Muestreo indirecto

Método de estimación generalizada

Andrés Gutiérrez
predictive@telmex.net.co

2008

Contenido

Contenido

Contenido

Contenido

Contenido

Idea

Objetivo:

- Producir estimaciones del total de la característica de interés.

¿Cómo?:

- En la práctica los marcos de muestreo de elementos no están siempre disponibles.
- Acceso a diferentes marcos de muestreo que, si bien no hacen parte de la población objetivo, sí están indirectamente relacionados a ésta.
- El muestreo indirecto está caracterizado porque la producción de estimaciones de simples totales es muy compleja. Lavallée (2001).
- El método de ponderación generalizada está caracterizado por su simplicidad. Lavallée (2001).

Idea

Objetivo:

- Producir estimaciones del total de la característica de interés.

¿Cómo?:

- En la práctica los marcos de muestreo de elementos no están siempre disponibles.
- Acceso a diferentes marcos de muestreo que, si bien no hacen parte de la población objetivo, sí están indirectamente relacionados a ésta.
- El muestreo indirecto está caracterizado porque la producción de estimaciones de simples totales es muy compleja. Lavallée (2001).
- El método de ponderación generalizada está caracterizado por su simplicidad. Lavallée (2001).

Idea

Objetivo:

- Producir estimaciones del total de la característica de interés.

¿Cómo?:

- En la práctica los marcos de muestreo de elementos no están siempre disponibles.
- Acceso a diferentes marcos de muestreo que, si bien no hacen parte de la población objetivo, sí están indirectamente relacionados a ésta.
- El muestreo indirecto está caracterizado porque la producción de estimaciones de simples totales es muy compleja. Lavallée (2001).
- El método de ponderación generalizada está caracterizado por su simplicidad. Lavallée (2001).

Idea

Objetivo:

- Producir estimaciones del total de la característica de interés.

¿Cómo?:

- En la práctica los marcos de muestreo de elementos no están siempre disponibles.
- Acceso a diferentes marcos de muestreo que, si bien no hacen parte de la población objetivo, sí están indirectamente relacionados a ésta.
- El muestreo indirecto está caracterizado porque la producción de estimaciones de simples totales es muy compleja. Lavallée (2001).
- El método de ponderación generalizada está caracterizado por su simplicidad. Lavallée (2001).

Idea

Objetivo:

- Producir estimaciones del total de la característica de interés.

¿Cómo?:

- En la práctica los marcos de muestreo de elementos no están siempre disponibles.
- Acceso a diferentes marcos de muestreo que, si bien no hacen parte de la población objetivo, sí están indirectamente relacionados a ésta.
- El muestreo indirecto está caracterizado porque la producción de estimaciones de simples totales es muy compleja. Lavallée (2001).
- El método de ponderación generalizada está caracterizado por su simplicidad. Lavallée (2001).

Fundamentos

- Estrategia de muestreo $(p(S), \hat{T}(S))$.
- Es imprescindible el acceso, al menos de manera implícita, a un marco de muestreo de la población objetivo, denotada como U_B.
- Es posible considerar la disponibilidad de un marco muestral de elementos de alguna otra población, U_A, vinculados con los elementos de U_B.
- Caso particular: muestreo de conglomerados, cuando el marco de muestreo de la población U_A es de conglomerados

Fundamentos

- Estrategia de muestreo $(p(S), \hat{T}(S))$.
- Es imprescindible el acceso, al menos de manera implícita, a un marco de muestreo de la población objetivo, denotada como U_B.
- Es posible considerar la disponibilidad de un marco muestral de elementos de alguna otra población, U_A, vinculados con los elementos de U_B.
- Caso particular: muestreo de conglomerados, cuando el marco de muestreo de la población U_A es de conglomerados

Fundamentos

- Estrategia de muestreo $(p(S), \hat{T}(S))$.
- Es imprescindible el acceso, al menos de manera implícita, a un marco de muestreo de la población objetivo, denotada como U_B.
- Es posible considerar la disponibilidad de un marco muestral de elementos de alguna otra población, U_A, vinculados con los elementos de U_B.
- Caso particular: muestreo de conglomerados, cuando el marco de muestreo de la población U_A es de conglomerados

