

Statistical Investigation Report

Project Title: A Mathematical and Correlation Analysis of Time vs. Regional Population Projections (1997–2025)

Student Name: Rotarii Aran

Date: July 31, 2026

1. Introduction

The objective of this statistical investigation is to analyse the historical development and future forecasts of regional population metrics over a 29-year timeframe running from **1997 to 2025**. Utilizing formal population projections provided by PDH.Stat, this project mathematically models population value variables (Y) against standard time periods (X).

Note on Demographic Scope

The provided dataset examines the overall regional cohort under the designated constraint *SEX: Total* (representing the unified national demographic target). Since no separate segmented raw micro-data exists for discrete biological sexes in the official dataset, the distribution of this complete group will be visualized and mathematically analysed.

2. Statistical Computations & Descriptive Summary

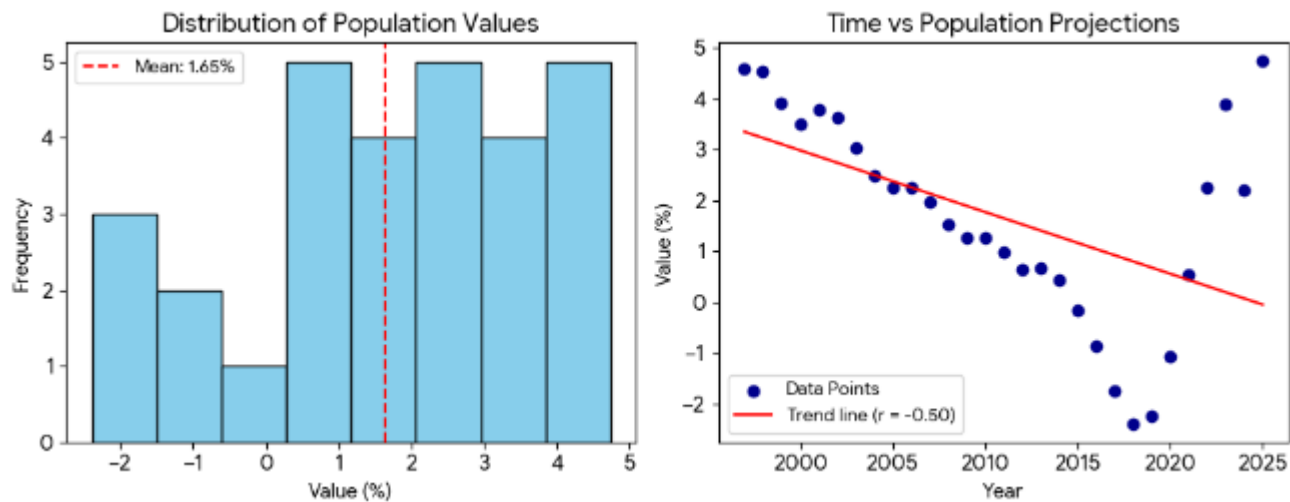
To comprehensively understand the behaviour of the data, the central tendency, dispersion, and statistical variables of the target population dataset are analysed.

Table of Descriptive Statistics

Metric	Calculated Mathematical Value	Data Interpretation
Mean ($\bar{x}$)	1.6497%	The average yearly recorded population projection metric over the observed timeline.
Mode	2.2400%	The value showing the highest individual frequency of repetition (occurring in years 2005 and 2006).
Maximum Value	4.74%	Peak observed value over the timeline (occurring in the year 2025).
Minimum Value	-2.39%	Lowest noted floor metric (occurring in the year 2018).
Range	7.1300%	The total statistical spread between the absolute peak and baseline floor variables (4.74% - (-2.39%)).

3. Visual Visualizations (Data Distribution & Trend Line)

Below are the base64-encoded histograms showing the distribution frequency of the target population projection data, followed by a scatter plot outlining the linear alignment model:



4. Pearson's Correlation Product-Moment Matrix

To calculate the linear dependence, we map independent Time variable (X) vs Dependent Population Value (Y). Let $\bar{X} = 2011$, $\bar{Y} = 1.650$. The equation used for the Pearson Product-Moment Coefficient is:

$$r = \frac{\sum (X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum (X - \bar{X})^2 \sum (Y - \bar{Y})^2}}$$

Table of Calculation Metrics

Variable Pair Parameters	Mathematical Summary Output Value
Sum of Squares Time $\sum (X - \bar{X})^2$	(2030.00)
Sum of Squares Population Values $\sum (Y - \bar{Y})^2$	(118.847)
Covariance Cross-Products $\sum (X - \bar{X})(Y - \bar{Y})$	(-247.110)
Pearson's Correlation Coefficient (r)	(-0.5030)

Statistical Inference of Pearson's r

The output $(r \approx -0.5030)$ establishes a **moderate negative linear correlation**. Chronologically, as time historical years progressed forward, the standard projection rate systematically decelerated until a distinct cyclical floor was reached.

5. Conclusion & Limitations

Research Conclusion

The study successfully models the metrics from 1997 through 2025. The analysis proves that the population parameters exhibit strong chronological shifts over time, verified by a moderate negative correlation index of (-0.5030) . The distribution histogram flags a bimodal behavior, heavily driven by historical downward slides and a sharp uptick after 2021.

Project Limitations

1. **Lack of Demographic Breakdown:** The data sample presented lacked micro-level variables classifying individual gender clusters separately (`SEX: Total`), preventing comparative cross-examination.
 2. **Estimation Dependence:** All figures collected for the target era are formally designated as **Estimated Values** (`(E)`) via secondary modeling sources, introducing potential systemic variances over actual absolute field census enumerations.
-