TU Dortmund

Introductory Case Studies

Project 1: Descriptive analysis of demographic data

Lecturers:
Prof. Dr. Katja Ickstadt
M. Sc. Zeyu Ding
M. Sc. Yassine Talleb

Author: Vishesh Srivastava

Group number: 3

Group members: Vikas Singh, Arindam Pal, Jaimin Prashantkumar Oza

May 12, 2023

Contents

1 Introduction

Demographic data are crucial in understanding a region, subregion, or country's population trends, health, and economic conditions. This project aims to perform a descriptive analysis of demographic data extracted from the International Data Base (IDB) of the U.S. Census Bureau (U.S. Census Bureau, 2022), containing life expectancy at birth and under-age five mortality for both sexes, males and females, in 227 countries for the years 2002 and 2022, further divided into five regions and 21 subregions. The goal of the project is to explore the demographic data to gain more insights into the size, growth rate, distribution, and composition of a population and to understand better demography, which helps policymakers to make informed decisions about resource allocation, planning, and services, which leads to an increase in life expectancy at birth and reduces under-age five mortality.

The different statistical plots present the demographic information of the dataset; the first three analyses use data from the year 2022. A histogram conducts a univariate analysis to observe the frequency of life expectancy at birth and under-age five mortality for both sexes, males and females. In contrast, a scatter plot presents the difference between the sexes and regions. The homogeneity and heterogeneity of life expectancy at birth and under-age five mortality for both sexes, males and females, are analyzed within the individual subregion and between different subregions using a boxplot, providing a brief overview of the spread of data. Afterward, a bivariate correlation is performed to identify a possible relationship between life expectancy and under-age five mortality for both sexes, males and females. At last, a change in value over 20 years is analyzed for life expectancy at birth and under-age five mortality using a scatterplot.

In 2022, the life expectancy at birth for females is higher than for males, while the under-age five mortality for males is higher than for females. Most countries in the subregion of Middle Africa show marginal differences in life expectancy and under-age five mortality, while all the countries belonging to different subregions of Africa show a difference in their life expectancy at birth and under-age five mortality. With the increase in life expectancy for males, females observe an increase in life expectancy, and under-age five mortality observes a decrease in both sexes, males and females. Life expectancy at birth increased while under-age five mortality decreased for all the regions during the 20 years, with a few exceptions.

In addition to the introductory section, there are four other sections. Section 2 deals with a detailed description of the data set and the corresponding project objectives. Section 3 describes all statistical methods used in the analysis, including their mathematical formulas. Section 4 covers a detailed analysis of the results with the relevant tables and plots, and Section 5 summarizes the central results.

2 Problem Statement

2.1 Description of Data Set

The data set used in this project is an extract from the International Data Base (IDB) of the U.S. Census Bureau (U.S. Census Bureau, 2022). It contains demographic information for 227 countries for 2002 and 2022, categorized by five regions and 21 subregions. This data set is collected mainly from hundreds of sources, such as censuses, surveys, administrative records, and vital statistics. In contrast, the national statistical offices of various countries are the main ones.

The demographic data set contains a total of 454 observations and ten variables. The first three are of nominal category and provide information about the countries, subregions, and regions. The fourth variable year is of discrete data type and contains only two values, which are 2002 and 2022. The remaining six are continuous and contain information regarding life expectancy and under-age five mortality for both sexes, males and females. The information regarding the life expectancy at birth is in years, while the under-age five mortality is per 1,000 births. Table 1 gives a brief overview of all the variables.

In the case of a categorical variable, no region and subregion are present for two observations in 2002 and 2022; this information is provided manually by looking at their respective country names. No life expectancy at birth and under-age five mortality in 2002 was provided for the countries Libya, Puerto Rico, South Sudan, Sudan, Syria, and the United States; the average mean value of the respective column substitutes these missing values. The possible reasons for these missing values include a lack of resources to conduct surveys due to their geographical location or corrupt data. As the data set includes countries with a population of 5,000 or more and is collected by a reputational organization as a leading authority on demographic data, i.e., the National Statistical Offices of different countries, thus the collected data can be considered of top quality.

Variable	**Type**	**Scale**
Country Name	Categorical	Nominal
Region	Categorical	Nominal
Subregion	Categorical	Nominal
Year	Discrete	Numerical (Interval Scaled)
Life Expectancy at Birth Both sexes	Continuous	Numerical
Life Expectancy at Birth Males	Continuous	Numerical
Life Expectancy at Birth Females	Continuous	Numerical
Under-age five Mortality Both Sexes	Continuous	Numerical
Under-age Five Mortality Males	Continuous	Numerical
Under-age Five Mortality Females	Continuous	Numerical

Table 1: Description of Variables

As the data is collected from the top institutes and the missing values have also been taken care of, our sample dataset is also representative of the population; since there is no randomness, we can consider our dataset to have high measurement accuracy (U.S. Census Bureau, 2022).

2.2 Project Objectives

The main aim of this project is to provide a comprehensive and detailed analysis of the demographic data using various statistical methods and techniques. The content-related objectives include an overview of life expectancy at birth and under-age five mortality for both sexes, males and females, and a comparison between males' and females' life expectancy at birth and under-age five mortality with the region information. In addition, the homogeneity and heterogeneity of life expectancy at birth and under-age five mortality within and between subregions are determined. Subsequently, life expectancy at birth and under-age five mortality for both sexes, males and females, are checked for any bivariate correlations. The above analysis uses data from the year 2022. At last, the change in life expectancy at birth and under-age five mortality for both sexes, males and females, is analyzed over 20 years.

