STAT 403 Tutorial · Week 10

Topics of this tutorial includes simple random sampling (without replacement) and properties of Horvitz-Thompson estimator (not required).

1. Simple random sampling.

Example. An investigator took an SRS of 240 children aged from 2 to 6 who visited a pediatric outpatient clinic. They found the following frequency distribution for free (unassisted) walking among the children: (suppose population size N=1000).

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). Find the mean, standard error, and a 95% CI for the average age for onset of free walking.

b). Suppose the researchers want to do another study in the same region and want 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 take?

Solution:

a). Denote the SRS sample by $u \in \mathcal{U}$ and the variable of interest by y_i, $i \in \mathcal{U}$. To estimate population mean $\bar{Y} = \frac{1}{N}\sum_{i\in\mathcal{U}} y_i$, we use the sample mean $\bar{y} = \frac{1}{n}\sum_{i\in u} y_i$, where n = sample size = 240.

$$\bar{y} = \frac{1}{240}\left(9\times 13 + 10\times 35 + 11\times 44 + \cdots + 20\times 1\right) = \frac{2899}{240} = 12.08.$$

The sample variance

$$s^2 = \frac{1}{n-1}\sum_{i\in u}(y_i-\bar{y})^2 = \frac{1}{240-1}\left[13\times(9-12.08)^2 + 35\times(10-12.08)^2 + \cdots + 1\times(20-12.08)^2\right]$$

$$= \frac{1}{239}\times 885.496 = 3.705$$

Therefore the Stand error of $\bar{y}$:

$$se(\bar{y}) = \sqrt{\left(1-\frac{n}{N}\right)\cdot\frac{s^2}{n}} = \sqrt{\left(1-\frac{240}{1000}\right)\times\frac{3.705}{240}} = \cancel{0.0117}.$$

0.1083

Thus a 95% CI for the average age for onset of free walking is given by

$$\bar{y} \pm z_{0.025} \times se(\bar{y}) = 12.08 \pm 1.96 \times \overset{0.1083}{\cancel{0.0117 = 12.08 \pm 0.023}} = \overset{[11.87,12.29]}{\cancel{[12.057,\ 12.103]}}.$$

b) The margin error is $z_{0.025} \times se(\bar{y}) = z_{0.025} \times \sqrt{\left(1-\frac{n'}{N}\right)\cdot\frac{s^2}{n'}} = 0.5$

$$\Rightarrow \quad \left(1-\frac{n'}{N}\right)\cdot\frac{1}{n'} = \left(\frac{0.5}{z_{0.025}}\right)^2 \times \frac{1}{s^2} = \left(\frac{0.5}{1.96}\right)^2 \Big/ 3.705 = 0.0176$$

$$\Rightarrow \quad \frac{1}{n'} - \frac{1}{1000} = 0.0176$$

$$\Rightarrow \quad n' = 53.76$$

$\Rightarrow$ Therefore, the sample size n' has to be greater or equal than 54.

2. Properties of Horvitz-Thompson estimator. (not required)

①. $\pi_1 + \pi_2 + \dots + \pi_N = \mathbb{E}(n)$, where n is the sample size.

Pf. Let $I_i = \begin{cases} 1, & \text{if unit } i \text{ is included in the sample.} \\ 0, & \text{otherwise} \end{cases}$

Then $\sum_{i=1}^{N} I_i$ = number of units included in the sample = n

Take expectation on both sides

$$\mathbb{E}\left(\sum_{i=1}^{N} I_i\right) = \sum_{i=1}^{N} \mathbb{E}(I_i) = \sum_{i=1}^{N} \pi_i = \mathbb{E}(n).$$

②. Horvitz-Thompson estimator is unbiased! $\hat{T} = \sum_{i\in u} \frac{y_i}{\pi_i}$

Write $\hat{T}$ as $\hat{T} = \sum_{i\in u} \frac{y_i}{\pi_i} = \sum_{i=1}^{N} \frac{y_i}{\pi_i} I_i$, then $\mathbb{E}(\hat{T}) = \mathbb{E}\left(\sum_{i=1}^{N} \frac{y_i}{\pi_i} I_i\right) = \sum_{i=1}^{N} \mathbb{E}\left(\frac{y_i}{\pi_i} I_i\right)$

$$= \sum_{i=1}^{N} \frac{y_i}{\pi_i} \mathbb{E} I_i = \sum_{i=1}^{N} \frac{y_i}{\pi_i} \pi_i = \sum_{i=1}^{N} y_i = T$$