Chapter 12
Sampling on Successive Occasions

Many times, we are interested in measuring a characteristic of a population on several occasions to estimate the trend in time of population means as a time series of the current value of population mean or the value of population mean over several points of time.

When the **same** population is sampled repeatedly, the opportunities for a flexible sampling scheme are greatly enhanced. For example, on the hth occasion we may have a part of sample that are matched with (common to) the sample at $(h-1)^{th}$ occasion, parts matching with both $(h-1)^{th}$ and $(h-2)^{th}$ occasions, etc.

Such a partial matching is termed as **sampling on successive occasions with partial replacement of units** or **rotation sampling** or **sampling for a time series**.

Notations:

Let P be the fixed population with N units.

y_t : value of certain dynamic character which changes with time t and can be measured for each unit on a number of occasions, $t = 1, 2, .., n$.

y_{ij} : value of y on j^{th} unit in the population at the ith occasion, $i = 1, 2, ..., h,\ j = 1, ..., N$.

$\bar{Y}_i = \frac{1}{N}\sum_j y_{ij}$: population mean for the ith occasion

$S_i^2 = \frac{1}{N-1}\sum_{j=1}^{N}(y_{ij} - \bar{Y}_i)^2$: population variance for the ith occasion.

Generally, we assume $S_1^2 = S_2^2 = ... = S^2$.

$$\rho_{ii*} = \frac{1}{N-1}\sum_{j=1}^{N}(y_{ij} - \bar{Y}_i)(y_{i*j} - \bar{Y}_{i*}).$$

is the population correlation coefficient between the observations at the occasions i and i^* $(i < i* = 1, 2, ..., h)$.

$\rho = \rho_{12}$

s_i^* : sample of size n_i selected at the ith occasion

s_{im}^* : part of s_i^* which is common to (i.e. matched with) s_{i-1}^*,

$s_{im}^* = s_i^* \cap s_{i-1}^*,\ i = 2, 3, ..., h \quad (s_{1m} = s_{2m})$

Note that s_{1m} and s_{2m} are of the sizes n_1^* and n_2^* respectively.

s_{iu}^* : set of units in s_i^* not obtained by the selection in s_{im}^*.

Often

$$s_{iu}^* = s_{i-1}^{*c} \bigcap s_i \quad (i = 2,...,h)(s_{1u}^* = P - s_{1m}^*).$$

Note that s_{iu}^* is of size $n_i^{**}(= n_i - n_i^*)$.

$\bar{y}_i =$ sample mean of units in i^{th} occasion.

$\bar{y}_i^* =$ sample mean of the units in s_{im}^* on the ith occasion.

$\bar{y}_i^{**} =$ sample mean of units in s_{iu} on the ith occasion.

$\bar{y}_i^{***} =$ sample mean of units in s_{im} on the $(i-1)^{th}$ occasion, $i = 2,3,...,h$

$\left(\bar{y}_2^{***} = \bar{y}_1^*, \bar{y}_i^{***}\right.$ depends on $\bar{y}_{i-1}$ *and* $\left.\bar{y}_{i-1}^{**}\right)$

Sampling on two occasions

Assume that $n_i = n$

$$n_i^* = m$$

$$n_i^{**} = u(= n - m), \quad i = 1,2$$

Suppose that the sample s_1^* is an SRSWOR from P. The sample

$$s_2^* = s_{2m}^* \bigcup s_{2u}^*$$

where s_{2m}^* is an SRSWOR sample of size m from s_1^* and

s_{2u}^* is an SRSWOR sample of size u from $\left(P - s_1^*\right)$.

Estimation of Population mean

Two types of estimators are available for the estimation of a population mean:

1. **Type 1 estimators:** They are obtained by taking a linear combination of estimators obtained from s_{2u}^* and s_{2m}^*.
2. **Type 2 estimators:** They are obtained by considering the best linear combination of sample means.