Fundamentos

- Estrategia de muestreo $(p(S), \hat{T}(S))$.
- Es imprescindible el acceso, al menos de manera implícita, a un marco de muestreo de la población objetivo, denotada como U_B.
- Es posible considerar la disponibilidad de un marco muestral de elementos de alguna otra población, U_A, vinculados con los elementos de U_B.
- Caso particular: muestreo de conglomerados, cuando el marco de muestreo de la población U_A es de conglomerados

Fundamentos

¿Cómo hacerlo?

- Seleccionar una muestra probabilística s_A de la población U_A para obtener estimaciones para la población U_B usando la correspondencia entre las dos poblaciones.
- Obtener estimaciones de una población de niños, aunque sólo se tiene acceso a una lista de padres conteniendo la respectiva identificación y ubicación de cada uno de ellos.
- La población objetivo, es la población de niños, pero es necesario seleccionar una muestra de padres para poder entrevistar a los niños.

Fundamentos

¿Cómo hacerlo?

- Seleccionar una muestra probabilística s_A de la población U_A para obtener estimaciones para la población U_B usando la correspondencia entre las dos poblaciones.
- Obtener estimaciones de una población de niños, aunque sólo se tiene acceso a una lista de padres conteniendo la respectiva identificación y ubicación de cada uno de ellos.
- La población objetivo, es la población de niños, pero es necesario seleccionar una muestra de padres para poder entrevistar a los niños.

Fundamentos

¿Cómo hacerlo?

- Seleccionar una muestra probabilística s_A de la población U_A para obtener estimaciones para la población U_B usando la correspondencia entre las dos poblaciones.
- Obtener estimaciones de una población de niños, aunque sólo se tiene acceso a una lista de padres conteniendo la respectiva identificación y ubicación de cada uno de ellos.
- La población objetivo, es la población de niños, pero es necesario seleccionar una muestra de padres para poder entrevistar a los niños.

Notación

- La población U_A contiene N_A unidades.
- Cada unidad perteneciente a la población de U_A será rotulada con la letra j. Cada unidad perteneciente a la población objetivo U_B será rotulada con la letra i.
- La correspondencia entre cada las dos poblaciones U_A y U_B pueden ser representadas por una matriz de vínculos, denotada por $\Theta_{AB} = [\theta_{ij}^{AB}]$ de tamaño $N_A \times N_B$.

Notación

- La población U_A contiene N_A unidades.
- Cada unidad perteneciente a la población de U_A será rotulada con la letra j. Cada unidad perteneciente a la población objetivo U_B será rotulada con la letra i.
- La correspondencia entre cada las dos poblaciones U_A y U_B pueden ser representadas por una matriz de vínculos, denotada por $\Theta_{AB} = [\theta_{ij}^{AB}]$ de tamaño $N_A \times N_B$.

Notación

- La población U_A contiene N_A unidades.
- Cada unidad perteneciente a la población de U_A será rotulada con la letra j. Cada unidad perteneciente a la población objetivo U_B será rotulada con la letra i.
- La correspondencia entre cada las dos poblaciones U_A y U_B pueden ser representadas por una matriz de vínculos, denotada por $\Theta_{AB} = [\theta_{ij}^{AB}]$ de tamaño $N_A \times N_B$.

Matriz de vínculos

Los posibles valores de la matriz están dados de la siguiente manera

$$\theta_{ij}^{AB} \begin{cases} > 0, & \text{si } j \text{ está relacionado con } i; \\ = 0, & \text{en otro caso.} \end{cases} \tag{1}$$

- Usualmente, si hay vínculo entre el elemento j-ésimo de la población U_A y el i-ésimo elemento de la población U_B, $\theta_{ij}^{AB} = 1$.

