Checkpoint 1 Report - Group 6

Topic: Comprehensive population and demographic trends simulation

Group members:

Jiaqi Yan, Wenxi Zhang, Yuehan Zhang, Yushuhong Lin, Zeyu Wang, Zicong He.

Git Repository Link: https://github.gatech.edu/yzhang4046/CSE6730

Abstract

Reliable population data is essential for tracking progress, designing policies, and implementing programs effectively. This project focuses on developing a comprehensive population simulation system that models demographic trends, including multiple impacts. The goal is to construct a model incorporating key demographic factors such as fertility, mortality, etc., which could potentially allow for the exploration of policy interventions such as education investments and child support programs. By simulating these dynamic variables over long-term horizons, we also aim to provide valuable insights into how population structures evolve under different conditions and policies.

1. System Description

The population dynamics simulation models demographic change at both individual and population levels. The system employs agent-based modeling techniques to capture emergent population behaviors that arise from individual decisions and life events.

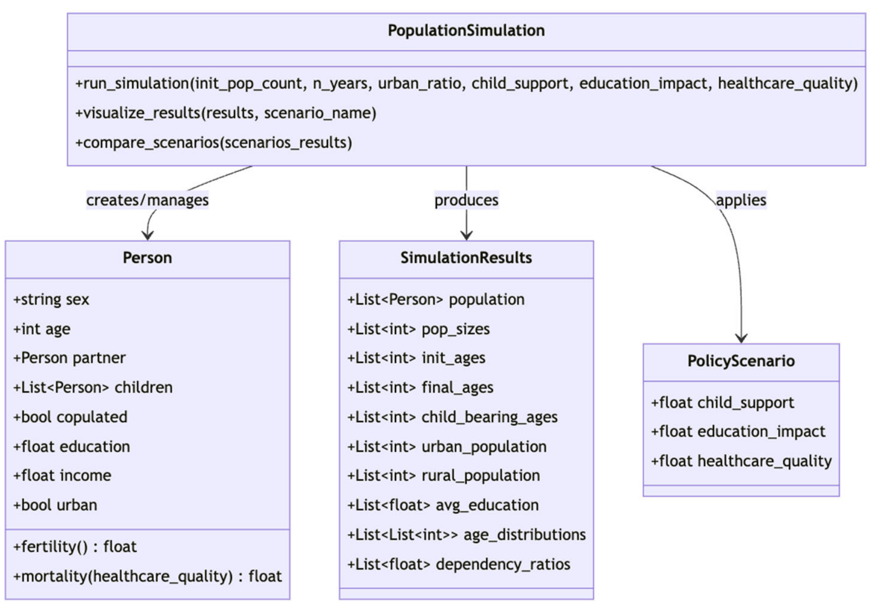

Fig.1. Class diagram of population simulation system

The Person class represents individuals as the fundamental units in our population dynamics simulation. Each person possesses core demographic attributes (sex, age, partner relationships, children) and socioeconomic characteristics (education and income levels, urban/rural residence). The class implements two critical biological functions: fertility calculates reproductive capacity based on age and sex, while mortality determines death probability using age, healthcare quality, and socioeconomic factors. At the population level, the simulation models partnership formation based on age compatibility and gender, reproduction decisions influenced by multiple factors such as partner fertility, existing children, education level, residence type, income, and government policies, and mortality following a "bathtub curve" pattern. As the simulation progresses, individuals age annually and experience socioeconomic changes, particularly income fluctuations during working years, reflecting realistic life trajectories.

2. Conceptual Model

2.1 Conceptual model of population simulation

The simulated workflow follows these steps:

