## Feral Pigs Sampling Frame and Sampling Design

NASS's list frame is a roster of known farmers and ranchers and includes a profile of each operation indicating the size of the operation and the commodities they produce. NASS continually works to keep the list frame as complete as possible, especially for the larger producers, by obtaining records for new or omitted operations from other USDA lists, producer association lists, and many other sources. The list frame is an efficient method to select samples and data can be collected using less expensive methods such as mail, telephone, or internet. However, a list frame does not provide complete coverage of all farms since farming arrangements are constantly changing. There will always be some farms that are out of business.

The Feral Pigs Sampling Frame (population) consist of farm operators in Alabama, Arkansas, California, Florida, Georgia, Louisiana, Mississippi, Missouri, North Carolina, South Carolina, and Texas with positive control data for corn, soybeans, wheat, rice or peanuts; or sorghum only for Texas. The Multivariate Probability Proportional to Size (MPPS) sampling design[1] will be used for the Feral Pigs Survey. The MPPS design takes advantage of the efficiencies of a Probability Proportional to Size (PPS) design, while adding the dimension of utilizing multiple variables from partially overlapping frames in the sample allocation. The target sample size for each commodity was derived using a simple random sample size formula that accounts for power.

The probability of selection for an MPPS design is:

$$\pi_{ij}=\min\{\,1,\max\left(n_{j1}\times\frac{x_{ij1}^{0.75}}{\sum_{i=1}^{N_h}x_{ij1}^{0.75}},\ldots,n_{jH}\times\frac{x_{ijH}^{0.75}}{\sum_{i=1}^{N_h}x_{ijH}^{0.75}}\right)\}$$

Where:
$\pi_{ij}$ is the maximum probability of selection for farm operator i in state j.
i is the farm operator.
h is the target commodity (h=1,…,H).
j is the state.
$N_h$ is the number of farm operators in the sampling frame for target commodity h, and
$x_{ijh}$ is the value of target commodity h for the $i^{th}$ farm operator in state j.

[1]Bailey, Jeff and Kott, Phillip (1997),"An Application of Multipurpose List Frame Sampling For Multi-Purpose Surveys, Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 496-500.

The target sample size for each targeted commodity is:

$$n_{jh} = \frac{1}{\left(\frac{1}{1+\frac{Z_{1-\beta}}{Z_{1-\frac{\alpha}{2}}}}\right)^2 \left(\frac{\left((CV_{jh})T_{jh}\right)}{100\left(N_{jh}s_{jh}\right)}\right)^2 + \frac{1}{N_{jh}}}$$

Where:
$n_{jh}$ is the sample size for target commodity h in state j,
$N_{jh}$ is the population for target commodity h in state j,
$Sj_h$ is the standard deviation for target commodity h in state j,
α: Type I error.
β: Type II error
$CV_{jh}$ is the coefficient of variation for target commodity h in state j, and
$T_{jh}$ is the total for target commodity h in state j.

A CV of 15, alpha of .05 and beta of 0.20 were used derive sample sizes for all state commodity combinations. The resulting sample sizes were used to derive the maximum probability of selection, $\pi$. A sample is chosen by sorting the sampling frame by the sum of targeted crop acres, generating a uniform random number (RN), calculating the cumulative probability for unit i as $Cumulative_i = Cumulative_{i+1} + \pi_i$, and selecting unit i if

$Cumulative_{i+1} < RN+k <= Cumulative_i$ for any k= 0,1,2,…n where n is the total number of units selected for the sample[2]. The final sample size was adjusted by the historical response rate (0.80) to 10,800.

[2] Hicks, Susan, Amrhein, John and Kott, Phil (1996), Methods to Meet Target Sample Sizes Under a Multivariate PPS Sampling Strategy, Proceedings of the Section on Survey Research Methods, American Statistical Association, pp. 234-238.