Thus the standard error for the estimated mean in domain 1 is:

$$\text{SE }(\bar{y}_1) = \sqrt{\left(1 - \frac{300}{3078}\right)\left(\frac{300}{299}\right)\left(\frac{128}{129}\right)}\frac{258{,}249.74}{\sqrt{129}} = 21{,}553.$$

An approximate 95% CI for the mean farm acreage for counties in domain 1, using the t critical value with 128 df, is $316{,}565.65 \pm 1.979\,(21{,}553)$, or [273,919, 359,212]. A similar calculation for domain 2 yields $\text{SE}\,(\bar{y}_2) = 28{,}852.24$ and an approximate 95% CI of [226,859, 340,769].

Suppose that we do not know how many counties in the population are in each domain. To estimate the total in domain 1, define

$$x_i = \begin{cases} 1, & \text{if county } i \text{ is in domain 1} \\ 0, & \text{otherwise} \end{cases}$$

and $u_i = y_i x_i$. Then

$$\hat{t}_{y1} = \hat{t}_u = \sum_{i \in \mathcal{S}} \frac{3078}{300} u_i = 418{,}987{,}302. \tag{4.25}$$

The standard error is

$$\text{SE}(\hat{t}_{y1}) = N\sqrt{1 - \frac{n}{N}}\frac{s_u}{\sqrt{n}} = 3078\sqrt{\left(1 - \frac{300}{3078}\right)}\frac{230{,}641.22}{\sqrt{300}} = 38{,}938{,}277$$

and a 95% CI for the population total in domain 1, using a t critical value with 128 df, is $418{,}987{,}302 \pm 1.979(38{,}938{,}277) = [341{,}941{,}269,\ 496{,}033{,}335]$. Similarly, a 95% CI for the population total in domain 2 is

$$497{,}939{,}808 \pm 1.974(3078)\sqrt{\left(1 - \frac{300}{3078}\right)}\frac{331{,}225.43}{\sqrt{300}} = [387{,}553{,}731,\ 608{,}325{,}884]\,. \blacksquare$$

In this section, we have shown that estimating domain means is a special case of ratio estimation because the sample size in the domain varies from sample to sample. If the sample size for the domain in an SRS is sufficiently large, we can use SRS formulas for inference about the domain mean.

Inference about totals depends on whether the population size of the domain, N_d, is known. If N_d is known, then the estimated total is $N_d\bar{y}_d$. If N_d is unknown, then define a new variable u_i that equals y_i for observations in the domain and 0 for observations not in the domain; then use $\hat{t}_u$ to estimate the domain total.

The results of this section are only for SRSs, and the approximations depend on having a sufficiently large sample so that $E(n_d)$ is large. Domain estimates for general survey designs are discussed in Section 11.3 and the problem of domain estimates when n_d is small is discussed in Section 14.2.

Domains and strata. Domains and strata are both subsets of the population. What is the difference between them?

Strata are used in the sample design. A stratified random sampling design specifies selecting a predetermined number of observations, n_h, from stratum h. Every sample that could possibly be drawn using that sampling design will have n_h observations from stratum h.

A domain is any population subset of interest, and calculating estimates for domains of interest is part of the analysis. Sometimes it is desired to calculate separate estimates for