Stat 576
5-20-25

Design effect

$$\text{deff} = \frac{V(\text{estimator} \mid \text{complex design})}{V(\text{estimator} \mid \text{SRS of same size})}$$ ①

$$\text{deft} = \frac{\text{S.E.}(\text{estimator} \mid \text{complex design})}{\text{S.E}(\text{estimator} \mid \text{SRS of same size})}$$

Example: Stratified sampling with proportional allocation

$$V[\bar{y}_{str}] = \frac{1}{N^2} \sum_{h=1}^{H} N_h^2 \frac{S_h^2}{n_h}\left(1 - \frac{n_h}{N_h}\right)$$

and $n_h = \frac{N_h}{N} n$

②

$$V[\bar{y}_{str}] = \frac{1}{\cancel{N^2}} \sum_{h=1}^{H} \cancel{N_h^2} \frac{S_h^2}{\left(\frac{\cancel{N_h} n}{\cancel{N}}\right)}\left(1 - \frac{\frac{\cancel{N_h} n}{N}}{\cancel{N_h}}\right)$$

$$= \sum_{h=1}^{H} W_h \frac{S_h^2}{n}\left(1 - \frac{n}{N}\right)$$

Also $V[\bar{y}] = \frac{S^2}{n}\left(1 - \frac{n}{N}\right)$

$$\Rightarrow \quad \text{deff} = \frac{\sum_{h=1}^{H} W_h S_h^2}{S^2}$$

$$\widehat{\text{deff}} = \sum_{h=1}^{H} W_h s_h^2 \Big/ s^2$$

Estimating a population size
(4 methods)

Method 1: Direct sampling (capture/recapture)

Collect and tag t items
Release and allow to remix
Collect a new sample of size n.
Count the # of tagged items in the sample,
+ call it s.

s has a hypergeometric distribution

Provided that N is large, the hypergeometric distribution can be approximated by the binomial distribution.

Assume that $s \sim \text{Bino}(n, \frac{t}{N})$

So $E(s) = \frac{nt}{N}$. We want to estimate N.

Try $\hat{N} = \frac{nt}{s}$. This is a ratio estimator.

So $\hat{N}$ will be biased, but asymptotically unbiased.

Recall $V\left[\frac{\bar{y}}{\bar{x}}\right] = \frac{1}{\bar{X}^2}\frac{S_e^2}{n}\left(1-\frac{1}{N}\right)$, where

$$S_e^2 = S_y^2 + \left(\frac{\bar{Y}}{\bar{X}}\right)^2 S_x^2 - 2\left(\frac{\bar{Y}}{\bar{X}}\right) S_{xy}$$

Write $\hat{N} = \frac{nt}{s}$ as $\frac{t}{s/n} = \frac{t}{\hat{p}}$ $\quad$ t plays the role of $\bar{y}$

$\hat{p}$ " " " " $\bar{x}$

$S_y^2 = 0$, $S_{xy} = 0$ since t is a constant

$S_x^2 = V[\hat{p}] = \frac{Npq}{N-1}$

So $S_e^2 = \left(\frac{t}{t/N}\right)^2 \frac{Npq}{N-1}$

$$\text{Now } V[\hat{N}] = \frac{1}{(t/N)^2}\left(\frac{t}{t/N}\right)^2 \frac{N\overset{t/N}{p}q}{N-1}\frac{1}{n}\left(1-\frac{1}{N}\right)$$

$$= \frac{1}{t/N}\left(\frac{t}{t/N}\right)^2 \frac{N}{N-1}\left(1-\frac{t}{N}\right)\frac{1}{n}\left(1-\frac{1}{N}\right)$$

$$= \frac{N^4}{N-1}\frac{\left(1-t/N\right)}{t}\frac{1}{n}\underbrace{\left(1-\frac{1}{N}\right)}_{\text{assume negligible}}$$

$$\hat{V}[\hat{N}] = \hat{N}^3 \frac{\left(1-t/\hat{N}\right)}{t}\frac{1}{n}$$

$$= \left(\frac{nt}{s}\right)^3 \frac{1}{t}\left(1-\frac{t}{nt/s}\right)\frac{1}{n}$$

$$\hat{V}[\hat{N}] = \frac{n^2 t^2}{s^3}\left(1 - \frac{s}{n}\right)$$

$$= \frac{n t^2 (n-s)}{s^3}$$

Method 2: Inverse sampling (capture/recapture)

Collect and tag t items

Release & allow to remix

Resample until s tags are found

Now n is the random variable

n has a negative hypergeometric distribution

n will have, approximately, a negative binomial (Pascal) distribution

$$n \sim NB(s, t/N) \qquad E(n) = \frac{s}{t/N}$$

$$V(n) = \frac{s\left(1 - \frac{t}{N}\right)}{(t/N)^2}$$

$$E(n) = \frac{sN}{t}$$

$$\text{Try } \hat{N} = \frac{nt}{s} \qquad E[\hat{N}] = \frac{t}{s} E(n) = N$$

unbiased!!

$$V[\hat{N}] = V\left[\frac{nt}{s}\right] = \frac{t^2}{s^2} V[n] = \frac{t^2}{s^2} \frac{s\left(1-\frac{t}{N}\right)}{\left(t/N\right)^2}$$

$$= \frac{N^2}{s}\left(1-\frac{t}{N}\right)$$

$$\hat{V}[\hat{N}] = \frac{\hat{N}^2}{s}\left(1-\frac{t}{\hat{N}}\right) = \frac{\left(\frac{nt}{s}\right)^2}{s}\left(1-\frac{t}{nt/s}\right)$$

$$= \frac{n^2t^2}{s^3}\left(1-\frac{s}{n}\right)$$

$$= \frac{n\,t^2(n-s)}{s^3}$$ (same as for direct sampling)

Method 3: Quadrat sampling

Divide the total area into N equal-sized sections, call quadrats

Select n of the quadrats using SRSWOR

Let $\lambda = \text{true density} = \frac{M \;\leftarrow \text{pop. size}}{A \;\nwarrow \text{total area}}$

Count the # of item in each of the sampled quadrats. Let X_i = count in quadrat i

⑪

X_i has a Poisson distribution with parameter $\frac{M}{N}$

$E[X_i] = \frac{M}{N}$, $V[X_i] = \frac{M}{N}$

So $E[\bar{x}] = \frac{M}{N}$, $V[\bar{x}] = \frac{M/N}{n}\left(1 - \frac{1}{N}\right)$

Let $\hat{M} = N\bar{x}$

Then $E[\hat{M}] = N\,E(\bar{x}) = M$ <u>unbiased</u>

⑫

$$V[\hat{M}] = N^2 V[\bar{x}] = \frac{N^2 \frac{M}{N}}{n}\left(1 - \frac{1}{N}\right)$$

$$= \frac{NM}{n}\left(1 - \frac{1}{N}\right)$$

$$\hat{V}[\hat{M}] = \frac{N\hat{M}}{n}\left(1 - \frac{1}{N}\right)$$

$$= \frac{N(N\bar{x})}{n}\left(1 - \frac{1}{N}\right)$$

$$= \frac{N^2\bar{x}}{n}\left(1 - \frac{1}{N}\right)$$