8 More Complex Sampling Designs

8.1 Ratio Estimation with a Stratified SRS

- Situation: A SRS of (x, y) pairs of data are taken within each of L strata.
- If scatterplots are made for each stratum and each plot suggests proportionality between x and y (i.e., linearity that passes through (0,0)), then we can generate an estimate of $\overline{y}_U$ or t based on ratio estimation that is an improvement to an estimate based solely on the Y values.
- There are two methods for constructing ratio estimators based on a stratified SRS.

8.1.1 Method 1: The Separate Ratio Estimator

- The **separate ratio estimator** is based on calculating L stratum ratio estimates and then form a weighted average of these separate ratio estimates to form a single estimate of the population ratio B.
- Suppose we want to estimate the population mean $\overline{y}_U$. For stratum h, the estimated stratum ratio is $\widehat{B}_h = \overline{y}_h / \overline{x}_h$. Thus, the stratum estimate of the stratum mean $\overline{y}_{Uh}$ is

$$\widehat{\overline{y}_{Uh}} = \frac{\overline{y}_h}{\overline{x}_h} \overline{x}_{Uh}.$$

- The **separate ratio estimate** of $\overline{y}_U$ is

$$\widehat{\overline{y}_{U}}_{sst} = \sum_{h=1}^{L} \frac{N_h}{N} \widehat{B}_h \overline{x}_{Uh} = \sum_{h=1}^{L} \frac{N_h}{N} \frac{\overline{y}_h}{\overline{x}_h} \overline{x}_{Uh} \tag{89}$$

with estimated variance

$$\begin{aligned} \widehat{V}(\widehat{\overline{y}_{U}}_{sst}) &= \sum_{h=1}^{L} \left[\left(\frac{N_h}{N} \right)^2 \frac{N_h - n_h}{N_h n_h (n_h - 1)} \sum_{i=1}^{n_h} (y_{hi} - \widehat{B}_h x_{hi})^2 \right] \\ &= \sum_{h=1}^{L} \left[\left(\frac{N_h}{N} \right)^2 \frac{N_h - n_h}{N_h n_h (n_h - 1)} \left(\sum_{i=1}^{n_h} y_{hi}^2 + \widehat{B}_h^2 \sum_{i=1}^{n_h} x_{hi}^2 - 2\widehat{B}_h \sum_{i=1}^{n_h} x_{hi} y_{hi} \right) \right] \end{aligned} \tag{90}$$

where x_{hi} and y_{hi} are the j^{th} sample values in stratum h.

- Thus, the standard error of $\widehat{\overline{y}_{U}}_{sst} = \sqrt{\widehat{V}(\widehat{\overline{y}_{U}}_{sst})}$.
- Note that the estimated variance in (90) is the sum of the separate stratum-based estimated variances.
  1. This formula is recommended only if the sample within each stratum is large enough to apply the estimated variance formula. See the handout on SRS ratio estimation for details.
  2. If the sample sizes n_h are small and the number of strata L is large, the estimates can be seriously biased.

- Typically, if the n_h are large enough to use a stratified ratio estimator, some texts recommending using z^* in the confidence intervals:

$$\widehat{\overline{y}_{U}}_{sst} \ \pm \ z^* \sqrt{\widehat{V}(\widehat{\overline{y}_{U}}_{sst})}$$

- For smaller n_h, replace z^* with t^* which is based on a t-distribution with $\sum h = 1^L(n_h - 1) \ = \ n - L$ where n is the total sample size.

- To get an estimate of the population total t, multiply the right-hand side of (89) by N. The estimated variance is gotten by multiplying the right-hand side of (90) by N^2.

