Analyzing World Population Growth

Lina Huang, SEAS '16 Systems Engineering

Abstract

Background: World overpopulation due to societal advances has always been a worry due to its implications for political policies, climate change, and natural resources. While historical models have worried about overpopulation, the 7th billion mark took only an additional 13 years after the 6th billion, indicating the first decline in marginal growth. In predicting future growth, most population models lack agreement or consistency, making population an interesting potential for data analysis and understanding.

Objective: While this study is not lofty enough to claim to have the answers to population predictions in 2100, it seeks to add to current studies which are rather static and mathematical. However, population is an every changing, complex system affected by various factors like economics, technology, and life expectancy. This study seeks to take a different approach by taking these factors into consideration and also recognizing how dynamic the equations used for predicting population should be.

Method and Results: Data was analyzed using regression analysis in JMP and basic data science libraries in Python like numpy, scipy, and skit-learn. The results disagrees with historical models like the Dooms Day model and offers less variability than predictions claimed by the UN.

Background

- **Past Historical Models:**

1. Deterministic Population Model

$$\frac{dN}{dt} = BN - MN = (B - M)N = PN,$$

B: Birth Rate
M: Mortality Rate
P: Production Rate
N: Population total

2. Malthusian Exponential Model: Assumes P, production rate, is a

$$N = N_0 e^{Pt},$$

3. Verhulst or Logistic Model: Assumes only B, birth rate, is a constant

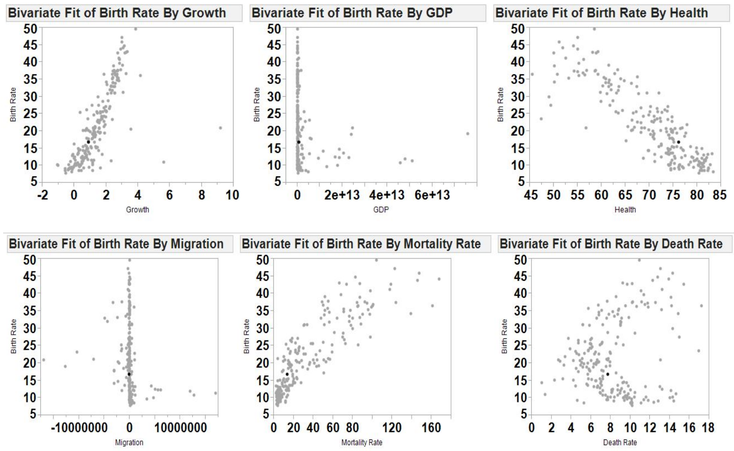
$$\frac{dN}{dt} = (B - mN)N.$$

4. Von Foerster Doomsday Model: Assumes P is an increasing function of N

$$\frac{dN}{dt} = P_0 N^{1+1/k}.$$

Results

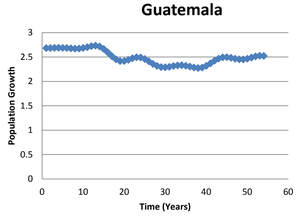
Step-Wise Analysis Against Birth Rate in JMP



Lock	Entered	Parameter	Estimate	nDF	SS	"F Ratio"	"Prob>F"
☑	☑	Intercept	76.2192367	1	0	0.000	1
☐	☑	GDP	-6.396e-14	1	54.77531	3.569	0.06031
☐	☑	Health	-0.7110939	1	1174.312	76.505	8.4e-16
☐	☐	Rural	0	1	35.15068	2.305	0.13054
☐	☐	Gender	0	1	22.20459	1.450	0.22997
☐	☑	Migration	3.01397e-7	1	113.4933	7.394	0.00711
☐	☑	Mortality Rate	0.13663254	1	769.8422	50.154	2.3e-11
☐	☑	Death Rate	-1.1210885	1	1674.594	109.097	1e-20
☐	☐	Urban Population %	0	1	32.48881	2.128	0.14615
☐	☐	Unemployment	0	1	25.50719	1.667	0.19811

Python Growth Regression Results

Ex of Pop Growth Plot 1960-2013



Prediction:
10.4 Billion by 2050

81/204 countries display R^2 significant population growth plots that can be fitted to linear regression

2040-2050 Predictions

2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050
9.36E+09	9.45E+09	9.55E+09	9.65E+09	9.75E+09	9.85E+09	9.95E+09	1.01E+10	1.02E+10	1.03E+10	**1.04E+10**

Methods

- **Linear Regression Assumptions in JMP**
  1. **Relationship between the independent and dependent variables is linear**
  2. **Independent variables are multivariate normal**
  3. **Little or no multicollinearity among independent variables**
  4. **Residuals are independent of each other**
  5. **Data has homoscedasticity**

Analyzed population factors correlated with data on 210 countries across 40+ years:

GDP	%Females in Work Force	Unemployment Rate	Net Migration
Life Expectancy	%Rural Population	%Urban Population	Total Population
Birth Rate	Mortality Rate	Death Rate	Population Growth

- **Regression in Python Using Numpy, SciPy, and Sklearn libraries**
  - **Population growth tends to match linear regression well**
  - **Used basic machine learning libraries mentioned above and all past population growth data to predict populations until 2050**

Implications

- **Conclusions**
- **While the DoomsDay Model is correct in the general increasing trend of production rates, population growth is consistently decreasing for most countries in a linear fashion**
- **Economic factors do not seem to affect the birth rate model currently, but this could be due to missing factors that were not considered.**

- **Proposed New Approach for future research:**
- **Survey Individuals in different countries about factors they consider when having children**
- **Take those factors into consideration when creating a regression model for birth rate to search for said missing factors**
- **Incorporate Machine Learning and Agent-Based Modeling to predict population growth as tools that enhance pure, deterministic, sensitivity-tested statistics**

Funding: Class of 1971 Robert J. Holtz Endowed Fund Faculty Mentor: Professor Barry Silverman, ESE