7.3 Lecture 7 Wednesday 01/31/01

Sectionning: see the logictics section

7.4 Review of last time, if you missed see: Lecture 6

7.4.1 A Ratio as a tool for estimating a mean

$$\bar{Y}_R = \mu_x \frac{\bar{Y}}{\bar{X}} = \mu_x R$$

This is just a constant times R so that :

Theorem 1 (4.C)

$$Var(\bar{Y}_R) = \frac{1}{n}(1 - \frac{n-1}{N-1})(r^2\sigma_x^2 + \sigma_y^2 - 2r\rho\sigma_x\sigma_y)$$

and we DO have a little bias :

Theorem 2 (4.D)

$$E(R) - r \doteq \frac{1}{\mu_x}\frac{1}{n}(1 - \frac{n-1}{N-1})(r\sigma_x^2 - \rho\sigma_x\sigma_y)$$

In order to do a simple comparison with the ordinary estimate of μ: $\bar{Y}$.
We will leave out all the finite population corrections.

$$Var(\bar{Y}) = \frac{\sigma_y^2}{n}$$

to compare with

$$Var(\bar{Y}_R) = \frac{\sigma_y^2}{n} + \frac{1}{n}(r^2\sigma_x^2 - 2r\rho\sigma_x\sigma_y)$$

The latter will be smaller if the () is negative, i.e. the ratio estimator will be better (in the sense of having a lower variance) if

$$r^2\sigma_x^2 - 2r\rho\sigma_x\sigma_y < 0$$

if r is positive this is equivalent to

$$\begin{aligned} r\sigma_x &< 2\rho\sigma_y \\ \frac{\sigma_x}{\mu_x} &< 2\rho\frac{\sigma_y}{\mu_y} \\ \rho &> \frac{C_x}{2C_y} \end{aligned}$$

The C's are the coefficients of variation, in some sense they are rescaled standard deviations it is true that a standard deviation of 10 means something totally different whether the mean is 100 or 100,000.

As before, we won't have on hand all the parameters and we "plug in" their estimates.

$100(1-\alpha)$ % confidence interval for μ_y: $(\bar{Y}_R + / - z_{\frac{\alpha}{2}} S_{\bar{Y}_R})$

Example :

As we actually KNOW the parameters in the hospital case, let's do the computation of $\sigma_{\bar{Y}_R}$.

We see that for the same vaiance we only need a sample of about 1/5 of the size.

There is a small bias. -we could go on to actually compute the MSE error in both cases, it is better in the second case.

7.5 Stratified Random Sampling

There are many ways of reducing varaibility of an estimator, the one we are going to talk abbout now is based on the following fundamental relationship (intra/inter):

$$\text{Total Variability} = \text{Between Group Var.} + \text{Within Group Var.}$$

Replace in this case group by strata and you have the principle behind stratified sampling designs, they try to make the var. as small as possible within each strate.

Often one has a natural categorical variable that provides the strata : sex, areas,..

But we can also build them artificially from continuous varaibles : big/medium/small.

From our box we see that good strata are with homogeneity within and variation between.

7.5.1 Notations and Properties

The Population parameters:

We suppose we have an intial population of N observations we are to decompose into L strata, each is of size N_ℓ, so that :

$$N = N_1 + N_2 + \cdots + N_L$$

The subpopulation of N_ℓ units is called the the ℓth strata. It's mean will be denoted μ_ℓ and it's variance σ_ℓ^2. We will use $x_{i\ell}$ to denote the ith population value of the ℓth stratum.

Wehave the overall mean

$$\mu = \sum_{\ell=1}^{L} \sum_{i=1}^{N_\ell} x_{i\ell} = \frac{1}{N} \sum_{i=1}^{N_\ell} \mu_\ell = \sum_{i=1}^{N_\ell} W_\ell \mu_\ell$$

where $W_\ell = \frac{N_\ell}{N}$

The estimates:

$$\bar{X}_\ell = \frac{1}{n_\ell} \sum_{i=1}^{n_\ell} X_{i\ell}$$

We see immediately that

$$E(\bar{X}_\ell) = \mu_\ell$$

We will take

$$\bar{X}_S = \sum W_\ell X_\ell$$

Theorem 3 *$\bar{X}_S$ is an unbiased estimate of μ*

We assume that the samples from the different populations are independent of one another and each stratum undergos a simple random sampling whose variance we know to be :

$$\frac{1}{n_\ell}(1 - \frac{n_\ell - 1}{N_\ell - 1})\sigma_\ell^2$$

this leads immediately to

Theorem 4

$$Var(\bar{X}_S) = \sum_{\ell=1}^{n_\ell} W_\ell^2 (\frac{1}{n_\ell})(1 - \frac{n_\ell - 1}{N_\ell - 1})\sigma_\ell^2$$

