The Statistics of Poverty and Inequality

Shyamal Borgohain
Sathvik H S
Gudavalli Sathya Tej
Abhishek Pradhan

Indian Statistical Institute, Bangalore Centre
26th March, 2025

Abstract

This report presents a comprehensive statistical analysis of socio-economic and demographic indicators across 97 countries, categorized into six geographical groups. The study examines key variables, including life expectancy (male and female), birth and death rates, infant mortality, Gross National Product (GNP) per capita, and poverty rates, to uncover global disparities and trends.
The report integrates global comparisons with localized UK trends, offering actionable insights into poverty alleviation, healthcare equity, and economic policy. Visualizations and statistical tests underscore the interplay between economic prosperity and quality of life, providing a robust foundation for cross-national policy analysis.

Acknowledgement

We would like to express our sincere gratitude to Professor Dr. Rituparna Sen, Indian Statistical Institute, Bangalore, for her insightful guidance and support throughout the course and in the preparation of this report on *The Statistics of Poverty and Inequality*. Her lectures and feedback were instrumental in deepening our understanding of the subject and in shaping the direction of this work.

Contents

Data Description

The data consists of the following variables for 97 Countries.

- Live birth rate per 1,000 of population
- Death rate per 1,000 of population
- Infant deaths per 1,000 of population(under 1 year old)
- Life expectancy at birth for males
- Life expectancy at birth for females
- Gross National Product(GNP) per capita
- Group – The countries are grouped so that comparisons between groups can be made.
  1. Eastern Europe
  2. Southern America and Mexico
  3. Western Europe, North America, Japan, Australia, New Zealand
  4. Middle East
  5. Asia
  6. Africa
- Country

The data was obtained from the U.N.E.S.C.O Demographic Yearbook 1990 and The Annual Register 1992.

Statistical Tests and Methods Used

2.1 Wilcoxon rank-sum Test

This test is a non-parametric alternative to two-sample t-test.
Let D_1 and D_2 be the distributions of the two samples respectively.

Null hypothesis (H_0): D_1 and D_2 are identical
Alternative hypothesis (H_a): D_1 is shifted to right or left of D_2

Rank all the observations jointly by ignoring the groups.

$$\textit{Test Statistic, } T = min(T_1, T_2)$$

where T_1 and T_2 are the sum of the ranks of the given samples respectively.
Decision: Reject H_0 if $T < T_L$ or $T > T_R$ where T_L and T_R are the critical values (at the required significance level α) obtained from the critical value table for Wilcoxon rank-sum test.
Wilcoxon rank-sum test is available in R as `wilcox.test(x, y)` where the arguments are the sample of two groups.

2.2 Levene's Test and Brown-Forsythe test

Null hypothesis (H_0): All groups have equal variances
Alternative hypothesis (H_a): At least one group has different variance than the rest

Let X_{ij}'s be the sample given, where X_{ij} is the jth observation in ith group.
Make simple transformation on each of the observations as

$$Z_{ij} = \begin{cases} |X_{ij} - \overline{X}_{i.}| & \text{for Levene's test} \\ |X_{ij} - \widetilde{X}_{i.}| & \text{for Brown-Forsythe test} \end{cases}$$

where:

- $\overline{X}_{i.}$ = Sample mean of ith group
- $\widetilde{X}_{i.}$ = Sample median of ith group

The mean of the transformed data (Z_{ij}) is a measure of variance. Thus in order to check homoscedasticity of the original data, check if the sample mean of transformed data is equal in each group or not. Therefore, perform ANOVA test on the transformed data.

$$\textit{Test Statistic, } T = \frac{(N-k)}{(k-1)} \cdot \frac{\sum\limits_{i=1}^{k} n_i(\overline{Z}_{i.} - \overline{Z}_{..})^2}{\sum\limits_{i=1}^{k} \sum\limits_{j=1}^{n_i} (Z_{ij} - \overline{Z}_{i.})^2} \qquad \text{Under } H_0, T \sim F(k-1, N-k)$$

where:

- N = total sample size
- k = number of groups
- n_i = sample size of ith group
- $\overline{Z}_{i.}$ = mean of deviations for ith group
- $\overline{Z}_{..}$ = grand mean of all deviations

Decision: Reject H_0 if $T > F_{\alpha,k-1,N-k}$
Both Levene's test and Brown-Forsythe test are available in car package of R as `leveneTest()` with arguments `centre = mean` and `centre = median` respectively.
Note: Brown-Forsythe test is more robust than Levene's test as it considers deviation around median rather than mean (which is more suspetible to outliers).

2.3 Welch's ANOVA Test

This test is used when normality assumption of classical ANOVA test is met but homoscedasticity assumption is not met.
This is a test which takes into account the sample variance of each group by considering them as weights for the test statistic.
Welch's test is available in R as `oneway.test()` with argument `var.equal = FALSE`.

2.4 Kruskal-Wallis Test

This test is a non-parametric alternative to ANOVA test.

Null hypothesis (H_0): All groups have identical distributions
Alternative hypothesis (H_a): At least one group differs in distribution from the rest

Rank all the observations jointly by ignoring the groups. Let R_{ij} be the rank of jth observation in ith group.

$$\textit{Test Statistic, } H = (N-1) \cdot \frac{\sum_{i=1}^{k} n_i(\overline{R}_{i.} - \overline{R}_{..})^2}{\sum_{i=1}^{k}\sum_{j=1}^{n_i}(R_{ij} - \overline{R}_{..})^2} \quad \text{Under } H_0, H \sim \chi^2_{k-1}$$

where:

- N = total sample size
- k = number of groups
- n_i = sample size of ith group
- $\overline{R}_{i.}$ = average rank of all observations in group i
- $\overline{R}_{..} = \dfrac{N+1}{2}$ = mean of all ranks

Decision: Reject H_0 if $H > \chi^2_{\alpha,k-1}$
Kruskal-Wallis test is availabe in R as `kruskal.test()`.

2.5 Dunn's Test and Bonferroni Correction

Dunn's test is a non-parametric pairwise multiple comparisons procedure based on rank sum. When Kruskal-Wallis test is rejected, then Dunn's test is used to determine exactly which groups are different.

$$\textit{Test Statistic, } z_{ij} = \frac{\bar{R}_i - \bar{R}_j}{\sqrt{\left[\frac{N(N+1)}{12} - \frac{\sum(T_s^3 - T_s)}{12(N-1)}\right] \cdot \left(\frac{1}{n_i} - \frac{1}{n_j}\right)}}$$

where:

- N = total sample size,
- n_i = size of i^{th} sample
- R_i = rank of i^{th} group
- $\bar{R}_i = \frac{R_i}{n_i}$ average rank i^{th} group
- T_s = number of tied observations at a particular rank

If multiple hypotheses are tested, the probability of observing a rare event increases and, therefore the likelihood of rejecting a null hypothesis (making a Type-I error) increases.

The Bonferroni Correction compensates for this increase by testing each individual hypothesis at a significance level of $\frac{\alpha}{m}$, where α is the desired overall level of significance and m is the number of hypotheses.

The Bonferroni Correction can also be applied as a p-value adjustment. Instead of adjusting level of α, the p-value is multiplied by m and α is left unchanged.

Decision: Reject H_0 if $p - value \leq \frac{\alpha}{2}$, where α is the level of significance.

2.6 F-Test for Regression

F-tests are frequently used to compare different statistical models.

Null hypothesis (H_0): There is no relationship between the variables.
Alternative hypothesis (H_a): The variables are significantly related.

Assumptions: Distribution is normal.

Source	**SS = Sum of Squares**	***df***	**MS = Mean Square**	***F***
Regression	$SSR = \frac{(ss_{xy})^2}{ss_{xx}}$	p	$MSR = \frac{SSR}{p}$	$F = \frac{MSR}{MSE}$
Error	$SSE = SS_{yy} - SSR$	$n-p-1$	$MSE = \frac{SSE}{n-p-1}$	
Total	$SST = SS_{yy}$	$n-1$		

Table 2.1: ANOVA Table for Linear Regression

where

- n = number of pairs in the sample,
- p = number of predictor (independent) variables

Use the F-distribution with degrees of freedom for regression $= df_R = p$, and degrees of freedom for error $= df_E = n - p - 1$. This F-test is always a right-tailed test since ANOVA is testing the variation in the regression model is larger than the variation in the error.

Decision: Reject H_0 if $p - value \leq \alpha$, where α is the level of significance.

2.7 $t-$Test for Regression

Null hypothesis (H_0): The predictor has no relationship with the dependent variable.
Alternative hypothesis (H_a): The predictor has a significant relationship.

The formula for the $t-$test statistic is:

$$t = \frac{\beta_1}{\sqrt{\frac{\text{MSE}}{SS_{xx}}}}$$

where

- β_1 = estimated slope coefficient
- MSE = Mean Square Error $= \dfrac{SSE}{n-p-1}$
- SS_{xx} = sum of squares of the deviations of the predictor variable $= \sum(x_i - \bar{x})^2$
- n = number of observations
- p = number of predictor (independent) variables

Use the $t-$distribution with degrees of freedom equal to $n - p - 1$.

