UNIVERSIDAD DE GRANADA

ENCUESTAS CON MARCOS MÚLTIPLES

Máster en Estadística Aplicada

Directores:
Prof. Dr. D. Antonio Arcos Cebrián
Prof. Dra. Dña. María del Mar Rueda García

Alumno:
David Molina Muñoz

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Granada 2012

ENCUESTAS CON MARCOS MÚLTIPLES

Memoria presentada por David Molina Muñoz para aspirar al Máster en Estadística Aplicada por la Universidad de Granada

DEPARTAMENTO DE ESTADÍSTICA E INVESTIGACIÓN OPERATIVA

Fdo. David Molina Muñoz

V. B. Directores del trabajo fin de Máster:

Fdo. Antonio Arcos Cebrián   Fdo. María del Mar Rueda García

UNIVERSIDAD DE GRANADA

Granada 2012

Índice general

Encuestas con marcos múltiples

0.1. ¿Qué son las encuestas con marcos múltiples, y por qué se usan?

Supongamos que se quiere realizar una encuesta sobre los estadísticos de Estados Unidos. Una aproximación podría ser usar el directorio de miembros de la Asociación Americana de Estadística (ASA) como marco muestral y tomar una muestra probabilística de personas de la lista de miembros. La mayor parte de las personas con las que se contactaría mediante esta muestra serían, probablemente, estadísticos, de manera que el marco sería rentable para ponerse en contacto con estadísticos. Sin embargo, muchos estadísticos en Estados Unidos no pertenecen a la ASA, y se tendrá una cobertura insuficiente de la población de estadísticos si solo se muestrea a partir del directorio de miembros de la ASA.

Se puede mejorar la cobertura tomando una muestra adicional de un segundo marco muestral, como el directorio de miembros del Instituto de Estadística Matemática (IMS). Así, se podría tomar una muestra probabilística del marco A (el directorio de la ASA), e independientemente tomar una muestra probabilística del marco B (el directorio del IMS). Los dos marcos se solapan, ya que muchos estadísticos pertenecen a ambas organizaciones. Como se muestra en la Fig. 1, hay tres ámbitos de población: el ámbito a se compone de las unidades poblacionales que están en el marco A pero no en el marco B, el ámbito b se compone de las unidades poblacionales que están en el marco B pero no en el marco A, y el ámbito ab se compone de las unidades poblacionales que están tanto en el marco A como en el B.

Los marcos A (directorio de miembros de la ASA) y B (directorio de miembros del IMS) producen una mejor cobertura de la población pero siguen sin incluir a todos los estadísticos. En este caso, se podría usar un muestreo con tres marcos, donde el marco C podría ser un marco con toda la población adulta. La estructura de la población con tres marcos se muestra en la Fig. 2. Este diseño con marcos múltiples tiene cuatros ámbitos: el ámbito c es la parte de la población de todo el marco C que no está en ninguno de los dos directorios.

Puesto que el marco C contiene a toda la población, podría preguntarse el por qué del trabajo adicional que supone tomar una muestra de marcos múltiples: ¿por qué no se podría tomar simplemente una muestra del marco C? La respuesta es el costo: el marco C, puesto que contiene la totalidad de la población, es muy caro de muestrear. Dado que la mayoría de los adultos no son estadísticos, muchos diseños muestrales del marco C albergarán muy pocas personas estadísticas. Los marcos A y B, por el contrario, son mucho menos caros de muestrear. Pero no incluyen la totalidad de la población de interés. Combinando las muestras de los marcos A, B y C, se puede aprovechar la economicidad de los marcos A y B a la vez que se evita el sesgo incluyendo la parte que falta de la población -ámbito *c*- en el marco C.

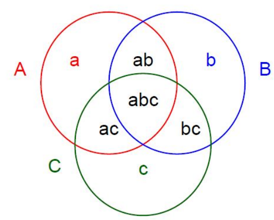

Figura 1: Los marcos A y B, con superposición y tres ámbitos.

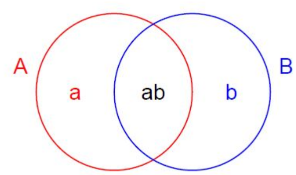

Figura 2: El marco C contiene la totalidad de la población; los marcos A y B se superponen y ambos están contenidos en el marco C.

