Anticipating the 2020 Census Data: A User Perspective

NCSL Census Data Update Meeting
December 3, 2020

Karin Mac Donald
Director
California Statewide Database
karin@statewidedatabase.org
Discussion Points:

• Data Quality Issues

• Disclosure Avoidance System (DAS)

• When can we finally get started?

• Other data
Why are we concerned about Data Quality?

• Data collection issues:
  • First online census
    • 6% of population lacks broadband access – in rural areas 25%
    • Concerns about unsecured computers, potential hacking, phishing, etc.
  • COVID
  • Various issues reported:
    • problems with data/assignments pushed to enumerators via app
    • Student populations on the move – duplications?
  • Fewer Non Response Follow Up (NRFU) visits

• Changing/abbreviated timelines
  • Significantly less time for data processing:
    • Planned data processing schedule pre-pandemic: 5 months; census bureau request due to COVID: 6 months; actual time allotted: < 3 months
  • Commerce Department Inspector General: “The streamlined data processing under the accelerated census plan poses a myriad of risks to accuracy and completeness.”
Disclosure Avoidance and Differential Privacy

• What is a Disclosure Avoidance System (DAS)?
  • Methodologies/Techniques applied to data to protect respondent’s privacy
  • Also referred to as Privacy Protection System

• What is Differential Privacy?
  • New methodology being implemented now
  • Part of the new Disclosure Avoidance System
  • The ‘noise’ that is injected into the collected data
  • ‘Noise’ makes data less accurate (to protect privacy)

• What is Postprocessing?
  • A statistical process that corrects counts by geography
  • It eliminates negative values

• Differential Privacy + Postprocessing = Disclosure Avoidance System
So what’s the problem? Or: How big is the problem?

- The new DAS is a work in progress
- Census released 4 datasets for evaluation and made changes after feedback was received from the research community
- Datasets took 2010 census data and applied new privacy measures
- Datasets are referred to as DP (Demonstration Product) 1 - 4
- DP 1 release: October 2019 (see Caliper Corporation’s maps on website)
  - Has bias that moves total population from urban to rural areas
  - Effect is less in urban areas but significant in rural areas.
  - Distributes minority populations from high concentration urban areas to low concentration rural areas
- DP 2 release: July 2020
  - Is better in total population distribution
  - Is still similar to release 1 in minority distribution
The DAS problem continued:

• DP 3 release: September 2020
  • Error in dataset: held zero population blocks invariant (plan was only to keep state population invariant); i.e. zero population blocks not available to distribute data to.
  • Made high concentration minority areas more minority and lower ones less so

• DP 4 release: 11/27/2020
  • Not much time to evaluate but appears to do better on total population
  • Increases minorities in high concentration minority areas; decreases minorities in lower minority population areas similar to release 3

• In all releases, the smaller the geography, the larger the changes in total population. This may have significant effects on redistricting of small jurisdictions

• The new DAS has certain biases that are being worked out: can they be fixed in time? How will we know?

• Census will not release another test dataset

• New DAS will be applied to the new PL94 and we will not have anything to compare the methodology to, i.e. we will not know whether there is a problem or how big of a problem there may be
Timelines, Schedules, Other data

• PL94 will be released between March 31 and July 31, 2020
• There are things that can be done BEFORE the data are released
  • Prisoner reallocation (collect and clean datasets etc.)
  • If you don’t have election and voter registration data already, start to collect it
  • Set up hearing schedules, website, draft informational materials, translations and accessibility tools
  • Figure out how to conduct hearings during COVID
  • Start hearings to collect public input
  • If ‘Communities of Interest’ are a criterion, start to collect them
  • If you have a commission: provide trainings, hire staff and contractors

• Utilize the Citizen Voting Age Population (CVAP) special tabulation
  • Will be released in Feb 2021
  • Preliminary Voting Rights Assessments
Take-Aways:

• The Census is not over!
• Stay informed and involved and ask questions!
• Support census staff and census operations by requesting they be given sufficient time to produce the best possible data.
• The delays in the census data release do not mean that you cannot get started with your redistricting activities:
  • Prepare other datasets: prisoner data, election returns, voter registration, etc.
  • Collect supplemental data (Communities of Interest) as needed
  • Utilize the CVAP dataset for preliminary voting rights assessment
Stay safe and Happy Redistricting!