2.8 Polynomial Regression

Polynomial regression is a type of regression analysis where the relationship between the independent variable x and the dependent variable y is modeled as an n-th degree polynomial. Unlike linear regression, which assumes a straight-line relationship, polynomial regression can capture non-linear patterns by fitting a curve to the data.
The model takes the form:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \cdots + \beta_n x^n + \epsilon$$

where:

- $\beta_0, \beta_1, \ldots, \beta_n$ are the coefficients.
- n is the degree of the polynomial.
- ϵ is the error term.

Key points:

- It is useful for data with curvilinear trends (e.g., parabolic or cubic patterns).
- Higher-degree polynomials increase flexibility but risk overfitting, especially with limited data.
- It is still a linear model in terms of the coefficients, so standard least-squares methods can be used to estimate them.
- Common applications include trend analysis, physics, and economics.

Choosing the right polynomial degree is critical to balance model fit and generalization.

Wealth Distribution

3.1 Plots and Tables

We start by presenting the data through plots and tables to illustrate distributions, followed by descriptive analysis and statistical tests.

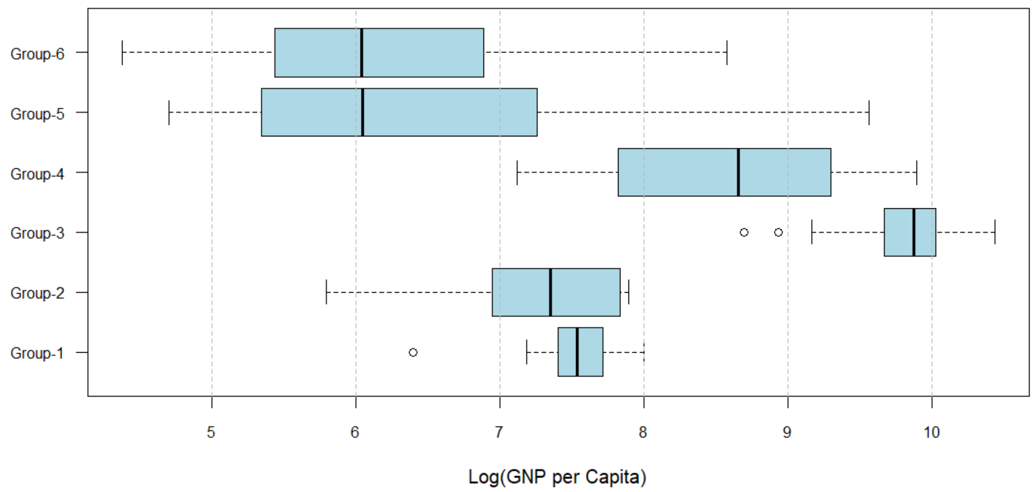


Figure 3.1: Boxplot of log of GNP per capita for each Country Group

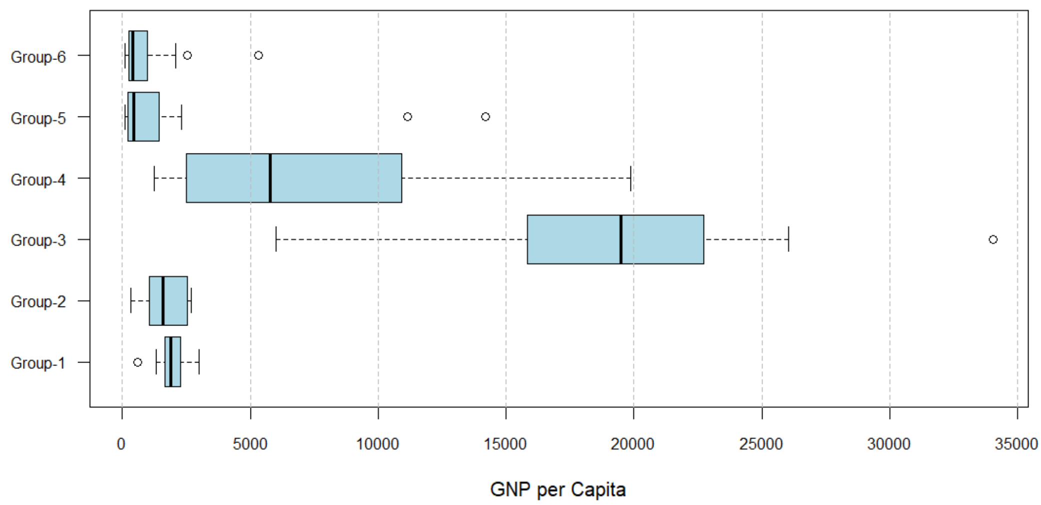


Figure 3.2: Boxplot of GNP per capita for each Country Group

Group	#Country	Q1	Median	Q3	IQR	Mean	SD
Group-1	9	7.40	7.54	7.72	0.32	7.48	0.48
Group-2	12	6.98	7.35	7.83	0.85	7.26	0.66
Group-3	19	9.67	9.88	10.00	0.36	9.76	0.44
Group-4	10	7.87	8.66	9.19	1.32	8.54	0.94
Group-5	14	5.47	6.05	7.09	1.62	6.50	1.49
Group-6	27	5.44	6.04	6.89	1.46	6.20	1.04

Table 3.1: Summary Statistics in log(GNP per capita) for each Country Group

Group	Skewness	Kurtosis	IQR/SD	$\pm 1\sigma$	$\pm 2\sigma$	$\pm 3\sigma$
Group-1	-1.03	0.10	0.66	77.78%	88.89%	100%
Group-2	-0.78	-0.58	1.28	83.33%	91.67%	100%
Group-3	-0.89	-0.08	0.81	73.68%	94.74%	100%
Group-4	-0.05	-1.54	1.40	60.00%	100%	100%
Group-5	0.86	-0.51	1.08	78.57%	92.86%	100%
Group-6	0.28	-0.65	1.39	66.67%	96.30%	100%

Table 3.2: Summary Statistics in log(GNP per capita) for each Country Group

Note: Distribution having skewness close to 0 (-0.5 to 0.5), kurtosis close to 3 (excess kurtosis between 2 and 4), IQR/SD ratio approximately 1.34(1.2 to 1.5), and data coverage around 68%, 95%, and 99.7% within $\pm 1\sigma$, $\pm 2\sigma$, and $\pm 3\sigma$ respectively, indicate approximate normality.