The statistical objectives required the use of different statistical methods. For univariate analysis of continuous variables, uses histograms with the mean and median as the measures of central tendency. The homogeneity and heterogeneity of the variables within and between the subregions use the interquartile range (IQR) and median information.

Pearson correlation coefficients perform a bivariate correlation between the variables. At last, a comparison of the values from 2002 to 2022 uses a scatterplot.

3 Statistical Methods

Statistical methods are crucial in data analysis, providing a framework for understanding complex data and allowing us to draw meaningful conclusions and make informed decisions. Statistical methods categorize into two classes: statistical measures used for the mathematical interpretation of a data set and statistical plots used for the graphical interpretation of a data set. The analysis uses software R (Version 4.2.1, R Core Team, 2022), including the packages ggpubr (Kassambara, 2023), RColorBrewer (Neuwirth, 2022), corrplot (Wei et al., 2021), gridExtra (Auguie and Antonov, 2017), dplyr (Wickham, 2023) and cowplot (O. Wilke, 2020).

3.1 Statistical Measures

3.1.1 Mean

The arithmetic mean is considered an essential tool for finding the central point of a numerical data set. The motivation for using the mean is that it provides a single value representing the data, making it easier to compare and analyze different data sets. It is calculated based on all the observations and is most affected by the sample fluctuations. The mean is calculated by adding all the values in the data set and dividing the sum by the total number of values in the set. Given n data-points $x_1, x_2, \ldots, x_n$ of a variable X, the mean (denoted by $\overline{x}$) is calculated as follows:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

(Black, 2010, p. 49)

3.1.2 Median

A median is a crucial tool for measuring the center of a given data set. It is often used with other statistical measures to provide a complete understanding of the data.

However, unlike the mean, the median is unaffected by extreme values, making it a more robust measure of central tendency.

The data set must first be arranged in ascending or descending order to calculate the median. The median of a data set with an odd number of data points is the middle value, while it is the average of the two middle values for a data set with an even number of data points. These middle value(s) separates the data set into the upper and lower halves. Let $x_1, x_2, \ldots, x_n$ be a list of data for the variable X, arranged in ascending or descending order, then the median (denoted by $\overline{x}$) is calculated as follows:

$$\overline{x} = \begin{cases} x_{\frac{n+1}{2}}, & \text{if } n \text{ is odd,} \\ \frac{1}{2}(x_{\frac{n}{2}} + x_{\frac{n}{2}+1}), & \text{otherwise.} \end{cases}$$

(Black, 2010, p. 48).

3.1.3 Range

The range measures the variability in a data set and provides information on the spread of the data. Given a data set, the range calculates the difference between the extreme (highest and lowest) values in the data set. Suppose $x_1, x_2, \ldots, x_n$ be a list of data for the variable X arranged in ascending order, then the range is given by:

$$\text{Range} = x_n - x_1$$

The range is a helpful tool for understanding the data set; however, it is not a suitable measure when extreme values are present (Black, 2010, p. 55).

3.1.4 Pearson Correlation Coefficient

The Pearson correlation coefficient is a statistical measure that quantifies the strength and direction of the relationship between two variables. When the increment or decrement of one variable consistently reflects in another variable, they are positively correlated, while if the opposite occurs, they are negatively correlated. The two variables are uncorrelated if there is no consistent relationship between them. The value of the coefficient ranges from -1 to $+1$, where -1 indicates a perfectly negative correlation, 0 indicates no correlation, and $+1$ indicates a perfect positive correlation. Given two data sets $x_1, x_2, \ldots, x_n$ for variable X and $y_1, y_2, \ldots, y_n$ for variable Y, the correlation coefficient between the variables x and y denoted by r_{xy} is formulated as:

$$r_{xy} = \frac{\sum\limits_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum\limits_{i=1}^{n}(y_i - \bar{y})^2}}$$

Where $\bar{x}$ and $\bar{y}$ denote the arithmetic means of the variables X and Y, respectively, and the number of observations is denoted as n (Profillidis and Botzoris, 2018, p. 186).

3.2 Statistical Plots

3.2.1 Boxplot

Boxplots are widely used in data analysis to provide a graphical depiction of a data set's distribution, summarizing large data sets in a compact format. They are often used as an alternative to the range because of less sensitivity to extreme values. They visually represent the minimum and maximum values, the median, and the first (Q1) and third (Q3) quartiles. Given a sorted data set, the median separates the data into two halves. The first quartile (Q1) and the third quartile (Q3) are the medians of the lower and upper halves of the data, respectively, and the minimum value in the data set is the smallest observation. In contrast, the maximum value in the data set is the largest observation.

A boxplot consists of a rectangular box extending from the first to the third quartile, with a vertical line inside the box indicating the median. The whiskers extend from the box to the minimum and maximum values of the data, respectively. It allows for quick comparisons of the median and quartiles of the data between different groups and is robust to extreme values.

The difference between the data set's third quartile (Q3) and the first quartile (Q1) is called the interquartile range, or IQR. It is a statistical tool used to measure a data set's spread. In addition to it, the data points that lie outside 1.5 times the IQR above the Q3 and below the Q1 are also known as "extreme values" (Potter et al., 2006, p. 100).