Matriz de vínculos

En el ejemplo de los padres, si la matriz de vínculos está dada por

$$\Theta_{AB} = \begin{pmatrix} \theta_{11}^{AB} & \theta_{12}^{AB} & 0 \\ 0 & \theta_{22}^{AB} & 0 \\ 0 & 0 & \theta_{33}^{AB} \\ 0 & 0 & \theta_{43}^{AB} \end{pmatrix} \tag{2}$$

- La primera pareja, dada por los elementos 1 y 2 de la población U_A tiene un hijo:.
- El elemento 1 de la población U_A tiene un hijo por fuera del matrimonio.
- La segunda pareja, dada por los elementos 3 y 4 de la población U_A tiene sólo un hijo.

Matriz de vínculos

En el ejemplo de los padres, si la matriz de vínculos está dada por

$$\Theta_{AB} = \begin{pmatrix} \theta_{11}^{AB} & \theta_{12}^{AB} & 0 \\ 0 & \theta_{22}^{AB} & 0 \\ 0 & 0 & \theta_{33}^{AB} \\ 0 & 0 & \theta_{43}^{AB} \end{pmatrix} \tag{2}$$

- La primera pareja, dada por los elementos 1 y 2 de la población U_A tiene un hijo:.
- El elemento 1 de la población U_A tiene un hijo por fuera del matrimonio.
- La segunda pareja, dada por los elementos 3 y 4 de la población U_A tiene sólo un hijo.

Matriz de vínculos

En el ejemplo de los padres, si la matriz de vínculos está dada por

$$\Theta_{AB} = \begin{pmatrix} \theta_{11}^{AB} & \theta_{12}^{AB} & 0 \\ 0 & \theta_{22}^{AB} & 0 \\ 0 & 0 & \theta_{33}^{AB} \\ 0 & 0 & \theta_{43}^{AB} \end{pmatrix} \tag{2}$$

- La primera pareja, dada por los elementos 1 y 2 de la población U_A tiene un hijo:.
- El elemento 1 de la población U_A tiene un hijo por fuera del matrimonio.
- La segunda pareja, dada por los elementos 3 y 4 de la población U_A tiene sólo un hijo.

Notación

- Se selecciona una muestra s_A de tamaño n_A mediante el uso de un diseño muestral $p_A(s_A)$.
- Sean $\pi_j^A > 0 \ \forall j \in U_A$ la probabilidad de inclusión del j-ésimo elemento.
- Para cada elemento en s_A se identifican las unidades en U_B cuya correspondencia con U_A es no nula. Es decir $\theta_{ij}^{AB} > 0$.

Notación

- Se selecciona una muestra s_A de tamaño n_A mediante el uso de un diseño muestral $p_A(s_A)$.
- Sean $\pi_j^A > 0 \ \forall j \in U_A$ la probabilidad de inclusión del j-ésimo elemento.
- Para cada elemento en s_A se identifican las unidades en U_B cuya correspondencia con U_A es no nula. Es decir $\theta_{ij}^{AB} > 0$.

Notación

- Se selecciona una muestra s_A de tamaño n_A mediante el uso de un diseño muestral $p_A(s_A)$.
- Sean $\pi_j^A > 0 \ \forall j \in U_A$ la probabilidad de inclusión del j-ésimo elemento.
- Para cada elemento en s_A se identifican las unidades en U_B cuya correspondencia con U_A es no nula. Es decir $\theta_{ij}^{AB} > 0$.

Notación

- Sea s_B el conjunto de n_B unidades de la población objetivo que se lograron identificar con ayuda de los elemento pertenecientes a la población U_A.

$$s_B = \{i \in B \mid \exists j \in s_A \text{ y } \theta_{ij}^{AB} > 0\} \quad (3)$$

- Se realiza el proceso de medición de la característica de interés y en s_B.
- El tamaño de muestra es aleatorio y es difícil establecer un presupuesto.
- Es posible predecir el número de vínculos entre las poblaciones (por ejemplo, un padre tiene uno o dos hijos).