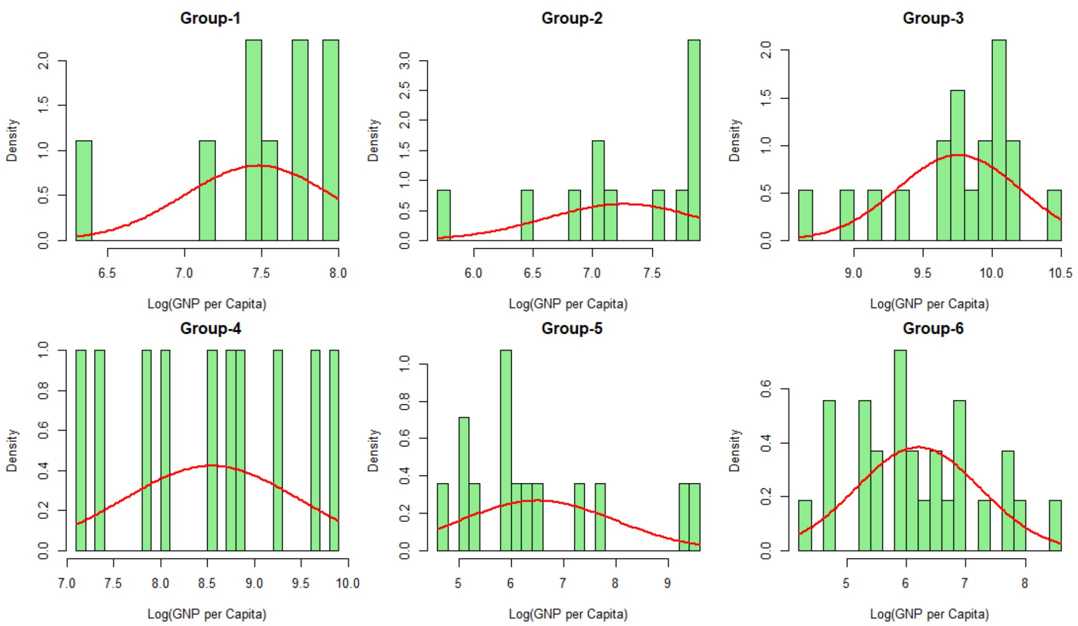


Figure 3.3: Histogram of log of GNP per capita for each Country Group

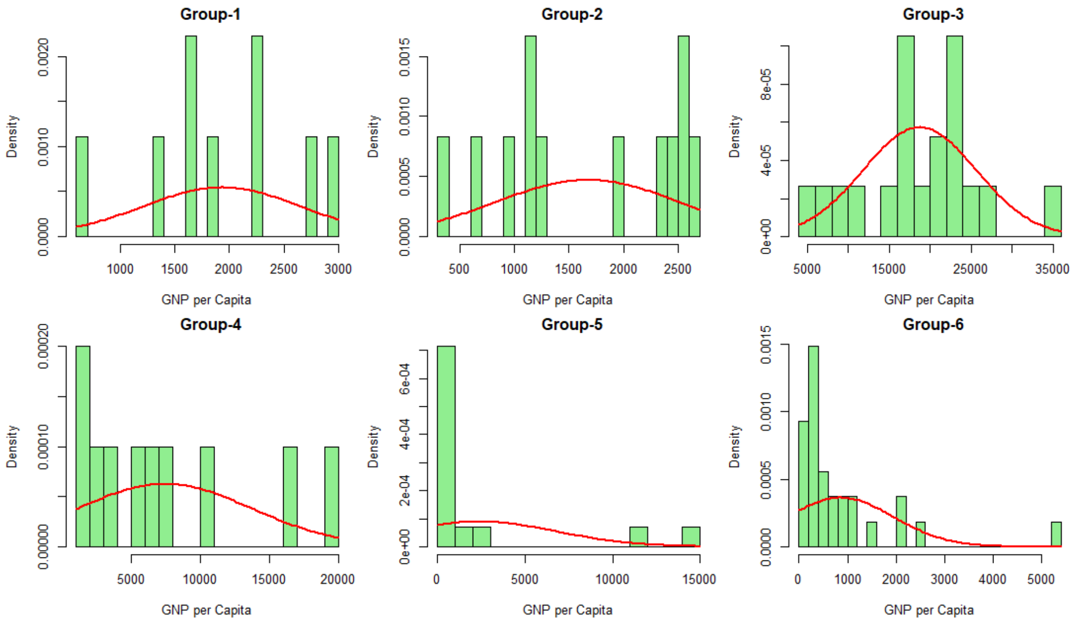


Figure 3.4: Histogram of GNP per capita for each Country Group

3.2 Groupwise Analysis of Wealth Distribution

Eastern Europe

- This group has moderate log(GNP per capita) values with relatively somewhat consistent income levels among its countries. Relatively low standard deviation and IQR imply that there is not much disparity in wealth. Negative skewness indicates that a few countries fall below the median income level, but the overall distribution is quite balanced.
- Albania appears as a low outlier in the Eastern Europe. The reasons may be economic isolation, inefficient central planning, lack of industrialization and so on.

Southern America and Mexico

- This group has a slightly lower median log(GNP) compared to Eastern Europe and shows greater spread, indicating more variation in relative income. The IQR and SD are relatively high, resulting in a broader range of wealth levels. However, the kurtosis and empirical rule implies moderate but noticeable inequality.

Western Europe, North America, Japan, Australia, New Zealand

- This group has uniformly high levels of wealth and very low relative inequality. The small SD and tight clustering of values suggest economic stability across the countries. in comparison to any other group, this group has countries with high economies and there is equitable wealth distribution.
- For this group, the log transformation has two low outliers(not too far away from the whiskers), while the GNP per capita boxplot has Switzerland as a significantly high outlier.
- The reasons may be due to Switzerland's strong banking and finance sector, which attracts global wealth, due to more industrialization, stable political and economic environment and so on. An important reason maybe due to small population with high-skilled workforce, boosting per capita wealth.

Middle East

- The Middle East has high variation in relative wealth with a wide IQR. There are some countries in the group with extreme wealth while others behind. The distribution is roughly symmetric, but the large spread values implies significant economic disparity within the region.

Asia

- Asia exhibits substantial inequality, with relatively low median log(GNP) and a very high SD and IQR. The distribution is skewed. The wide dispersion indicates the presence of high economies and countries struggling with poverty, suggesting high levels of internal inequality.

- Despite the log transformation, which tends to compress extreme values, the data still reveal a stark contrast between the poor and high economies in the region. However, log transformation compressed extreme economies Hong Kong and Singapore when they are actually significantly high outliers in Asia if we talk about GNP per capita.

- In 1993, a World Bank report The East Asian Miracle credited neoliberal policies with the economic boom, including the maintenance of export-oriented policies, low taxes and minimal welfare states. Institutional analyses found that some level of state intervention was involved. Some analysts argued that industrial policy and state intervention had a much greater influence than the World Bank report suggested.

Africa

- Africa has the lowest median and mean log(GNP) among all groups. The distribution is roughly symmetric but wide spread reflects a high degree of inequality, with a few relatively better-off countries standing out against a backdrop of widespread economic hardship. The high IQR and SD indicates broad and significant economic disparity.

- Despite the log transformation, which tends to compress extreme values, the data still reveal a stark contrast between the poor and high economies in the region. However, Libya can be considered as a significantly high outlier in Africa if we talk about GNP per capita.

- The reasons may be due to better governance and high revenues obtained from the petroleum sector. Another reason maybe due to small population.(*Source: Google*)

3.3 Analysis of Wealth Distribution among the Groups

We wish to see if the country groups have identical distributions. We observe that the country groups do not have normally distributed log of GNP per capita(also for GNP per capita.) So, we can apply Kruskal-Wallis test to see if the distribution of the groups differ statistically. The test yields a p-value of $1.109 \times 10^{-11} < 0.05$. Thus, there is enough evidence to conclude that distribution of some groups are statistically significant.

Now, in order to determine exactly which pair of groups are statistically significant, we perform Dunn-Bonferroni test. At 5% level of significance we reject if p-value < 0.025. The results of the test are given in the table below:

Group (i, j)	Adjusted p-value	Inference
(1, 2)	1.0000	Do not reject
(1, 3)	0.0175*	Reject
(1, 4)	1.0000	Do not reject
(1, 5)	0.6238	Do not reject
(1, 6)	0.1734	Do not reject
(2, 3)	0.0013*	Reject
(2, 4)	0.6748	Do not reject
(2, 5)	1.0000	Do not reject
(2, 6)	0.2843	Do not reject
(3, 4)	0.6848	Do not reject
(3, 5)	0.0000*	Reject
(3, 6)	0.0000*	Reject
(4, 5)	0.0115*	Reject
(4, 6)	0.0007*	Reject
(5, 6)	1.0000	Do not reject

Table 3.3: Dunn-Bonferroni inference for distribution of log of GNP per Capita of groups.

Note: The results for log of GNP per capita and GNP per capita for the country Groups will be same as the results are based on the ranks, not the values of the variables.

3.4 Conclusion

The analysis of wealth distribution across six geographical groups revealed significant disparities, with Group 3 exhibiting the highest median wealth and most equitable distribution. Group 1 showed moderate but stable wealth levels with low inequality. Group 4 also displayed wide variability. Group 2 demonstrated intermediate wealth with greater spread than Group 1. Groups 5 and 6 had both the lowest median wealth and highest inequality.
Statistical tests (Kruskal-Wallis, Dunn-Bonferroni) resulted in significant differences between groups ($p < 0.05$), particularly Group 3 with other Groups, namely 1, 2, 5 and 6. Moreover, significant differences were observed for Group 4 with Group 5 and 6.

Life Expectancy

4.1 Analysis of the 1990 Data

4.1.1 Comparison Between Male and Female Life Expectancy

The correlation coefficient between male and female life expectancy is found to be 0.98, indicating a strong linear relationship (Figure 4.1).

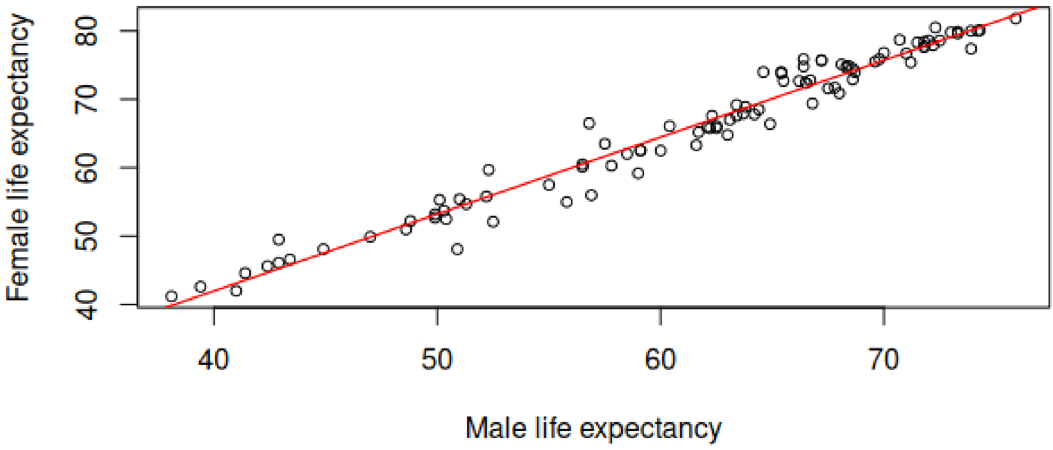


Figure 4.1: Scatterplot of Male vs Female Life Expectancy at Birth (1990)

Both interquartile range (IQR) and standard deviation of female life expectancy (17.9 years, 11.00 years) were respectively greater than that of male life expectancy (12.8 years, 9.62 years), indicating greater spread in female life expectancy. This is seen visually from the boxplot of male and female life expectancy (Figure 4.2).