Type 1 estimators:

Two estimators are available for estimating $\bar{Y}_2$

(i) $t_{2u} = \bar{y}_2^{**}$

with $Var(\bar{y}_2^{**}) = \dfrac{S_2^2}{u} = \dfrac{1}{W_u}$ (say)

(ii) t_{2m} = linear regression estimate of $\bar{Y}_2$ based on the regression of y_{2j} on y_{1j}

$= \bar{y}_2^* + b(\bar{y}_1 - \bar{y}_1^*)$

where $b = \dfrac{\sum_{i \in s_{2m}^*} (y_{1j} - \bar{y}_1^*)(y_{2j} - \bar{y}_2^*)}{\sum_{j \in s_{1m}^*} (y_{1j} - \bar{y}_1^*)^2}$ is the sample regression coefficient.

Recall in case of double sampling, we had

$$Var(\hat{\bar{Y}}_{regd}) = S_y^2\left(\frac{1}{n} - \frac{1}{N}\right) - \rho^2 s_y^2\left(\frac{1}{n} - \frac{1}{n^*}\right)$$

$$= \frac{S_y^2}{n} - \rho^2 S_y^2\left(\frac{1}{n} - \frac{1}{n^*}\right)$$

$$= \frac{(1-\rho^2)}{n} S_y^2 + \frac{\rho^2 S_y^2}{n^*} \qquad \text{(ignoring term of order } \frac{1}{N}\text{)}.$$

So in this case

$$Var\left(t_{2m}\right) = \frac{S_2^2(1-\rho^2)}{m} + \frac{\rho^2 S_2^2}{n}$$

$$= \frac{1}{W_m} \text{ (say).}$$

If there are two uncorrelated unbiased estimators of a parameter, then the best linear unbiased estimator of parameter can be obtained by combining them using a linear combination with suitably chosen weights. Now we discuss how to choose weights in such a linear combination of estimators.

Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two uncorrelated and unbiased estimators of θ, i.e., $E(\hat{\theta}_1) = E(\hat{\theta}_2) = \theta$ and $Var(\hat{\theta}_1) = \sigma_1^2$, $Var(\hat{\theta}_2) = \sigma_2^2, Cov(\hat{\theta}_1, \hat{\theta}_2) = 0.$

Consider $\hat{\theta} = \omega\hat{\theta}_1 + (1-\omega)\hat{\theta}_2$ where $0 \le \omega \le 1$ is the weight. Now choose ω such that $Var(\hat{\theta})$ is minimum.

$$Var(\hat{\theta}) = \omega^2\sigma_1^2 + (1-\omega)^2\sigma_2^2$$

$$\frac{\partial Var(\hat{\theta})}{\partial \omega} = 0$$

$$\Rightarrow 2\omega\sigma_1^2 - 2(1-\omega)\sigma_2^2 = 0$$

$$\Rightarrow \omega = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \omega^*, \text{ say}$$

$$\left.\frac{\partial^2 Var(\hat{\theta})}{\partial \omega^2}\right|_{\omega=\omega^*} > 0.$$

The minimum variance achieved by $\hat{\theta}$ is

$$Var(\hat{\theta})_{Min} = \omega^{*2}\sigma_1^2 + (1-\omega^*)^2\sigma_2^2$$
$$= \frac{\sigma_2^4\sigma_1^2}{\left(\sigma_1^2+\sigma_2^2\right)^2} + \frac{\sigma_1^4\sigma_2^2}{\left(\sigma_1^2+\sigma_2^2\right)^2}$$
$$= \frac{\sigma_2^2\sigma_1^2}{\sigma_1^2+\sigma_2^2} = \frac{1}{\frac{1}{\sigma_2^2}+\frac{1}{\sigma_1^2}}.$$

Now we implement this result in our case.

Consider the linear combination of t_{2u} and t_{2m} as

$$\hat{\bar{Y}}_2 = \omega t_{2u} + (1-\omega)t_{2m}$$

where the weights ω are obtained as

$$\omega = \frac{W_u}{W_u + W_m}$$

so that $\hat{\bar{Y}}_2$ is the best combined estimate.

The minimum variance with this choice of ω is

$$Var(\hat{\bar{Y}}_2) = \frac{1}{W_u + W_m} = \frac{S_2^2(n-u\rho^2)}{(n^2-u^2\rho^2)}.$$

For $u=0$ (complete matching), $Var(\hat{\bar{Y}}_2) = \frac{S_2^2}{n}$.

For $u=n$ (no matching), $Var(\hat{\bar{Y}}_2) = \frac{S_2^2}{n}$.