Notación

- Sea s_B el conjunto de n_B unidades de la población objetivo que se lograron identificar con ayuda de los elemento pertenecientes a la población U_A.

$$s_B = \{i \in B \mid \exists j \in s_A \text{ y } \theta_{ij}^{AB} > 0\} \tag{3}$$

- Se realiza el proceso de medición de la característica de interés y en s_B.
- El tamaño de muestra es aleatorio y es difícil establecer un presupuesto.
- Es posible predecir el número de vínculos entre las poblaciones (por ejemplo, un padre tiene uno o dos hijos).

Notación

- Sea s_B el conjunto de n_B unidades de la población objetivo que se lograron identificar con ayuda de los elemento pertenecientes a la población U_A.

$$s_B = \{i \in B \mid \exists j \in s_A \text{ y } \theta_{ij}^{AB} > 0\} \tag{3}$$

- Se realiza el proceso de medición de la característica de interés y en s_B.
- El tamaño de muestra es aleatorio y es difícil establecer un presupuesto.
- Es posible predecir el número de vínculos entre las poblaciones (por ejemplo, un padre tiene uno o dos hijos).

Notación

- Sea s_B el conjunto de n_B unidades de la población objetivo que se lograron identificar con ayuda de los elemento pertenecientes a la población U_A.

$$s_B = \{i \in B \mid \exists j \in s_A \text{ y } \theta_{ij}^{AB} > 0\} \tag{3}$$

- Se realiza el proceso de medición de la característica de interés y en s_B.
- El tamaño de muestra es aleatorio y es difícil establecer un presupuesto.
- Es posible predecir el número de vínculos entre las poblaciones (por ejemplo, un padre tiene uno o dos hijos).

Estimación del total

Definition

La **matriz estandarizada de vínculos** está dada por

$$\tilde{\Theta}_{AB} = \Theta_{AB}[\mathrm{diag}(\mathbf{1}_A \Theta_{AB})]^{-1} \tag{4}$$

Nótese que!!!

$$\mathbf{1}'_A \Theta_{AB} = (\theta^{AB}_{+1}, \theta^{AB}_{+2}, \ldots, \theta^{AB}_{+N_B}) \tag{5}$$

donde $\theta^{AB}_{+i} = \sum_{j \in U_A} \theta^{AB}_{ji}$ debe ser no nula para todo $i \in U_B$. Significaría que todo hijo debe estar vinculado al menos a un padre.

Estimación del total

Definition

La **matriz estandarizada de vínculos** está dada por

$$\tilde{\Theta}_{AB} = \Theta_{AB}[\mathrm{diag}(\mathbf{1}_A\Theta_{AB})]^{-1} \tag{4}$$

Nótese que!!!

$$\mathbf{1}'_A\Theta_{AB} = (\theta^{AB}_{+1}, \theta^{AB}_{+2}, \ldots, \theta^{AB}_{+N_B}) \tag{5}$$

donde $\theta^{AB}_{+i} = \sum_{j\in U_A} \theta^{AB}_{ji}$ debe ser no nula para todo $i \in U_B$. Significaría que todo hijo debe estar vinculado al menos a un padre.

Resultados

Theorem

Si $\tilde{\Theta}_{AB}$ *es una matriz de vínculo estándar, entonces*

$$\tilde{\Theta}'_{AB}\mathbf{1}_A = \mathbf{1}_B \tag{6}$$

Theorem

$$t_y = \sum_{j \in U_A} \sum_{i \in U_B} \frac{\theta_{ji}^{AB}}{\theta_{+i}^{AB}} y_i \tag{7}$$

Resultados

Theorem

Si $\tilde{\Theta}_{AB}$ es una matriz de vínculo estándar, entonces

$$\tilde{\Theta}'_{AB}\mathbf{1}_A = \mathbf{1}_B \tag{6}$$

Theorem

$$t_y = \sum_{j \in U_A} \sum_{i \in U_B} \frac{\theta_{ji}^{AB}}{\theta_{+i}^{AB}} y_i \tag{7}$$

MPG

- Partiendo de que

$$
\begin{aligned}
t_y &= \mathbf{1}'_A \tilde{\Theta}_{AB} \mathbf{y} \\
&= \mathbf{1}'_A \mathbf{Z}
\end{aligned}
$$