Figure 4.2: Boxplot of Male and Female Life Expectancy

It was also seen by calculation and the boxplot that the median of female life expectancy (67.8 years) was higher than that of male life expectancy (63.7 years).
Wilcoxon rank-sum test was performed (to check whether the medians of male and female life expectancy were same) yielding a p-value of 0.0005849. Thus there is enough evidence to suggest that the medians of male and female life expectancy were different.

4.1.2 Male Life Expectancy Across The Groups

Boxplots by group of male life expectancy (Figure 4.3) shows clear differences across geographical groups (For example, some groups have similar medians while some others have medians which are too low or too high than the rest).

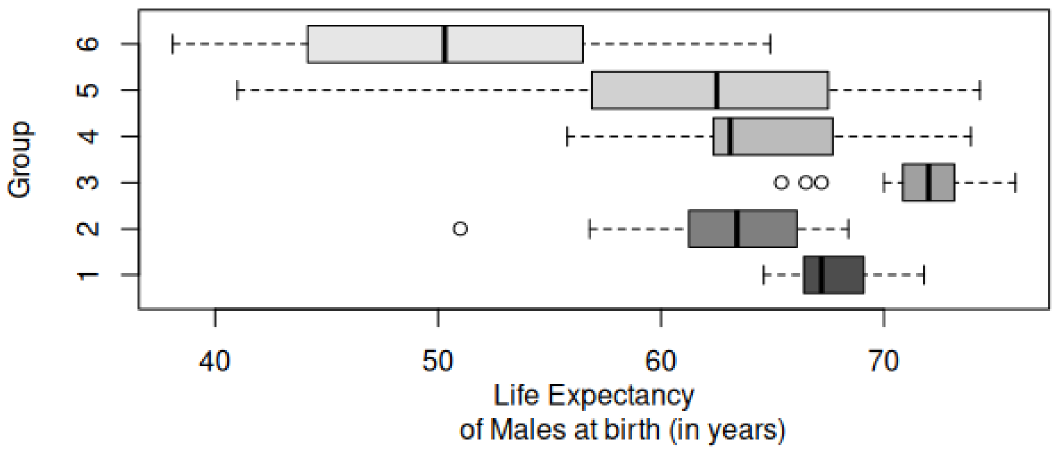


Figure 4.3: Boxplot of Male Life Expectancy by Group

- Tests (histogram and qq plot, empirical percentages, skewness,...) were conducted to check if male life expectancy in each group was normally distributed. The tests gave enough evidence to suggest that male life expectancy in each group was normally distributed.
- Brown-Forsythe test was conducted to check if male life expectancy was homoscedastic across the groups yielding a p-value of 0.0006872. Thus there is enough evidence to conclude that male life expectancy is not homoscedastic across the groups i.e, variances of male life expectancy in each group are different.

- In order to check if the means of male life expectancy in each group are same, ANOVA test needs to be performed. But since the homoscedasticity assumption of classical ANOVA test is violated, Welch's ANOVA test was performed which produced a p-value of 5.227e-14. Thus there is enough evidence to conclude that mean of male life expectancy in each group is not the same. In other words, the effect of geographical grouping on male life expectancy is statistically significant.

4.1.3 Female Life Expectancy Across The Groups

The patterns observed before are seen in female life expectancy as well as is seen in the boxplots by group of female life expectancy(Figure 4.4).

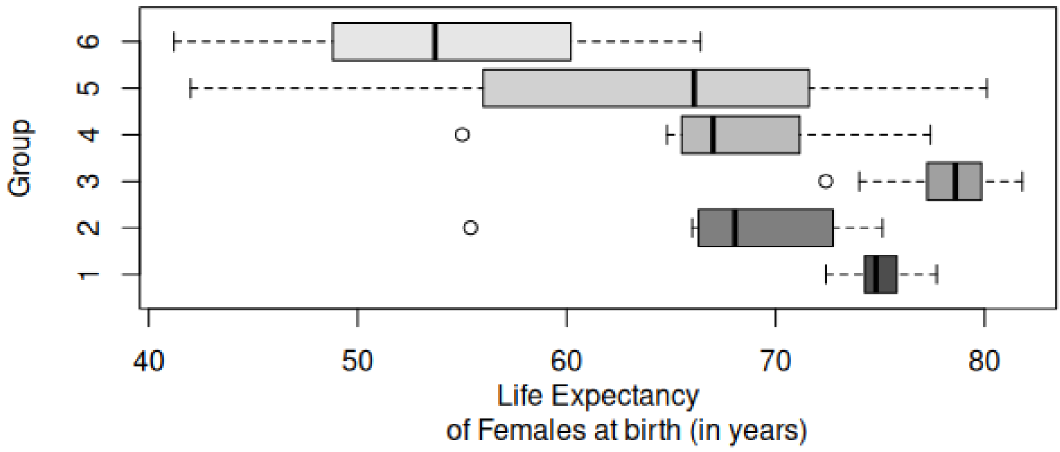


Figure 4.4: Boxplot of Female Life Expectancy by Group

The same tests which were done on male life expectancy by group are done on female life expectancy. The results are not different here and are as expected.

- Tests for normality give satisfactory evidence to conclude that the female life expectancy in each group is normally distributed.
- Brown-Forsythe test to check homoscedasticity across the groups yields a p-value of 3.752e-06. Hence there is enough evidence to conclude that female life expectancy is not homoscedastic across the groups.
- Welch's ANOVA test produces a p-value less than 2.2e-16. Therefore there is enough evidence to conclude that mean of female life expectancy in each group is not the same.

4.1.4 Difference in Life Expectancy

The difference in life expectancy (female - male) shows the following observations:

1. Minimum difference: -2.8 years (Nepal)
2. 1st Quartile: 3.2 years
3. Median difference: 4.2 years
4. Mean difference: 4.67 years
5. 3rd Quartile: 6.1 years

6. Maximum difference: 9.7 years (Peru)

The disparity in Nepal was due to high burden of CMNN (communicable, maternal, neonatal, and nutritional) diseases.Maternal malnutrition was a significant factor for death in Nepal. These factors might have affected female life expectancy more than male life expectancy causing the male life expectancy to be higher than that of females (i.e, negative difference).
The histogram (Figure 4.5) shows most countries (≈ 60%) have differences between 2-6 years, while approximately 5% show higher male life expectancy than female (i.e, negative differences).

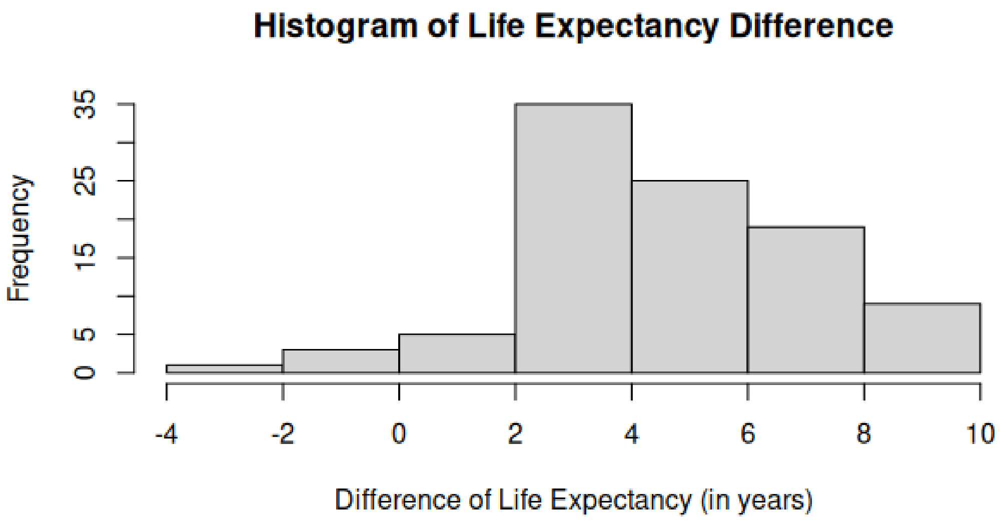


Figure 4.5: Histogram of Life Expectancy Differences (Female - Male)

- Tests for normality do not give satisfactory evidence to conclude that life expectancy differences are normally distributed. Thus there is enough evidence to conclude that life expectancy differences are not normally distributed.
- To check if the means of difference in life expectancy are same across the groups, Kruskal-Wallis test is conducted (classical and Welch's ANOVA cannot be done because assumption of normality is not met). This yields a p-value of 4.094e-11. Thus there is enough evidence to conclude that mean of difference in life expectancy across the groups is same. i.e, there is statistically significant impact of grouping on difference in life expectancy.