3.2.2 Histogram

Histograms are an essential tool in data analysis, as they provide a graphical representation of the distribution of a data set that is crucial for analyzing continuous numerical data. The histogram is constructed by dividing the range of the data into a set of contiguous intervals of equal size called bins. It is constructed by drawing rectangles of equal width and connecting horizontal line segments from the endpoint to the endpoint of a bin. The number of observations falling into each bin is then counted and represented as each bin's height. This height is typically plotted on the Y-axis, which contains the frequency information, while the bar's width represents the bin's range. Histograms are very useful for detecting which bin has the highest frequency (Black, 2010, p. 21).

3.2.3 Scatterplot

A scatterplot visually represents the relationship between two numerical variables, X and Y, in a two-dimensional space. It is represented by a series of points, each representing a pair of values, one for each variable. The primary purpose of a scatterplot is to show the extent to which two variables are related and to identify any patterns or trends that may exist between them. These two variables can exhibit three types of relationships, i.e., positive, negative, or zero correlation, as discussed in the Pearson Correlation Coefficient of Subsection 3.1.4. Scatter plots are very useful in exploratory data analysis and are often used to visualize the relationships between variables. They are simple to create and understand, effectively communicating findings (Black, 2010, p. 33).

A diagonal line is included in the scatterplot to facilitate the comparison of two variables. All the data points on the diagonal line have the same value for the two variables. The deviation of the data points from the diagonal line shows a change in the values of the variables. Specifically, a shift upward or downward of the data points from the diagonal line represents an increase or decrease in the values of the variables respectively (van Aartsengel and Kurtoglu, 2013, p. 223).

4 Statistical Analyses

The following section defines the statistical analysis performed to analyze the life expectancy at birth and under-age five mortality of both sexes, males and females. As a preprocessing step, the missing values for the continuous numeric variables are substituted by the respective column mean value. In contrast, in the case of categorical variables, the values are manually substituted, looking into the data. The frequency distribution, homogeneity, heterogeneity analysis, and bivariate correlations use data from 2022. The data from 2002 and 2022 are incorporated to examine the change in life expectancy at birth and under-age five mortality values over 20 years.

4.1 Frequency Distribution of Variables

In this subsection, a frequency distribution of life expectancy at birth and under-age five mortality for both sexes, males and females, is shown as histograms with respective mean and median information. The life expectancies at birth and under-age five mortality for both sexes, males and females, are shown in Figure 1.

The mean life expectancy for both sexes, males and females, is 74.58, 72.10, and 77.18 years, while the median values are 75.82, 73.26, and 78.69 years, respectively. The minimum life expectancy at birth for males is 52.10 years, while for females, it is 55.28 years. The maximum life expectancy at birth is 85.70 years for males and 93.5 years for females. On average, males live between 67.93 and 77.19 years, and females live between 72.63 and 82.56 years.

The under-age five mortality for both sexes, males and females, show a mean value of 26.68, 29.23, and 24.01 per 1,000 births, respectively, while the median values are 15.08, 17.55, and 13.62 per 1,000 births, respectively. Under-age five mortality for males shows a range of 159.75 per 1,000 births, while females show a range of 144.45 per 1,000 births. The African region has the highest under-age five mortality, with a mean value of 60.70 per 1,000 births, while the European region has the lowest, with a mean value of 5.81 per 1,000 births. The analysis suggests that the under-age five mortality for females is lower than males for all regions, while it is the lowest in the European region. On the other hand, females have a higher life expectancy than males. Overall, females have better health outcomes regarding life expectancy at birth and under-age five mortality.

The frequency distribution considering the differences between sexes and regions is shown in Figure 2 as a scatter plot. For life expectancy at birth, as shown in Figure 2 (a), females tend to live longer compared to males, with a single exception from the country of Montserrat. However, as most data points follow a similar trend, this exception will be ignored. In addition, the life expectancy for females is highest in the European