8.1.2 Method 2: The Combined Ratio Estimator

- The **combined ratio estimator** is based on first calculating the stratified SRS estimates of the means and then correcting for the relationship between X and Y. That is:

  1. Calculate (i) the stratified SRS estimate $\overline{x}_{st}$ of $\overline{x}_U$ from the x-data, (ii) the stratified SRS estimate $\overline{y}_{st}$ of $\overline{y}_U$ from the y-data, and (iii) the estimated ratio $\widehat{B}_c$:

$$\overline{y}_{st} \ = \ \sum_{i=1}^{L} \frac{N_h}{N} \overline{y}_h \qquad \overline{x}_{st} \ = \ \sum_{i=1}^{L} \frac{N_h}{N} \overline{x}_h \qquad \widehat{B}_c = \frac{\overline{y}_{st}}{\overline{x}_{st}}$$

  2. Then the **combined ratio estimate** of $\overline{y}_U$ is

$$\widehat{\overline{y}_{U}}_{cst} \ = \ \widehat{B}_c \, \overline{x}_U \tag{91}$$

  Let $f_h = N_h/N$. Then, the estimated variance

$$\begin{aligned} \widehat{V}(\widehat{\overline{y}_{U}}_{cst}) \ &= \ \sum_{i=1}^{L} \left[\left(\frac{N_h}{N} \right)^2 \frac{(1-f_h)}{n_h(n_h-1)} \sum_{j=1}^{n_h} (y_{hi} - \widehat{R}_c x_{hi})^2 \right] \qquad (92) \\ &= \ \sum_{i=1}^{L} \left[\left(\frac{N_h}{N} \right)^2 \frac{(1-f_h)}{n_h(n_h-1)} \left(\sum_{j=1}^{n_h} y_{hi}^2 + \widehat{R}_c^2 \sum_{j=1}^{n_h} x_{hi}^2 - 2\widehat{R}_c \sum_{j=1}^{n_h} x_{hi} y_{hi} \right) \right] \end{aligned}$$

- Note that the estimated variance in (92) is the sum of the variances calculated about a common ratio estimate $\widehat{R}_c$. This formula is valid only if the total sample size is large.

- To get an estimate of the population total t, multiply the right-hand side of (91) by N. The estimated variance is determined by multiplying the rhs of (92) by N^2.

- Confidence interval estimation is the same as in Method 1 but using $\widehat{\overline{y}_{U}}_{cst}$ and $\widehat{V}(\widehat{\overline{y}_{U}}_{cst})$.

<u>A Comparison of the Two Ratio Methods</u>

- Both methods produce biased estimated of $\overline{y}_U$ or t with the bias decreasing as the sample size increases. Specifically,

  – For the separate ratio estimator we require that the sample sizes within each stratum be large for the bias to be small.

- For the combined ratio estimator we require that the total sample size be large for the bias to be small.
- Thus, the bias will tend to be less serious for the combined estimator than for the separate estimator.

- Unless the ratio relationship is similar across the strata, the separate estimator will be more efficient (have smaller variability) than the combined estimator.
- The lower efficiency of the combined estimator, however, is often offset by smaller bias and the fact that we do not need to know the separate $\overline{x}_{Uh}$ stratum means.

8.2 Regression Estimation with a Stratified SRS

- Situation: SRSs of (x, y) values are taken within each of the L strata.
- If scatterplots are made for each stratum and each plot suggests linearity between x and y, then we can generate an estimate of $\overline{y}_U$ or t based on regression estimation that is an improvement to an estimate based solely on the y values.
- There are two methods for constructing regression estimators based on a stratified SRS.

8.2.1 Method 1: The Separate Regression Estimator

- The **separate regression estimator** is based on calculating L stratum regression estimates and then form a weighted average of these separate stratum regression estimates to form a single estimate of the population $\overline{y}_U$ or t.
- This method is appropriate when is assumed that the linear relationships are not the same across all of the strata.
- Suppose we want to estimate the population mean $\overline{y}_U$. For stratum h, the estimated stratum regression line is

$$\widehat{y} = a_h + b_h x$$

  where a_h and b_h are the intercept and slope of a simple linear regression estimated using least squares.