4.2 Analysis of the 2000 Data

Additional life expectancy data of males and females was collected for the 97 countries in the given data from the Demographic Yearbook 2000 and analyzed further. The summary of the observations which were very similar to those of 1990 are given below:

- The correlation coefficient between male and female life expectancy was 0.98.
- Interquartile range(IQR) and standard deviation of female life expectancy (17.3 years, 12.36 years) were respectively higher than that of male life expectancy (13.4 years, 10.84 years).
- The median of female life expectancy (73.1 years) was also higher than median of female life expectancy (67.3 years).

- Wilcoxon rank-sum test (conducted to check if the median of female and male life expectancy was same) produced a p-value of 0.0004457. Thus there is enough evidence to suggest that the medians of male and female life expectancy were different.

- Normality tests gave enough evidence suggest that both male and female life expectancy were normally distributed in each group.

- Brown-Forsythe test (conducted to check homoscedasticity of male and female life expectancy across the groups) gave enough evidence to conclude that both male and female life expectancy were not homoscedastic across the groups. (p-values 4.776e-06 and 2.7e-07 respectively)

- Welch's ANOVA test (conducted to check if the means of male and female life expecanty was same across the groups) gave enough evidence to conclude that the effect of geographical grouping was statistically significant on both male and female life expectancy.(p-values 3.17e-14 and 1.029e-15 respectively)

- Minimum difference in life expectancy was -0.9 years while the maximum was 12.50 years.

- Majority ($\approx 55\%$) of the difference lied the 2-6 years interval. $\approx 5\%$ of the countries had higher male life expectancy than female life expectancy.

- Normality tests did not give satisfactory evidence to conclude that difference in life expectancy was normally distributed in each group.

- Kruskal-Wallis test (conducted to check if means of difference in life expectancy are same across the groups) produced a p-value of 1.169e-11. Thus there is enough evidence to conclude that geographical grouping has statistically significant effect on life expectancy difference.

4.3 Change in Life Expectancy after a Decade

The difference of 2000 data and 1990 data was taken countrywise for both male and female life expectancy and analyzed further.

4.3.1 Comparison Between Male and Female Life Expectancy

The interquartile range (IQR) and standard deviation of change in life expectancy was higher in females (4.8 years, 5.23 years) than males(4.5 years, 4.68 years). This indicates a higher spread and the boxplot (Figure 4.6) indicates the same. The boxplot also shows outliers in female life expectancy change than male life expectancy change.

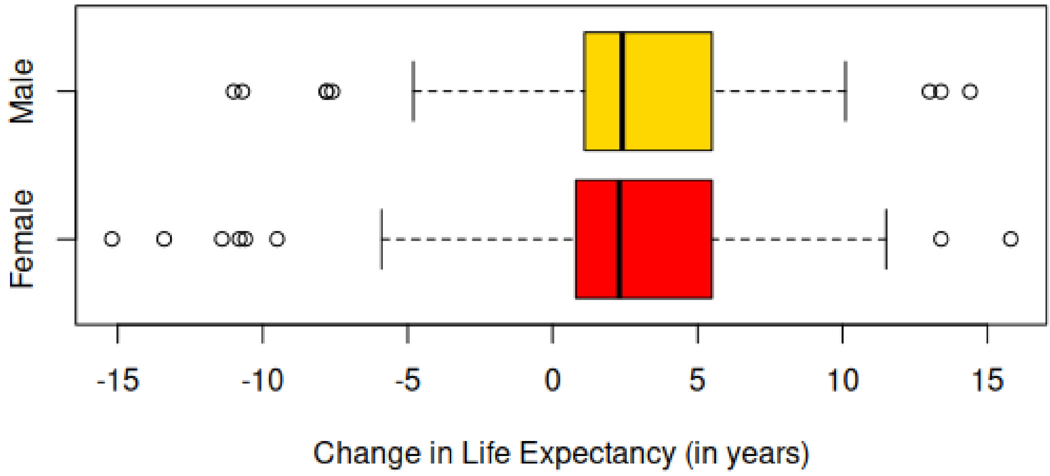


Figure 4.6: Boxplot of Male and Female Life Expectancy

The medians of life expectancy change in males and females is found to be 2.4 years and 2.3 years respectively.
Wilcoxon rank-test yields a p-value of 0.869, thus giving enough evidence to conclude that the change in life expectancy was similar in both males and females.

4.3.2 Change in Male Life Expectancy

- The maximum increase in male life expectancy was 14.4 years observed in Iraq.
- The maximum decrease (minimum difference) in male life expectancy was 11 years observed in South Africa.
- 18 of the countries saw a decrease in male life expectancy and the rest of 79 countries saw an increase.
- As can be seen from the boxplot of change in male life expectancy by group (Figure 4.7), all the countries in group 3 and group 4 saw an increase in male life expectancy.

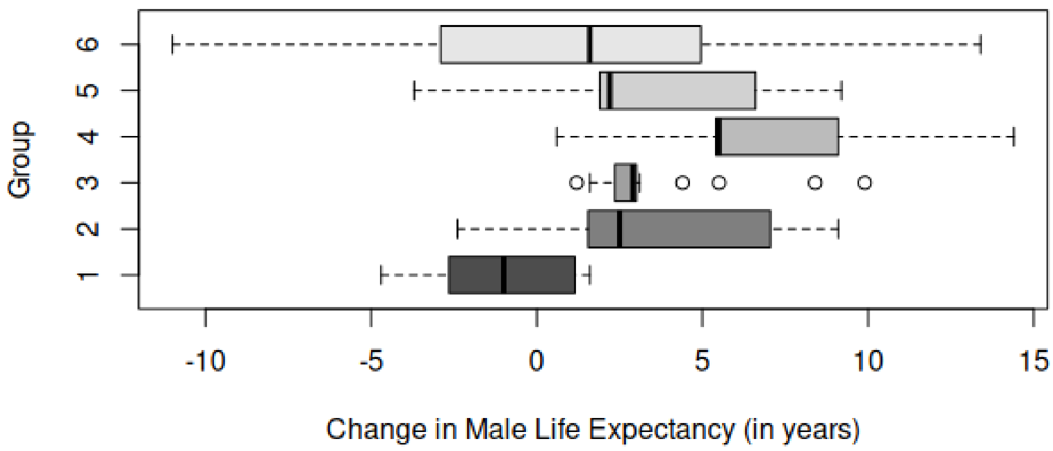


Figure 4.7: Boxplot of Change in Male Life Expectancy by Group

- It is also seen from the boxplot that group 6 has the highest IQR among the groups while group 3 has the least leading to many outliers.

4.3.3 Change in Female Life Expectancy

- The maximum increase in female life expectancy was 15.8 years observed in Iran.
- The maximum decrease (minimum difference) in female life expectancy was 15.2 years observed in South Africa.
- 19 of the countries saw a decrease in female life expectancy, 77 countries saw an increase and Nigeria did not observe any changes.
- As can be seen from the boxplot of change in female life expectancy by group (Figure 4.8), all the countries in group 2 and group 3 saw an increase in female life expectancy.

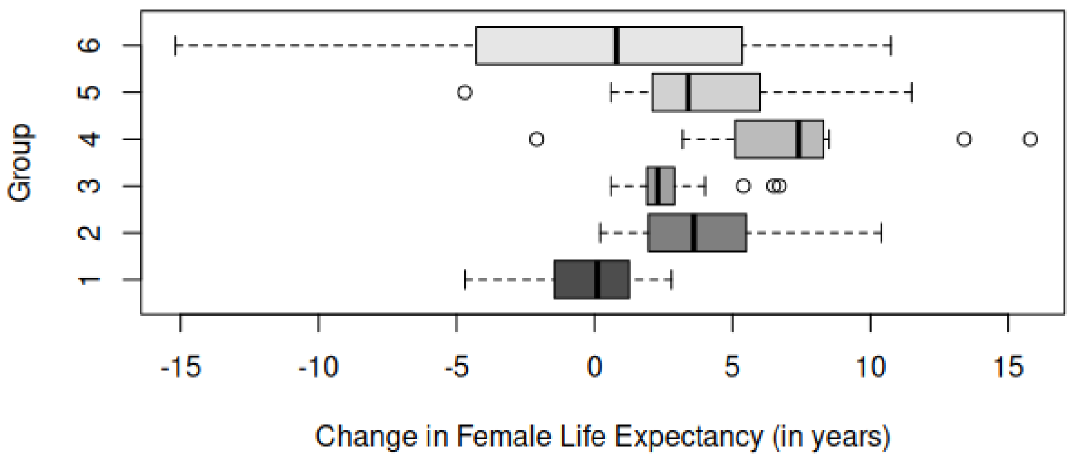


Figure 4.8: Boxplot of Change in Female Life Expectancy by Group

- It is seen from the boxplot that group 6 and group 3 have the highest and least interquartile ranges respectively as before.

4.4 Conclusion

- Most of the countries have higher female life expectancy than male life expectancy in both the data.
- Female life expectancy has higher spread than male life expectancy across in both the data.
- Geographical grouping plays a significant role in male, female and difference in life expectancy.
- The change in male and female life expectancy in the decade (1990-2000) was similar and majority of the countries saw an increase in life expectancy of both the genders.

