

Question was: How many strata?

Stat 576
4-29-25

Assume that Dalenius-Hudges demarcation
and Neyman allocation will be used.

①

$$V(\bar{y}_{str}) \approx \frac{1}{n} \left(\sum_{h=1}^H W_h S_h \right)^2$$
$$\approx \frac{1}{n} \left(\sum_{h=1}^H \frac{a_h^2}{\sqrt{12}} \right)^2 \quad \text{where } a_h = \sqrt{W_h} (y_h - y_{hr})$$

$$\text{let } \bar{a} = \frac{1}{H} \sum_{h=1}^H a_h$$

If D-H demarcation is carried out
correctly, all of the a_h 's will equal
a constant $\bar{a}$

②

$$\text{So } V[\bar{y}_{str}] \approx \frac{1}{n} \left(\sum_{h=1}^H \frac{\bar{a}^2}{\sqrt{12}} \right)^2$$
$$= \frac{1}{n} \left(H \frac{\bar{a}^2}{\sqrt{12}} \right)^2 = \frac{H^2 \bar{a}^4}{12n}$$
$$= \frac{H^2 \left(\frac{1}{H} \sum_{h=1}^H a_h \right)^4}{12n} = \frac{1}{H^2 12n} \left(\sum_{h=1}^H \sqrt{W_h} (y_h - y_{hr}) \right)^4$$

For planning purposes, suppose that the stratum widths will be constant = 1

(3)

$$\text{Now } V(\bar{y}_{st}) \approx \frac{1}{12nH^2} l^4 \left(\sum_{h=1}^H \sqrt{f_h} \right)^4$$

Say that we are given a desired margin of error, E

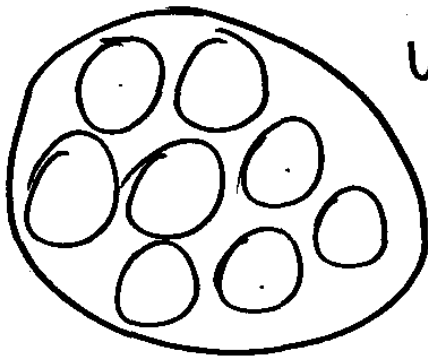
$$\text{Set } E = z_{\alpha/2} \sqrt{\frac{l^4}{12nH^2} \left(\sum_{h=1}^H \sqrt{f_h} \right)^4}$$

+ solve for H

$$H = \frac{z_{\alpha/2} l^2 \left(\sum_{h=1}^H \sqrt{f_h} \right)^2}{E \sqrt{12n}}$$

(4)

Cluster Sampling



Use a SRSWOR of clusters, and take every item in those clusters

PSU = primary sampling units = clusters (5)

SSU = secondary sampling units = items in the clusters

N = # of PSUs in the population
= # clusters in population

n = # clusters in sample

M_i = # of items in cluster i

$K = \sum_{i=1}^N M_i$ = # items in population

t_i = total of the Y -values in i th cluster (6)

$t = \sum_{i=1}^N t_i$ = population total of Y -values

$\bar{Y} = \frac{t}{K}$, $\bar{T} = \frac{t}{N}$ = Average cluster total
= population mean

$\bar{y}_i = \frac{t_i}{M_i}$ = cluster mean

$S^2 = \frac{\sum_{i=1}^N \sum_{j=1}^{M_i} (y_{ij} - \bar{Y})^2}{K-1}$

Say we want to estimate t .

⑦

$$\begin{aligned}\text{Try } \hat{t} &= N \bar{t} \\ &= N \frac{1}{n} \sum_{i=1}^n t_i \\ &= \frac{N}{n} \sum_{i=1}^n \sum_{j=1}^{m_i} y_{ij}\end{aligned}$$

$$\begin{aligned}E[\hat{t}] &= E[N \bar{t}] = N E[\bar{t}] = \\ &= N \bar{T} = N \left(\frac{t}{N} \right) = t\end{aligned}$$

$$\begin{aligned}V[\hat{t}] &= V[N \bar{t}] = N^2 V[\bar{t}] \\ &= N^2 \frac{S_t^2}{n} \left(1 - \frac{n}{N} \right), \text{ where}\end{aligned}$$

$$S_t^2 = \frac{\sum_{i=1}^n (t_i - \bar{T})^2}{N-1}$$

$$\begin{aligned}\hat{V}[\hat{t}] &= N^2 \frac{s_t^2}{n} \left(1 - \frac{n}{N} \right), \text{ where} \\ s_t^2 &= \frac{\sum_{i=1}^n (t_i - \bar{t})^2}{n-1}\end{aligned}$$

⑧

ANOVA decomposition:

(9)

$$\begin{aligned} \sum_{i=1}^N \sum_{j=1}^{M_i} (Y_{ij} - \bar{Y})^2 &= \sum_{i=1}^N \sum_{j=1}^{M_i} (Y_{ij} - \bar{Y}_i + \bar{Y}_i - \bar{Y})^2 \\ &= \underbrace{\sum_{i=1}^N \sum_{j=1}^{M_i} (Y_{ij} - \bar{Y}_i)^2}_{SSW} + \underbrace{\sum_{i=1}^N M_i (\bar{Y}_i - \bar{Y})^2}_{SSB} + 0 \end{aligned}$$

Since $V[\hat{t}]$ involves a term similar to SSB ,
we want SSW to be large & SSB to be small.

That is, we want heterogeneity within
& homogeneity between clusters

(10)

To estimate $\bar{Y}$, use $\bar{y}_{clus} = \frac{\hat{t}}{K}$

$$E[\bar{y}_{clus}] = \frac{E[\hat{t}]}{K} = \frac{t}{K} = \bar{Y}$$

$$V[\bar{y}_{clus}] = \frac{1}{K^2} V[\hat{t}] = \frac{1}{K^2} N^2 \frac{S_t^2}{n} \left(1 - \frac{1}{N}\right)$$

$$\hat{V}[\bar{y}_{\text{clus}}] = \frac{1}{K^2} N^2 \frac{S_t^2}{n} \left(1 - \frac{1}{N}\right)$$

(11)

Special case: $M_i \equiv M$

Then $K = NM$

$$\bar{y}_{\text{clus}} = \frac{\hat{t}}{K} = \frac{N\bar{t}}{NM} = \frac{\bar{t}}{M}$$

$$V[\bar{y}_{\text{clus}}] = \frac{1}{\cancel{N^2} M^2} \cancel{N^2} \frac{S_t^2}{n} \left(1 - \frac{1}{N}\right) = \frac{1}{M^2} \frac{S_t^2}{n} \left(1 - \frac{1}{N}\right)$$