STA304/1003: Summer 2014 - EXAM FORMULA SHEET

Some General Results:

1. $SE(pt.est) = \sqrt{\hat{V}(pt.est)}$

2. Approximate $100(1-\alpha)\%$ CI for a location parameter: $pt.est \pm z_{\alpha/2} SE(pt.est)$

3. Standard Normal Critical Values: $z_{0.005} = 2.58$, $z_{0.01} = 2.33$, $z_{0.025} = 1.96$, $z_{0.05} = 1.65$

4. SRS, STRS, Systematic, Two Stage Cluster point estimates are unbiased

5. Ratio estimates, One Stage Cluster sampling estimates are asymptotically unbiased

Simple Random Sampling without replacement (SRS) :

Population Parameter	Point Estimate	Variance of Point Estimate	Estimated Variance
$S^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y}_U)^2$	$s^2 = \frac{1}{n-1}\sum_{i\in\mathcal{S}}(y_i - \bar{y})^2$	—	—
$\bar{y}_U = \frac{1}{N}\sum_{i=1}^{N} y_i$	$\bar{y} = \frac{1}{n}\sum_{i\in\mathcal{S}} y_i$	$\left(1-\frac{n}{N}\right)\frac{S^2}{n}$	$\left(1-\frac{n}{N}\right)\frac{s^2}{n}$
$\tau = \sum_{i=1}^{N} y_i$	$\hat{\tau} = N\bar{y}$	$N^2\left(1-\frac{n}{N}\right)\frac{S^2}{n}$	$N^2\left(1-\frac{n}{N}\right)\frac{s^2}{n}$
$p = \bar{y}_U$; $y_i = \begin{cases} 1, & with\ probability\ p \\ 0, & with\ probability\ 1-p \end{cases}$	$\hat{p} = \bar{y}$	$\frac{N-n}{N-1}\frac{p(1-p)}{n}$	$\left(1-\frac{n}{N}\right)\frac{\hat{p}(1-\hat{p})}{n-1}$

Sample size for $100(1-\alpha)\%$ CI with margin of error, e :

$$n = \frac{z_{\alpha/2}^2 S^{*2}}{e^2 + \frac{z_{\alpha/2}^2 S^{*2}}{N}} = \frac{n_0}{1+\frac{n_0}{N}} \text{ ; where } n_0 = \left(\frac{z_{\alpha/2}S^*}{e}\right)^2$$

S^* is an estimate of S : either s or $\sqrt{p^*(1-p^*)}$, where p^* maximizes $p - p^2$.

Stratified Random Sampling (STRS):

Population Parameter	Point Estimate	Variance of Point Estimate	Estimated Variance
$\tau_i = \sum_{j=1}^{N_i} y_{ij}$	$\hat{\tau}_i = \frac{N_i}{n_i} \sum_{j\in\mathcal{S}_i} y_{ij} = N_i \bar{y}_i$	$N_i^2 \left(1 - \frac{n_i}{N_i}\right) \frac{S_i^2}{n_i}$	$N_i^2 \left(1 - \frac{n_i}{N_i}\right) \frac{s_i^2}{n_i}$
$\bar{y}_{iU} = \frac{1}{N_i} \sum_{j=1}^{N_i} y_{ij}$	$\bar{y}_i = \frac{1}{n_i} \sum_{j\in\mathcal{S}_i} y_{ij}$	$\left(1 - \frac{n_i}{N_i}\right) \frac{S_i^2}{n_i}$	$\left(1 - \frac{n_i}{N_i}\right) \frac{s_i^2}{n_i}$
$S_i^2 = \frac{1}{N_i-1} \sum_{j=1}^{N_i} (y_{ij} - \bar{y}_{iU})^2$	$s_i^2 = \frac{1}{n_i-1} \sum_{j\in\mathcal{S}_i} (y_{ij} - \bar{y}_i)^2$	—	—
$\tau = \sum_{i=1}^{L} \tau_i$	$\hat{\tau}_{str} = \sum_{i=1}^{L} \hat{\tau}_i = \sum_{i=1}^{L} N_i \bar{y}_i$	$\sum_{i=1}^{L} \left(1 - \frac{n_i}{N_i}\right) N_i^2 \frac{S_i^2}{n_i}$	$\sum_{i=1}^{L} \left(1 - \frac{n_i}{N_i}\right) N_i^2 \frac{s_i^2}{n_i}$
$\bar{y}_U = \frac{\tau}{N} = \frac{\sum_{i=1}^{L}\sum_{j=1}^{N_i} y_{ij}}{N}$	$\bar{y}_{str} = \frac{\hat{\tau}_{str}}{N} = \sum_{i=1}^{L} \frac{N_i}{N} \bar{y}_i$	$\sum_{i=1}^{L} \left(1 - \frac{n_i}{N_i}\right) \left(\frac{N_i}{N}\right)^2 \frac{S_i^2}{n_i}$	$\sum_{i=1}^{L} \left(1 - \frac{n_i}{N_i}\right) \left(\frac{N_i}{N}\right)^2 \frac{s_i^2}{n_i}$
$p = \bar{y}_U$; $y = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } 1-p \end{cases}$	$\hat{p}_{str} = \sum_{i=1}^{L} \frac{N_i}{N} \hat{p}_i$; where $\hat{p}_i = \bar{y}_i$ & $s_i^2 = \frac{n_i}{n_i-1} \hat{p}_i(1-\hat{p}_i)$	$\sum_{i=1}^{L} \left(\frac{N_i}{N}\right)^2 \frac{p_i(1-p_i)}{n_i} \frac{N_i-n_i}{N_i-1}$	$\sum_{i=1}^{L} \left(1 - \frac{n_i}{N_i}\right) \left(\frac{N_i}{N}\right)^2 \frac{\hat{p}_i(1-\hat{p}_i)}{n_i-1}$