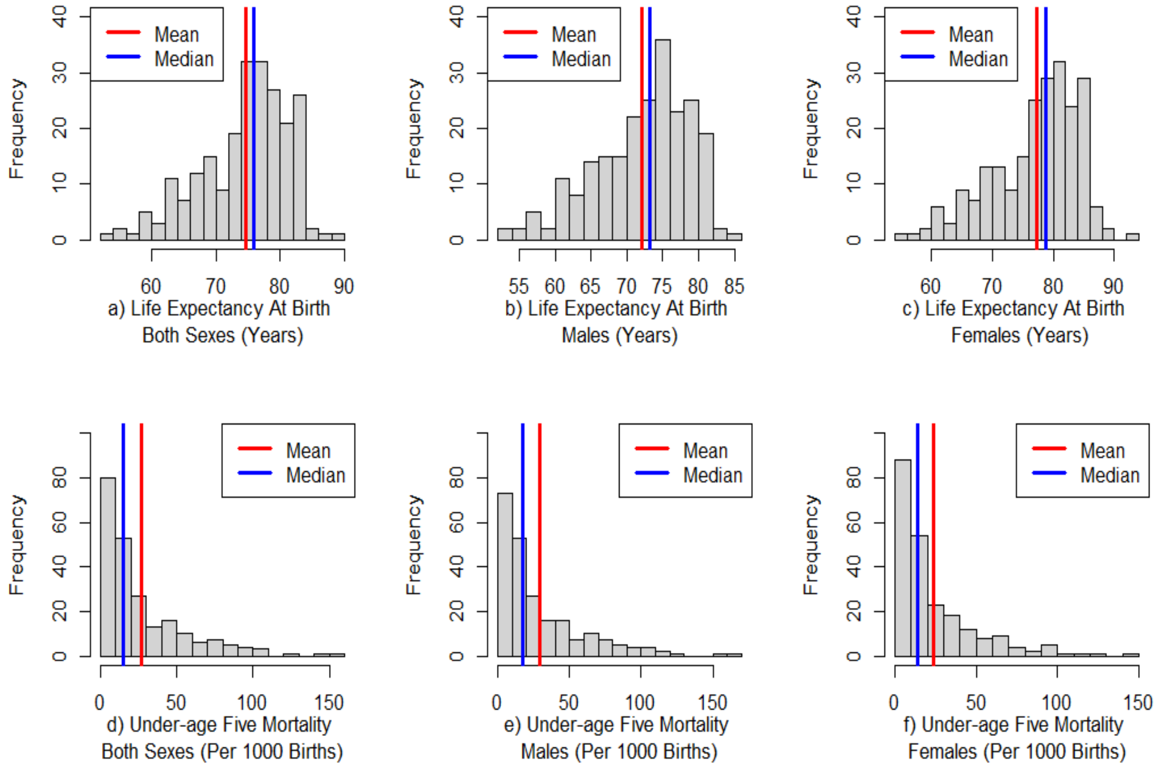
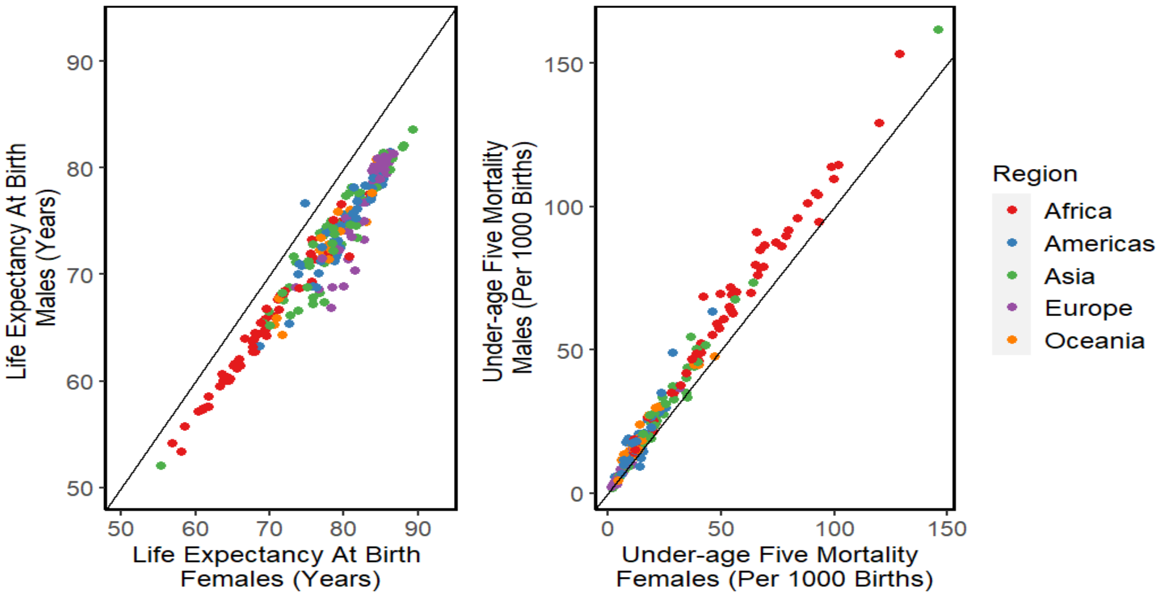


Figure 1: Frequency distribution of continuous variables

region and lowest in the African region. These differences cannot solely be attributed to sociological factors but are also influenced by specific biological characteristics. For instance, research has shown that men experience higher levels of oxidative damage to vital mitochondrial components than women and that H2O2 production is lower in females than males. These biological factors may contribute to the observed differences in life expectancy between males and females (Viña et al., 2005).

Males tend to experience higher under-age five mortality in all the regions compared to females, with eight exception values from India, 'Saint Kitts and Nevis', Saint Lucia, Montserrat, Croatia, 'Wallis and Futuna', Estonia, and Montenegro. This difference can be explained by various factors: for example, male infants tend to be born with a higher risk of birth complications like respiratory distress syndrome, making them more vulnerable to diseases. Male infants also have weaker immune systems in comparison to females. Additionally, males are more likely to be born prematurely, increasing their chances of death (Pongou, 2013).



