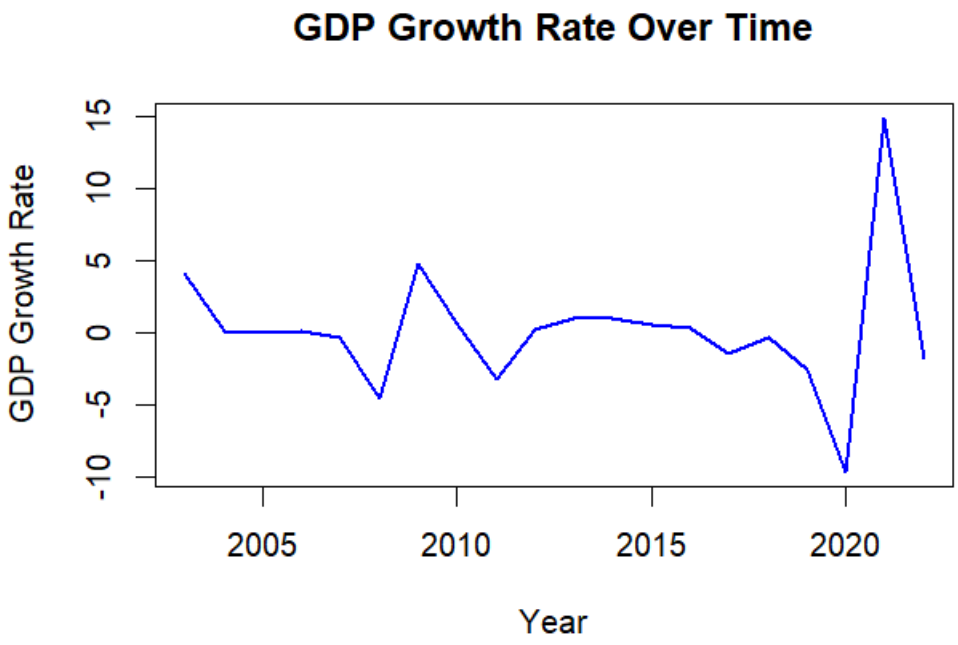
Results and discussions

Population Growth in Number VS GDP INCREASE



GDP PER CAPITA VS POPULATION GROWTH RATE (Percentage)





Output of the code(GDP PER CAPITA VS POPULATION GROWTH RATE (Percentage))

```
Residuals:
     Min       1Q   Median       3Q      Max
-10.0099  -1.8235   0.0535   0.5764  14.5023

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)         0.7167     5.0742   0.141    0.889
Population_Growth  -0.4238     3.8685  -0.110    0.914

Residual standard error: 4.718 on 18 degrees of freedom
Multiple R-squared:  0.0006663, Adjusted R-squared:  -0.05485
F-statistic: 0.012 on 1 and 18 DF,  p-value: 0.914
```

```
> data <- read.csv("C:/Users/yashg/Downloads/INDIA GDP AND POPULATION.csv")
> str(data)
'data.frame':   63 obs. of  3 variables:
 $ date             : chr  "12/31/1960" "12/31/1961" "12/31/1962" "12/31/1963" ...
 $ GDP_Growth       : num  0 3.72 2.93 5.99 7.45 ...
 $ Population_Growth: int  445954579 456351876 467024193 477933619 489059309 500114346 510992617 521987069 533
431909 545314670 ...
> model <- lm(GDP_Growth ~ Population_Growth, data = data)
> summary(model)

Call:
lm(formula = GDP_Growth ~ Population_Growth, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-12.5875  -0.9971   0.4251   2.0722   5.0800

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.903e+00  1.232e+00   1.544  0.12781
Population_Growth 3.476e-09  1.287e-09   2.701  0.00894 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.109 on 61 degrees of freedom
Multiple R-squared:  0.1068,    Adjusted R-squared:  0.09216
F-statistic: 7.294 on 1 and 61 DF,  p-value: 0.008942

> anova_result <- anova(model)
> print(anova_result)
Analysis of Variance Table

Response: GDP_Growth
                  Df Sum Sq Mean Sq F value   Pr(>F)
Population_Growth  1  70.49  70.491  7.2943 0.008942 **
Residuals         61 589.50   9.664
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary(model)$coefficients
                     Estimate   Std. Error  t value    Pr(>|t|)
(Intercept)       1.902588e+00 1.232420e+00 1.543782 0.127813595
Population_Growth 3.475974e-09 1.287022e-09 2.700787 0.008942145
>
> library(ggplot2)
```