Type II estimators:

We now consider the minimum variance linear unbiased estimator of $\bar{Y}_2$ under the same sampling scheme as under Type I estimator.

A best linear (linear in terms of observed means) unbiased estimator of $\hat{\bar{Y}}_2$ is of the form

$$\hat{\bar{Y}}_2^* = a\bar{y}_1^{**} + b\bar{y}_1^* + c\bar{y}_2^* + d\bar{y}_2^{**}$$

where constants a,b,c,d and matching fraction $\lambda\left(=\frac{m}{n}=\frac{n-1}{n}\right)$ are to be suitably chosen so as to minimize the variance.

Assume $S_1^2 = S_2^2$.

Now $E(\hat{\bar{Y}}_2^*) = (a+b)\bar{Y}_1 + (c+d)\bar{Y}_2$.

If $\hat{\bar{Y}}_2^*$ has to be an unbiased estimator of $\bar{Y}_2$, i.e.

$$E(\hat{\bar{Y}}_2^*) = \bar{Y}_2,$$

it requires

$$a+b=0$$
$$c+d=1.$$

Since a minimum variance unbiased estimator would be uncorrelated with any unbiased estimator of zero, we must have

$$Cov(\hat{\bar{Y}}_2^*, \bar{y}_1^{**} - \bar{y}_1^*) = 0 \qquad (1)$$
$$Cov(\hat{\bar{Y}}_2^*, \bar{y}_2^* - \bar{y}_2^{**}) = 0\,. \qquad (2)$$

Since

$$Cov(\bar{y}_2^*, \bar{y}_1^{**}) = 0 = Cov(\bar{y}_2^*, \bar{y}_2^{**})$$
$$Cov(\bar{y}_2^*, \bar{y}_1^*) = \frac{\rho S^2}{m}$$
$$Cov(\bar{y}_2^{**}, \bar{y}_1^{**}) = Cov(\bar{y}_2^{**}, \bar{y}_2^*) = 0$$
$$Var(\bar{y}_2^*) = \frac{S^2}{m}$$
$$Var(\bar{y}_2^{**}) = \frac{S^2}{u}.$$

Now solving (1) and (2) by neglecting terms of order $\frac{1}{N}$, we have

$$\begin{aligned} Cov(\hat{\bar{Y}}_2^*, \bar{y}_1^{**} - \bar{y}_1^*) &= Cov(a\bar{y}_1^{**} + b\bar{y}_1^* + c\bar{y}_2^* + d\bar{y}_2^{**}, \bar{y}_1^{**} - \bar{y}_1^*) \\ &= aVar(\bar{y}_1^{**}) + bC\text{ov}(\bar{y}_1^*, \bar{y}_1^{**}) + cCov(\bar{y}_2^*, \bar{y}_1^{**}) + dC\text{ov}(\bar{y}_2^{**}, \bar{y}_1^{**}) \\ &\quad - aC\text{ov}(\bar{y}_1^{**}, \bar{y}_1^*) - bVar(\bar{y}_1^*) - cC\text{ov}(\bar{y}_1^*, \bar{y}_2^*) - dCov(\bar{y}_2^{**}, \bar{y}_1^*) \end{aligned}$$

or

$$-a\frac{\rho S^2}{m} + \frac{cS^2}{m} = \frac{(1-c)S^2}{u}\,. \qquad (3)$$

Similarly, from (2), we have

$$Cov(\bar{y}_2^*, \bar{y}_2^* - \bar{y}_2^{**}) = 0$$
$$\Rightarrow -\frac{aS^2}{m} + \frac{c\rho S^2}{m} = \frac{aS^2}{u}. \qquad (4)$$

Solving (3) and (4) gives

$$a = \frac{\lambda\mu\rho}{1-\rho^2\mu^2}, \quad c = \frac{\lambda}{1-\rho^2\mu^2}$$

where $\mu = \frac{u}{n} = 1-\lambda, \ \lambda = \frac{n-u}{n}$

$$b = -a, \ d = 1-c.$$

Substituting $a,b,c,d,$ the best linear unbiased estimator of $\bar{Y}_2$ is

$$\hat{\bar{Y}}_2^* = \left[\lambda\mu\rho(\bar{y}_1^{**} - \bar{y}_1^*) + \lambda\bar{y}_2^* + \frac{\mu(1-\rho^2\mu)\bar{y}_2^{**}}{(1-\rho^2\mu^2)}\right].$$

For these values of a and c,

$$Var(\hat{\bar{Y}}_2^*) = \left(\frac{1-\rho^2\mu S^2}{1-\rho^2\mu^2 n}\right).$$

Alternatively, minimize $Var(\hat{\bar{Y}}_2^*)$ with respect to a and c and find optimum values of a and c. Then find the estimator and its variance.

