How Accurate Was California's 2020 Census?

Technical Appendix

CONTENTS

Eric McGhee

Appendix A. Iterative Proportional Fitting

The Census Bureau has released Post-Enumeration Survey estimates separately for each state and nationally by race and ethnicity, gender, age, and rental status. To obtain estimates for some of these subgroups by state, we combine the national HTC and state-level total PES estimates with iterative proportional fitting (IPF).

IPF is a process of adjusting cell counts in one data set to match total counts in another (Lomax and Norman 2016). In this case, the cell counts come from the official census results released for redistricting, and consist of the intersection of race/ethnicity, age, and state of residence. IPF then adjusts these counts to match the total values for each of the variables individually, as provided by the PES, progressively adjusting them until there is no longer an improvement in the match. We then compare these new cell values to the original ones for our estimate of under- or overcount.

The results of IPF have been shown to match the maximum likelihood estimate and so are the values most consistent with the data. While it is always possible for the actual values for any state to be significantly different from the IPF estimates, the results are highly constrained to match the state and national totals so significant deviations in large jurisdictions are less likely. However, because cells with small numbers of people have a smaller effect on the aggregate totals that serve as the constraint, the IPF estimated values for such cells might deviate more from the true values. To address this possibility, we only report the results for large groups in the main text.

The Census Bureau does not report PES values for group quarters populations, and the census data released thus far do not identify group quarters numbers for each subgroup. To address this problem, we downweighted the cell values in the census data with group quarters estimates from the 2020 5-year American Community Survey (ACS) for each group. The reassembled state totals closely matched the non-group-quarters numbers reported by the PES.

We ran the IPF process with each of the national variables turned into dummies. But the census reported the national PES net coverage error for the entire count and separately for each subgroup, leaving us with only one of the two net coverage estimates we needed for each dummy variable. For example, the net coverage estimate for African Americans did not include a companion estimate for all US residents who are *not* African American. To identify the out group in each case, we had to solve for it using the net coverage estimates the Census Bureau did provide. This required knowing the share of the national non-group quarters population in each category of the dummy—for example, the non-group quarters population who is and is not African American. We used the 2020 5-year ACS group quarters estimates for this purpose as well.

Because the redistricting data do not contain detailed estimates by age (beyond distinguishing adults and children) or rental status, we cannot perform IPF for certain HTC groups like young men or renters.

The Public Policy Institute of California is dedicated to informing and improving public policy in California through independent, objective, nonpartisan research.