SRSWOR

Stat 576
4-3-25

Simple random sampling without replacement (1)

Find $E(\bar{y})$, $V(\bar{y})$, an estimate of $V(\bar{y})$

Let $Z_j = \begin{cases} 1 & \text{if } Y_j \text{ is in the sample} \\ 0 & \text{otherwise} \end{cases}$

Then $\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{n}\sum_{j=1}^{N} Z_j Y_j$

Z_j has a Bernoulli distribution with parameter p, where $p = P[Z_j = 1] = P[Y_j \text{ is in sample}]$

So $p = \frac{\#\text{ of samples containing } Y_j}{\#\text{ of different samples possible}} = \frac{1 \cdot \binom{N-1}{n-1}}{\binom{N}{n}}$ (2)

$$= \frac{(N-1)!}{(n-1)!\,\cancel{(N-n)!}} \Big/ \frac{N!}{n!\,\cancel{(N-n)!}} = \frac{\cancel{(N-1)!}}{\cancel{N!}_{N}} \frac{\cancel{n!}\,n}{\cancel{(n-1)!}}$$

$$= \frac{n}{N}$$

$\therefore\ E[Z_j] = p = \frac{n}{N}$

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

3

$$\mathrm{Cov}(Z_j, Z_k) = E[Z_j Z_k] - E[Z_j]E[Z_k]$$

$$= \underbrace{E[Z_j Z_k]} - \left(\frac{n}{N}\right)^2$$

$$E[Z_j Z_k] = 1 \cdot P(Z_j Z_k = 1) + 0 \cdot P(Z_j Z_k = 0)$$

$$= P(Z_j Z_k = 1)$$

$$= P[\,Y_j \text{ and } Y_k \text{ are in the sample}\,]$$

$$= \frac{\binom{N-2}{n-2}}{\binom{N}{n}}$$

4

$$= \frac{(N-2)!}{(n-2)!\,\cancel{(N-n)!}} \Big/ \frac{N!}{n!\,\cancel{(N-n)!}}$$

$$= \frac{(N-2)!}{N!}\,\frac{n!}{(n-2)!} = \frac{n(n-1)}{N(N-1)}$$

$$\mathrm{Cov}(Z_j, Z_k) = E[Z_j Z_k] - \frac{n^2}{N^2}$$

$$= \frac{n(n-1)}{N(N-1)} - \frac{n^2}{N^2}$$

$$= \frac{n}{N}\left[\frac{n-1}{N-1} - \frac{n}{N}\right] = \frac{n}{N}\left[\frac{\cancel{Nn} - N - (\cancel{Nn} - n)}{N(N-1)}\right]$$

$$= \frac{-n(N-n)}{N^2(N-1)}$$ ⑤

$$E(\bar{y}) = E\left[\frac{1}{n}\sum_{j=1}^{N} Z_j Y_j\right]$$

$$= \frac{1}{n}\sum_{j=1}^{N} Y_j E(Z_j) = \frac{1}{\cancel{n}}\frac{\cancel{n}}{N}\sum_{j=1}^{N} Y_j$$

$$= \bar{Y}$$

$$V(\bar{y}) = V\left[\frac{1}{n}\sum_{j=1}^{N} Z_j Y_j\right]$$

$$= \frac{1}{n^2}\left[\sum_{j=1}^{N} Y_j^2 V(Z_j) + \sum_{j \neq k}\sum Cov(Z_j Y_j, Z_k Y_k)\right]$$

⑥

$$= \frac{1}{n^2}\left[\frac{n}{N}\left(1-\frac{n}{N}\right)\sum_{j=1}^{N} Y_j^2 + \frac{-n(N-n)}{N^2(N-1)}\sum_{j \neq k}\sum Y_j Y_k\right]$$

$$= \frac{1}{n^2}\frac{n}{N}\left(1-\frac{n}{N}\right)\left[\sum_{j=1}^{N} Y_j^2 - \frac{\cancel{N-n}}{\cancel{N}(N-1)}\frac{1}{\cancel{\left(1-\frac{n}{N}\right)}}\sum_{j \neq k}\sum Y_j Y_k\right]$$

$$= \frac{1}{nN}\left(1-\frac{n}{N}\right)\left[\sum_{j=1}^{N} Y_j^2 - \frac{1}{N-1}\sum_{j \neq k}\sum Y_j Y_k\right]$$

$$\left[\text{Note: } \left(\sum_{j=1}^{N} Y_j\right)^2 = \sum Y_i^2 + \sum_{j \neq k}\sum Y_j Y_k\right]$$

