



Some thoughts on Undercount Problem

Bikas K Sinha

Retired Professor, Indian Statistical Institute, Kolkata, India

Bikassinha1946@gmail.com

The undercount problem is well-known; so is also the ad-hoc solution provided in the literature. The "Text Book Series: Statistics for Social Science & Public Policy" lists a book entitled "Statistics for Lawyers" by Finkelstein and Levin [1989/2001]. Chapter 9 [Subsection 9.1.8] of the book is devoted to a discussion on Current Population Survey [CPS]. It reads 'Since the 1980 census, the US Census Bureau has been the subject of intense litigation. The root of the problem...net undercount of the population ...notably Blacks and Hispanics are undercounted ...the Bureau made public its plan to adjust for the 2000 Census'. It continues to say...."two types of post-census sampling...Nonresponse Follow-up Program AND Post Enumeration Survey [PES]...the estimation of the undercount is based on capture/recapture technology: the census is the capture and the PES is the recapture".

As usual, a 2×2 joint response table is prepared: Census vs. PES with the cell frequency counts denoted by $[a, b; c, d]$ where 'a' corresponds to the frequency count of those 'captured' during the Census and 'recaptured' during the PES; 'b' = frequency count of those captured during the census but not found during the PES; 'c' =frequency count of those missing during the census but found during the PES, and finally, 'd' =frequency count of those totally missing during the entire study. The problem is to provide a 'reasonable estimate' of 'd', given the other three frequency counts. The problem is deceptively simple and no wonder, we have an ad-hoc solution: ' $d=bc/a$ ' which is readily available. The fact is that most often this turns out to be an 'underestimate' of the actual missing count!

I propose to discuss this problem and present a direction towards achieving reasonable solution to it.

Keywords: Estimating equations; chi-square; Bayesian formulation.