- Thus, we will have L simple linear regression models.
- We combine the stratum mean estimates $\widehat{\overline{y}_{ULh}}$:

$$\begin{aligned} \widehat{\overline{y}_{ULh}} &= a_h + b_h \overline{x}_{Uh} \\ &= \overline{y}_h + b_h(\overline{x}_{Uh} - \overline{x}_h) \end{aligned} \tag{93}$$

  to get the **separate regression estimate** of the population mean $\overline{y}_U$

$$\begin{aligned} \widehat{\overline{y}_{ULsst}} &= \sum_{i=1}^{L} \frac{N_h}{N} (a_h + b_h \overline{x}_{Uh}) \\ &= \sum_{i=1}^{L} \frac{N_h}{N} (\overline{y}_h + b_h(\overline{x}_{Uh} - \overline{x}_h)) \end{aligned} \tag{94}$$

- Let $f_h = N_h/N$. The estimated variance of $\overline{y}_{ULsst}$ is

$$\widehat{V}(\widehat{\overline{y}_{ULsst}}) = \sum_{i=1}^{L} \left(\frac{N_h}{N}\right)^2 \frac{(1-f_h)}{n_h(n_h-2)} SSE_h = \sum_{i=1}^{L} \frac{(1-f_h)}{n_h} MSE_h \tag{95}$$

where SSE_h and MSE_h are the error sum of squares and the mean square error of simple linear regression for the h^{th} stratum data.

- Note that the estimated variance in (95) is the weighted sum of the separate stratum-based estimated variances.
- Typically, if the n_h are large enough to use a stratified regression estimator, the overall sample size is large enough to use z^* in the confidence intervals.
- To get an estimate of the population total t, multiply the right-hand side of (94) by N. The estimated variance is gotten by multiplying the rhs of (95) by N^2.

8.2.2 Method 2: The Combined Regression Estimator

- This method is appropriate when is assumed that the linear relationships are the same across all of the strata.
- First, we need to calculate the weighted average of the L slope coefficients. That is:
  1. Calculate b_h for each stratum.
  2. Determine the weighted slope estimate $b_c = \dfrac{\sum_{i=1}^{L} c_h b_h}{\sum_{i=1}^{L} c_h}$

     where $c_h = \left(\dfrac{N_h}{N}\right)^2 \dfrac{(1-f_h)}{n_h} s_{xi}^2$ and s_{xi}^2 is the sample variance of the sampled x-values in the h^{th} stratum.
  3. Then the **combined regression estimate** of $\overline{y}_U$ is

$$\widehat{\overline{y}_{ULcst}} = \widehat{\overline{y}_{Ust}} + b_c(\overline{x}_U - \overline{x}_{st}) \tag{96}$$

     where

$$\widehat{\overline{y}_{Ust}} = \sum_{i=1}^{L} \frac{N_h}{N} \overline{y}_h \qquad \overline{x}_{st} = \sum_{i=1}^{L} \frac{N_h}{N} \overline{x}_h.$$

  4. The estimated variance is

$$\widehat{V}(\overline{y}_U)_{Lcst} = \sum_{i=1}^{L} \left[\left(\frac{N_h}{N}\right)^2 \frac{(1-f_h)}{n_h(n_h-2)} \sum_{j=1}^{n_h} [(y_{hi} - \overline{y}_h) - b_c(x_{hi} - \overline{x}_h)]^2 \right] \tag{97}$$

- To get an estimate of the population total t, multiply the right-hand side of (96) by N. The estimated variance is gotten by multiplying the rhs of (97) by N^2.

A Comparison of the Two Regression Methods

- Both methods produce biased estimated of $\overline{y}_U$ or t with the bias decreasing as the sample size increases. Specifically,
  - For the separate regression estimator we require that the sample sizes within each stratum be large for the bias to be small.
  - For the combined regression estimator we require that the total sample size be large for the bias to be small.
- Thus, the bias will tend to be less serious for the combined estimator than for the separate estimator.
- If the linear relationships are not similar across the strata, the separate estimator is preferred.
- If we are confident that the linear relationships are similar across the strata, the combined estimator is preferred.

EXAMPLE: A manufacturing company supplies a product that is packaged under two brand names for marketing purposes. These two brands serve as strata for estimating potential sales volume for the next quarter. A SRS of customers for each brand is contacted and asked to provide a potential sales figure y (in number of units) for the upcoming quarter. Last year's sales figure x for the same quarter is also available for each sampled customer. For Brand 1 and Brand 2 there are $N_1 = 120$ and $N_2 = 180$ customers with total sales in the same quarter last year of $t_1 = 21500$ and $t_2 = 21200$ units. Based on the stratified sample below estimate the potential mean sales per customer for next quarter

- Using both methods of ratio estimation.
- Using both methods of regression estimation.