(a) Life expectancy at birth for males and females

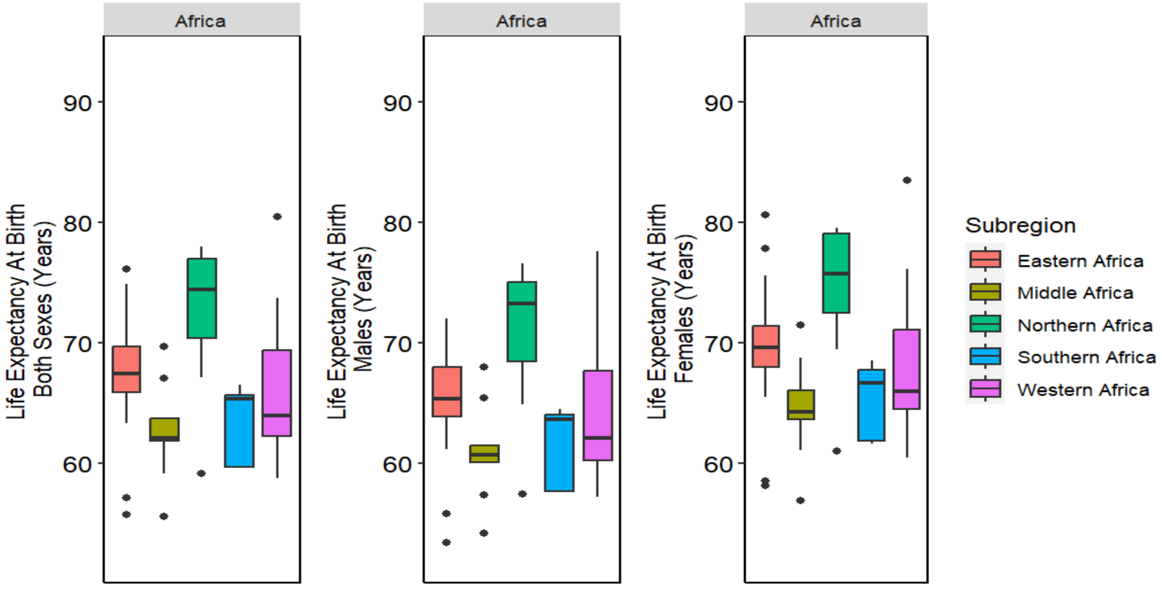
(b) Under-age five mortality for males and females

Figure 2: Differences between sexes and regions in terms of life expectancy and under-age five mortality for males and females

4.2 Homogeneity and Heterogeneity Analysis of Variables

In this subsection, the homogeneity and heterogeneity of the life expectancy at birth and under-age five mortality for both sexes, males and females, are analyzed using boxplots. The African region is chosen for this analysis because it has the highest IQR of 7.46 years for the life expectancy for both sexes and 37.94 per 1,000 births for under-age five mortality of both sexes, as shown in Table 2 and Table 3 on page 18 in the Appendix, this indicates higher variability within African subregions. The Q1, median, and Q3 in the African region show variability between subregions; thus, it is more interesting to analyze the data for this region.

The life expectancy at birth for both sexes, males and females, shows the same trend for homogeneity and heterogeneity, as shown in Figures 3(a), 3(b), and 3(c). For life expectancy at birth for both sexes, the median varies among different subregions, showing heterogeneity among them. The subregion of Middle Africa shows the smallest IQR of 1.87 years, while Western Africa shows the highest IQR of 7.11 years. The data suggests that Middle Africa is the most homogeneous subregion, while Western Africa is the most heterogeneous.



(a) Life expectancy at birth for both sexes (b) Life expectancy at birth for males (c) Life expectancy at birth for females

Figure 3: Variability of life expectancy at birth for both sexes, males, and females within and across subregions of African region

The under-age five mortality for both sexes, males and females, are shown in Figures 6(a), 6(b), and 6(c) on page 19 in the Appendix. The median varies for both sexes, males and females, between different subregions, indicating heterogeneity among them. The Eastern Africa subregion is the most homogeneous, with an IQR of 17.70 per 1,000 births for both sexes and 13.80 per 1,000 births for females. The next is Southern Africa, with an IQR of 18.40 per 1,000 births for both sexes and 16.80 per 1,000 births for females. In the case of males, the subregion of Southern Africa is the most homogeneous, with an IQR of 20.00 per 1,000 births, followed by Eastern Africa, with an IQR of 23.20 per 1,000 births. In contrast, the subregion of Middle Africa is the least homogeneous for both sexes, males and females, as it shows the highest IQR. In conclusion, Southern Africa and Eastern Africa are the most homogeneous within the individual subregion, while different subregions are heterogeneous among each other.

4.3 Bivariate Correlation Analysis between Variables

This subsection analyzes the relationship between life expectancy at birth and under-age five mortality for both sexes, males and females, for any bivariate correlation. The

Pearson method calculates the correlation coefficients, and a scatter plot displays the data points, allowing visualization of the distributions, as shown in Figure 4.

A positive correlation nearly equal to one exists between all three variables of life expectancy at birth, i.e., both sexes, males and females. The lowest value for the correlation coefficient, +0.97, is observed between life expectancy for males and females, showing that when the life expectancy for males increases, the life expectancy for females increases, and vice versa. Life expectancy at birth and under-age five mortality show a negative correlation between both sexes, males and females.

There is a positive correlation between the variables of under-age mortality for both sexes, males and females, indicating that with an increase in male mortality, females also observe an increase in mortality, and vice versa. The lowest correlation coefficient value, i.e., -0.91, is observed between life expectancy at birth and under-age five mortality in females. It is observed from this result that with an increase in life expectancy at birth, under-age five mortality decreases, and vice versa.