Hay varias razones para considerar el uso de una encuesta con marcos múltiples para una colección de datos. Las encuestas con marcos múltiples pueden mejorar notablemente la eficiencia del conjunto de datos cuando se muestrea una población rara. Una población rara es un subgrupo de interés que comprende solamente una pequeña parte, normalmente el 10 % o menos, del total de la población. En el ejemplo planteado, los estadísticos son un pequeño segmento de la población adulta; complementando la encuesta sobre la población general con las muestras de los directorios, el número de estadísticos en el conjunto de datos se incrementa. Una encuesta determinada con una población específica como las personas con una enfermedad rara puede obtener un mayor tamaño muestral muestreando en marcos adicionales que contengan una alta tasa de personas de la categoría deseada.

Las encuestas con marcos múltiples pueden emplearse cuando se deben utilizar diferentes técnicas para la recolección de los datos para alcanzar los distintos segmentos de la población; por ejemplo, el marco A en la Fig. 1 podria ser un marco de números de teléfono fijo y el marco B podria componerse de números de teléfono móvil. Hartley (1962), cuando introdujo la teoría de las encuestas con marcos múltiples, escribió que la principal razón para considerar una encuesta con marcos múltiples es reducir los costes de la recolección de datos al mismo tiempo que se muestrea toda la población objeto de estudio:

> En la metodología de las encuestas por muestreo uno encuentra, a menudo, que un marco conocido que cubre aproximadamente *todas* las unidades de la población resulta muy costoso de muestrear, mientras que hay otros marcos disponibles (por ejemplo, las listas de unidades especiales) para métodos de muestreo más económicos. Sin embargo, éstos últimos solo cubren una proporción aproximadamente conocida o incluso desconocida de la población. Este artículo desarrolla una metodología general sobre el empleo de cualquier número de tales marcos sin necesidad de ningún conocimiento previo sobre el alcance de su solapamiento mutuo. (Hartley, 1962, p. 203)

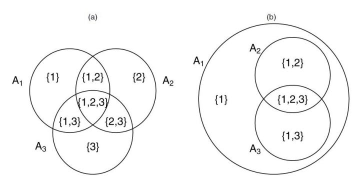

Figura 3: El marco B es un subconjunto del marco A.

La situación que consideró Hartley (1962) se representa en la Fig. 3. Esta es la situación más común para la que se utiliza el muestreo con marcos múltiples. A menudo, el marco B es un marco de lista; por ejemplo, en una encuesta agrícola, el marco B podría ser una lista de las granjas grandes de un país. La lista puede estar anticuada, y por tanto, no incluir granjas creadas recientemente. Además, la lista no incluirá las explotaciones agrícolas más pequeñas. Aunque en el marco B se pueden emplear métodos eficientes para la recolección de datos, la cobertura incompleta de la población se traduce en que las estimaciones de la cantidad total de tierra sembrada con soja usando únicamente la muestra del marco B serán, probablemente, muy pequeñas. Utilizando un marco de área procedente del marco A, conseguimos una cobertura total de las granjas del país. En un marco de área, el país se divide en áreas geográficas y se muestrean los segmentos de área. Los entrevistadores viajan a cada uno de los segmentos seleccionados y recolectan los datos de todas las explotaciones agrarias dentro de los límites del segmento. El marco de área, por tanto, incluye todas las granjas del país, pero la recolección de datos es mucho más cara que en el marco de lista.

La Encuesta de Finanzas del Consumidor de Estados Unidos (Bucks et al., 2006) es un ejemplo de encuesta con dos marcos que concuerda con la Fig. 3.

Una muestra (la muestra del marco A) se selecciona de entre la población de Estados Unidos usando un marco de área con un diseño muestral estratificado multietápico. Sin embargo, la Comisión de la Reserva Federal de Estados Unidos también está interesada en las características de las familias que poseen inversiones como bonos libres de impuestos, de modo que la muestra del marco completo contendrá, probablemente, pocos hogares que posean este tipo tan raro de activos. Por tanto, una segunda muestra se toma del marco de lista (marco B) de registros elaborados a partir de la información de la declaración de impuestos.

0.1.1. Encuestas con marcos múltiples disjuntos

Considérese la situación planteada en la Fig. 3, en la cual el marco A contiene la totalidad de la población y el marco B contiene únicamente un subconjunto de la población. En algunos casos, es posible determinar qué unidades poblacionales del marco A están en el dominio ab y eliminarlas del marco A antes de muestrear. En una encuesta agrícola con dos marcos, una granja del marco de lista no sería entrevistada en la muestra del marco de área. Esta situación, en la cual las unidades duplicadas en los marcos de muestreo se eliminan antes de la recolección de datos, de manera que los marcos de muestreo son disjuntos, se denomina encuesta con marcos múltiples disjuntos. González-Villalobos y Wallace (1996) proponen ejemplos y métodos de estimación para encuestas con marcos múltiples disjuntos.