Brand 1

x	204	143	82	256	275	198
y	210	160	75	280	300	190

Brand 2

x	137	189	119	63	103	107	159	63	87
y	150	200	125	60	110	100	180	75	90

```
The regression equation is    upcoming = - 16.9 + 1.14 lastyear

Predictor        Coef     SE Coef         T        P
Constant       -16.94       15.79     -1.07    0.344
lastyear      1.13698     0.07747     14.68    0.000

S = 12.41       R-Sq = 98.2%     R-Sq(adj) = 97.7%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         1       33171       33171    215.40     0.000
Residual Error     4         616         154
Total              5       33787

================================================================

The regression equation is    upcoming = - 5.16 + 1.11 lastyear

Predictor        Coef     SE Coef         T        P
Constant       -5.157       7.862     -0.66    0.533
lastyear      1.10654     0.06504     17.01    0.000

S = 7.779       R-Sq = 97.6%     R-Sq(adj) = 97.3%

Analysis of Variance

Source            DF          SS          MS         F         P
Regression         1       17515       17515    289.41     0.000
Residual Error     7         424          61
Total              8       17939
```

8.3 Multistage Sampling

- There are many ways to modify cluster sampling for more complex sampling situations. One common modification is to take a sample of secondary units from within sampled clusters (instead of inspecting every secondary unit within each sampled cluster). The process of drawing samples from the selected clusters is called **subsampling**.
- Suppose a SRS of n primary units is selected from a population of N primary units. Then SRSs of secondary units of sizes $m_1, m_2, \ldots, m_n$ are selected, respectively, from the $M_1, M_2, \ldots, M_n$ possible secondary sampling units contained within these n primary units.
- The total sample of $m = \sum_{i=1}^{n} m_i$ secondary units is called a **two-stage cluster sample**.
- In two-stage cluster sampling, we select m_i units from the M_i secondary units ($m_i \leq M_i$) in primary unit i. In one-stage cluster sampling, we inspected all M_i secondary units ($m_i = M_i$) in a cluster.
- If the secondary units (Stage 2) are the individuals of the study population, then sampling would stop. Suppose, however, the secondary units are also clusters of individuals (tertiary units) in the study population. We would then subsample tertiary units within the secondary units (Stage 3). This sample of tertiary units from the study population and is called a **three-stage cluster sample**.
- Any sampling design which requires multiple stages of subsampling is called a **multistage design**. Two-stage and three-stage cluster samples are examples of multistage designs.
- Multistage designs are used for a variety of practical reasons. One common motivation is the same as it was for one-stage cluster sampling — administrative convenience.
  - To obtain a sample of fish caught in a commercial fishery, it may be necessary to first take a sample of boats and then take a sample of fish from each selected boat.
  - To obtain a sample of plants of certain species, it may be convenient to first take a sample of plots and then take a sample of plants within each selected plot.
  - To obtain a sample of local voter preferences, a random sample of city blocks is taken. Then within each of these blocks a random sample of households is taken.
- Sources of variability in multistage sampling:
  - In one-stage cluster sampling, the estimate varies because due to one source: different samples of primary units yield different estimates.
  - In two-stage cluster sampling, the estimate varies due of two sources: different samples of primary units and then different samples of secondary units within primary units.
  - In general, if there a k stages of subsampling, there will be k sources of variability.
  - Thus, variances and variance estimators for multistage sampling with k-stages will contain the sum of k components of variability.

8.3.1 Two-Stage Cluster Sampling

The following notation is similar to one-stage cluster sampling:

N = the number of population primary sampling units (PSUs)