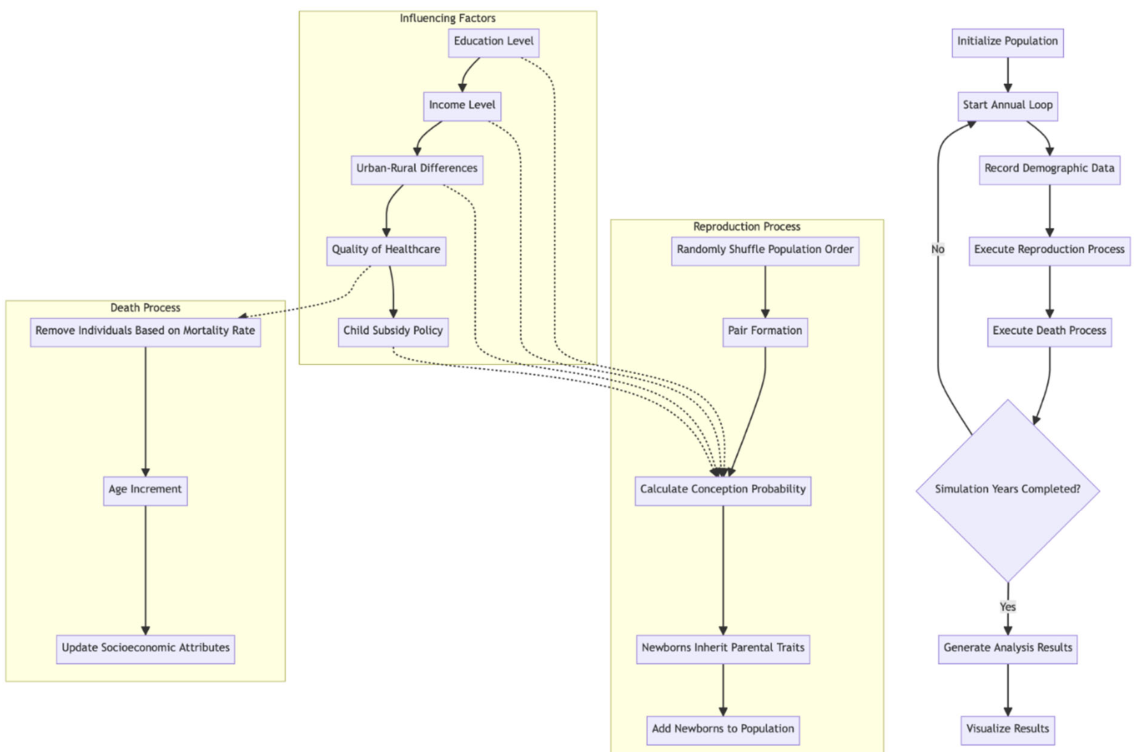

Fig.2. Population simulation workflow diagram

The simulation begins with an initialization phase, creating an initial population with a specified urban/rural ratio. The age distribution of the initial population is based on a normal distribution, ensuring a reasonable population structure at the start of the simulation. During this phase, the system also sets starting attributes for all individuals, including education levels, income levels, and other characteristics.

The core of the simulation is the annual cycle process. At the beginning, the system records current population statistics, including total population, urban-rural distribution, age structure, and dependency ratios as key indicators. Subsequently, the system executes partnership formation and reproduction processes, generating new individuals who join the population. The system calculates the death probability for each individual and removes those who "die." At the end, all surviving individuals age one year, and socioeconomic attributes are updated according to life stages.

After completing all simulation years, the system enters the results generation phase. The system compiles population statistics across the entire simulation timeline, revealing long-term trends and patterns. To facilitate interpretation, the system generates multiple visualizations for key indicators, such as population pyramids, age distribution changes, and dependency ratio trends. Finally, the system compares results across different policy scenarios, highlighting the relative effectiveness of various interventions.

2.2 Conceptual model of policy parameters and impacts

The simulation incorporates three key adjustable policy parameters (child_support, education_impact, healthcare_quality) used to create and evaluate different population policy scenarios:

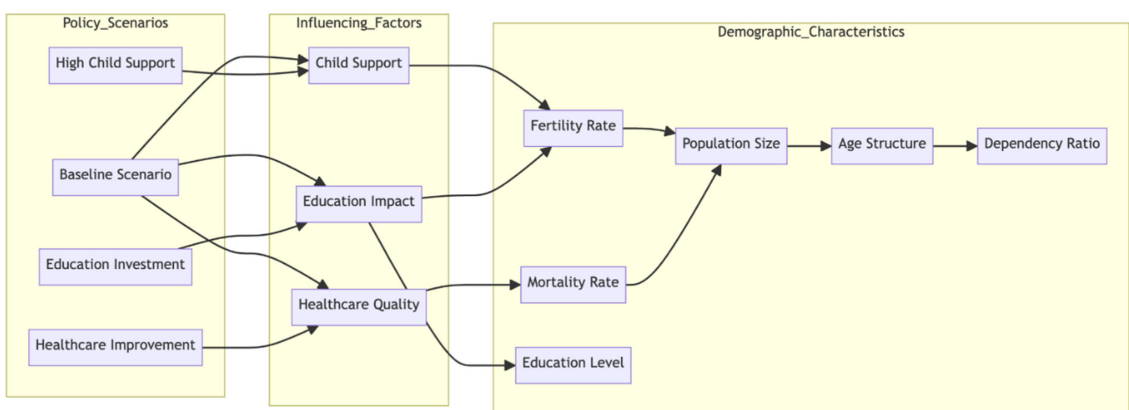

Fig.3. Policy parameters and impacts diagram

The child support parameter simulates government financial incentives for childbearing, increasing conception probability particularly for families with existing children. The education impact parameter represents education's influence on reproductive choices, with higher values indicating stronger fertility reduction effects. The healthcare quality parameter reflects healthcare system effectiveness, with improvements reducing mortality rates across all age groups, especially among vulnerable populations. These parameters interact through individual behaviors to produce complex population-level outcomes. For instance, better healthcare reduces deaths but may increase elderly dependency ratios, while education improvements might decrease short-term fertility but enhance long-term population quality and productivity. These non-linear interactions highlight the value of comprehensive simulation approaches for understanding population dynamics.

3. Platform(s) of Development

The simulation system is implemented using Python, using several libraries listed below:

NumPy and pandas: For efficient numerical operations and data manipulation
Matplotlib and Seaborn: For creating visualizations and plots
tqdm: For progress tracking during simulation execution

The approach follows a modular design, ensuring separate class definitions for individuals and simulation components. Each function is distinctly structured to handle simulation, visualization, and scenario comparison independently. Parameterization allows for flexibility, enabling configurable initial conditions such as population size, simulation years, and urban ratio. Additionally, policy parameters can be adjusted to facilitate scenario testing. For data visualization, a multi-faceted approach is employed using Mermaid, providing comparative visualization capabilities to evaluate different policy scenarios effectively.

