Census Adjustment

Philip B. Stark
Department of Statistics
University of California, Berkeley
www.stat.berkeley.edu/~stark

This article appeared in *Encyclopedia of Social Science Research Methods* (2004). *v. 1,* 112–13. M. Lewis-Beck, A. Bryman, & T.F. Liao, eds., Sage Publications.

Reference list updated 6 April 2007.

The U.S. census tries to enumerate all residents of the United States, block by block, every 10 years. (A block is the smallest unit of census geography; the area of blocks varies with POPULATION density: There are about 7 million blocks in the United States.) State and substate counts matter for apportioning the House of Representatives, allocating federal funds, congressional redistricting, urban planning, and so forth. Counting the population is difficult, and two kinds of ERROR occur: *gross omissions* (GOs) and *erroneous enumerations* (EEs). A GO results from failing to count a person; an EE results from counting a person in error. Counting a person in the wrong block creates both a GO and an EE. Generally, GOs slightly exceed EEs, producing an undercount that is uneven demographically and geographically. In 1980, 1990, and 2000, the U.S. Census Bureau tried unsuccessfully to adjust census counts to reduce differential undercount using *dual-systems estimation* (DSE), a method based on CAPTURE-RECAPTURE. (Some other countries adjust their censuses using different methods.) For discussion, see the special issues of *Survey Methodology* (1992), *Journal of the American Statistical Association* (1993), *Statistical Science* (1994), and *Society* (2001). DSE involves the following:

- taking a random sample of blocks (in 2000, DSE sampled 25,000 blocks);
- trying to enumerate the residents of those blocks after census day, in the Post-Enumeration Survey or PES;
- trying to match PES records to census records, on the basis of data that often are incomplete or erroneous;
- estimating the undercount within demographic groups called *post strata* (in 2000, DSE used 448 post strata) by comparing capture-recapture estimates with census counts.

Being counted in the census is considered capture; being counted in the PES is considered recapture. The PES attempts to identify EEs and to account for movers. Ultimately, this yields an adjustment factor for each post stratum: the estimated population of the post stratum divided by the census count for the post stratum. The adjustment for each block in the country is found by separating the census count in that block into its components, post stratum by post stratum; applying the adjustment factor for each post stratum to the corresponding count; and then summing the adjusted counts to get the adjusted total for the block. This procedure is justified by the SYNTHETIC ASSUMPTION that the undercount rate is constant within each post stratum, regardless of geography. Failure of the synthetic assumption is called HETEROGENEITY.

There is another way to estimate the population, called *demographic analysis* (DA). DA estimates the population from administrative records and estimates of immigration and emigration using the identity

$$\text{population} = \text{births} - \text{deaths} + \text{immigration} - \text{emigration}.$$

Historically, DA estimates were the primary evidence of net census undercount and of differential undercount by race—motivating census adjustment.

In 1980, the U.S. Census Bureau decided not to adjust the census: Too many data were missing. In 1990, the Bureau sought to adjust, but the administration overruled the Bureau, finding that adjustment was unlikely to improve accuracy. The Bureau planned to adjust the 2000 census but in the end decided not to because DSE disagreed with DA. According to the 2000 census, there were about 281.4 million people residing in the United States. Adjustment would have added 1.2% to this number, but DA indicated that the census had found 0.7% too many people.

The U.S. census is remarkably accurate; proposed adjustments are relatively small. For example, in 2000, adjustment would have increased the population share of Texas by 0.043%, from 7.4094% to 7.4524%. That would have been the biggest state share change. Adjustment must be extremely accurate for such tiny changes to improve the census. Response errors in the census or the PES lead to problems in matching records and produce *processing errors* in DSE that are large on the relevant scale. Heterogeneity and CORRELATION BIAS also produce large errors.

DSE begins with capture-recapture but has layer upon layer of complexity in which details have big effects on the population estimate. For example, there are procedures for getting data by proxy interviews, searching neighboring blocks for missing records, detecting duplicate records, accounting for people who move between census day and the PES, and imputing missing data. The keys to DSE are matching PES records to census records accurately, independence, and the synthetic assumption—not counting better.

DSE would have added 3.3 million people net to the 2000 census. As of October 2001, the Census Bureau estimated net processing error in DSE to be 5 to 6 million and gross error in the DSE to be more than 12 million. In comparison, gross census error was estimated to be about 10 million (U.S. Census Bureau, 2001b). Error in the adjustment is at least as large as the census error that DSE is intended to fix: Adjusting the U.S. census could easily make it worse.

Philip B. Stark

References

Freedman, D. A., & Wachter, K. W. (2003). *On the likelihood of improving the accuracy of the census through statistical adjustment.* In *Science and Statistics: A Festscrift for Terry Speed*. Institute of Mathematical Statistics Monograph 40, 197–230. D.R. Goldstein, editor.

Stark, P.B. (2004). Census Adjustment. In *Encyclopedia of Social Science Research Methods, v. 1,* 112–113. M. Lewis Beck, A. Bryman, and T.F. Liao, eds. Sage Publications.

Stark, P.B. (1999). Differences between the 1990 and 2000 Census Adjustment Plans, and their Impact on Error. Technical Report 550, Department of Statistics, University of California, Berkeley. Revised 3 May 2000. http://www.stat.berkeley.edu/~stark/Census/donner99.pdf

Wachter, K.W. & Freedman, D.A. (2000). The fifth cell: Correlation bias in U.S. census adjustment. *Evaluation Review*, *v. 24*, 191–211.

Journal of the American Statistical Association (1993). The 1990 U.S. census [Special issue]*, v. 88.*

Society (2001). The 2000 U.S. Census [Special issue]*, v. 39,* 2–53.

Statistical Science (1994). The 1990 U.S. census [Special issue].*, v. 9.*

Survey Methodology. (1992). The 1980 U.S. census [Special issue], v. *18*.

U.S. Census Bureau (2001a). *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy*. (With supporting documentation, Repts. B1–24). Washington, DC: Government Printing Office. http://www.census.gov/dmd/www/EscapRep.html.

U.S. Census Bureau (2001b). *Report of the Executive Steering Committee for Accuracy and Coverage Evaluation Policy on Adjustment for Non-Redistricting Uses* (With supporting documentation, Repts. 1–24). Washington, DC: Government Printing Office. http://www.census.gov/dmd/www/EscapRep2.html.

U.S. Census Bureau (2003). *Technical Assessment of A.C.E. Revision II*. http://www.census.gov/dmd/www/ace2.html.
