DAT565/DIT407 Assignment 1

Saif Sayed
gussayedfa@student.gu.se

Gona Ibrahim Abdulrahman
gusibrgo@student.gu.se

2024-03-27

This is a report for assignment 1 for the course *Introduction to Data Science & AI* from Chalmers and Gothenburg University.

Problem : Dependency ratio

1 Dependency Ratio (1860 to 2022)

In Task 1, we were supposed to read the "swedish population by year and sex 1860-2022.csv" file from SCB [1] and plot a figure showing the dependency ratio of Sweden from 1860 to 2022.

To accomplish this task, we utilized the pandas library to read and store the .csv file. We examined and processed the data, replacing invalid values with the correct ones. For instance:

$$df[\text{"}age\text{"}] = df[\text{"}age\text{"}].replace(\text{"}110+\text{"}, 110).astype(float)$$

Next, we categorized the data into three groups: (i) Children (0-14), (ii) Labor force (15-64), and (iii) Elderly (65+). After grouping the data, we used the 'transpose()' method to interchange the rows and columns, facilitating easier data access and manipulation. The resulting dataset is shown in Figure 1.

Finally, we calculated the dependency ratio for each year and added the resulting values to a new column named 'Ratio'. The dependency ratio was calculated using the following formula:

$$Dependency\_ratio = 100 * (children + elderly)/laborforce$$

Using the 'Ratio' column and the corresponding years, we generated a figure illustrating the dependency ratio from 1860 to 2022 (refer to Figure 2).

Our source code can be found in Appendix A of this document.

2 Fraction of Children, Elderly and Dependant population (1860-2022)

In Task 2, we were supposed to plot another figure illustrating the fractions of children, elderly individuals, and the total dependent population in relation to the overall Swedish population from 1860 to 2022.

To accomplish this task, we added additional columns to the dataset and calculated the respective fractions for children, elderly individuals, and the total dependent population of the total Swedish population from 1860-2022. The resulting dataset is shown in Figure 1.

Similarly to Task 1, we plotted a figure showcasing the three different fractions over the period spanning from 1860 to 2022 (refer to Figure 3).

Our source code can be found in Appendix A of this document.

age	(-1, 14]	(14, 64]	(64, 110]	Ratio	Total	\
1860	1292962	2365121	201645	63.193680	3.859791e+06	
1861	1318714	2385678	208106	63.999417	3.912562e+06	
1862	1338720	2405974	212807	64.486441	3.957565e+06	
1863	1362024	2431967	217661	64.955034	4.011717e+06	
1864	1380235	2460997	219574	65.006540	4.060871e+06	
...	...	...	...	...	...	
2018	1819729	6374745	2035711	60.479909	1.023025e+07	
2019	1834821	6427401	2065367	60.680639	1.032765e+07	
2020	1837798	6453411	2088086	60.834247	1.037936e+07	
2021	1839103	6494457	2118766	60.942262	1.045239e+07	
2022	1829093	6545326	2147137	60.749151	1.052162e+07	

age	Children Fraction	Elderly Fraction	Dependent Fraction
1860	0.334982	0.052242	0.387225
1861	0.337046	0.053189	0.390235
1862	0.338269	0.053772	0.392041
1863	0.339511	0.054256	0.393768
1864	0.339886	0.054071	0.393957
...	...	...	...
2018	0.177877	0.198989	0.376867
2019	0.177661	0.199984	0.377645
2020	0.177063	0.201177	0.378240
2021	0.175951	0.202706	0.378657
2022	0.173841	0.204069	0.377911

[163 rows x 8 columns]

Figure 1: Resulting Table with all the columns

3 Discussion

Understanding the Dependency Ratio

The dependency ratio is a crucial metric that reflects the balance between the economically active (independent) population and those who rely on their support (dependents). It's calculated as the ratio of dependents (children and elderly) to the working-age population (independent) per 100 people.

Understanding the Figures:

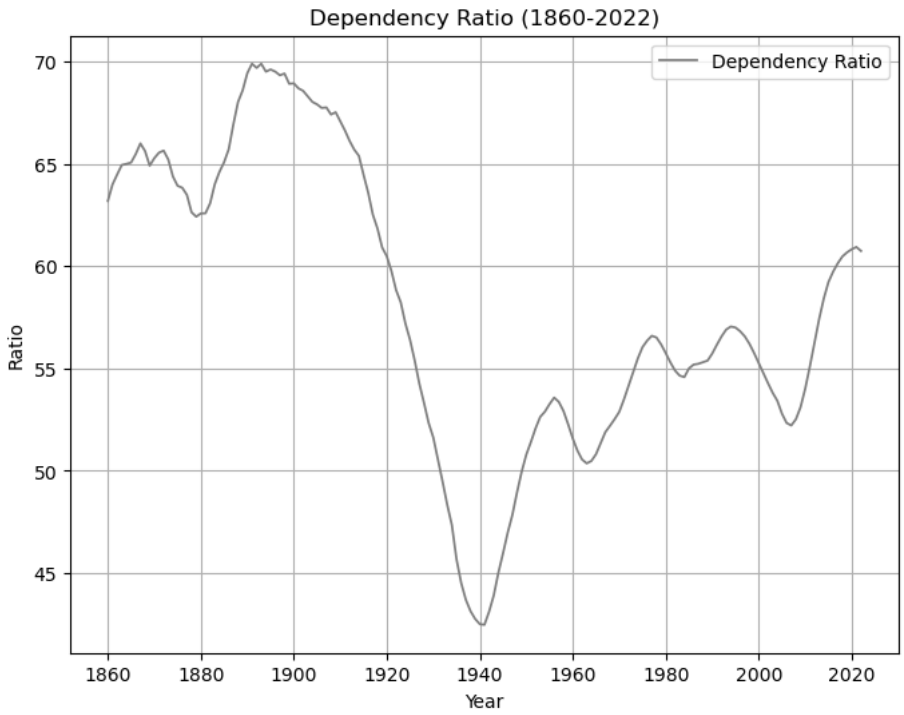

Figure 2: Dependency ratio in Sweden(1860-2022).

1860-1880: High Dependency Ratio

Poverty, high birth rates, and reliance on children for agricultural labor contributed to Sweden's high dependency ratio. The graph's initial peak aligns with this period of economic challenges and shorter life expectancy

1885-1940: Decreasing Dependency Ratio

Immigration: Sweden's significant immigration boosted the working-age population.

Labor Force Demand: Industrialization required more workers, further reducing the dependency ratio.

Healthcare Impact: Advances in healthcare led to longer life expectancy, especially for the elderly.

1940-1955: Rising Dependency Again

Healthcare Advances: Longer life expectancy due to better healthcare led to an aging population.

Societal Transition: Sweden shifted from an agrarian to an industrialized society

1960 Onward: Persistent Dependency Challenges

Social Welfare System: Sweden's robust support system aids the elderly and families and lower ages contribute to an aging population.

Changing Family Dynamics: Career-oriented women and lower birth rates play a role

2010-2022: Escalating Dependency

Dependency Ratio Rises: Population growth, improved healthcare, and retirement milestones contribute.

Elderly Impact: Longer life expectancy leads to an increasing proportion of elderly individuals.

Retirement Milestones: More people reached retirement age.

Workforce Planning: Balancing the workforce became critical.

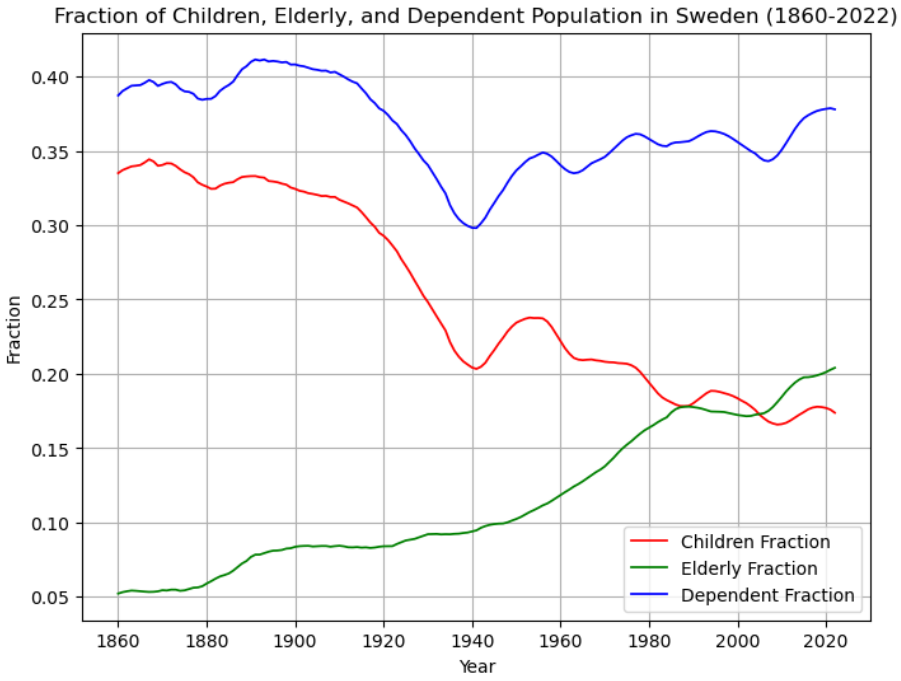

Figure 3: Fraction of Children, Elderly and Total Dependant Population.