Till now, we used SRSWOR for the two occasions. We now consider unequal probability sampling schemes on two occasions for estimating $\hat{\bar{Y}}_2$. We use the same notations as defined in varying probability scheme.

Des Raj Scheme:

Let s_1^* be the sample selected by PPSWR from P using x as a size (auxiliary) variable.

Then $p_i = \frac{x_i}{X_{tot}}$ is the size measure of i, where X_{tot} is the population total of auxiliary variable.

$$s_2^* = s_{2m}^* \cup s_{2u}^*$$

where s_{2m}^* is an SRSWR(m) from s_1^* and s_{2u}^* is an independent sample selected from P by PPSWR using u draws $(m+u=n)$.

The estimator is

$$\hat{Y}_{2des} = \omega t_{2m} + (1-\omega)t_{2u}; \quad 0 \le \omega \le 1$$

where

$$t_{2m} = \sum_{j \in s_{2m}^*} \left(\frac{y_{2j} - y_{1j}}{mp_j} \right) + \sum_{j \in s_1^*} \left(\frac{y_{1j}}{np_j} \right)$$

$$t_{2u} = \sum_{j \in s_{2u}^*} \left(\frac{y_{2j}}{up_j} \right).$$

Assuming

$$\sum_{j=1}^{N} P_j \left(\frac{Y_{1j}}{P_j} - Y_1 \right)^2 = \sum_{j=1}^{N} P_j \left(\frac{Y_{2j}}{P_j} - Y_2 \right)^2 = V_0 \text{ (say)}.$$

For the optimum sampling fraction

$$\lambda = \frac{m}{n},$$

$$Var(\bar{Y}_{2des}) = \frac{V_0(1 + \sqrt{2(1-\delta)})}{2n}$$

where

$$\delta = \frac{\sum_{i=1}^{N} P_i \left(\frac{Y_{1i}}{P_i} - Y_1 \right) \left(\frac{Y_{2i}}{P_i} - Y_2 \right)}{\sigma_{pps}(y_1)\sigma_{pps}(y_2)}$$

$$Var_{pps}(z) = \sum_{i=1}^{N} P_i \left(\frac{Z_i}{P_i} - Z \right)^2 = \sigma_{pps}^2(z)$$

$$Z = \sum_{i=1}^{N} Z_i.$$

(ii) Chaudhuri-Arnab sampling scheme

Let s_1^* be a sample selected by Midzuno's sampling scheme,

$$s_2^* = s_{2m}^* \cup s_{2u}^*$$

where s_{2m}^* = SRSWOR (m) sample from s_1^*

s_{2u}^* = sample of size u from P by Midzuno's sampling scheme.

Then an estimator of $\bar{Y}$ is

$$\hat{\bar{Y}}_{2ca} = \alpha t_{2m} + (1-\alpha) t_{2u}; \quad 0 \le \alpha \le 1$$

where

$$t_{2m} = \sum_{j \in s_{2m}^*} \left(\frac{(y_{2j} - y_1)n}{m\pi_j} \right) + \sum_{j \in s_1^*} \left(\frac{y_{1j}}{\pi_j} \right)$$

$$t_{2u} = \sum_{j \in s_{2u}} \frac{y_{2j}}{\pi_j^*}$$

$$\pi_j = np_j$$

$$\pi_j^* = up_j.$$

Similarly, other schemes are also there.

Sampling on more than two occasions

When there are more than two occasions, one has a large flexibility in using both sampling procedures and estimating the character.

Thus on occasion i

- one may have parts of the sample that are matched with occasion $(i-1)$
- parts that are matched with occasion $(i-2)_j$
- and so on.