Conclusions

Testing Hypothesis: Null Hypothesis (H0): The GDP growth rate and population growth rate do not significantly relate.

We are unable to reject the null hypothesis because the p-value for the population growth rate coefficient (0.914) is higher than the significance level (alpha = 0.05).

```
+   cat("Fail to reject the null hypothesis: There is no significant relationship between population growth rate and GDP
growth.\n")
+ }
Fail to reject the null hypothesis: There is no significant relationship between population growth rate and GDP growth.
>
```

Analysis of Regression:

The summary of the regression model indicates that the population growth rate has a non-significant coefficient (p-value = 0.914).

The population growth rate has a negligible effect on the GDP growth rate, as indicated by the coefficient estimate of -0.424.

```
Based on the regression analysis:
> if (summary(model)$coefficients["Population_Growth", "Pr(>|t|)"] < 0.05) {
+   cat("The effect of population growth rate on GDP growth rate is statistically significant.")
+ } else {
+   cat("There is no statistically significant effect of population growth rate on GDP growth rate.")
+ }
There is no statistically significant effect of population growth rate on GDP growth rate.> cat("\nThe coefficient of de
termination (R-squared) indicates that approximately", round(summary(model)$r.squared * 100, 2), "% of the variability i
n GDP growth rate can be explained by the population growth rate.")
```

Regression analysis and hypothesis testing show that, in the given context, there is not enough evidence to draw the conclusion that GDP growth rate is significantly impacted by population growth rate.

The findings imply that variables other than population growth rate might have a greater bearing on the explanation of changes in GDP growth rate over time.

Along with visualizing the data, the plots produced also depict the trend of the GDP growth rate over time and the relationship between GDP growth rate and population growth rate.

Future Scope:

- Cross-Country Comparison: Finding similarities and differences between the results and those of other nations dealing with comparable demographic and economic issues can help us better understand the state of the world economy.
- Multilevel Analysis: A more thorough understanding of the cumulative impact of various factors on GDP growth can be obtained by taking into account factors other than population growth rate, such as investment, government policies, education level, healthcare facilities, etc.

  This Indicates that GDP per capita and Population of India are not co related and exist independently. As we can see that the GDP kept decreasing and increasing randomly but the population was on a increase and stopped very rarely.

Acknowledgement

We express our sincere gratitude to all those who contributed to the successful completion of this group project. Our heartfelt thanks go to our team members for their unwavering dedication, collaborative spirit, and tireless efforts throughout the project.

We extend our appreciation to Yasvhardhan Singh Latwal and Priyanshu Kumar . for their valuable insights, support, and teamwork. Each member's unique skills and commitment played a pivotal role in the accomplishment of our shared goals.

Special thanks are due to our mentor, Mr. Aradhana Narang for his invaluable guidance, encouragement, and expertise. His mentorship significantly enriched our project, and we are grateful for the knowledge and insights he shared. This project has been a collaborative journey, and we are
grateful for the collective effort that made it a reality.

Reference

[1]Reserve Bank of India

[2]www.macrotrends.com

[3] Montgomery, D. C., Peck, E. A., & Vining, G. G. (2012). Introduction to Linear Regression Analysis (5th ed.). Wiley.