Example from book :page 216

Actually of course, in order to estimate the variance, we do NOT have the information about the parameters so we use the unbiased estiamte for the variance of each strata:

$$s_\ell^2 = \sum_{i=1}^{n_\ell} (X_{i\ell} - \bar{X}_\ell)^2$$

7.5.2 Design of Stratification: Allocation Methods

We will make a preliminary simplification, we are going to ignore the finite population correction, and to begin with we will do a theoretical analysis, suppose we knew the parameters σ_ℓ^2, how should we choose the n_ℓ's so to minimize the variance of $\bar{X}_S$. Of course if there is no limitation we know that increase in precision is obtained by increasing the n_ℓ's just make them enormous. In real situations however the amount of surveying possible is limited in advance so we will have to submit ourselves to the constraint:

$$n_1 + n_2 + \cdots + n_L = n$$

, so we want to minimize

$$Var(\bar{X}_S) = \sum_{\ell=1}^{L} \frac{W_\ell^2 \sigma_\ell^2}{nl}$$

Theorem 5 *The sample sizes $n_1, \ldots, n_L$ that minimize $var(\bar{X}_S)$ subject to the constraint $n_1 + n_2 + \cdots + n_L = n$ are given by :*

$$n_\ell = n \frac{W_\ell \sigma_\ell}{\sum_\ell W_\ell \sigma_\ell}$$

Proof: This uses something we have not, or you may not seen; Lagrange multipliers, here is a little reminder of how they work, without any proofs or unnecessary justifications:

One-dimensional case : Suppose we need:

1. To minimize f(x)
2. Have the constraint g(x)=0

We use a new function:

$$L(x, \lambda) = f(x) + \lambda g(x)$$

Solve for x as a function of λ by posing

$$\frac{\partial L}{\partial x} = 0 = f'(x) + \lambda g'(x)$$

and then solve $g(x) = 0$.

In more dimensions , it is the same, in our example the variables are $n_1, n_2, \ldots, n_L$ and the condition can be written $g(n_1, \ldots, n_L) = n_1 + n_2 + \cdots n_L - n$, the function we want to minimize is

$$f(n_1, n_2, \ldots, n_L) = \frac{\sum_{\ell=1}^{L} W_\ell^2 \sigma_\ell^2}{n_\ell}$$

so that our lagrange function is:

$$L(n_1, n_2, \ldots, n_\ell, \lambda) = \frac{\sum_{\ell=1}^{L} W_\ell^2 \sigma_\ell^2}{n_\ell} + \lambda(\sum_{\ell=1}^{L} n_\ell - n)$$

for $\ell = 1, \ldots L$,

$$\frac{\partial L}{\partial n_\ell} = -\frac{W_\ell^2 \sigma_\ell^2}{n_\ell^2} + \lambda$$

which gives $n_\ell = \frac{W_\ell \sigma_\ell}{\sqrt{\lambda}}$ the constraint gives:

$$n = \frac{1}{\sqrt{\lambda}} \sum W_\ell \sigma_\ell \Longrightarrow \frac{1}{\sqrt{\lambda}} = \frac{n}{\sum W_\ell \sigma_\ell}$$

$$n_\ell = n \frac{W_\ell \sigma_\ell}{\sum_\ell W_\ell \sigma_\ell}$$

What does it say ? More variability, sample more.

This optimal scheme is called Neyman allocation.

Corollary:

Denote by $\bar{X}_{SO}$ the estimate obtained by this optimal strategy, then

$$Var(\bar{X}_{SO}) = \frac{(\sum_\ell W_\ell \sigma_\ell)^2}{n}$$

Proof:

$$\begin{aligned} Var(\bar{X}_{SO}) &= \textstyle\sum_\ell \frac{W_\ell^2 \sigma_\ell^2}{n_\ell} \\ &= \frac{1}{n}(\textstyle\sum_\ell \frac{W_\ell^2\sigma_\ell^2}{W_\ell\sigma_\ell})(\sum_\ell W_\ell\sigma_\ell) \\ &= \frac{(\sum_\ell W_\ell\sigma_\ell)^2}{n} \end{aligned}$$

If we apply this to four states of the hospital example, then we get:

A	B	C	D
.106	.210	.250	.434

Minimize the variance (because of MSE), then we have a constrained optimization problem that we have solved with Lagrange multipliers.

However this is usually difficult to apply because we don't often have the information about the variances in each strata.

Proportional allocation is when :

$$\frac{n_1}{N_2} = \frac{n_2}{N_2} = \cdots = \frac{n_\ell}{N_\ell} = \cdots = \frac{n_L}{N_L}$$

This gives

$$n_\ell = n\frac{n_\ell}{N_\ell} = nW_\ell$$

Some algebra leads to

Theorem 6

$$Var(\bar{X}_{SP}) - Var(\bar{X}_{SO}) = \frac{1}{n}\sum_{\ell=1}^{L} W_\ell(\sigma_\ell - \bar{\sigma})^2$$

Where $\bar{\sigma} = \sum_{\ell=1}^{L} W_\ell\sigma_\ell$.

In the example it's about 20% larger.

7.5.3 Concluding remarks

We have considered srs for the estimation of a parameter μ.

WE obtained a random variable $\bar{X}$ which estimates μ,

which is an example of the general case where $\hat{\theta}$ estimates θ and we evaluate its precision by looking at the $SE(\hat{\theta})$ which is just another name for the standard deviation of the random variable $\hat{\theta}$.

Often by linearization or just because it is a sum of iid components :

$$\hat{\theta} \sim \mathcal{N}$$

this enables confidence statements, however for them to be useful we have to estimate $\hat{\theta}$'s SE, this is done using $s^2_{\hat{\theta}}$ and thus we produce confidence intervals for θ :

$$\hat{\theta} \pm s_{\hat{\theta}} \times z_{\alpha/2} \text{ is a } (1-\alpha)100\% \text{ confidence interval}$$