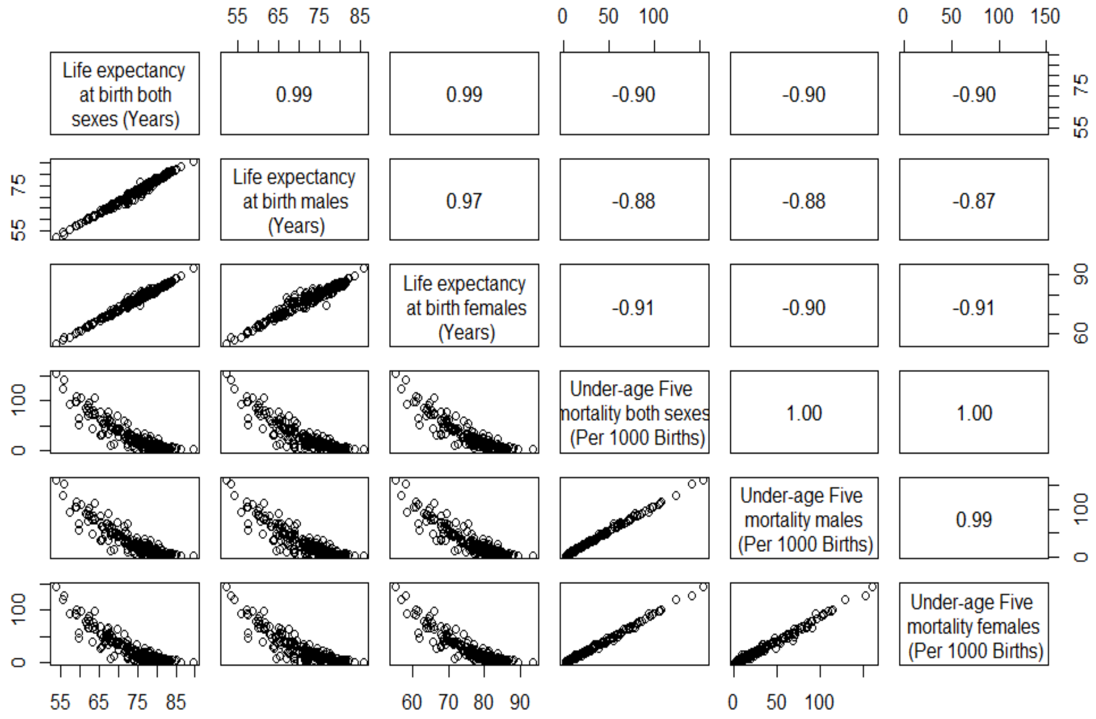
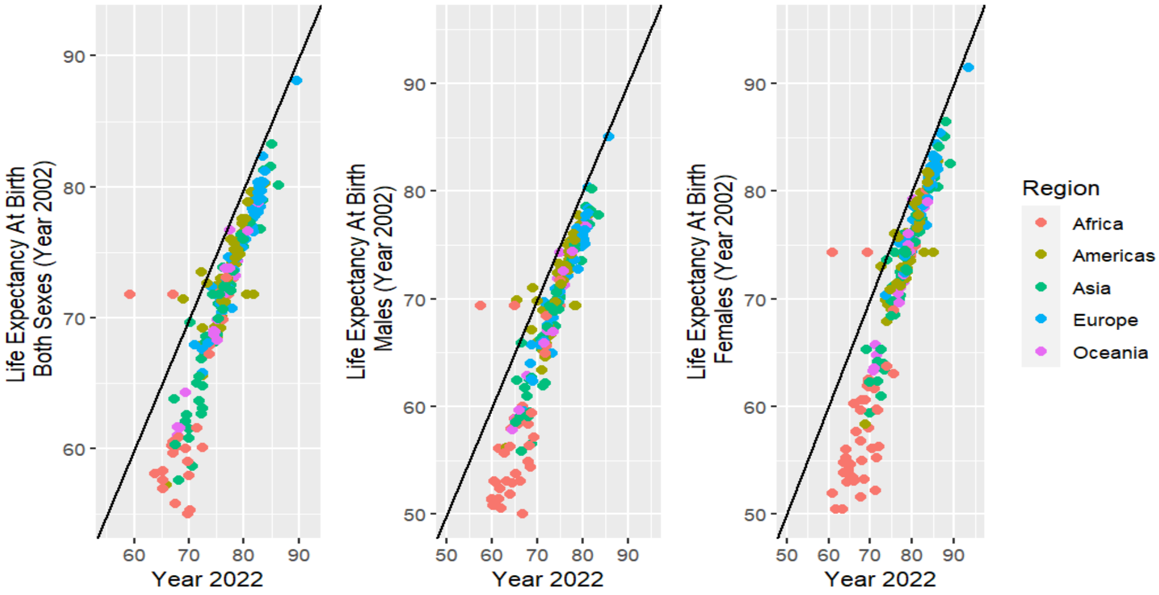


Figure 4: Bivariate correlation between the continuous variables

4.4 Comparison of Changes in Variable Values from 2002 to 2022

This subsection uses a scatter plot to examine the variation in life expectancy at birth and under-age five mortality for both sexes, males and females, from 2002 to 2022. The life expectancy at birth for both sexes, males and females, increased in the past two decades, as shown in Figure 5. The life expectancy at birth for both sexes in Figure 5(a) and males in Figure 5(b) shows four exceptional values from the countries of South Sudan, Sudan, Peru, and Mexico. Our current analyses will ignore these exceptions, as most data points follow the same trend. The European region has the highest life expectancy for both sexes, males and females, while the African region has the lowest life expectancy at birth.

Under-age five mortality for both sexes, males and females, shows a decreasing trend over 20 years, as shown in Figures 7(a), 7(b), and 7(c) on page 19 in the Appendix, with two exception values from Sudan and South Sudan of the African region. Europe has the lowest under-age five mortality while Africa has the highest. This analysis indicates that life expectancy at birth has increased while under-age five mortality has decreased over the past 20 years.



