Simple Random Sample

A simple random sample of ***n*** elements is taken from a population of ***N*** elements where each element has the same chance of being selected into the sample and each sample has the same chance of being selected.

	Estimate	Variance	Standard Error
Total	$\hat{t} = N\bar{y}$	$\hat{V}(\hat{t}) = N^2 V(\bar{y}) = N^2\left(1 - \frac{n}{N}\right)\frac{s^2}{n}$	$SE(\hat{t}) = \sqrt{\hat{V}(\hat{t})}$
Mean	$\bar{y} = \frac{1}{n}\sum y_i$	$\hat{V}(\bar{y}) = \left(1 - \frac{n}{N}\right)\frac{s^2}{n}$	$SE(\bar{y}) = \sqrt{\hat{V}(\bar{y})}$
Proportion	$\hat{p} = \bar{y}$	$\hat{V}(\hat{p}) = \left(1 - \frac{n}{N}\right)\frac{\hat{p}(1-\hat{p})}{n-1}$	$SE(\hat{p}) = \sqrt{\hat{V}(\hat{p})}$

Where the sample variance is $s^2 = \frac{1}{n-1}\sum (y_i - \bar{y})^2$,

With a 100(1-α)% confidence interval of $Estimate \pm z_{\frac{\alpha}{2}} SE(Estimate)$.

To estimate the total number of the population with a characteristic estimated by a proportion, $\hat{t} = \sum N\hat{p}$, with $\hat{V}(\hat{t}) = N^2\hat{V}(\hat{p})$ or $SE(\hat{t}) = N \times SE(\hat{p})$.

Stratified Sampling

Divide the population of ***N*** elements into ***H*** strata so that stratum ***h*** contains N_h elements. Independently take a simple random sample from each stratum, resulting in a sample of n_h from stratum ***h***.

	Stratified Estimate	Variance	Standard Error
Total	$\hat{t}_{str} = \sum \hat{t}_h = \sum N_h \bar{y}_h$	$\hat{V}(\hat{t}_{str}) = \sum V(\hat{t}_h) = \sum \left(1 - \frac{n_h}{N_h}\right) N_h^2 \frac{s_h^2}{n_h}$	$SE(\hat{t}_{str}) = \sqrt{\hat{V}(\hat{t}_{str})}$
Mean	$\bar{y}_{str} = \frac{\hat{t}_{str}}{N} = \sum \frac{N_h}{N} \bar{y}_h$	$\hat{V}(\bar{y}_{str}) = \frac{1}{N^2} \hat{V}(\hat{t}_{str}) = \sum \left(1 - \frac{n_h}{N_h}\right)\left(\frac{N_h}{N}\right)^2 \frac{s_h^2}{n_h}$	$SE(\bar{y}_{str}) = \sqrt{\hat{V}(\bar{y}_{str})}$
Proportion	$\hat{p}_{str} = \sum \frac{N_h}{N} \hat{p}_h$	$\hat{V}(\hat{p}_{str}) = \sum \left(1 - \frac{n_h}{N_h}\right)\left(\frac{N_h}{N}\right)^2 \frac{\hat{p}_h(1 - \hat{p}_h)}{n_h - 1}$	$SE(\hat{p}_{str}) = \sqrt{\hat{V}(\hat{p}_{str})}$

Where $\bar{y}_h = \frac{\sum y_{hj}}{n_h}$ and $\hat{p}_h = \bar{y}_h$;

Where $s_h^2 = \sum \frac{(y_{hj} - \bar{y}_h)^2}{n_n - 1}$, the sum of the simple variances from each stratum since all elements in each stratum are included in the measures (survey);

With a 100(1-α)% confidence interval of $Estimate \pm z_{\frac{\alpha}{2}} SE(Estimate)$.

To estimate the total number of the population with the characteristic estimated by a proportion, $\hat{t}_{str} = \sum N_h \hat{p}_h$, with $\hat{V}(\hat{t}_{str}) = N^2 \hat{V}(\hat{p}_{str})$.

Weights can be used to compute estimates but cannot be used to compute standard errors: $w_{ij} = \frac{1}{\pi_{hj}}$, where $\pi_{hj} = \frac{n_h}{N_h}$, the sampling fraction of stratum ***h*** (probability of selection).

One-Stage Cluster Sampling

A simple random sample of ***n*** clusters (primary sampling units, PSUs) is taken from a population of ***N*** PSUs. All of the M_i elements (secondary sampling units, SSUs) within PSU ***i*** are observed, where each PSU could have a different number of elements.

	Unbiased Estimate	Variance	Standard Error
Total	$\hat{t}_{unb} = \frac{N}{n}\sum t_i$	$\hat{V}(\hat{t}_{unb}) = N^2\left(1-\frac{n}{N}\right)\frac{s_t^2}{n}$	$SE(\hat{t}_{unb}) = \sqrt{\hat{V}(\hat{t}_{unb})}$
Mean	$\hat{\bar{y}}_{unb} = \frac{\hat{t}_{unb}}{K}$		$SE(\hat{\bar{y}}_{unb}) = \frac{SE(\hat{t}_{unb})}{K}$

Note: Estimates are identified as unbiased as a notation device to signify cluster sampling.

Where $K = \sum M_i$;

Where t_i is known since all elements within PSU ***i*** are sampled;

Where $s_t^2 = \frac{1}{n-1}\sum\left(t_i - \frac{\hat{t}}{N}\right)^2$.

Weights can be used to compute estimates but cannot be used to compute standard errors:

$$w_{ij} = \frac{NM_i}{nm_i}$$

A simple random sample of ***n*** clusters (primary sampling units, PSUs) is taken from a population of ***N*** PSUs. In stage two, select a random sample of $\boldsymbol{m_i}$ from the population of $\boldsymbol{M_i}$ elements (secondary sampling units, SSUs) within PSU ***i***.

	Unbiased Estimate	Variance	Standard Error
Total	$\hat{t}_{unb} = \frac{N}{n}\sum \hat{t}_i$	$\hat{V}(\hat{t}_{unb}) = N^2\left(1-\frac{n}{N}\right)\frac{s_t^2}{n} + \frac{N}{n}\sum\left(1-\frac{m_i}{M_i}\right)M_i^2\frac{s_i^2}{m_i}$	$SE(\hat{t}_{unb}) = \sqrt{\hat{V}(\hat{t}_{unb})}$
Mean	$\hat{\bar{y}}_{unb} = \frac{\hat{t}_{unb}}{K}$		$SE(\hat{\bar{y}}_{unb}) = \frac{SE(\hat{t}_{unb})}{K}$

Note: Estimates are identified as unbiased as a notation device to signify cluster sampling.

Where $K = \sum M_i$;

Where $\hat{t}_i = M_i \bar{y}_i$;

Where $s_t^2 = \frac{\sum\left(\hat{t}_i - \frac{\hat{t}_{unb}}{N}\right)^2}{n-1}$ and $s_i^2 = \frac{\sum(y_{ij} - \bar{y}_i)^2}{m_i - 1}$.

Weights can be used to compute estimates but cannot be used to compute standard errors:

$$w_{ij} = \frac{NM_i}{nm_i}$$