Monday, September 22

Note: Rather than creating extra homework problems on the topics from today's lecture, if there are any problems on the examination from today's lecture I will use the examples from this lecture (although I may change the numbers).

Inferences Concerning Stratum Parameters

When strata correspond to one or more domains of interest, additional inferences concerning those domains can easily be done when using stratified random sampling.

1. Estimation of μ_i or τ_i in one stratum.
2. Estimation of mean or total for several strata combined.
3. Estimation of the difference in the mean or total between two strata.

Estimation of μ_i or τ_i for One Stratum

With stratified random sampling, the sampling design for obtaining the sample from *each stratum* is *simple random sampling*. So inferences concerning a stratum mean or total use the results from simple random sampling. We have

$$\hat{\mu}_i = \bar{y}_i \quad \text{and} \quad V(\hat{\mu}_i) = \left(1 - \frac{n_i}{N_i}\right)\frac{\sigma_i^2}{n_i},$$

and $\hat{\tau}_i = N_i\bar{y}_i$ and $V(\hat{\tau}_i) = N_i^2 V(\hat{\mu}_i)$, noting that if we need to estimate the variance we replace σ_i^2 with s_i^2 and V with $\hat{V}$.

Example: Suppose we have the following results from a survey that used stratified random sampling.

i	N_i	n_i	$\bar{y}_i$	s_i
1	1000	100	36	6
2	2000	200	25	5
3	3000	300	16	4

What is $\hat{\mu}_1$ and the estimate of the variance of that estimator? What is $\hat{\tau}_1$ and the estimate of the variance of that estimator?

Estimation of the Mean or Total for Several Combined Strata

Suppose we want to estimate $\mu_{i,j}$, the mean of strata i and j *combined*. The estimator is

$$\hat{\mu}_{i,j} = \frac{N_i}{N_i + N_j}\bar{y}_i + \frac{N_j}{N_i + N_j}\bar{y}_j$$

which has variance

$$V(\hat{\mu}_{i,j}) = \left(\frac{N_i}{N_i + N_j}\right)^2 V(\hat{\mu}_i) + \left(\frac{N_j}{N_i + N_j}\right)^2 V(\hat{\mu}_j).$$

To compute $\hat{\tau}_{i,j}$ we would use

$$\hat{\tau}_{i,j} = \hat{\tau}_i + \hat{\tau}_j = N_i\bar{y}_i + N_j\bar{y}_j,$$

which has variance

$$V(\hat{\tau}_{i,j}) = V(\hat{\tau}_i) + V(\hat{\tau}_j) = N_i^2 V(\hat{\mu}_i) + N_j^2 V(\hat{\mu}_j).$$

Example: Suppose we have the following results from a survey that used stratified random sampling.

i	N_i	n_i	$\bar{y}_i$	s_i
1	1000	100	36	6
2	2000	200	25	5
3	3000	300	16	4

What is the estimate of $\mu_{2,3}$ and the estimated variance of that estimator?

More generally, we can do this for any number of strata. For example, to estimate to mean of strata i, j, and k combined, we can use the estimator

$$\hat{\mu}_{i,j,k} = \frac{N_i}{N_i + N_j + N_k}\bar{y}_i + \frac{N_j}{N_i + N_j + N_k}\bar{y}_j + \frac{N_k}{N_i + N_j + N_k}\bar{y}_k,$$

which as variance

$$V(\hat{\mu}_{i,j,k}) = \left(\frac{N_i}{N_i + N_j + N_k}\right)^2 V(\hat{\mu}_i) + \left(\frac{N_j}{N_i + N_j + N_k}\right)^2 V(\hat{\mu}_j) + \left(\frac{N_k}{N_i + N_j + N_k}\right)^2 V(\hat{\mu}_k).$$

And the total for those strata combined, $\tau_{i,j,k}$, is estimated as

$$\hat{\tau}_{i,j,k} = \hat{\tau}_i + \hat{\tau}_j + \hat{\tau}_k$$

which has variance

$$V(\hat{\tau}_{i,j,k}) = V(\hat{\tau}_i) + V(\hat{\tau}_j) + V(\hat{\tau}_j).$$

Estimation of a Difference Between Strata

Suppose we want to estimate $\mu_i - \mu_j$. The estimator is simply

$$\hat{\mu}_i - \hat{\mu}_j = \bar{y}_i - \bar{y}_j,$$

which has variance

$$V(\hat{\mu}_i - \hat{\mu}_j) = V(\hat{\mu}_i) + V(\hat{\mu}_j).$$

Similarly if we want to estimate $\tau_i - \tau_j$ the estimator is

$$\hat{\tau}_i - \hat{\tau}_j = N_i\bar{y}_i - N_j\bar{y}_j,$$

which has variance

$$V(\hat{\tau}_i - \hat{\tau}_j) = V(\hat{\tau}_i) + V(\hat{\tau}_j).$$

Note that although we are *subtracting* estimators, the variances are still *additive*.

Example: Suppose we have the following results from a survey that used stratified random sampling.

i	N_i	n_i	$\bar{y}_i$	s_i
1	1000	100	36	6
2	2000	200	25	5
3	3000	300	16	4

What is the estimate of $\mu_1 - \mu_3$ and what is the estimated variance of the estimator $\hat{\mu}_1 - \hat{\mu}_3$?

Note: We can also do these kinds of inferences with post-stratification, or for stratified random sampling where the domains of interest do not correspond to the strata. The estimators are the same but the variances are different.