One may consider a single multiple regression of all previous matchings on the current occasion. However, it has been seen that the loss of efficiency incurred by using the information from the latest two or three occasions only is fairly small in many occasions.

Consider the simple sampling design where

$$s_i^* = s_{im}^* \bigcup s_{iu}^*,$$

where s_{im}^* is a sample by SRSWOR of size m_i from $s_{(i-1)}^*$,

s_{iu}^* is a sample by SRSWOR of size $u_i (= n - m_i)$ from the units not already sampled.

Assume $n_i = n,\ S_2^2 = S^2$ for all i.

On the ith occasion, we have therefore two estimators

$$t_{iu} = \bar{y}_i^{**} \text{ with } Var(t_{iu}) = \frac{S^2}{u_i} = \frac{1}{W_{iu}}$$

$$t_{im} = \bar{y}_i^* + b_{(i-1)i}(\hat{\bar{Y}}_{(i-1)} - \bar{y}_i^{***})$$

where $b_{(i-1),i}$ is the regression of y_{ij} on $y_{(i-1)j}$

$$b_{(i-1),i} = \frac{\sum_{s_i^{**}} (y_{(i-1)j} - \bar{y}_i^{***})(y_{ij} - \bar{y}_i^*)}{\sum_{s_i^{**}} (y_{(i-1)j} - \bar{y}_i^{***})^2}$$

and

$$Var(t_{im}) = \frac{S^2(1-\rho^2)}{m_i} + \rho^2 Var(\hat{\bar{Y}}_{(i-1)}) = \frac{1}{W_{im}}$$

assuming that $\rho_{(i-1),i} = \rho,\ i = 2,3,...,.$ and terms of order $\frac{1}{N}$ are negligible.

The expression of $Var(t_{im})$ has been obtained from the variance of regression estimator under the double sampling

$$Var(\hat{\bar{y}}_{regd}) = S_u^2\left(\frac{1}{n}-\frac{1}{N}\right)-\rho^2 S_y^2\left(\frac{1}{n}-\frac{1}{n^*}\right)$$
$$= \frac{S_y^2(1-\rho^2)}{n}-\frac{\rho^2 S_y^2}{n^*}$$

which is obtained after ignoring the terms of $\frac{1}{N}$ by using m_i for n and replacing

$\frac{\rho^2 S_y^2}{n^*}\left(=\beta^2 V(\bar{x}*)\right)$ by $\rho^2 Var(\hat{\bar{Y}}_{(i-1)})$ since $\beta=\rho$ and S_i^2 is constant. Using weights as the inverse of variance, the best weighted estimator from t_{iu} and t_{im} is

$$\hat{\bar{Y}}_{iu} = \omega_i + t_{iu} + (1-\omega_i) + t_{im}$$

where

$$\omega_i = \frac{W_{iu}}{W_{iu}+W_{im}}.$$

Then

$$Var(\hat{\bar{y}}_i) = \frac{1}{W_{iu}+W_{im}} = \frac{g_i S^2}{n} \quad \text{(say)},\ i=1,2,\ldots,(g_1=1).$$

Substituting $\frac{1}{W_{iu}} = \frac{S^2}{u_i}$

in $\frac{1}{W_{iu}+W_{im}} = \frac{g_i S^2}{n}$,

we have

$$\frac{n}{g_i} = u_i + \frac{1}{\frac{1-\rho^2}{m_i}+\frac{\rho^2 g_{i-1}}{n}}.$$

Now maximize $\frac{n}{g_i}$ with respect to m_i so as to minimize $Var(\hat{\bar{y}}_i)$. So differentiate $\frac{n}{g_i}$ with respect to m_i and substituting it to be zero, we get

$$\frac{(1-\rho^2)}{m_i^2} = \left(\frac{1-\rho^2}{m_i}+\frac{\rho^2 g_{i-1}}{n}\right)^2$$
$$\Rightarrow \hat{m}_i = \frac{n\sqrt{1-\rho^2}}{g_{i-1}(1+\sqrt{1-\rho^2})}.$$

