Population Projections Methodology (2023 Baseline)

California State Department of Finance
Demographic Research Unit

Release notes

Vintage 2024 (2024.09.23): initial public data release

2023 Baseline Projections/2024 Vintage

The Demographic Research Unit within the California Department of Finance (DOF) is responsible for the production and release of the population projections for the state and counties of California[1]. The 2023 Baseline (Vintage 2024) projections incorporate the latest historical population, birth, death, and migration data available as of July 2023 and are informed by 2020 Census data products.

The population projections are produced using a cohort-component model. For each projection year, the model updates the population by applying age-specific fertility, mortality, and migration rates, known as the components of change, to the existing population. These are calculated using the basic demographic equation [1]:

Current Population = Previous Population + (Birth-Deaths) + Net Migration

The cohort-component method tracks people born in a specific year throughout their lives, with each cohort evolving as assumptions about the components of change are applied. New cohorts are formed by applying the forecast fertility rates to women 15-to-45 years of age and adding the resulting births as a new population starting at age zero. The population events are simulated using Stata in the following sequence: births, deaths, in-migration, and out-migration [2]. Afterward, all individuals, except newborns and special populations, are aged by one year to reflect the aging process. Special populations[2] are adjusted to align with the DOF's estimated changes at the county level. The group quarters (GQ) population is then combined with the household population to establish the starting total population for the next year's projection.

---

[1] The population projections were prepared under the mandate of California Government Code (Cal. Gov't Code § 13073, 13073.5). It is state policy that all state plans make use of the ". . . population projections and demographic data that is provided by the State's Demographic Research Unit" (Cal. State Admin. Manual § 1100).

[2] As described in section Special Populations.

Data and Methods

1. Base Population

The basis of the DOF 2024 population projection series is the 2020 Census April 1, 2020 population count of 39,538,223 and informed by the Population Estimates Program for age, sex, and race/ethnicity data for California. DOF produced a projections base file, to allocate "Other" race from Census results to comply with U.S. Office of Management and Budget (OMB) race and ethnicity standards. This April 1, 2020 dataset is aged forward to July 1, 2020, using the model described below, which serves as the initial estimated population of 39,541,722.

2. Fertility and Mortality

Records of births and deaths, including selected demographic characteristics, for 1990 to 2023 are obtained under an agreement with the California Department of Public Health. Data were evaluated for quality. Missing values in birth and death records were filled in using hot deck imputation where the missing value was pulled from a randomly selected similar record. Mortality and fertility rates are calculated by the number of events (deaths or births) during the year divided by the population at the midpoint of the year.

2.1. Fertility

Using historical birth data since 1990, fertility rates are calculated for females by age in each county by fitting the following spatial Poisson model:

$$\ln(\mu_{ijt}) = \ln(N_{ijt}) + (\beta_1 + \eta_j)T_t + BX_{it} + \rho_{geoi} + \eta_j + \eta_i + \varepsilon_{ijt} ,$$

Where the log fertility rate μ is modeled as a function of the log population (N_{ijt}) plus a fixed coefficient (β_1) on time (T_t) and a county-level random coefficient (η_j) on time, a vector of fixed effects B on county-year specific covariates X, a spatial residual ρ, a county and group-specific random intercepts η_i and η_j, and an error term ε. This model generates a set of coefficients that are used to predict future rates.

In projecting future births, there are two key considerations for each year of the projection period: (1) the age of the female population and (2) the fertility rate that will be applied to each cohort in a given year. The number of births after July 1, 2023, is determined by applying the projected fertility rate to the projected population for females by age. Children are assigned a slightly higher probability of male sex (105 males per 100 females).

During the projections review process, decisions were made to override the median projected rate in favor of higher or lower rates to achieve a county-level

target for the total fertility rate. In cases of very recent changes in fertility, the series used a decaying linear or exponential weight as a function of *N* years back from the most current year to give more, or less weight to the latest data.

2.2. Mortality

The Lee-Carter methodology is applied to produce the mortality forecast with the R[3] package "Demography version 2.0" [3], [4]. For counties with few events, the data is pooled with neighboring counties. Rates for ages above age 90 are generated by an extrapolation of the mortality rate above age 30 using a logistic model [5].

The Lee-Carter method, developed in 1992, is a popular approach for modeling and predicting mortality rates over time. It breaks down mortality into three components: an age-specific average, the sensitivity of each age group to changes, and a trend that captures mortality changes across all ages. The method uses mathematical tools like singular value decomposition to estimate these components and applies time series models to forecast future mortality rates [6].

3. Migration

Historical migration trends are based on the DOF estimated county net domestic and foreign migration totals from the July 1 components of change in the estimates series published in December 2023 [7]. Each net flow is then disaggregated into three different flows - domestic in and out and foreign in. The foreign out population is calculated as the residual between estimated foreign in-migration and net foreign migration. Each flow was then forecast to reach a flow consistent with historical migration. This analysis was done at the county level on a case-by-case basis. Domestic and foreign migration were expected to converge to a stable figure separately.

The age distribution of the migrant population is generated using the American Community Survey (ACS) Tables B07001 and B07401 for the years 2010-2015 to 2018-2022 [8]. Demographic characteristics of the migrant population are randomly selected from ACS Public Use Microdata Sample data for years 2015-2022 for each migration flow/age cohort and are added to the projections dataset [9].