⑦

$$= \frac{1}{nN}\left(1-\frac{n}{N}\right)\left[\sum_{j=1}^{N} y_j^2 - \frac{1}{N-1}\left(\left(\sum y_j\right)^2 - \sum y_j^2\right)\right]$$

$$= \frac{1}{nN}\left(1-\frac{n}{N}\right)\left[\sum_{j=1}^{N} y_j^2 \underbrace{\left(1+\frac{1}{N-1}\right)}_{\frac{N}{N-1}} - \frac{1}{N-1}\left(\sum y_j\right)^2\right]$$

$$= \frac{1}{n\cancel{N}}\left(1-\frac{n}{N}\right)\frac{\cancel{N}}{N-1}\underbrace{\left[\sum_{j=1}^{N} y_j^2 - \frac{\left(\sum y_j\right)^2}{N}\right]}_{N\sigma^2}$$

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

⑧

Define $S^2 = \frac{N}{N-1}\sigma^2$

Define $\left(1-\frac{n}{N}\right) = fpc =$ finite population correction

Then $V(\bar{y}) = \frac{S^2}{n} \cdot fpc$

We still need an estimator of S^2

Try s^2

$$E(s^2) = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (\underbrace{y_i - \bar{Y}} + \underbrace{\bar{Y} - \bar{y}})^2\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (y_i - \bar{Y})^2 + n(\bar{Y} - \bar{y})^2 + 2(\bar{Y} - \bar{y})\underbrace{\sum_{i=1}^{n}(y_i - \bar{Y})}_{n\bar{y} - n\bar{Y}}\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} (y_i - \bar{Y})^2 - n(\bar{Y} - \bar{y})^2\right]$$

$$= \underbrace{\frac{1}{n-1} E\left[\sum_{i=1}^{n} (y_i - \bar{Y})^2\right]}_{①} - \underbrace{\frac{n}{n-1} E(\bar{y} - \bar{Y})^2}_{②}$$

10

$$② : \frac{n}{n-1} V(\bar{y}) = \frac{n}{n-1} \frac{S^2}{n} f_{pc}$$

$$① \quad \frac{1}{n-1} E\left[\sum_{j=1}^{N} Z_j (Y_j - \bar{Y})^2\right]$$

$$= \frac{1}{n-1} \sum_{j=1}^{N} \frac{n}{N} (Y_j - \bar{Y})^2 = \frac{n}{n-1} \underbrace{\frac{1}{N} \sum_{j=1}^{N} (Y_j - \bar{Y})^2}_{\sigma^2}$$

$$E(s^2) = ① - ②$$

$$= \frac{n}{n-1} \sigma^2 - \frac{n}{n-1} \frac{S^2}{n} f_{pc}$$

(11)

$$= \frac{n}{n-1}\left[\frac{N-1}{N} S^2 - \frac{1}{n} S^2 fpc\right]$$

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

$$1 - \cancel{\tfrac{1}{N}} - \tfrac{1}{n} + \cancel{\tfrac{1}{N}}$$

$$\frac{n-1}{n}$$

$$E(s^2) = S^2$$

(12)

SRS WR	SRS WOR
$E(\bar{y}) = \bar{Y}$	$E(\bar{y}) = \bar{Y}$
$V(\bar{y}) = \frac{\sigma^2}{n}$	$V(\bar{y}) = \frac{S^2}{n} fpc$
$E(s^2) = \sigma^2$	$E(s^2) = S^2$
$\hat{V}(\bar{y}) = \frac{s^2}{n}$	$\hat{V}(\bar{y}) = \frac{s^2}{n} fpc$
	$fpc = \left(1 - \frac{n}{N}\right)$
	$S^2 = \frac{N}{N-1}\sigma^2$

6 A university has 807 faculty members. For each faculty member, the number of refereed publications was recorded. This number is not directly available on the database, so requires the investigator to examine each record separately. A frequency table for number of refereed publications is given below for an SRS of 50 faculty members.

Refereed Publications	0	1	2	3	4	5	6	7	8	9	10
Faculty Members	28	4	3	4	4	2	1	0	2	1	1

a Plot the data using a histogram. Describe the shape of the data.

b Estimate the mean number of publications per faculty member, and give the SE for your estimate.

c Do you think that $\bar{y}$ from (b) will be approximately normally distributed? Why or why not?

10 Which of the following SRS designs will give the most precision for estimating a population mean? Assume that each population has the same value of the population variance S^2.

1. An SRS of size 400 from a population of size 4000
2. An SRS of size 30 from a population of size 300
3. An SRS of size 3000 from a population of size 300,000,000