Now the optimum sampling fracture $\frac{\hat{m}_i}{n}$ can be determined successively for $i = 2,3,...$ for given values of ρ. Substituting this in the expression of $\frac{n}{g_i}$, we have

$$\frac{1}{g_i} = 1 + \frac{(1-\sqrt{1-\rho^2})^2}{g_{i-1}\rho^2}$$

or $q_i = 1 + aq_{i-1}$

where

$$q_i = \frac{1}{g_i},\ q_1 = 1,\ a = \frac{(1-\sqrt{1-\rho^2})}{(1+\sqrt{1-\rho^2})};\ 0 < a < 1.$$

Repeated use of this relation gives

$$\begin{aligned}
q_{i-1} &= 1 + aq_{i-2} \\
\Rightarrow q_i &= 1 + a(1 + aq_{i-1}) \\
&= 1 + a + a^2\, q_{i-1} \\
q_{i-2} &= 1 + aq_{i-3} \\
\Rightarrow q_i &= 1 + a + a^2(1 + aq_{i-2}) \\
&\vdots \\
&= \frac{(1-a^i)}{(1-a)} = \frac{1}{1-a} \text{ as } i \to \infty.
\end{aligned}$$

For sampling an infinite number of times, the limiting variance factor g_∞ is

$$g_\infty = 1 - a = \frac{2\sqrt{1-\rho^2}}{1+\sqrt{1-\rho^2}}.$$

The limiting value of $V(\hat{\bar{Y}}_i)$ as $i \to \infty$ is

$$\lim_{i\to\infty} Var(\hat{\bar{Y}}_i) = Var(\hat{\bar{Y}}_\infty) = \frac{2S^2\sqrt{1-\rho^2}}{n\left(1+\sqrt{1-\rho^2}\right)}.$$

The limiting value of optimum sampling fraction as $i \to \infty$ is

$$\lim_{i\to\infty} \frac{\hat{m}_i}{n} = \frac{\hat{m}_\infty}{n} = \frac{\sqrt{1-\rho^2}}{g_\infty\left(1+\sqrt{1-\rho^2}\right)} = \frac{1}{2}.$$

Thus for the estimation of current population mean by this procedure, one would not have to match more than 50% of the sample drawn on the last occasion.

Unless ρ is very high, say more than 0.8, the reduction in variance $(1 - g_h)$ is only modest.

Type II estimation

Consider

$$\hat{\bar{Y}}_i = a_i \hat{\bar{Y}}_{i-1} + b_i \bar{y}_{i-1}^{**} + c\bar{y}_i^{***} + d_i \bar{y}_i^{**} + e_i \bar{y}_i^{*}$$

Now $\quad E(\hat{\bar{Y}}_i) = (a_i + b_i + c_i)\bar{Y}_{(i-1)} + (d_i + e_i)\bar{Y}_i.$

So for unbiasedness,

$$\begin{aligned} c_i &= -(a_i + b_i) \\ d_i &= 1 - e_{i.}. \end{aligned}$$

An unbiased estimator is of the form

$$\hat{\bar{Y}}_i = a_i \hat{\bar{Y}}_{(i-1)} + b_i \bar{y}_i^{**} - (a_i + b_i)\bar{y}_i^{***} + d_i \bar{y}_i^{**} + (1 - d_i)\bar{y}_i^{*}.$$

To find optimum weights, minimize

$Var(\hat{\bar{Y}}_i)$ with respect to a_i, b_i, d_i.

Alternatively, one can consider that

$$\hat{\bar{Y}}_i = a_i \hat{\bar{Y}}_{i-1} + b_i \bar{y}_i^{**} - (a_i + b_i)\bar{y}_i^{***} + d_i \bar{y}_i^{**} + (1 - a_i)\bar{y}_i^{*}.$$

must be uncorrelated with all unbiased estimators of zero. Thus

$$\begin{aligned} &Cov(\hat{\bar{Y}}_i, \bar{y}_{i-1}^{**} - \bar{y}_i^{***}) = 0 \\ &Cov(\hat{\bar{Y}}_{i-1}, \bar{y}_{i-1}^{**} - \bar{y}_i^{***}) = 0 \\ &Cov(\hat{\bar{Y}}_i, \bar{y}_{i-2}^{**} - \bar{y}_{i-1}^{***}) = 0 \end{aligned}$$

Using these restrictions, find the constants and get the estimator.