Dado que los marcos de muestreo no se superponen, los totales poblacionales en las encuestas con marcos múltiples disjuntos se estiman fácilmente sumando los totales poblacionales estimados de cada marco. Hansen et al. (1953) mencionaron un uso precoz de una encuesta con dos marcos disjuntos. La denominaron "muestreo conjunto a partir de una lista y de un área de muestreo en una o dos etapas" (p. 327), en la cual contaban con una lista incompleta, y posiblemente anticuada, de miembros de la población objeto de estudio. El marco de lista se consideró como estrato 1. Las unidades del marco de lista eran eliminadas del marco de área, de manera que el estrato 2 consistía en las unidades poblacionales que no estaban en el estrato 1. Una muestra multietápica se tomó del estrato 2. Como Hansen et al. (1953) puntualizaron, este diseño muestral es un caso especial de muestreo estratificado: cada unidad de la población está únicamente en un estrato, y el muestreo se lleva a cabo independientemente en los dos estratos. Los métodos de afijación del muestreo estratificado pueden usarse para determinar los tamaños muestrales en los dos estratos.

El proceso de screening puede hacerse en cualquier etapa de la recolección de datos. Una de las primeras encuestas con marco dual fue la Encuesta de Ventas al Por Menor, llevada a cabo por la Oficina del Censo de Estados Unidos en 1949 (Hansen et al., 1953, p.516). Las unidades primarias de muestreo (upm's) se eligieron usando una muestra estratificada de grupos de condados. Dentro de cada upm, no era factible obtener un listado actual completo de los negocios de venta al por menor dada la alta volatibilidad del pequeño negocio. Dentro de cada upm, un censo de empresas minoristas se tomó a partir de una lista de los registros de la Oficina de Seguros de la Tercera Edad y los Supervivientes; y se tomó una muestra de empresas que no estaban en la lista. En este caso, un diseño con dos marcos se usó dentro de cada upm seleccionada. Los diseños muestrales para los dos marcos no eran independientes, sin embargo, dado que las dos muestras compartian las mismas upm's. Así, el estimador del total de ventas se obtuvo como resultado de sumar los dos estimadores dentro de las upm's.

La Encuesta Nacional de Familias de América (Brick et al., 2002) tiene dos componentes: una encuesta telefónica de hogares con teléfono usando el método de marcado aleatorio y una muestra de área de hogares sin teléfono. Dado que el objetivo principal de la encuesta es recolectar datos sobre niños que viven en hogares de bajos ingresos, es importante incluir los hogares que no tienen teléfono en la muestra: en los hogares con ingresos bajos es altamente probable que se carezca de teléfono. Después de seleccionar la muestra de área, los hogares son examinados en busca de teléfono: sólo los hogares sin teléfono son entrevistados en el marco A. Como resultado del examen, el marco A está formado por aquellos hogares que no se encuentran en el marco B.

Incluso cuando los marcos se superponen, una encuesta con marcos múltiples puede analizarse como una encuesta con marcos múltiples disjuntos eliminando los registros del área de solapamiento. Las encuestas de la Fundación Nacional de Ciencia de Estados Unidos (2003) que se emplean en el Sistema Científico y de Ingenieria de Datos Estadísticos (SESTAT) se realizan como encuestas con tres marcos con solapamiento pero se analizan como si fueran tres marcos disjuntos. La población objeto de estudio son los residentes en Estados Unidos no pertenecientes a ninguna institución con menos de 75 años que (1) tienen al menos el título de bachiller y (2) o bien tienen una titulación universitaria de ciencias o ingeniería, o bien están trabajando como científicos o ingenieros. Los datos del SESTAT se basan en tres tipos de encuestas: (1) la Encuesta Nacional de Graduados de Bachiller (ENGB), con un marco de muestreo obtenido del Censo de Estados Unidos, el cual incluye las personas con, al menos, el título de bachillerato; (2) la Encuesta Nacional de Nuevos Graduados Universitarios (ENNGU), con un marco de muestreo de instituciones educativas, cuyas muestras recogen personas que obtuvieron un título universitario en ciencias o ingeniería dentro de los últimos dos años; y (3) la Encuesta de Estudiantes de Doctorado (EED), en la que se muestrea a gente que está estudiando un doctorado de ciencias o ingeniería en una institución de Estados Unidos. El marco muestral del EED se construye a partir de la Encuesta de Doctorandos, la cual es un censo de todos los doctorandos de Estados Unidos.