Appendix

```
data <- data.frame(
  Date = as.Date(c("2003-12-31", "2004-12-31", "2005-12-31", "2006-12-31", "2007-12-31", "2008-12-31", "2009-12-31", "2010-12-31", "2011-12-31", "2012-12-31", "2013-12-31", "2014-12-31", "2015-12-31", "2016-12-31", "2017-12-31", "2018-12-31", "2019-12-31", "2020-12-31", "2021-12-31", "2022-12-31")),
  Population_Growth = c(1.74, 1.69, 1.62, 1.54, 1.48, 1.43, 1.4, 1.39, 1.37, 1.34, 1.31, 1.25, 1.19, 1.19, 1.16, 1.09, 1.03, 0.96, 0.8, 0.68),
  GDP_Growth = c(4.06, 0.06, 0, 0.14, -0.4, -4.57, 4.78, 0.64, -3.26, 0.22, 0.93, 1.02, 0.59, 0.26, -1.46, -0.34, -2.58, -9.7, 14.88, -1.81)
)

# Perform hypothesis testing
population_growth_coeff <- summary(lm(GDP_Growth ~ Population_Growth, data = data))$coefficients["Population_Growth", "Pr(>|t|)"]

# Set significance level
alpha <- 0.05

# Test the null hypothesis
if (population_growth_coeff < alpha) {
  cat("Reject the null hypothesis: There is a significant relationship between population growth rate and GDP growth.\n")
} else {
  cat("Fail to reject the null hypothesis: There is no significant relationship between population growth rate and GDP growth.\n")
}

model <- lm(GDP_Growth ~ Population_Growth, data = data)

# Summary of regression results
summary(model)

cat("\n--- Regression Analysis Conclusion ---\n")
cat("\nCoefficient Estimate for Population Growth:", coef(model)[2])
cat("\nP-value for Population Growth:", summary(model)$coefficients["Population_Growth", "Pr(>|t|)"])
cat("\nR-squared:", summary(model)$r.squared)
```

```
cat("\n\nBased on the regression analysis:\n")
if (summary(model)$coefficients["Population_Growth", "Pr(>|t|)"] < 0.05) {
  cat("The effect of population growth rate on GDP growth rate is statistically significant.")
} else {
  cat("There is no statistically significant effect of population growth rate on GDP growth rate.")
}
cat("\nThe coefficient of determination (R-squared) indicates that approximately",
round(summary(model)$r.squared * 100, 2), "% of the variability in GDP growth rate can be explained by the population growth rate.")

plot(data$Population_Growth, data$GDP_Growth, main = "GDP per capita growth vs Population Growth Rate", xlab = "Population Growth Rate", ylab = "GDP Growth Rate", pch = 19)

# Add regression line
abline(model, col = "red")



# Convert Date to numeric year for plotting
data$Year <- as.numeric(format(data$Date, "%Y"))

# Plot GDP growth rate over time
plot(data$Year, data$GDP_Growth, type = "l", col = "blue", lwd = 2,
     main = "GDP Growth Rate Over Time", xlab = "Year", ylab = "GDP Growth Rate"
```