Birth Rate and Death Rate

Demographic metrics like birth and death rates are essential for understanding population growth, health standards, and economic status. These indicators, when examined alongside economic variables such as GNP, provide a deeper insight into global development patterns.

5.1 Birth Rate and Death Rate

Birth rate refers to the number of live births per 1,000 people in a year. It reflects population growth trends and societal development stages.
Death rate (mortality rate) represents the number of deaths per 1,000 individuals annually. It helps evaluate healthcare quality and overall living conditions.

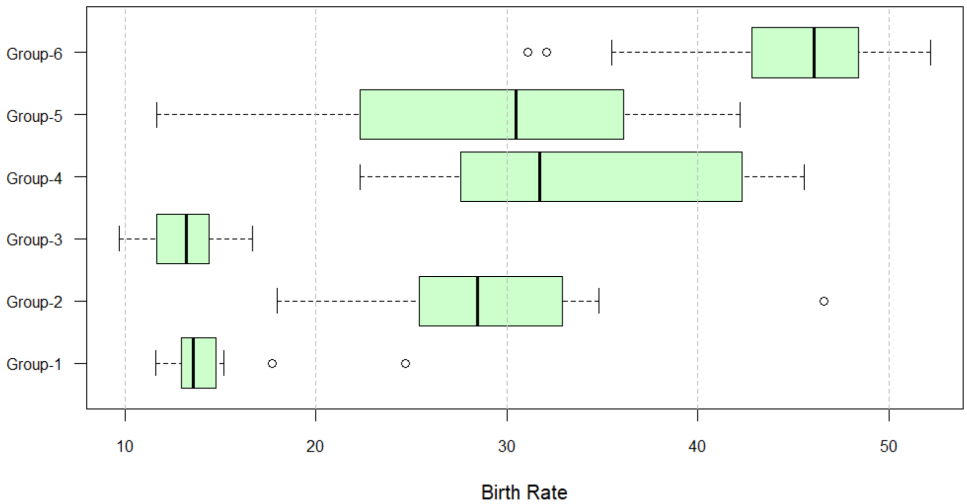


Figure 5.1: Boxplot of Birth Rate by Country Group

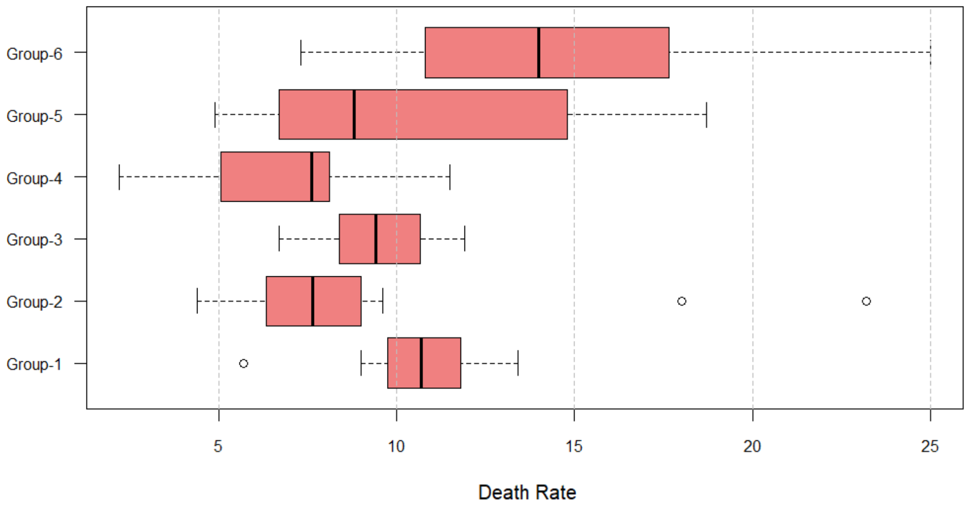


Figure 5.2: Boxplot of Death Rate by Country Group

5.1.1 Groupwise Analysis of Birth Rate and Death Rate

- The box plot for birth rates across the six groups reveals significant variation. Groups 6 exhibit the highest median birth rates, with Group 5 showing the widest spread. In contrast, Groups 1, and 3 have lowest median birth rates with less variability, while Group 2, 4 and 5 have comparable median birth rates and also have wide spread.

- The death rate box plot indicates distinct differences among the groups. Groups 6 have the highest median death rates, while Group 4,5 and 6 show the large range, spread and variability. Groups 1, 2, 3, 4 and 5 exhibit comparable and low median death rates. Group 4 is highly negative skewed, while Group 5 is highly right skewed.

5.2 Analysis of Relationship Between Birth and Death Rates

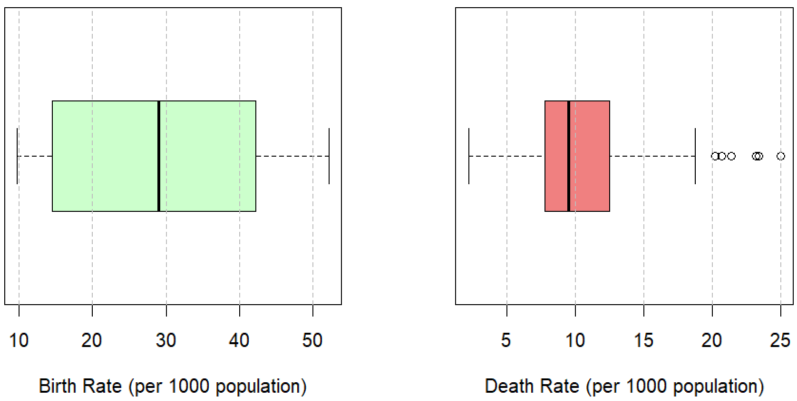


Figure 5.3: Polynomial Regression of Birth Rate and Death Rate

Suppose the distributions are normal, we see if we can fit any model.

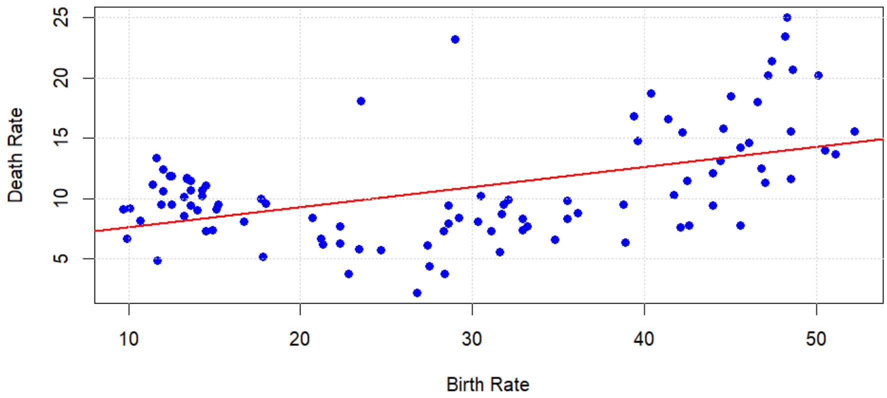


Figure 5.4: Scatterplot of Birth Rate and Death Rate with Linear Regression

- In the linear regression model, for every one-unit increase in birth rate, the death rate increases by approximately 0.167 units. The F-statistic of 29.41 with $p = 4.448 \times 10^{-7}$, confirming that the model explains a meaningful portion of the variability in death rate. The residual standard error is 4.083, i.e, the actual death rates deviate from the predicted values by about 4.08 units. While the model is significant, the R-squared value of 0.2364 suggests that only about 23.64% of the variability in death rate is explained by birth rate alone, indicating that other factors likely contribute to death rate variation.

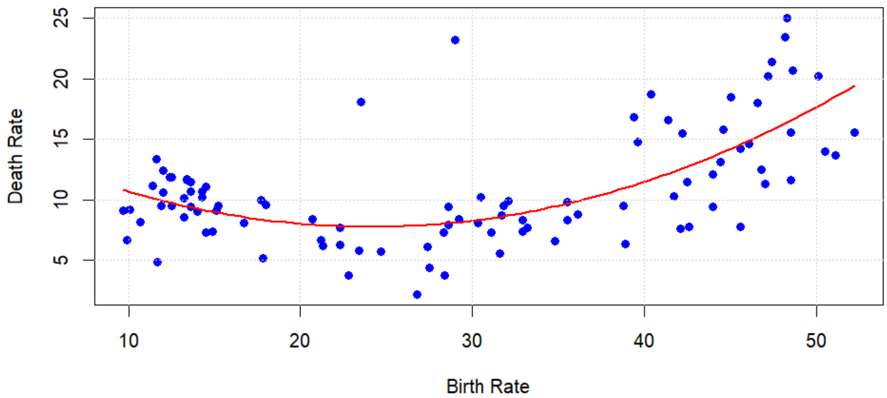


Figure 5.5: Polynomial Regression of Birth Rate and Death Rate

- The quadratic regression model analyzes the relationship between BirthRate and DeathRate using a second-degree polynomial. Both polynomial terms are statistically significant: the first component has a t-value of 6.221 with a p-value of 1.36e-08, and the second has a t-value of 5.570 with a p-value of 2.41e-07, as shown by the t-tests for individual coefficients. This indicates strong evidence that the relationship between birth and death rates is nonlinear.
- The F-test for overall model significance yields an F-statistic of 34.87 (p = 4.704e-12), confirming that the quadratic model explains a significant portion of the variability in death rate. The residual standard error has decreased to 3.559, suggesting improved prediction accuracy compared to the linear model. Moreover, the R-squared value has increased to 0.4259, meaning about 42.6% of the variation in death rate is now explained by birth rate, indicating a better fit than the linear model.