n = the number of sampled PSUs

M_i = number of secondary sampling units (SSUs) in PSU i

m_i = number of SSus sampled in PSU i

$M = \sum_{i=1}^{N} M_i$ = the number of SSUs in the population

y_{ij} = the y-value associated with SSU j in PSU i

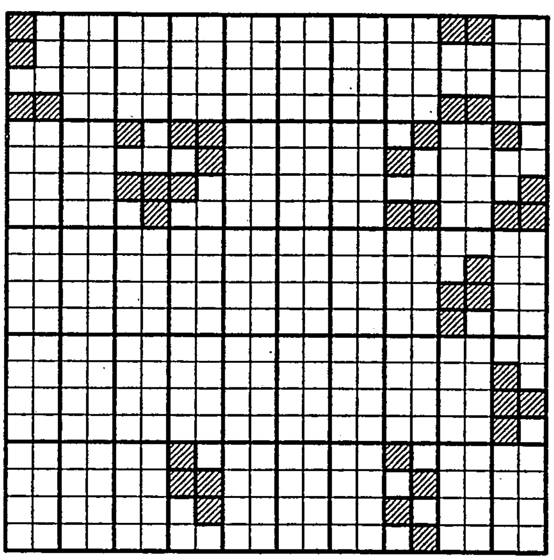
$t_i = \sum_{j=1}^{M_i} y_{ij}$ = PSU i total $\qquad$ $t = \sum_{i=1}^{N}\sum_{j=1}^{M_i} y_{ij} = \sum_{i=1}^{N} t_i$ = population total

$\overline{y}_{iU} = \dfrac{t_i}{M_i}$ = mean per SSU in the i^{th} PSU $\qquad$ $\overline{y}_U = \dfrac{t}{M}$ = population mean per SSU

$\overline{t}_i = \dfrac{t}{N}$ = population mean of PSU totals

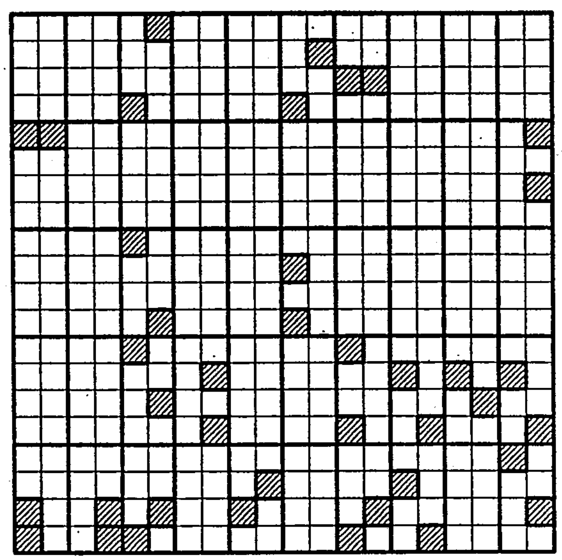
$y_i = \dfrac{1}{m_i}\sum_{j=1}^{m_i} y_{ij}$ = the total of sampled SSUs in PSU i

$\overline{y}_i = \dfrac{1}{m_i}\sum_{j=1}^{m_i} y_{ij} = y_i/m_i$ = the sample mean of sampled SSUs in PSU i



A two-stage sample of ten primary units and four secondary units per primary unit.

$N = 50,\ n = 10,\ M_0 = 400,$
$M_i = 8,\ m_i = 4$



A two-stage sample of twenty primary units and two secondary units per primary unit.

$N = 50,\ n = 20,\ M_0 = 400$
$M_i = 8,\ m_i = 2$

- Because a SRS of size m_i is taken within primary unit i, an unbiased estimator of the primary unit total y_i is

$$\widehat{t}_i \;=\; \frac{M_i}{m_i}\sum_{j=1}^{m_i} y_{ij} \;=\; M_i\overline{y}_i \tag{98}$$

- Combining and weighting the n independent unbiased $\widehat{y}_i$ estimates provides us with an unbiased estimate of the population total t:

$$\widehat{t}_{unb} \;=\; \frac{N}{n}\sum_{i=1}^{n}\widehat{t}_i \tag{99}$$

- The variability of $\widehat{t}_{unb}$ has two sources: sampling across primary units and sampling within primary units. This is reflected in the variance of $\widehat{t}_{unb}$:

$$V(\widehat{t}_{unb}) \;=\; N(N-n)\,\frac{S_t^2}{n} \;+\; \frac{N}{n}\sum_{i=1}^{N} M_i(M_i-m_i)\,\frac{S_i^2}{m_i} \tag{100}$$

$$\text{where } S_t^2 \;=\; \frac{1}{N-1}\sum_{i=1}^{N}(t_i-\overline{t}_i)^2 = \text{the variance of PSU totals and}$$

$$S_i^2 \;=\; \frac{1}{M_i-1}\sum_{j=1}^{M_i}(y_{ij}-\overline{y}_{iU})^2 = \text{ the variance of the SSU } (y_{ij} \text{ values within PSU}i.$$

