Stat 576
10-5-17

①

Systematic Sampling

N = population size
n = sample size

Let $k = \frac{N}{n}$, rounded up to an integer

Let R be a random number chosen with equal probabilities from $1, \ldots, k$

Select items $R, R+k, R+2k, \ldots, R+(n-1)k$

②

Example. $N = 20$, need $6 = n$

$$k = \frac{20}{6} = 3\tfrac{1}{3} \to 4$$

Choose R from $1, 2, 3, 4$

If $R = 1$: $1, 5, 9, 13, 17$
$R = 2$: $2, 6, 10, 14, 18$
$R = 3$: $3, 7, 11, 15, 19$
$R = 4$: $4, 8, 12, 16, 20$

③

Another variation: round k down

$k = 3\frac{1}{3} \to 3$

$R=1$: 1, 4, 7, 10, 13, 16, 19

$R=2$: 2, 5, 8, 11, 14, 17, 20

$R=3$: 3, 6, 9, 12, 15, 18

Another variation: round k up

Select R from 1 to N

Use k as your skip length, and wrap around

④

In our example, select R from 1 to 20

$k = 3\frac{1}{3} \to 4$

$R=1$: 1, 5, 9, 13, 17

$R=2$: 2, 6, 10, 14, 18

$\vdots$

$R=20$: 20, 4, 8, 12, 16

This is called <u>circular systematic sampling</u>

Note: If you could assume that the population was randomly ordered, then a systematic sample would be SRS.

(5)

Ratio Estimation

Suppose that each observation consists of an ordered pair (x_i, y_i)

Goal: Estimate the "population ratio"

$$B = \frac{\overline{Y}}{\overline{X}}$$

Try $\hat{B} = \frac{\bar{y}}{\bar{x}}$

Need to find $E(\hat{B})$ and $V(\hat{B})$

(6)

Use the Delta method, which uses a first or second order Taylor series

$$f(x,y) \approx f(x_0,y_0) + \left.\frac{\partial f}{\partial x}\right|_{(x_0,y_0)}(x-x_0) + \left.\frac{\partial f}{\partial y}\right|_{(x_0,y_0)}(y-y_0)$$

$$+ \left.\frac{\partial^2 f}{\partial x^2}\right|_{(x_0,y_0)} \frac{(x-x_0)^2}{2} + \left.\frac{\partial^2 f}{\partial y^2}\right|_{(x_0,y_0)} \frac{(y-y_0)^2}{2} + \left.\frac{\partial^2 f}{\partial x \partial y}\right|_{(x_0,y_0)} (x-x_0)(y-y_0)$$

Apply this to $f(x,y) = y/x$ (7)

$$\frac{\partial f}{\partial x} = -\frac{y}{x^2} \qquad \frac{\partial^2 f}{\partial x^2} = \frac{2y}{x^3}$$

$$\frac{\partial f}{\partial y} = \frac{1}{x} \qquad \frac{\partial^2 f}{\partial y^2} = 0$$

$$\frac{\partial^2 f}{\partial x \, \partial y} = -\frac{1}{x^2}$$

$$\frac{y}{x} \approx \frac{y_0}{x_0} - \frac{y_0}{x_0^2}(x-x_0) + \frac{1}{x_0}(y-y_0)$$
$$+ \frac{2y_0}{x_0^3}\frac{(x-x_0)^2}{2} + 0 - \frac{1}{x_0^2}(x-x_0)(y-y_0)$$

So (8)

$$\hat{B} = \frac{\bar{y}}{\bar{x}} \approx \frac{\bar{Y}}{\bar{X}} - \frac{\bar{Y}}{\bar{X}^2}(\bar{x}-\bar{X}) + \frac{1}{\bar{X}}(\bar{y}-\bar{Y})$$
$$+ \frac{2\bar{Y}}{\bar{X}^3}\frac{(\bar{x}-\bar{X})^2}{2} - \frac{1}{\bar{X}^2}(\bar{x}-\bar{X})(\bar{y}-\bar{Y})$$

$$E(\hat{B}) \approx \frac{\bar{Y}}{\bar{X}} - 0 + 0 + \frac{\bar{Y}}{\bar{X}^3}V(\bar{x})$$
$$- \frac{1}{\bar{X}^2}\mathrm{Cov}(\bar{x},\bar{y})$$

use the 1st order terms

$$V(\hat{B}) \approx \frac{\bar{Y}^2}{\bar{X}^4} V(\bar{x}) + \frac{1}{\bar{X}^2} V(\bar{y}) - 2\frac{\bar{Y}}{\bar{X}^3} \mathrm{Cov}(\bar{x},\bar{y})$$

Summary: $E(\hat{B}) \approx B + \frac{1}{\bar{X}^2}\left[B V(\bar{x}) - \mathrm{Cov}(\bar{x},\bar{y})\right]$

$$V(\hat{B}) \approx \frac{1}{\bar{X}^2}\left[B^2 V(\bar{x}) + V(\bar{y}) - 2B\,\mathrm{Cov}(\bar{x},\bar{y})\right]$$

11 Mayr et al. (1994) took an SRS of 240 children who visited their pediatric outpatient clinic. They found the following frequency distribution for the age (in months) of free (unassisted) walking among the children:

Age (months)	9	10	11	12	13	14	15	16	17	18	19	20
Number of Children	13	35	44	69	36	24	7	3	2	5	1	1

a Construct a histogram of the distribution of age at walking. Is the shape normally distributed? Do you think the sampling distribution of the sample average will be normally distributed? Why, or why not?

b Find the mean, SE, and a 95% CI for the average age for onset of free walking.

c Suppose the researchers wanted to do another study in a different region, and wanted a 95% CI for the mean age of onset of walking to have margin of error 0.5. Using the estimated standard deviation for these data, what sample size would they need to take?

12 The percentage of patients overdue for a vaccination is often of interest for a medical clinic Some clinics examine every record to determine that percentage; in a large practice, though, taking a census of the records can be time-consuming. Cullen (1994) took a sample of the 580 children served by an Auckland family practice to estimate the proportion of interest.

a What sample size in an SRS (without replacement) would be necessary to estimate the proportion with 95% confidence and margin of error 0.10?

b Cullen actually took an SRS *with* replacement of size 120, of whom 27 were *not* overdue for vaccination. Give a 95% CI for the proportion of children not overdue for vaccination.