Strata Sample Sizes:

Proportional Allocation	Optimal Allocation	Neyman Allocation
$n_i = n \frac{N_i}{\sum_{\ell=1}^{L} N_\ell}$	$n_i = \left(\frac{\frac{N_i S_i}{\sqrt{c_i}}}{\sum_{\ell=1}^{L} \frac{N_\ell S_\ell}{\sqrt{c_\ell}}} \right) n$	$n_i = \left(\frac{N_i S_i}{\sum_{\ell=1}^{L} N_\ell S_\ell} \right) n$

Sample size for $100(1-\alpha)\%$ CI with margin of error, e :

$n = \frac{z_{\alpha/2}^2 v}{e^2}$; where $v = \sum_{i=1}^{L} \frac{n}{n_i} \left(\frac{N_i}{N}\right)^2 S_i^{*2}$ and S_i^* is an estimate of S_i

Measures of Homogeneity within Groups/Clusters:

$R_a^2 = 1 - \frac{MSW}{S^2}$ $\qquad$ $ICC = 1 - \frac{m}{m-1} \frac{SSW}{SSTO}$; where $\frac{-1}{m-1} \le ICC \le 1$

Ratio Estimation:

$$e_i = y_i - r x_i$$

$$s_r^2 = \frac{1}{n-1} \sum_{i \in \mathcal{S}} e_i^2$$

Parameter	**Point Estimate**	**Estimated Variance**
$R = \frac{\bar{y}_U}{\bar{x}_U} = \frac{\tau_y}{\tau_x}$	$r = \hat{R} = \frac{\bar{y}}{\bar{x}} = \frac{\hat{\tau}_y}{\hat{\tau}_x}$	$\left(1 - \frac{n}{N}\right) \frac{s_r^2}{n\bar{x}^2}$
τ_y	$\hat{\tau}_{yr} = r\ \tau_x$	$\left(1 - \frac{n}{N}\right) \left(\frac{\tau_x}{\bar{x}}\right)^2 \frac{s_r^2}{n}$
$\bar{y}_U$	$\hat{\bar{y}}_r = r\ \bar{x}_U$	$\left(1 - \frac{n}{N}\right) \left(\frac{\bar{x}_U}{\bar{x}}\right)^2 \frac{s_r^2}{n}$

Cluster Sampling:

Parameter	**Point Estimate**	**Variance of Point Estimate**	**Estimated Variance**
τ_y	$\hat{\tau}_{unb} = \frac{N}{n} \sum_{i \in \mathcal{S}} \hat{\tau}_i = \frac{N}{n} \sum_{i \in \mathcal{S}} M_i \bar{y}_i$	$N^2 \left(1 - \frac{n}{N}\right) \frac{S_b^2}{n} + \frac{N}{n} \sum_{i=1}^{N} \left(1 - \frac{m_i}{M_i}\right) M_i^2 \frac{S_i^2}{m_i}$	$N^2 \left(1 - \frac{n}{N}\right) \frac{s_b^2}{n} + \frac{N}{n} \sum_{i \in \mathcal{S}} \left(1 - \frac{m_i}{M_i}\right) M_i^2 \frac{s_i^2}{m_i}$
$\bar{y}_U$ if M is known	$\hat{\bar{y}}_{unb} = \frac{\hat{\tau}_{unb}}{M}$	$V(\hat{\tau}_{unb})/M^2$	$\frac{1}{n\overline{M}^2} \left(1 - \frac{n}{N}\right) s_b^2 + \frac{1}{nN\overline{M}^2} \sum_{i \in \mathcal{S}} M_i^2 \left(1 - \frac{m_i}{M_i}\right) \frac{s_i^2}{m_i}$
$\bar{y}_U$ if M is unknown	$\hat{\bar{y}}_r = \frac{\sum_{i \in \mathcal{S}} \hat{\tau}_i}{\sum_{i \in \mathcal{S}} M_i} = \frac{\sum_{i \in \mathcal{S}} M_i \bar{y}_i}{\sum_{i \in \mathcal{S}} M_i}$	—	$\frac{1}{n\overline{M}^2} \left(1 - \frac{n}{N}\right) s_r^2 + \frac{1}{nN\overline{M}^2} \sum_{i \in \mathcal{S}} M_i^2 \left(1 - \frac{m_i}{M_i}\right) \frac{s_i^2}{m_i}$ where $s_r^2 = \frac{1}{n-1} \sum_{i \in \mathcal{S}} (M_i \bar{y}_i - M_i \hat{\bar{y}}_r)^2$

$S_b^2 = \frac{1}{N} \sum_{i=1}^{N} (\tau_i - \frac{\tau}{N})^2$ and $s_b^2 = \frac{1}{n-1} \sum_{i \in \mathcal{S}} (M_i \bar{y}_i - \overline{M} \hat{\bar{y}}_{unb})^2$

Systematic Sampling Variance: $V(\hat{\bar{y}}_{sys}) = \left(1 - \frac{1}{k}\right) \frac{S_t^2}{n^2} = \left(1 - \frac{1}{k}\right) \frac{MSB}{n} \approx \frac{S^2}{n} [1 + (n-1) ICC]$