- The first component in (100) is the variance across clusters. The second component is the variance due to estimating the y_i values from subsamples of SSUs within PSUs.
- Note: if all SSUs were sampled in the sampled PSUs ($m_i = M_i$), then the second component would be zero yielding the one-stage cluster sample variance.
- Because S_t^2 and the S_i^2 values are not known, we substitute data estimates into (100) to get an estimated variance:

$$\widehat{V}(\widehat{t}_{sub}) \;=\; N(N-n)\,\frac{s_t^2}{n} \;+\; \frac{N}{n}\sum_{i=1}^{n} M_i(M_i-m_i)\,\frac{s_i^2}{m_i} \tag{101}$$

$$\begin{aligned}
\text{where } s_t^2 \;&=\; \frac{1}{n-1}\sum_{i=1}^{n}\left(\widehat{t}_i-\frac{\widehat{t}_{unb}}{N}\right)^2 \\
&=\; \text{the sample variance of the estimated PSU } (\widehat{t}_i) \text{ totals} \\
s_i^2 \;&=\; \frac{1}{m_i-1}\sum_{j=1}^{m_i}(y_{ij}-\overline{y}_i)^2 = \text{ variance of sampled units within PSU}i.
\end{aligned}$$

- Divide $\widehat{t}_{unb}$ by M_0 and $\widehat{V}(\widehat{t}_{unb})$ by M_0^2, respectively, to get estimates $\widehat{\overline{y}}_{U,unb}$ of the population mean $\overline{y}_U$ per SSU and its estimated variance $\widehat{V}(\widehat{\overline{y}}_{U,unb})$.
- Thompson (2002) also discusses (i) how to apply ratio estimation to two-stage cluster sampling and (ii) estimation when PSUs are selected proportional to size.
- On page 186 of the textbook, Lohr provides a ratio estimator for two-stage cluster sampling.

8.3.2 Equal Sized Primary Units and Equal Sized Within-Cluster Samples

- Suppose that each of the n clusters have the same number M of secondary units ($M_1 = M_2 = \cdots = M_N = M$). Then, $M_0 = NM$.
- Suppose a SRS of n clusters is taken. Within each of the n clusters a SRS of m secondary units is taken. Then the total number of secondary units selected is nm.
- There is a total of $\binom{N}{n}\binom{M}{m}$ possible two-stage cluster samples and each one has the same probability of being selected. Thus, the probability that any particular two-stage cluster sample will be selected $= \dfrac{1}{\binom{N}{n}\binom{M}{m}}$.

Estimation of $\overline{y}_U$ and t

The unbiased estimators of $\overline{y}_U$ and t are

$$\widehat{t}_{unb} = \frac{M_0}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} = \frac{M_0}{n}\sum_{i=1}^{n}\overline{y}_i \qquad (102)$$

$$\widehat{\overline{y}}_{U,unb} = \frac{\widehat{t}_{unb}}{M_0} = \frac{1}{nm}\sum_{i=1}^{n}\sum_{j=1}^{m} y_{ij} = \frac{1}{n}\sum_{i=1}^{n}\overline{y}_i \qquad (103)$$

The unbiased variance estimators are

$$\widehat{V}(\widehat{t}_{unb}) = N(N-n)\frac{s_u^2}{n} + \frac{M_0(M-m)}{nm}\sum_{i=1}^{n} s_i^2 \qquad (104)$$

$$\widehat{V}(\widehat{\overline{y}}_{U,unb}) = \frac{N(N-n)}{M_0^2}\frac{s_u^2}{n} + \frac{M-m}{M_0 nm}\sum_{i=1}^{n} s_i^2 \qquad (105)$$

Conservative Two-Sided Confidence Intervals

The conservative confidence intervals for t and $\overline{y}_U$ are:

$$\widehat{t}_{unb} \pm t^*\sqrt{\widehat{V}(\widehat{t}_{unb})} \qquad \text{and} \qquad \widehat{\overline{y}}_{U,unb} \pm t^*\sqrt{\widehat{V}(\widehat{\overline{y}}_{U,unb})} \qquad (106)$$

where t^* is the the upper $\alpha/2$ critical value from the $t(n-1)$ distribution.