- Es posible construir una expresión para el estimador de Narain-Horvitz-Thompson en términos del vector $\mathbf{Z}$. Por lo tanto

$$
\begin{aligned}
\hat{t}_y = \hat{t}_{z,NHT} &= \mathbf{1}'_A \mathbf{I_A} \mathbf{\Pi}_A^{-1} \mathbf{Z} && (8) \\
&= \mathbf{1}'_A \mathbf{I_A} \mathbf{\Pi}_A^{-1} \tilde{\Theta}_{AB} \mathbf{y} && (9)
\end{aligned}
$$

MPG

- Partiendo de que

$$\begin{aligned} t_y &= \mathbf{1}'_A \tilde{\Theta}_{AB} \mathbf{y} \\ &= \mathbf{1}'_A \mathbf{Z} \end{aligned}$$

- Es posible construir una expresión para el estimador de Narain-Horvitz-Thompson en términos del vector $\mathbf{Z}$. Por lo tanto

$$\hat{t}_y = \hat{t}_{z,NHT} = \mathbf{1}'_A \mathbf{I_A} \mathbf{\Pi}_A^{-1} \mathbf{Z} \tag{8}$$

$$= \mathbf{1}'_A \mathbf{I_A} \mathbf{\Pi}_A^{-1} \tilde{\Theta}_{AB} \mathbf{y} \tag{9}$$

MPG

Luego, el vector de ponderaciones toma la forma

$$\mathbf{w} = \mathbf{1}'_A \mathbf{I_A} \mathbf{\Pi}_A^{-1} \tilde{\Theta}_{AB} \tag{10}$$

- Cada elemento de **w** está definido como

$$w_i = \sum_{j \in U_A} I_j \frac{\tilde{\Theta}_{ji}^{AB}}{\pi_j^A} \qquad \text{para todo } i \in U_B \tag{11}$$

- De esta forma, se dice que los pesos w_i han sido obtenidos mediante el método de ponderación generalizada

MPG

Luego, el vector de ponderaciones toma la forma

$$\mathbf{w} = \mathbf{1}'_A \mathbf{I_A} \mathbf{\Pi}_A^{-1} \tilde{\Theta}_{AB} \tag{10}$$

- Cada elemento de **w** está definido como

$$w_i = \sum_{j \in U_A} I_j \frac{\tilde{\Theta}_{ji}^{AB}}{\pi_j^A} \qquad \text{para todo } i \in U_B \tag{11}$$

- De esta forma, se dice que los pesos w_i han sido obtenidos mediante el método de ponderación generalizada

Resultados

Theorem

El estimador de t_y es insesgado.

Theorem

El vector $\mathbf{w}$ provee estimaciones insesgadas si y sólo si la matriz $\tilde{\Theta}_{AB}$ es una matriz de vínculo estándar.

Resultados

Theorem

El estimador de t_y es insesgado.

Theorem

El vector $\mathbf{w}$ provee estimaciones insesgadas si y sólo si la matriz $\tilde{\Theta}_{AB}$ es una matriz de vínculo estándar.

Resultados

Theorem

La varianza de $\hat{t}_y$ está dada por

$$Var(\hat{t}_y) = \mathbf{Z}'\mathbf{\Delta}_A\mathbf{Z} \tag{12}$$
$$= \mathbf{y}'\Delta_B\mathbf{y} \tag{13}$$

donde Δ_A es la matriz de varianzas y covarianzas de tamaño $N_A \times N_A$ de las variables indicadoras de los elementos de la población U_A y $\Delta_B = \tilde{\Theta}'_{AB}\Delta_B\tilde{\Theta}_{AB}$.