5.3 Relationship of log of GNP per capita with Birth and Death Rate

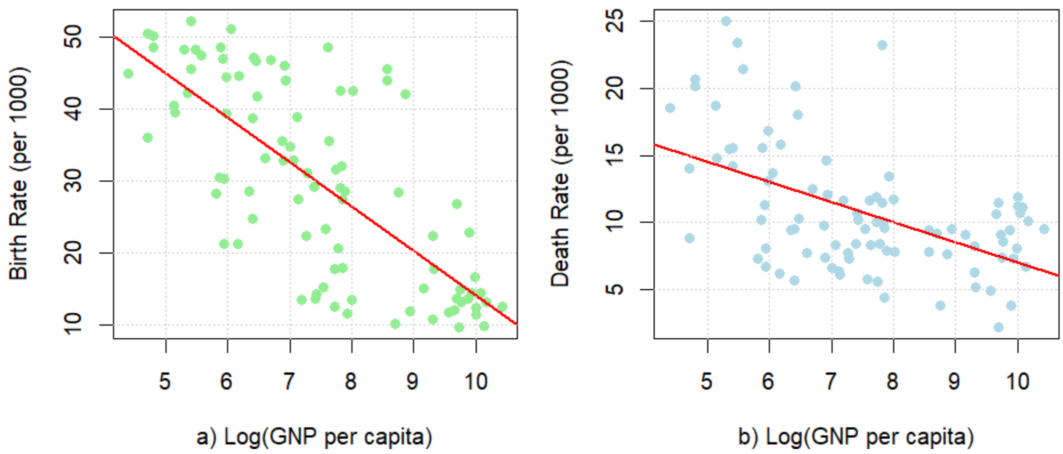


Figure 5.6: Scatterplot of log(GNP per capita) with Birth and Death Rate

Suppose the distributions are normal, we see if we can fit any model.

- Fig 5.6(a) The slope estimate is -6.1319 with a t-value of -10.26 and a p-value less than 2e-16, indicating that as the log of GNP per capita increases, the birth rate decreases significantly. The model's F-statistic of 105.3 ($p \leq 2.2e-16$) confirms overall significance. The residual standard error is 9.324, and the R-squared value is 0.5419, suggesting that about 54.2% of the variation in birth rate is explained by the log of GNP per capita.

- Fig 5.6 (b) The slope estimate is -1.495 with a t-value of -5.817 and a p-value of 9.27e-08, indicating that higher log of GNP per capita is associated with lower death rates. The F-test confirms the model's significance with an F-statistic of 33.84 ($p \leq 9.268e-08$). The residual standard error is 4.009, suggesting good accuracy. However, the R-squared is 0.2755, so only about 27.55% of the variation in death rate is explained by log of GNP per capita, indicating a weaker model fit compared to the birth rate model.

5.4 Conclusion

- Log of GNP per capita is a strong predictor of both BirthRate and DeathRate, with a stronger influence on birth rate than death rate. Modeling DeathRate as a quadratic function of BirthRate improves the fit significantly over the linear model, suggesting a nonlinear relationship. All models are statistically significant, but the BirthRate vs Log of GNP per capita is a good model.

Population Growth Rate

6.1 Introduction

This report examines two critical socio-economic issues:

1. **Population Growth** – How quickly populations grow, how we measure this, and what it means for the UK.
2. **Poverty in the UK** – A comparative and trend-based analysis of poverty rates, their causes, and government interventions.

6.2 Population Growth Analysis

Population Growth Rate in the UK

Population growth is determined by the difference between **birth rate** and **death rate**.
UK Data from the Dataset:

- Birth Rate: 13.6 per 1,000 people
- Death Rate: 11.5 per 1,000 people
- Population Growth Rate = Birth Rate − Death Rate = 13.6 − 11.5 = 2.1 per 1,000 people

What Does This Mean?

- A growth rate of **2.1 per 1,000** indicates slow population expansion. Compared to developing nations where growth rates can exceed 20 per 1,000, this suggests a **stable or aging population** rather than rapid expansion.

6.2.1 Population Doubling Time

Doubling Time = 70 / Annual Growth Rate (%)

Annual Growth Rate (%)	Doubling Time (Years)
2.0%	35 years
1.0%	70 years
0.2% (UK)	350 years

- The UK's **slow growth** means its population is **stabilizing** rather than rapidly expanding.
- Developing countries like **India and Nigeria** have **growth rates above 2%**, meaning their populations double much faster.

6.2.2 Understanding Gradient of the Gradient

The **Gradient of the Gradient** refers to how fast the **rate of inflation is changing over time**. It is essentially the **second derivative** of the population growth with respect to time.

1. Inflation rate → How much population is increasing over time.
2. First derivative (Gradient of Inflation) → The rate at which inflation is increasing or decreasing (i.e., is inflation getting worse or improving?).
3. Second derivative (Gradient of the Gradient) → The acceleration or deceleration of inflation.

If the **Gradient of the Gradient is positive**, it means inflation is **accelerating** over time. If it is **negative**, inflation is **decelerating** (slowing down).
So, the population growth rate for the UK was **0.27% in 1990**, increased to **0.84% by 2011**, and then declined to **0.32% by 2025**.
Therefore, the **Gradient of the Gradient was positive until 2011**, indicating an accelerating growth rate, but **negative after 2011**, reflecting a deceleration in population growth.

6.3 Poverty in the UK: Trends and Comparisons

6.3.1 UK Poverty Compared to Other Nations

To understand UK poverty, we compare it to both **wealthier nations (USA, Germany)** and **poorer nations (India, Nigeria)**.

Country	**UK**	**USA**	**Germany**	**India**	**Nigeria**
Birth Rate (per 1,000)	13.6	16.7	11.4	30.5	48.5
Death Rate (per 1,000)	11.5	8.1	11.2	10.2	15.6
Growth Rate (%)	2.1	8.6	0.2	20.3	32.9
GNP per capita	16,100	21,790	22,320	350	360
Life Expectancy (Male)	72.2	71.5	71.8	52.5	48.8
Life Expectancy (Female)	77.9	78.3	78.4	52.1	52.2
Infant Mortality (per 1,000)	8.4	9.1	7.4	91	105

Table 6.1: Comparison of socio-economic indicators across selected countries

6.4 Key Observations:

1. **Low Population Growth:**
   (a) The UK's **2.1 per 1,000** growth rate is much lower than countries like **India (20.3)** and **Nigeria (32.9)**.
   (b) **Germany (0.2)** is increasing at a very slow pace, indicating aging issues.
2. **Economic Standing and Poverty Levels:**
   (a) The UK's **GNP per capita ($16,100)** is much higher than India ($350) and Nigeria ($360), showing its wealth.

(b) However, **GNP is lower than the USA ($21,790)** and **Germany ($22,320)**, meaning economic growth might be slower.

3. **Infant Mortality as a Poverty Indicator:**

   (a) The UK's **infant mortality rate (8.4 per 1,000)** is low compared to **India (91.0)** and **Nigeria (105)**, showing better healthcare.

   (b) **Germany (7.4)** has slightly better numbers, meaning the UK still faces health inequality issues.

6.5 Visualizing Trends

To make the data clearer, let's analyze **how life expectancy correlates with economic prosperity**.
A scatter plot helps us understand whether **richer countries have longer life expectancy**.

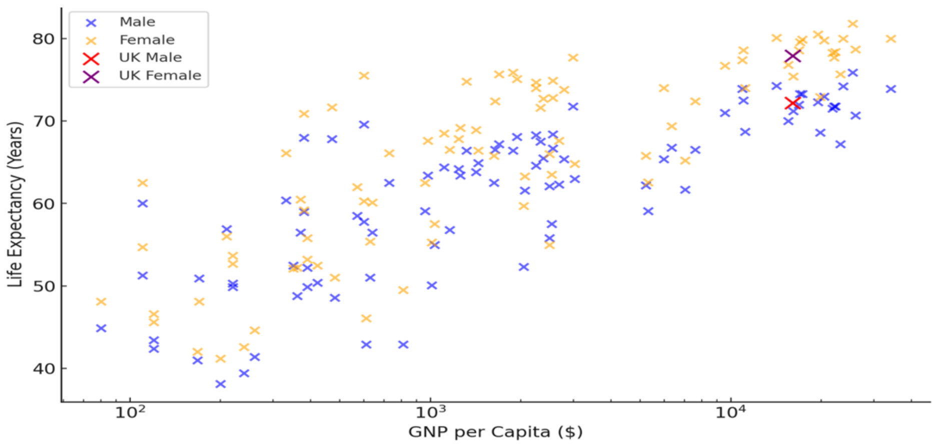


Figure 6.1: Life Expectancy vs. GNP per Capita (All Countries)

6.6 Interpretation of the Visualization:-

- There is a **clear trend** showing that as **GNP per capita increases, life expectancy also rises**.
- The UK (marked in red and purple) falls **in the middle-to-high income range**, with a **moderate life expectancy compared to its wealth level**.
- Countries with very low GNP per capita (e.g., Nigeria, India) have significantly **shorter life expectancy**, highlighting the impact of poverty on health.
- Some wealthier nations (e.g., Germany, USA) **achieve higher life expectancy than the UK**, indicating potential room for improvement in the UK's healthcare and social policies.