Figure 13: Two-Stage Cluster Sampling Example

The population mean $\overline{y}_U = 33.385$ per secondary unit with a total abundance $t = 13354$. There are $M_0 = 400$ SSUs and $N = 50$ PSUs of size $M = 8$. A SRS of $n = 20$ clusters is randomly taken. Then, SRSs of size $m = 2$ are taken within each of the 20 sampled clusters. Thus, a total of $nm = 40$ SSUs were sampled and are indicted in ( ). The sample variance of the twenty $\widehat{t}_{unb}$ values is 4817.64.

(18)	20	15	20	20	15	19	18	24	23	20	26	29	28	28	**(31)**	31	34	28	32
13	20	16	20	15	23	19	26	21	21	**(24)**	**(30)**	23	26	**(25)**	33	31	28	32	38
16	18	20	24	25	26	22	23	26	26	22	27	25	25	34	28	37	36	**(38)**	**(31)**
17	**(17)**	16	22	21	23	22	27	27	24	28	32	29	33	27	37	37	38	35	33
15	19	23	17	**(21)**	23	21	23	24	25	31	**(26)**	32	34	32	33	31	31	36	37
21	24	20	21	28	26	30	22	31	25	29	29	27	30	29	37	35	32	38	43
23	17	24	25	**(24)**	27	31	29	31	**(34)**	**(27)**	36	29	29	34	**(39)**	**(37)**	37	40	36
18	24	21	25	27	22	32	32	**(31)**	26	28	34	34	37	35	**(34)**	**(38)**	38	37	40
22	26	28	26	24	29	33	26	27	27	34	31	39	32	**(36)**	38	37	40	44	43
(23)	27	28	29	26	32	25	**(31)**	35	34	32	33	37	32	42	**(40)**	40	37	42	44
23	**(21)**	**(31)**	23	30	27	31	30	32	35	30	40	32	37	37	36	40	44	44	40
26	29	**(31)**	26	30	31	**(34)**	36	30	38	36	32	38	38	37	42	42	41	40	49
(28)	24	28	27	26	31	32	29	32	33	38	34	39	38	40	37	41	43	42	43
(32)	25	31	32	29	29	35	38	38	32	36	35	**(39)**	**(42)**	39	40	44	42	41	45
27	29	35	28	35	35	31	**(40)**	35	37	**(38)**	**(44)**	40	40	47	39	**(49)**	48	51	49
30	29	32	32	33	30	**(36)**	38	42	36	35	38	44	47	45	49	**(41)**	43	44	51
28	35	35	34	34	33	41	33	34	35	39	44	44	48	44	50	49	48	**(53)**	54
29	33	32	36	39	33	33	34	35	**(42)**	46	47	48	47	46	45	44	52	54	**(55)**
28	37	38	37	33	33	34	37	45	**(40)**	39	42	42	46	47	48	52	47	46	53
38	39	39	37	34	38	39	45	39	42	45	41	44	51	46	50	52	51	51	53

i	$\sum_{i=1}^{2} y_{ij}$	$\overline{y}_i$	s_i^2	$\widehat{t}_i$
1	35	17.5	0.5	140
2	54	27.0	18.0	216
3	56	28.0	18.0	224
4	69	34.5	24.5	276
5	45	22.5	4.5	180
6	65	32.5	4.5	260
7	53	26.5	0.5	212
8	73	36.5	12.5	292
9	75	37.5	0.5	300
10	44	22.0	2.0	176
11	62	31.0	0.0	248
12	65	32.5	4.5	260
13	76	38.0	8.0	304
14	60	30.0	8.0	240
15	76	38.0	8.0	304
16	82	41.0	18.0	328
17	81	40.5	4.5	324
18	90	45.0	32.0	360
19	82	41.0	2.0	328
20	108	54.0	2.0	432
Totals	1351		172.5	5404

8.3.3 Stratified Two-Stage Cluster Sampling Example

The population mean $\overline{y}_U = 33.385$ per SSU with a total abundance $t = 13354$. The population is stratified into 2 strata. There are $M_0 = 200$ SSUs and $N = 25$ PSUs of size $M = 8$ within each stratum. A SRS of $n = 9$ clusters is randomly taken from stratum 1 and a SRS of $n = 11$ clusters is randomly taken from stratum 2. Then, SRSs of size $m = 2$ are taken within each of the sampled clusters. Thus, $nm = 18$ SSUs were sampled in stratum 1 and $nm = 22$ SSUs were sampled in stratum 2. The sampled units and are indicted in ( ).