Children Fraction (Red Line): Between 1860 and 1920, Sweden had a high fraction of children due to agricultural labor needs and birth rates. Later, this stabilized as societal norms shifted.

Elderly Fraction (Green Line): Over time, the elderly fraction consistently rises due to improved healthcare and longer life expectancy, reflecting an aging society.

Dependent Population Fraction (Blue Line):
1860-1880: High dependency mainly due to higher fractions of children.
1885-1940: Dependency decreases as immigration boosts the working-age population.
1940-1955: Dependency rises again due to better elderly care and societal transitions.
1960 Onward: Persistent dependency challenges, influenced by social welfare and retirement age.
2010-2022: Escalating dependency as the population grows and more people reach retirement age.

The Context
Pre-Industrial Era (1860-1880): High birth rates and shorter life expectancy and families relied on children for farm work.
Industrialization (1885-1940): Immigration and labor force demand reduced dependency and improved healthcare extended life expectancy.
Post-World War II (1940-1955): Elderly care improved, leading to an aging population and transition from agriculture to industry.
Social Welfare Era (1960 Onward): Robust social welfare system, lower retirement ages and career-oriented women and lower birth rates.
Recent Decades (2010-2022): Population still growth,longer life expectancy for the elderly and workforce planning challenges.

Summary

Sweden's demographic landscape mirrors broader global trends. People are living longer, resulting in an aging population. Simultaneously, family sizes are decreasing due to various factors. These changes are influenced by social dynamics, economic conditions, and advances in healthcare

References

[1] Statistiska centralbyrån. *Folkmängden efter ålder och kön. År 1860 - 2022.* Retrieved 2023-10-20. 2023. URL: https://www.statistikdatabasen.scb.se/pxweb/sv/ssd/START__BE__BE0101__BE0101A/BefolkningR1860N/.

A Python code

This is the code we used to read and plot the data of the Swedish population.

```
1 import pandas as pd
2 import numpy as np
3 import matplotlib.pyplot as plt
4
5 # Read the CSV file
6 df = pd.read_csv('
      swedish_population_by_year_and_sex_1860-2022.csv')
7
8 # Convert age to numeric (replace '110+' with 110)
```

```
df["age"] = df["age"].replace("110+", 110).astype(float)

# Group by age classes: children(0-14), labor force (15-64), elderly(65+)
classes = df.groupby(pd.cut(df["age"], [-1, 14, 64, 110])).sum(numeric_only=True)

# We no longer need the 'age' column
classes = classes.drop(columns=['age'])

# Switch the row elements with column elements and vice versa
classes = classes.transpose()

# Calculates the dependency ratio and adds the resulting value to a new column 'Ratio' for each year
classes["Ratio"] = 100 * ((classes.iloc[:, 0] + classes.iloc[:, 2] )/ classes.iloc[:, 1])

# Stores the years from 1860-2022 in a numpy array
years = np.asarray(classes.index.values, float)

# Plots the figure to show the dependency ratio in Sweden from 1860-2022
fig1, ax1 = plt.subplots(figsize=(8, 6))
ax1.set_title("Dependency_Ratio_(1860-2022)")
ax1.set_xlabel('Year')
ax1.set_ylabel('Ratio')
ax1.plot(years, classes['Ratio'], linestyle="-", linewidth=1.25, color="gray", label='Dependency_Ratio')
ax1.grid(True)
ax1.legend()


# Calculates the fraction of children, elderly, and total dependent population from 1860-2022 and adds the resulting values to their respective new columns
classes["Total"] = classes.sum(axis=1)
classes["ChildrenFraction"] = classes.iloc[:, 0] / classes["Total"]
classes["ElderlyFraction"] = classes.iloc[:, 2] / classes["Total"]
classes["DependentFraction"] = (classes.iloc[:, 0] + classes.iloc[:, 2]) / classes["Total"]

# Plots the figure to show the fractions(1860-2022)
```

```
43 fig2, ax2 = plt.subplots(figsize=(8, 6))
44 ax2.plot(years, classes["ChildrenFraction"], linestyle
      = "-", linewidth=1.25, color="r", label="
      Children_Fraction")
45 ax2.plot(years, classes["ElderlyFraction"], linestyle=
       "-", linewidth=1.25, color="g", label="
      Elderly_Fraction")
46 ax2.plot(years, classes["DependentFraction"],
      linestyle= "-", linewidth=1.25, color="b", label="
      Dependent_Fraction")
47 ax2.set_title("Fraction_of_Children,_Elderly,
      _and_Dependent_Population_in_Sweden_(1860-2022)")
48 ax2.set_xlabel("Year")
49 ax2.set_ylabel("Fraction")
50 ax2.legend()
51 ax2.grid(True)
```