CODE using GDP Growth (numbers) vs GDP PER CAPITA

```r
#analysing

data <- data.frame(
  Date = as.Date(c("2003-12-31", "2004-12-31", "2005-12-31", "2006-12-31", "2007-12-31", "2008-12-31", "2009-12-31", "2010-12-31", "2011-12-31", "2012-12-31", "2013-12-31", "2014-
  Population_Growth = c(1.74, 1.69, 1.62, 1.54, 1.48, 1.43, 1.4, 1.39, 1.37, 1.34, 1.31, 1.25, 1.19, 1.19, 1.16, 1.09, 1.03, 0.96, 0.8, 0.68),
  GDP_Growth = c(4.06, 0.06, 0, 0.14, -0.4, -4.57, 4.78, 0.64, -3.26, 0.22, 0.93, 1.02, 0.59, 0.26, -1.46, -0.34, -2.58, -9.7, 14.88, -1.81)
)

# Perform hypothesis testing
population_growth_coeff <- summary(lm(GDP_Growth ~ Population_Growth, data = data))$coefficients["Population_Growth", "Pr(>|t|)"]

# Set significance level
alpha <- 0.05

# Test the null hypothesis
if (population_growth_coeff < alpha) {
  cat("Reject the null hypothesis: There is a significant relationship between population growth rate and GDP growth.\n")
} else {
  cat("Fail to reject the null hypothesis: There is no significant relationship between population growth rate and GDP growth.\n")
}

model <- lm(GDP_Growth ~ Population_Growth, data = data)

# Summary of regression results
summary(model)

cat("\n--- Regression Analysis Conclusion ---\n")
cat("\nCoefficient Estimate for Population Growth:", coef(model)[2])
cat("\nP-value for Population Growth:", summary(model)$coefficients["Population_Growth", "Pr(>|t|)"])
cat("\nR-Squared:", summary(model)$r.squared)
cat("\n\nBased on the regression analysis:\n")
if (summary(model)$coefficients["Population_Growth", "Pr(>|t|)"] < 0.05) {
  cat("The effect of population growth rate on GDP growth rate is statistically significant.")
} else {
  cat("There is no statistically significant effect of population growth rate on GDP growth rate.")
}
cat("\nThe coefficient of determination (R-squared) indicates that approximately", round(summary(model)$r.squared * 100, 2), "% of the variability in GDP growth rate can be explai

plot(data$Population_Growth, data$GDP_Growth, main = "GDP per capita growth vs Population Growth Rate", xlab = "Population Growth Rate", ylab = "GDP Growth Rate", pch = 19)

# Add regression line
abline(model, col = "red")
```

```r
# Convert Date to numeric year for plotting
data$Year <- as.numeric(format(data$Date, "%Y"))

# Plot GDP growth rate over time
plot(data$Year, data$GDP_Growth, type = "l", col = "blue", lwd = 2,
     main = "GDP Growth Rate Over Time", xlab = "Year", ylab = "GDP Growth Rate")
```

```
> data <- read.csv("C:/Users/yashg/Downloads/INDIA GDP AND POPULATION.csv")
> str(data)
'data.frame':   63 obs. of  3 variables:
 $ date             : chr  "12/31/1960" "12/31/1961" "12/31/1962" "12/31/1963" ...
 $ GDP_Growth       : num  0 3.72 2.93 5.99 7.45 ...
 $ Population_Growth: int  445954579 456351876 467024193 477933619 489059309 500114346 510992617 521987069 533
431909 545314670 ...
> model <- lm(GDP_Growth ~ Population_Growth, data = data)
> summary(model)

Call:
lm(formula = GDP_Growth ~ Population_Growth, data = data)

Residuals:
     Min       1Q   Median       3Q      Max
-12.5875  -0.9971   0.4251   2.0722   5.0800

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)
(Intercept)       1.903e+00  1.232e+00   1.544  0.12781
Population_Growth 3.476e-09  1.287e-09   2.701  0.00894 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.109 on 61 degrees of freedom
Multiple R-squared:  0.1068,    Adjusted R-squared:  0.09216
F-statistic: 7.294 on 1 and 61 DF,  p-value: 0.008942

> anova_result <- anova(model)
> print(anova_result)
Analysis of Variance Table

Response: GDP_Growth
                  Df Sum Sq Mean Sq F value   Pr(>F)
Population_Growth  1  70.49  70.491  7.2943 0.008942 **
Residuals         61 589.50   9.664
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> summary(model)$coefficients
                     Estimate   Std. Error  t value    Pr(>|t|)
(Intercept)       1.902588e+00 1.232420e+00 1.543782 0.127813595
Population_Growth 3.475974e-09 1.287022e-09 2.700787 0.008942145
>
> library(ggplot2)
```