

JPSM/MPSM

Survey Methodology Seminar Series

Managing Data Collection with the End in Mind: Adaptive Sample Design and Management at NASS

JAKI MCCARTHY, PH.D.

Chief, Survey Methodology and Technology Branch
USDA's National Agricultural Statistics Service

Thursday, October 4, 2012
4:00-5:00pm

Speaker will be at the University of Michigan
Video Room 368 - ISR South Basement
Room 1208 LeFrak Hall– Maryland

Abstract

With limited time or resources, what is the best way to handle data collection for a given sample? The answer to this question depends heavily on the survey in question. The best data collection procedures for a sample may differ depending on the sample design and estimation scheme, the data collection period, the paradata available, information known about the sample units or the quality standards for the survey estimate. And of course, cost is always an additional constraint. In this talk I will give three different examples of NASS surveys where data collection procedures were altered to improve the survey results given cost constraints. What constitutes an “improvement” to the survey results can be different for every survey and each of these survey samples was managed in a different way.

The Quarterly Agricultural Survey (QAS) is used to produce crop and livestock inventory production statistics at the state and national level. For this survey, NASS has developed predictive models to identify, prior to data collection, those sample units most likely to be nonrespondents. Data collection strategies for these sample units can be different from those used to contact “easy” respondents – for example using specific interviewers, or in person interviews. However, high cost strategies are likely only necessary for potential nonrespondents who can substantially impact survey estimates. As in many establishment surveys, some respondents may be extremely large and dominate survey estimates. In the past, individual field offices have each determined how to collect QAS data from each operation in their state. NASS is now exploring how to apply special handling in a consistent way across all records that have both high nonresponse propensities AND are large (based on frame data) relative to key survey estimates.

NASS has also created another predictive model to pre-identify highly likely nonrespondents in the Agricultural Resource Management Survey (ARMS) which collects farm finance and cost of production data. ARMS uses calibration weighting to adjust for nonresponse and undercoverage. The likely nonrespondent group is quite large and special handling strategies cannot be applied to all of them. In this case, NASS is exploring the idea of identifying important subsets of likely nonrespondents. A promising strategy is to target records within this group that will contribute the most in coverage for the calibration targets. Finally, the County Agricultural Production Survey (CAPS) is a large survey used to produce county level acreage and production estimates. NASS will publish a yield estimate for a commodity in a county if either there are a minimum number of reports OR fewer reports representing a minimum percentage of the harvested acres in the county. For this survey the data collection was managed to reach this threshold in the maximum number of counties.

In each of these examples, the approach taken in data collection is heavily dependent on the needs of the survey. Data collection strategies that go beyond simply maximizing overall response rates will make the most improvements to data quality given the resources available. However, the specific needs and characteristics of the individual survey must be considered to best manage that survey's data collection.