Con estos tres marcos de muestreo, una persona puede estar incluída en más de un marco. Por ejemplo, una persona que haya obtenido el título de bachiller antes del censo anual y después haya hecho un docorado estaría incluído en los tres marcos de muestreo. Los marcos muestrales de la SESTAT son, por tanto, los representados en la Fig. 2, donde C es el marco para la ENGB, B el marco para la ENNGU, y A el marco para la EED. Sin embargo, la multiplicidad no se usa en el análisis de los datos del SESTAT. En su lugar, se establece una jerarquía pra la inclusión en uno de los marcos. Una persona muestreada en la ENGB se elimina del análisis de datos (asignándole un peso de 0) si él o ella hubiera sido muestreado o en la ENNGU o en la EED. De forma similar, a un encuestado en la ENNGU se le otorga un peso de 0 si el encuestado está también en el marco de muestreo de la EED. Como resultado, la base de datos conjunta del SESTAT es, esencialmente, un muestreo estratificado con tres estratos: el marco de la EED, la parte del marco de la ENNGU que está fuera del marco de la EED, y la parte del marco de la ENGB que no se superpone con ninguno de los otros dos marcos.

Borrando registros de forma jerárquica, podemos analizar datos procedentes de encuestas con marcos múltiples con solapamiento, como los del SESTAT, como muestras estratificadas. Este enfoque, sin embargo, reduce el tamaño muestral y desecha datos. En las encuestas de la SESTAT, relativamente pocas unidades muestreadas en la ENGB se encuentran en las áreas de solapamiento, de modo que la pérdida de eficiencia es pequeña. En las encuestas con un solapamiento mayor, sin embargo, es mucho más eficiente usar toda la información recolectada de las áreas de superposición a la hora de estimar las características de la población. Las encuestas con marcos múltiples con solapamiento se tratan en la siguiente sección.

0.1.2. Encuestas con marcos múltiples con solapamiento

En muchas encuestas con marcos múltiples, no se sabe de antemano la pertenencia de una unidad poblacional a uno u otro dominio. Por ejemplo, si el marco A se compone de los números de teléfono fijo y el marco B esta formado de los números de teléfono móvil, no se conoce a priori si un miembro de un hogar muestreado usando un marco pertenece también al otro marco (Brick et al., 2006). La pertenencia a uno u otro dominia de una persona muestreada debe ser comprobado a partir de las cuestiones de la encuesta. En otros casos, puede ser difícil emparejas unidades en los marcos y comprobar si una persona del marco a es realmente la misma que una persona con el mismo nombre en el marco B.

Haines y Pollock (1998) describieron el uso de las encuestas con marcos múltiples para estimar el tamaño de una población y comentaron la correspondencia entre las encuestas con marcos múltiples y los métodos de captura-recaptura para estimar el tamaño de la población. Combinaron información procedente de marcos de lista incompletos y de marco de áreas para estimar el número de nidos de águila en una región. Iachan y Dennis (1993) usaron un diseño con marcos múltiples para muestrear a la población de indigentes, donde el marco A estaba formado por los albergues para indigentes, el marco B estaba compuesto por los comedores sociales y el marco C consistía en algunos lugares de la calle. Esta situación se representa en la Fig. 4. Aunque la unión de los tres marcos aún omite a parte de la población de indigentes, la cobertura se mejora, ya que se tiene más de un marco.

Algunas veces se toman muestras independientes del mismo marco, por lo que los marcos A y B coinciden. Esto puede ocurrir si se desea tener dos diseños muestrales. En una encuesta sobre fauna salvaje, una muestra puede ser tomada usando un muestreo estratificado polietápico y otra puede ser tomada mediante un diseño muestral secuencial o adaptativo. Cuando los marcos se

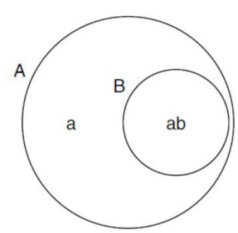

Figura 4: Los marcos A, B y C, incompletos y con superposición.

solapan, se debe tener cuidado al estimar cualquier cantidad. Los individuos de las áreas de solapamiento, como *ab* pueden ser seleccionados en otra u otras encuestas, de forma que los estimadores de estas cantidades poblacionales deben ajustarse para compensar esta multiplicidad. En la sección 2, se resumen los estimadores puntuales que se han propuesto para las encuestas con marcos múltiples.