(a) Life expectancy at birth for both sexes

(b) Life expectancy at birth for males

(c) Life expectancy at birth for females

Figure 5: Comparison of the changes in life expectancy at birth values over 20 years for both sexes, males and females

5 Summary

This project involved a descriptive analysis of a demographic data set extracted from the International Data Base of the U.S. Census Bureau by the instructors of the course Introductory Case Studies at TU Dortmund University in the summer term of 2023. The data set contains 454 observations from 227 countries, grouped by regions and subregions. The variables country, region, and subregion are categorical, the year is a discrete interval scaled, and life expectancy at birth and under-age five mortality for both sexes, males and females, are continuous.

Initially, the frequency distributions of the life expectancy at birth and under-age five mortality for both sexes, males and females, were analyzed for 2022 while considering differences between sexes and regions. The data showed that the average life expectancy at birth for females (77.18 years) is higher than for males (72.10 years), while the under-age five mortality for females (24.01 per 1,000 births) is lower compared to males (29.23 per 1,000 births). These results suggest that, on average, females live longer than males and have lower under-age five mortality than males; this can be due to less exposure to physical work, reduced stress levels, or biological differences.

Next, the homogeneity and heterogeneity of life expectancy at birth and under-age five mortality for both sexes, males and females, were analyzed within and between subregions for 2022. The African region was chosen here, as high IQRs are observed for life expectancy at birth and under-age five mortality of both sexes. The life expectancy at birth for both sexes, males and females, was heterogeneous between subregions. At the same time, the subregions of Middle Africa and Eastern Africa were found to be the most homogeneous. The under-age five mortality for both sexes, males and females, was also found to be heterogeneous among each other and homogeneous for the individual subregions of Eastern Africa and Southern Africa. This analysis showed that the life expectancy at birth and under-age five mortality vary between the different subregions of Africa. At the same time, it was almost equal in the subregion of Eastern Africa.

The analysis of homogeneity and heterogeneity was followed by a bivariate correlation analysis between life expectancy at birth and under-age five mortality for both sexes, males and females, for the year 2022, showing a negative correlation between life expectancy at birth and under-age five mortality. Furthermore, life expectancy at birth for both sexes, males and females, was positively correlated. Under-age five mortality between both sexes, males and females, was also positively correlated. It was observed that an increase in males' life expectancy at birth resulted in an increase in females' life expectancy at birth. At the same time, there is an inverse relationship between life expectancy at birth and under-age five mortality. Similarly, an increase in under-age five mortality for males led to an increase in under-age five mortality for females, and vice versa. At last, a comparison of the change in life expectancy at birth and under-age five mortality between the years 2002 and 2022 was analyzed using a scatterplot. The results showed that life expectancy at birth had increased while under-age five mortality had decreased over the past 20 years; this can be due to improved health conditions and increased social awareness.

In conclusion, this project provided insights into the demographic data of various countries and highlighted the importance of appropriate statistical measures and graphical methods for exploratory and descriptive analysis. Future investigations include examining the impact of various factors on life expectancy at birth and mortality and identifying potential solutions to improve health conditions.

Bibliography

Auguie, B. (2017): gridExtra: Miscellaneous Functions for "Grid" Graphics, R package version 2.3. URL: https://CRAN.R-project.org/package=gridExtra.

Black K. (2010): Business Statistics: For Contemporary Decision Making, Sixth Edition, John Wiley Sons, Inc., Hoboken, NJ.

Kassambara, A. (2023): ggpubr: 'ggplot2' Based Publication Ready Plots, R package version 0.6.0. URL: https://CRAN.R-project.org/package=ggpubr.

Neuwirth, E. (2022). RColorBrewer: ColorBrewer Palettes (Version 1.1-3) [Software]. Retrieved from https://CRAN.R-project.org/package=RColorBrewer.

Pongou, R. (2013). Why Is Infant Mortality Higher in Boys Than in Girls? A New Hypothesis Based on Preconception Environment and Evidence From a Large Sample of Twins. Demography, 50(2), 421-444.

Potter K., Hagen H., Kerren A., and Dannenmann P. (2006): Methods for presenting statistical information: The box plot. *Visualization of large and unstructured data sets*, 4:97–106.

Profillidis V.A. and Botzoris G.N. (2018): *Modeling of Transport Demand: Analyzing, Calculating, and Forecasting Transport Demand*, Elsevier.

R Core Team (2022). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. Vienna, Austria. url: https://www.R-project.org/.

The U.S. Census Bureau (2022): The International Data Base (IDB). URL: https://www.census.gov/programs-surveys/international-programs/about/idb.html (visited on 2nd May 2023).

van Aartsengel A. and Kurtoglu S. (2013): *Handbook on Continuous Improvement Transformation*, Springer, Berlin, Heidelberg.

Viña, J., Borrás, C., Gambini, J., Sastre, J., Pallardó, F. V. (2005). Why females live longer than males? Importance of the upregulation of longevity-associated genes by oestrogenic compounds. *FEBS Letters*, 579(12), 2541-2545.

Wei, T. Simko, V. (2021). R package 'corrplot': Visualization of a Correlation Matrix, Version 0.92. Retrieved from https://github.com/taiyun/corrplot.

Wickham, H., François, R., Henry, L., Müller, K., Vaughan, D. (2023). dplyr: A Grammar of Data Manipulation, R package version 1.1.2. URL: https://CRAN.R-project.org/package=dplyr.