3 version 4.3.1 (2023-06-16)

4. Special Populations

Special treatment is required for the population living in group quarters, including prisons, school dormitories, military barracks, residential hospitals, nursing homes, monasteries, and other group accommodations. Thus, this population is excluded from the simulation: the GQ population does not age, have children, move, or die. GQ characteristics are based on the 2020 Census and held constant through the forecast period.

Similarly, most university students living in households (rather than dorms) exhibit similar population dynamics to the GQ population. For this reason, a large portion of college students are treated as GQ instead of the regular household population.

Assumptions and Limitations

The projection models rely heavily on trends and relationships observed in the past. The projection is based on a model that has persistent below-replacement fertility consistent with the second demographic transition stage of low fertility [10], [11]. Furthermore, gains in life expectancy are expected during the projection period but the projections reflect a slowdown in improvements in mortality. We also assume that mortality observed during COVID-19 will not hold constant, and most of the increase in death counts result from an older age structure of the population.

These trends assume a consistent evolution, but societal changes and attitudes to demographic behaviors can happen at any time and lead to a different "end stage" [11]. The results assume no radical change in the economic, policy, or natural environments. The proposed projections assume economic stability throughout the forecast period, and that resource constraints (i.e. water, housing, and transportation capacity grow at a sufficient pace.) Changes in migration, education, housing, or transportation policy would have significant effects and are not considered here; likewise, there is a risk of unforeseen changes in technology and productivity.

The model is subject to several sources of bias in addition to those mentioned above: among them, bias from the use of discrete time scales and the reliability of the census data, including the ACS. These deficiencies could affect the age, race, ethnic, geographic, and gender distributions of the population.

Deliverables

The complete public use dataset (P-3) contains counts of the population for each California county for July 1 of every year from 2020 through 2070, by individual year of age (0-110+), sex, and race/ethnicity groups. The race/ethnicity groups presented in this report are the Non-Hispanic Races: White; Black of African

American; American Indian or Alaska Native; Asian; Native Hawaiian or Other Pacific Islander; Multiracial; and Total Hispanic ethnicity of any race.

Summarized data are published as P-1 series (statewide) and P-2 series (county) projections. DOF also publishes a document including the total county population and components of change. See the DOF website at dof.ca.gov/Forecasting/Demographics/Projections/.

Review

Preliminary projections are offered for review to several regional planning agencies in California. Their feedback is evaluated by DOF and adjustments are made for specific counties before final production with the latest available data.

An internal review was conducted within the DOF to ensure consistency with economic and other forecast assumptions.

Acknowledgments

Research design, data collection, analysis and interpretation, technical report, and dataset by Andrés Gallardo, Ethan Sharygin, and Walter Schwarm. Review: James Miller.

Suggested citation.

California Department of Finance. Demographic Research Unit. 2024. *State And County Population Projections 2020-2070* [computer file]. Sacramento: California Department of Finance. September 2024.

References

[1] S. H. Preston, P. Heuveline, and M. Guillot, *Demography: Measuring And Modeling Population Processes*. Oxford, UK: Publishers, Blackwell, 2001.

[2] StataCorp, 'Stata Statistical Software: Release 14', 2015, *StataPress, CollegeStation,TX*: 14.

[3] R Core Team, 'R: A Language and Environment for Statistical Computing', 2023, *R Foundation for Statistical Computing, Austria*: 4.3.1 (2023-06-16).

[4] R. J. Hyndman, H. Booth, L. Tickle, and J. Maindonald, 'demography: Forecasting Mortality, Fertility, Migration and Population Data', 2023. [Online]. Available: https://CRAN.R-project.org/package=demography

[5] V. Kannisto, J. Lauritsen, A. R. Thatcher, and J. W. Vaupel, 'Reductions in Mortality at Advanced Ages: Several Decades of Evidence from 27 Countries', *Popul Dev Rev*, vol. 20, no. 4, pp. 793–810, 1994, doi: 10.2307/2137662.

[6] R. D. Lee and L. R. Carter, 'Modeling and Forecasting U.S. Mortality', *J Am Stat Assoc*, vol. 87, no. 419, pp. 659–671, Sep. 1992, doi: 10.1080/01621459.1992.10475265.

[7] State of California Department of Finance, 'E-6. Population Estimates and Components of Change by County — July 1, 2020-2023', Sacramento, California, Dec. 2023.

[8] U.S. Census Bureau, 'American Community Survey 5-Year Estimates, Tables B07001 and B07401', Jun. 2024. [Online]. Available: https://data.census.gov/cedsci/

[9] S. Ruggles *et al.*, 'IPUMS USA: Version 15.0 [dataset]', Jun. 2024, *IPUMS, Minneapolis, MN*. doi: 10.3886/ICPSR36854.v1.

[10] R. Lesthaeghe, 'The second demographic transition: A concise overview of its development', *Proceedings of the National Academy of Sciences*, vol. 111, no. 51, pp. 18112–18115, 2014, doi: 10.1073/pnas.1420441111.

[11] B. Zaidi and S. P. Morgan, 'The second demographic transition theory: A review and appraisal', Jul. 31, 2017, *Annual Reviews Inc.* doi: 10.1146/annurev-soc-060116-053442.