0.2. Estimadores puntuales en encuestas con marcos múltiples

La teoría clásica del muestreo considera un único marco. El universo $\mathcal{U}$ tiene N unidades. El total poblacional para una característica y es $Y = \sum_{i=1}^{N} y_i$. Se usa un diseño muestral probabilístico para seleccionar una muestra $\mathcal{S}$ del marco. Denotamos por $\pi_i = P(i \in \mathcal{S})$ a las probabilidades de inclusión para el marco. El estimador de Horvitz-Thompson para el total de la población es

$$\hat{Y}_{HT} = \sum_{i \in \mathcal{S}} w_i y_i,$$

donde $w_i = 1/\pi_i$ son los pesos muestrales.

En esta sección, se propondrán métodos de estimación para encuestas con dos marcos cuando se toman muestras independientes de los dos marcos. Se considera aquí la situación de la Fig. 1, el caso más general. Seguimos teniendo que y_i es una característica de la unidad poblacional i, y el total poblacional es $Y = \sum_{i=1}^{N} y_i$. Pero ahora se toman dos muestras, una de las N_A unidades del marco A y otra muestra de las N_B unidades del marco B. Las unidades poblacionales que están en el área de solapamiento ab pueden ser muestreadas en cualquiera de las dos encuestas o en ambas encuestas.

Sea

$$\delta_i(a) = \begin{cases} 1 & \text{si la unidad } i \text{ está en el dominio } a \\ 0 & \text{en otro caso} \end{cases}$$

$$\delta_i(b) = \begin{cases} 1 & \text{si la unidad } i \text{ está en el dominio } b \\ 0 & \text{en otro caso} \end{cases}$$

y

$$\delta_i(ab) = \begin{cases} 1 & \text{si la unidad } i \text{ está en el dominio } ab \\ 0 & \text{en otro caso} \end{cases}$$

Para la encuestas general con dos marcos con solapamiento representada en la Fig. 1, podemos escribir el total poblacional como la suma de los totales de los tres dominios:

$$Y = Y_a + Y_{ab} + Y_b = \sum_{i=1}^{N} \delta_i(a) y_i + \sum_{i=1}^{N} \delta_i(ab) y_i + \sum_{i=1}^{N} \delta_i(b) y_i.$$

Un estimador de Y se puede obtener sumando los estimadores de Y_a, Y_{ab} e Y_b. Sea $\mathcal{S}_a$ la muestra del marco A, cuya probabilidad de inclusión denotaremos por $\pi_i^A = P\{i \in \mathcal{S}_A\}; \mathcal{S}_A$ contiene n_A unidades observadas. Las cantidades equivalentes para el marco B son $\mathcal{S}_B, \pi_i^B = P\{i \in \mathcal{S}_B\}$, y n_B. Sean w_i^A los pesos muestrales del marco A y w_i^B los pesos muestrales del marco B. Estos pesos pueden ser las inversas de las probabilidades de inclusión π_i^A y π_i^B o pueden ser los pesos de Hájek (1981), con $w_i^A = N_A[\pi_i^A \sum_{j \in \mathcal{S}_A}(1/\pi_j^A)]^{-1}$ y $w_i^B = N_B[\pi_i^B \sum_{j \in \mathcal{S}_B}(1/\pi_j^B)]^{-1}$. Usando estos pesos, se definen los estimadores de los dominios de la muestra del marco A:

$$\hat{Y}_a^A = \sum_{i \in \mathcal{S}_A} w_i^A \delta_i(a) y_i$$

$$\hat{Y}_{ab}^A = \sum_{i \in \mathcal{S}_A} w_i^A \delta_i(ab) y_i$$

Los estimadores correspondientes para la muestra del marco B son

$$\hat{Y}_b^B = \sum_{i \in \mathcal{S}_B} w_i^B \delta_i(b) y_i$$

$$\hat{Y}_{ab}^B = \sum_{i \in \mathcal{S}_B} w_i^B \delta_i(ab) y_i$$

Según los resultados clásicos de la teoría del muestreo, cada estimador para cada dominio es aproximadamente insesgados para la cantidad poblacional correspondiente. Los tamaños poblacionales se estiman tomando $y_i =!$ para todas las unidades, obteniendo como resultado los estimadores $\hat{N}_a^A, \hat{N}_{ab}^A, \hat{N}_{ab}^B$ y $\hat{N}_b^B$