6.7 Poverty Rate

The poverty rate represents the **percentage of people living below the national poverty line** based on income and household surveys.

- **How Poverty is Measured**:
  - **National Estimates**: Countries define their own poverty lines based on household income and living standards.
  - **World Bank's Global Standard**: In **September 2022**, the World Bank updated the **global poverty line**:
    * **Extreme Poverty Line**: $2.15 per day (previously $1.90)
    * **Lower-Middle-Income Countries' Poverty Line**: $3.65 per day

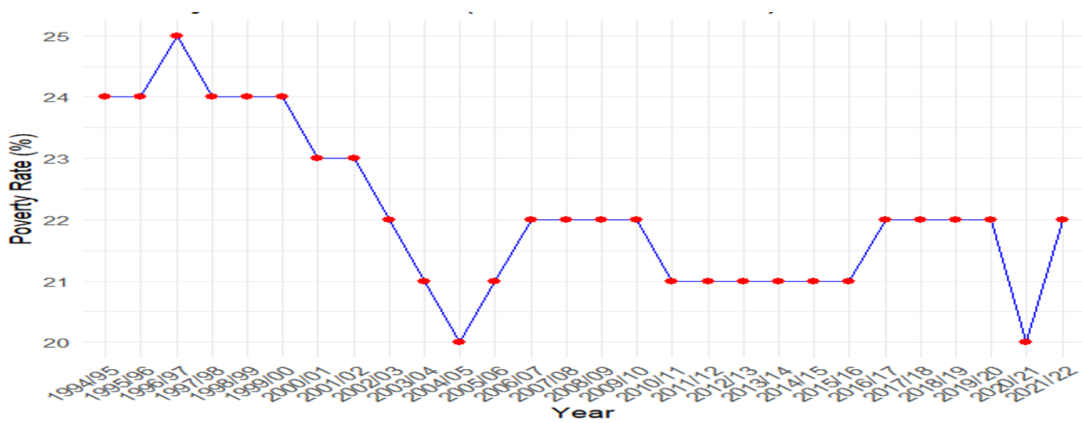
6.8 Poverty Rate Comparison by Country

Country	Poverty Rate (%)
UK	18.6
USA	18.0
Germany	14.8
India	21.9
Nigeria	40.1

- The UK's **18.6% poverty rate is higher than Germany's (14.8%)** but similar to the **USA's (18%)**.
- While the UK has a **strong welfare system**, rising living costs, wage stagnation, and inflation continue to **keep poverty levels relatively high**.
- The UK performs better than **India (21.9%)** and **Nigeria (40.1%)**, where economic instability and lack of social support systems contribute to much higher poverty rates.

6.9 UK Poverty Trends Over Time: Is the Situation Improving?

To examine UK poverty trends, we will analyze **poverty rates** over the years.



Poverty Rate in the UK (1994/95 – 2021/22)

Key Observations:

Table 6.2: **Poverty Rate by Region in the UK**

Region	**Poverty Rate (%)**
Scotland	21
Northern Ireland	17
North West	23
North East	25
Yorkshire and the Humber	23
Wales	22
West Midlands	27
East Midlands	23
East	18
London	25
South East	19
South West	20

- **Decline in Early 2000s:** The poverty rate **dropped from 24% in the late 1990s to 20% in 2004/05**. This could be due to **economic growth, social welfare policies, or labor market improvements** during that time.
- **Stable Period (2005 - 2021):** The poverty rate remained around **21–22%**, with no significant long-term improvement. This suggests that while extreme poverty may have been reduced, **structural issues such as income inequality and cost of living pressures kept the rate steady**.
- **Recent Years (2020/21 - 2021/22):** The rate dipped to **20% in 2020/21** before rising back to **22% in 2021/22**. This fluctuation might be related to the **economic impact of the COVID-19 pandemic**, government interventions, and recovery efforts.

Here is the Bar chart of above data

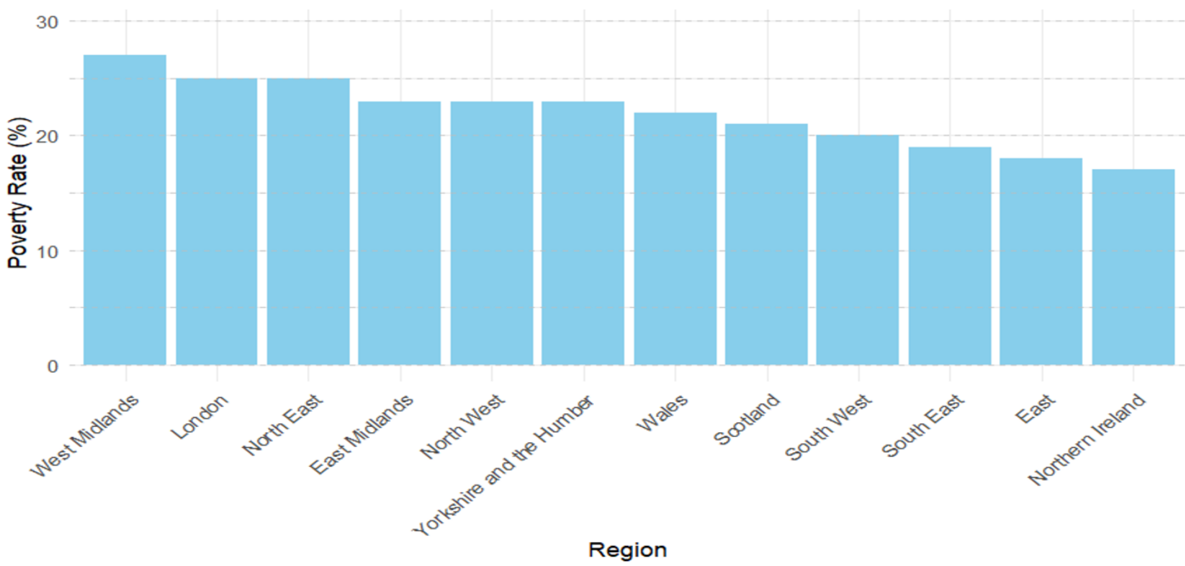


Figure 6.2: Poverty Rate by Region in the UK

Key Observations:

- **West Midlands (27%)** stands out as having the highest rate, which could indicate structural economic challenges in that area, such as industrial decline or regional disparities in job opportunities.
- **London (25%)** has one of the highest poverty rates despite being the UK's economic hub.
  This suggests that **high living costs and income inequality** contribute significantly to poverty.
- The **North East (25%)** and **North West (23%)** regions also have high poverty rates, likely due to post-industrial economic struggles and high unemployment.
- **Yorkshire and the Humber (23%)** follows a similar trend, as many towns in this area suffered economic decline after deindustrialization.
- **Scotland (21%)** and **Wales (22%)** have similar poverty rates, possibly influenced by regional economic policies and rural poverty.
- **South East (19%)** and **South West (20%)** have **relatively lower poverty rates** compared to regions like the Midlands and the North.
- These regions benefit from a stronger economy, higher wages, and better employment opportunities.

In conclusion, poverty in the UK remains a significant issue, despite being a wealthy nation.

Compared to developing countries, the UK has strong social support systems, but economic inequality, high costs, and wage stagnation continue to drive poverty.

References

- The Statistics of Poverty and Inequality (Journal of Statistics Eduacation)
- United Nations Demographic Yearbook Collection
- Mortality and risk factors of disease in Nepal: Trend and projections from 1990 to 2040 (PubMed Central)
- Why do women live longer than men? (Our World in Data)
- Joseph Rowntree Foundation, UK Poverty 2024 Report (Jan 23, 2024)
- Global poverty rates by 2025 from the World Population Review website
- U.K. Population Growth Rate 1961-2025 (macrotrends)
- Dunn's Test (Stack Exchange)
- Bonferroni Correction (Wikipedia)
- Levene's Test and Brown-Forsythe Test (Wikipedia)
- Kruskal-Wallis test (Wikipedia)
- Why are Hong-Kong and Singapore Outliers?(Wikipedia)
- F-Test and $t-$ Test for Regression
- Welch, B. L. (1947). "The generalization of "Student's" problem when several different population variances are involved" (Biometrika)
- Day, A. (ed.) (1992), *The Annual Register 1992*, 234, London:Longmans