Figure 14

Stratum 1

(18)	20	15	20	20	15	19	18	24	23
13	20	16	20	15	23	19	26	21	21
16	18	20	24	25	26	22	23	26	26
17	**(17)**	16	22	21	23	22	27	27	24
15	19	23	17	**(21)**	23	21	23	24	25
21	24	20	21	28	26	30	22	31	25
23	17	24	25	**(24)**	27	31	29	31	**(34)**
18	24	21	25	27	22	32	32	**(31)**	26
22	26	28	26	24	29	33	26	27	27
(23)	27	28	29	26	32	25	**(31)**	35	34
23	**(21)**	**(31)**	23	30	27	31	30	32	35
26	29	**(31)**	26	30	31	**(34)**	36	30	38
(28)	24	28	27	26	31	32	29	32	33
(32)	25	31	32	29	29	35	38	38	32
27	29	35	28	35	35	31	**(40)**	35	37
30	29	32	32	33	30	**(36)**	38	42	36
28	35	35	34	34	33	41	33	34	35
29	33	32	36	39	33	33	34	35	**(42)**
28	37	38	37	33	33	34	37	45	**(40)**
38	39	39	37	34	38	39	45	39	42

Stratum 2

20	26	29	28	28	**(31)**	31	34	28	32
(24)	**(30)**	23	26	**(25)**	33	31	28	32	38
22	27	25	25	34	28	37	36	**(38)**	**(31)**
28	32	29	33	27	37	37	38	35	33
31	**(26)**	32	34	32	33	31	31	36	37
29	29	27	30	29	37	35	32	38	43
(27)	36	29	29	34	**(39)**	**(37)**	37	40	36
28	34	34	37	35	**(34)**	**(38)**	38	37	40
34	31	39	32	**(36)**	38	37	40	44	43
32	33	37	32	42	**(40)**	40	37	42	44
30	40	32	37	37	36	40	44	44	40
36	32	38	38	37	42	42	41	40	49
38	34	39	38	40	37	41	43	42	43
36	35	**(39)**	**(42)**	39	40	44	42	41	45
(38)	**(44)**	40	40	47	39	**(49)**	48	51	49
35	38	44	47	45	49	**(41)**	43	44	51
39	44	44	48	44	50	49	48	**(53)**	54
46	47	48	47	46	45	44	52	54	**(55)**
39	42	42	46	47	48	52	47	46	53
45	41	44	51	46	50	52	51	51	53

Stratum 1

i	$\sum_{i=1}^{2} y_{ij}$	$\overline{y}_i$	s_i^2	$\widehat{t}_i$
1	35	17.5	0.5	140
2	45	22.5	4.5	180
3	65	32.5	4.5	260
4	44	22.0	2.0	176
5	62	31.0	0.0	248
6	65	32.5	4.5	260
7	60	30.0	8.0	240
8	76	38.0	8.0	304
9	82	41.0	2.0	328
Totals	534		34.0	2136

Stratum 2

i	$\sum_{i=1}^{2} y_{ij}$	$\overline{y}_i$	s_i^2	$\widehat{t}_i$
1	54	27.0	18.0	216
2	56	28.0	18.0	224
3	69	34.5	24.5	276
4	53	26.5	0.5	212
5	73	36.5	12.5	292
6	75	37.5	0.5	300
7	76	38.0	8.0	304
8	82	41.0	18.0	328
9	81	40.5	4.5	324
10	90	45.0	32.0	360
11	108	54.0	2.0	432
Totals	817		138.5	3268

For stratum 1, $s_t^2 = 3762$ and for stratum 2, $s_t^2 = 4352.29$.