Population size estimates

Stat 576

11-21-17

①

Method 4: Stocked quadrats

Divide the total area A into N equal sections

Select n of these by SRSWOR

Record $x_i = \begin{cases} 1 & \text{present} \\ 0 & \text{absent} \end{cases}$ for each quadrat

Assume that X_i actually has a Poisson distribution.

Its parameter $\mu = \frac{M}{N}$

②

$$P(X_i = 0) = \frac{\mu^0 e^{-\mu}}{0!} = e^{-\mu}$$

$$\overset{\shortparallel}{q}$$

$$\ln q = -\mu = -\frac{M}{N}$$

so $M = -N \ln q$

Let $\hat{M} = -N \ln \hat{q}$,

where $\hat{q} = \frac{\#\text{ unstocked quadrats}}{n}$

③

Let $f(x) = -N\ln x$

$$f'(x) = -\frac{N}{x} \qquad f''(x) = \frac{N}{x^2}$$

$$f(x) \approx f(x_0) + f'(x_0)(x-x_0) + \frac{f''(x_0)}{2}(x-x_0)^2$$

$$-N\ln x \approx -N\ln x_0 - \frac{N}{x_0}(x-x_0) + \frac{N}{2x_0^2}(x-x_0)^2$$

$$\therefore\ \hat{M} = -N\ln\hat{q} \approx -N\ln q - \frac{N}{q}(\hat{q}-q) + \frac{N}{2q^2}(\hat{q}-q)^2$$

④

$$E[\hat{M}] \approx -N\ln q - 0 + \frac{N}{2q^2}V[\hat{q}]$$

$$= M + \frac{N}{2q^2}V[1-\hat{p}]$$

$$= M + \frac{N}{2q^2}V[\hat{p}]$$

$$E[\hat{M}] \approx M + \underbrace{\frac{N}{2q^2}\frac{pq}{n}\left(1-\frac{1}{N}\right)\frac{N}{N-1}}_{\text{Bias}}$$

$$V[\hat{M}] \approx V\left[-N\ln q - \frac{N}{q}(\hat{q}-q)\right]$$

$$= \frac{N^2}{q^2} V[\hat{q}]$$

$$= \frac{N^2}{q^2} V[\hat{p}] = \frac{N^2}{q^2}\,\frac{pq}{n}\left(1-\frac{n}{N}\right)\frac{N}{N-1}$$

$$= \frac{N^3 p}{nq\,(N-1)}\left(1-\frac{n}{N}\right)$$

$$\hat{V}[\hat{M}] = \frac{N^3\hat{p}}{n\hat{q}\,(N-1)}\left(1-\frac{n}{N}\right)$$

Nonresponse

Think of the population as being stratified

Stratum	size	total	mean	variance
Respondents	N_R	t_R	$\bar{Y}_R$	S_R^2
Nonrespond.	N_M	t_M	$\bar{Y}_M$	S_M^2

We wish to estimate $\bar{Y} = \frac{t}{N} = \frac{t_R + t_M}{N}$

$$= \frac{N_R\bar{Y}_R + N_M\bar{Y}_M}{N} = W_R\bar{Y}_R + W_M\bar{Y}_M$$

7

If nonresponse is ignored, then

$\bar{y}$ is estimating $\bar{Y}_R$ and

s^2 is estimating S_R^2

Solution: Do a second round: call-back

Suppose that some proportion λ of the nonrespondents now respond.

Summary: 1 Randomly select n items from N

Find n_R respondents & n_M nonrespondents

$n_R + n_M = n$

8

2 Do a callback on the n_M nonrespondents and λn_M now respond

Final sample size is $n_R + \lambda n_M$

Construct the Horvitz-Thompson estimator

$$\hat{t}_{HT} = \sum_{i=1}^{\hat{n}} \frac{y_i}{\pi_i}$$

For the respondents: $\pi_i = \dfrac{n_R}{N_R}$

(9)

For the nonrespondents captured in the 2^{nd} wave,

$$\pi_i = P(\text{selected in } 1^{st} \text{ sample} \cap \text{responded in callback})$$

$$= \frac{n_M}{N_M} \cdot \lambda$$

$$\hat{t}_{HT} = \sum_{resp} \frac{y_i}{(n_R/N_R)} + \sum_{non\ res} \frac{y_i}{\left(\frac{n_M}{N_M}\lambda\right)}$$

$$= N_R \frac{1}{n_R} \sum_{resp} y_i + N_M \frac{1}{\lambda} \frac{1}{n_M} \sum_{\substack{non \\ resp}} y_i$$

(10)

$$\hat{t}_{HT} = N_R \bar{y}_R + N_M \bar{y}_M$$

$$\bar{y}_{HT} = \frac{\hat{t}_{HT}}{N} = W_R \bar{y}_R + W_M \bar{y}_M$$

But N_R & N_M are unknown

$$\hat{N}_R = \frac{n_R}{n} N \quad \text{and} \quad \hat{N}_M = \frac{n_M}{n} N$$

So $\hat{W}_R = \dfrac{n_R}{n}$ and $\hat{W}_M = \dfrac{n_M}{n}$

11

$$\bar{y}_{adj} = \frac{n_R}{n}\bar{y}_R + \frac{n_M}{n}\bar{y}_M$$

$$= \frac{n_R}{n}\frac{1}{n_R}\sum_{resp} y_i + \frac{n_M}{n}\frac{1}{\lambda n_M}\sum_{\substack{non\\resp}} y_i$$

$$= \frac{1}{n}\left[\sum_{resp} y_i + \frac{1}{\lambda}\sum_{\substack{non\\resp}} y_i\right]$$

Stat 576 HW #7

10.4 A particular sportsmen's club is concerned about the number of brook trout in a certain stream. During a period of several days, $t = 100$ trout are caught, tagged, and then returned to the stream. Note that the sample represents 100 different fish; hence, any fish caught on these days that had already been tagged is immediately released. Several weeks later a second sample of $n = 120$ trout is caught and observed. Suppose 27 in the second sample are tagged ($s = 27$). Estimate N, the total size of the population, and place a bound on the error of estimation.

10.10 A zoologist wishes to estimate the size of the turtle population in a given geographical area. She believes that the turtle population size is between 500 and 1000; hence, an initial sample of 100 (10%) appears to be sufficient. The $t = 100$ turtles are caught, tagged, and released. A second sampling is begun one month later, and she decides to continue sampling until $s = 15$ tagged turtles are recaptured. She catches 160 turtles before obtaining 15 tagged turtles ($n = 160$, $s = 15$). Estimate N and place a bound on the error of estimation.

10.16 Cars passing through an intersection are counted during randomly selected ten-minute intervals throughout the working day. Twenty such samples show an average of 40 cars per interval. Estimate, with a bound on the error, the number of cars that you expect to go through the intersection in an eight-hour period.