Wilke, C.O. (2020). cowplot: Streamlined Plot Theme and Plot Annotations for 'ggplot2'. R package version 1.1.1. URL: https://CRAN.R-project.org/package=cowplot.

Appendix

A Additional Tables

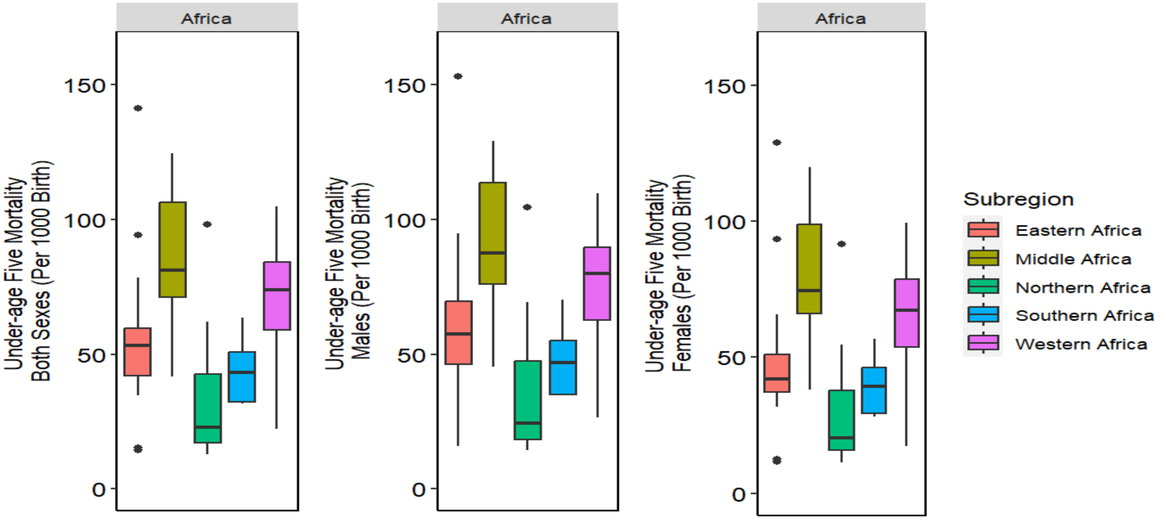
Region	Q1	Median	Q3	IQR
Africa	62.24	65.85	69.69	7.46
Americas	75.22	77.90	79.66	4.44
Asia	72.19	75.76	78.47	6.28
Europe	77.22	81.51	82.56	5.34
Oceania	74.44	75.32	77.53	3.09

Table 2: Summary statistics by region for life expectancy at birth of both sexes

Region	Q1	Median	Q3	IQR
Africa	41.87	59.16	79.82	37.94
Americas	9.07	13.43	18.04	8.97
Asia	9.28	18.62	31.60	22.32
Europe	3.74	4.32	6.05	2.31
Oceania	12.37	14.42	25.50	13.13

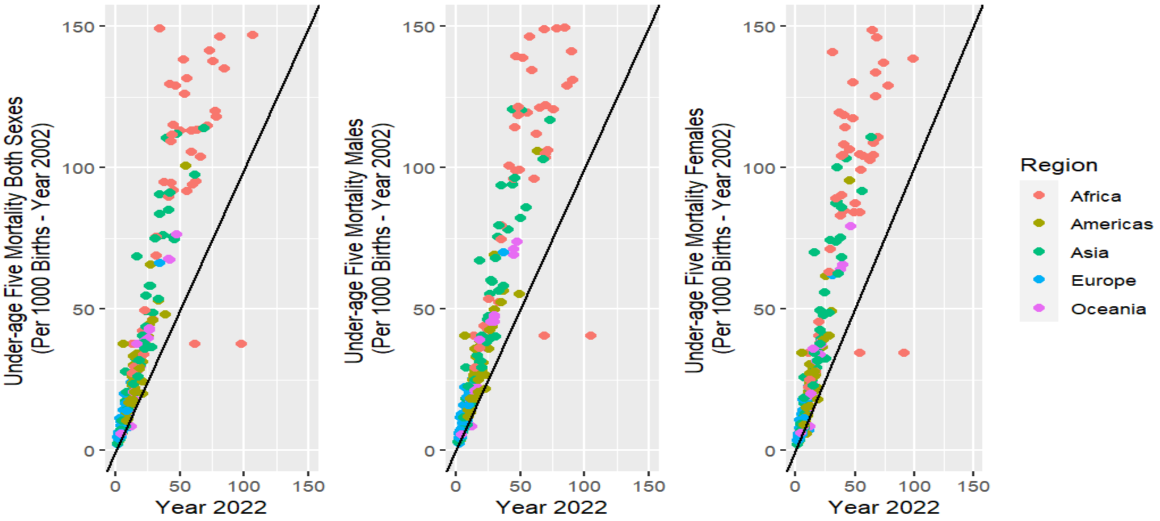
Table 3: Summary statistics by region for under-age five mortality of both sexes

B Additional Figures



(a) Under-age five mortality for both sexes (b) Under-age five mortality for males (c) Under-age five mortality for females

Figure 6: Variability of under-age five mortality for both sexes, males and females within and between subregions of African region



(a) Under-age five mortality for both sexes (b) Under-age five mortality for males (c) Under-age five mortality for females

Figure 7: Comparison of the changes in under-age five mortality values over 20 years for both sexes, males and females