4. Literature Review

Currently, the world faces challenges such as rapid urbanization, population aging, and large-scale regional migration. These phenomena have far-reaching impacts on socioeconomic conditions, public resource allocation, and policymaking. Reliable population data is essential for tracking progress, designing policies, and implementing programs effectively. However, in low- and middle-income countries (LMICs), data collection is often inconsistent due to resource limitations, political manipulation, or lack of prioritization. [1] Over-reliance on estimates instead of empirical data weakens policy decisions, highlighting the need for accurate and country-specific population data. Therefore, constructing a detailed and comprehensive simulation system is of paramount importance.

In previous population simulation studies, researchers have primarily relied on several classical methods to describe and forecast population dynamics. First, classical population growth models are used. For example, the Malthusian model, based on Thomas Malthus's theory of population growth proposed in 1798, posits that in the absence of resource constraints, the population grows exponentially [2]. In contrast, the logistic model, introduced by Pierre-François Verhulst in 1838, describes how population growth eventually saturates when environmental resources and carrying capacity are considered [3]. Additionally, to capture the nuances of population structure changes, researchers have widely employed the Leslie matrix model, which focuses on the survival and fertility rates of different age groups to predict future age distributions and structural evolution [4]. There have been models proposed earlier that Population pharmacodynamic (PD) models describe how drug effects change over time by linking drug exposure to response, offering a deeper understanding than single assessments. These models enable the simulation of different dosing regimens, aiding in dose optimization and study design. As the third paper in a series, this work introduces methods for developing and evaluating population PD models.[5]

Meanwhile, to provide reliable initial conditions and boundary constraints for macro-level system dynamics models or agent-based models, scholars have utilized synthetic population generation techniques. By integrating census data with sample survey data, methods such as Iterative Proportional Fitting (IPF/IPU) and genetic algorithms are used to generate virtual populations with detailed household and individual attributes [6]. Our research plan will build upon these established methods by combining them with agent-based simulation techniques. Through calibration with historical data, sensitivity analysis, and scenario simulations, we aim to further explore the comprehensive impact of policy, economic, and social factors on population change and validate the practicality and predictive accuracy of our model.

First, we identify the key population dynamics to be simulated: natural population growth (birth and death rates), migration (both internal and external), changes in family structure (marriage, divorce, and remarriage), and shifts in age distribution. Policy, economic, and social factors directly affect these dynamics. For example, favorable economic conditions and high-quality healthcare and education services can increase birth rates and reduce death rates, while strict immigration policies can limit migration. These factors collectively determine the overall trend of population development.

We will utilize detailed data provided by the Data USA [7] and the U.S. Census Bureau [8]. These datasets cover population statistics, age distributions, family structures, and regional mobility, and will serve as the initial conditions and boundary constraints for our model.

In terms of methodology, we will construct a multi-layer hybrid simulation system. The macro-model module employs system dynamics methods to establish overall population growth trends based on historical data. The micro-model module is built on an agent-based modeling approach, where each agent has clearly defined attributes and behavioral rules to simulate processes such as marriage, childbirth, migration, and death. Synthetic population generation techniques—using methods such as IPF/IPU algorithms and genetic algorithms—will generate virtual population data with detailed household and individual attributes. The parameters produced by the macro module will guide the micro-model, while the statistical results from the micro-model feed back into the macro module for parameter calibration, ensuring the overall system's accuracy and stability.

Reference:

[1] J. P. Ioannidis, "The mass production of redundant, misleading, and conflicted systematic reviews and meta-analyses," International Journal of Epidemiology, vol. 46, no. 1, pp. 4–11, 2017. [Online]. Available:https://doi.org/10.1093/ije/dyw328.

[2] Malthus, T.R. (1798). An Essay on the Principle of Population. London: J. Johnson.

[3] Verhulst, P.-F. (1838). Recherches mathématiques sur la loi d'accroissement de la population. Correspondance Mathématique et Physique, 10, 113-121.

[4] Leslie, P.H. (1945). On the use of matrices in certain population mathematics. Biometrika, 33, 183-212.

[5] P. D. M. Maciejewski, B. Wang, J. M. Mockenhaupt, and G. S. H. Yeo, "Title of the Article," CPT: Pharmacometrics & Systems Pharmacology, vol. X, no. Y, pp. Z, 2013. [Online]. Available:https://doi.org/10.1038/psp.2013.71.

[6] Beckman, R.J., Baggerly, K.A.,&McKay, M.D. (1996). Creating Synthetic Baseline Populations. Transportation Research Part A: Policy and Practice, 30(6), 415-429.

[7] Data USA. Retrieved from https://datausa.io/profile/geo/atlanta-ga

[8] U.S. Census Bureau. Atlanta Population Data. Retrieved from https://data.census.gov/all?g=160XX00US1304000