Matriz Identidad

La población U_A y la población U_B tienen una correspondencia uno a uno. Así, $N_A = N_B = N$ y la matriz de vínculo está dada por

$$\Theta_{AB} = \mathbf{I}_{N\times N} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{pmatrix}$$

Por tanto, $\mathbf{w} = (\frac{I_1^A}{\pi_1^A}, \cdots, \frac{I_{N_A}^A}{\pi_{N_A}^A})$ y $\mathbf{Z} = \mathbf{y}$. Luego, el estimador $\hat{t}_y$ será

$$\hat{t}_y = \hat{t}_{y,NHT} = \mathbf{1}_A'\mathbf{I_A}\mathbf{\Pi}_A^{-1}\mathbf{y} \tag{14}$$

Uno para todos

La población objetivo se encuentra particionada en N_I^B conglomerados, cada uno de tamaño N_i^B. Cada conglomerados de U_I^B está asociado exactamente con un elemento j de U_A. Nótese que $N_I^B = N^A$. Por tanto la matriz de vinculo está dada por

$$\Theta_{AB} = \begin{pmatrix} \mathbf{1}'_{B1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}'_{B2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{1}'_{BN_I^B} \end{pmatrix} \tag{15}$$

Uno para todos

- El vector de pesos $\mathbf{w}$ está definido como

$$\mathbf{w} = \left(\frac{I_1^A}{\pi_1^A}\mathbf{1}_{B1}, \ldots, \frac{I_{N_I^A}^A}{\pi_1^A}\mathbf{1}_{BN_I^B}\right)' \tag{16}$$

- El estimador se puede escribir como

$$\hat{t}_y = \sum_{i=1}^{N_I^B} \frac{I_1^A}{\pi_1^A} t_{y,U_i} \tag{17}$$

donde $t_{y,U_i} = \sum_{k \in U_i^B} y_k$ es el total de i-ésimo conglomerado de la población U_I^B

Uno para todos

- El vector de pesos $\mathbf{w}$ está definido como

$$\mathbf{w} = \left(\frac{I_1^A}{\pi_1^A}\mathbf{1}_{B1}, \ldots, \frac{I_{N_I^A}^A}{\pi_1^A}\mathbf{1}_{BN_I^B}\right)' \tag{16}$$

- El estimador se puede escribir como

$$\hat{t}_y = \sum_{i=1}^{N_I^B} \frac{I_1^A}{\pi_1^A} t_{y,U_i} \tag{17}$$

donde $t_{y,U_i} = \sum_{k \in U_i^B} y_k$ es el total de i-ésimo conglomerado de la población U_I^B

Todos para uno

En este caso se considera que la población U^A está particionada en N_I^A conglomerados, cada uno de tamaño N_j^A $j = 1, \ldots, N_I^A$. Cada conglomerados de U_I^A está asociado exactamente con un elemento i de U_B. Nótese que $N_I^A = N^B$. Por tanto la matriz de vinculo está dada por

$$\Theta_{AB} = \begin{pmatrix} \mathbf{1}_{A1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{1}_{A2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \ldots & \mathbf{1}'_{AN_I^A} \end{pmatrix} \tag{18}$$

Todos para uno

- El vector de pesos **w** está definido como

$$\mathbf{w} = \left(\frac{1}{N_1^A}\sum_{j\in U_I^A}\frac{I_j^A}{\pi_j^A}, \ldots, \frac{1}{N_{N_I^A}^A}\sum_{j\in U_{N_I^A}^A}\frac{I_j^A}{\pi_j^A}\right)' \tag{19}$$

- El estimador es

$$\hat{t}_y = \sum_{i=1}^{N_I^A}\frac{y_i}{N_i^A}\sum_{j\in U_i^A}\frac{I_j^A}{\pi_j^A} \tag{20}$$

Todos para uno

- El vector de pesos **w** está definido como

$$\mathbf{w} = \left(\frac{1}{N_1^A}\sum_{j\in U_I^A}\frac{I_j^A}{\pi_j^A}, \ldots, \frac{1}{N_{N_I^A}^A}\sum_{j\in U_{N_I^A}^A}\frac{I_j^A}{\pi_j^A}\right)' \qquad (19)$$

- El estimador es

$$\hat{t}_y = \sum_{i=1}^{N_I^A}\frac{y_i}{N_i^A}\sum_{j\in U_i^A}\frac{I_j^A}{\pi_j^A} \qquad (20)$$

Agradecimientos

Muchas gracias !!!

http://predictive.wordpress.com