En una escuesta con dos marcos disjuntos, el dominio ab es un conjunto vacío y los dominios a y b pueden verse como estratos. El total estimado para la población es una encuesta con dos marcos disjuntos es

$$\hat{Y} = \hat{Y}_a^A + \hat{Y}_b^B.$$

Si los marcos se superponen y las observaciones del dominio ab pueden ser seleccionadas en cualquiera de las dos muestras o en ambas muestras, necesitamos hacer algunos ajustes dada la multiplicidad en ab. Combinando ambas muestras, sin ajuste, obtendriamos un estimador sesgado:

$$E[\hat{Y}_a^A + \hat{Y}_b^B + \hat{Y}_{ab}^A + \hat{Y}_{ab}^B] \approx Y + Y_{ab}$$

Por tanto, se necesita combinar la información de los estimadores independientes $\hat{Y}_{ab}^A$ y $\hat{Y}_{ab}^B$ para estimar Y_{ab}. Las siguientes secciones describen algunos de los métodos que pueden usarse para estimar el total de la población Y.

0.2.1. Promediando las estimaciones de los dominios de solapamiento

El método más simple para esimar el total de la población es promediar los estimadores de los dominios que se muestrean en más de un marco. Para la encuesta con dos marcos de la Fig. 1, tanto $\hat{Y}_{ab}^A$ como $\hat{Y}_{ab}^B$ estiman Y_{ab}. así, para evitar los problemas de multiplicidad, podemos estimar Y por

$$\hat{Y}_{ave} = \hat{Y}_a^A + \frac{1}{2}(\hat{Y}_{ab}^A + \hat{Y}_{ab}^B) + \hat{Y}_b^B.$$

De forma más general, se puede usar una media ponderada, como la que propuso Hartley (1962):

$$\hat{Y}(\theta) = \hat{Y}_a^A + \theta\hat{Y}_{ab}^A + (1-\theta)\hat{Y}_{ab}^B + \hat{Y}_b^B, \tag{1}$$

donde $0 \leq \theta \leq 1$. Este estimador reduce el peso de cada unidad muestreada en el dominio de solapamiento ab para compensar la multiplicidad.Se definen los nuevos pesos

$$\tilde{w}_i^A = \delta_i(a)w_i^A + \theta\delta_i(ab)w_i^A$$

y

$$\tilde{w}_i^B = \delta_i(b)w_i^B + (1-\theta)\delta_i(ab)w_i^B.$$

Por tanto,

$$\hat{Y}(\theta) = \sum_{i\in\mathcal{S}_A} \tilde{w}_i^A y_i + \sum_{i\in\mathcal{S}_B} \tilde{w}_i^B y_i$$

Cada estimador para cada dominio es, aproximadamente insesgado para la cantidad poblacional que estima, por lo tanto $\hat{Y}(\theta)$ es un estimador aproximadamente insesgado para el total de la población Y. Dado que los marcos A y B se muestrean independientemente y θ está fijo, la varianza del estimador es

$$V[\hat{Y}(\theta)] = V(\hat{Y}_a^A + \theta\hat{Y}_{ab}^A) + V[(1-\theta)\hat{Y}_{ab}^B + \hat{Y}_b^B]. \tag{2}$$

Nótese que $\hat{Y}(0) = \hat{Y}_a + \hat{Y}_{ab}^B + \hat{Y}_b^B$; lo cual corresponde a eliminar las observaciones del dominio ab de la muestra del marco A. Por tanto, la estimación en encuestas con marcos múltiples disjuntos puede considerarse como un caso especial de la situación general.

0.2.2. El estimador de Hartley

El estimador de la sección 2.1 es sencillo de calcular pero puede ser poco eficiente en comparación con otros estimadores. Si el estimador $\hat{Y}_{ab}^B$ del marco

B tiene mucha más precisión que $\hat{Y}^A_{ab}$ para estimar el dominio total Y_{ab}, tendría sentido otorgar más peso a $\hat{Y}^B_{ab}$ para estimar Y_{ab}.

Hartley (1962, 1974) propuso elegir θ en (1) para minimizar la varianza de $\hat{Y}(\theta)$. Ya que los marcos se muestrean independientemente, la varianza de $\hat{Y}(\theta)$ es la dada en (2). Por ello, para los diseños generales de encuestas, la valor de θ que minimiza la varianza es

$$\theta_{opt} = \frac{V(\hat{Y}^B_{ab}) + Cov(\hat{Y}^B_b, \hat{Y}^B_{ab}) - Cov(\hat{Y}^A_a, \hat{Y}^A_{ab})}{V(\hat{Y}^A_{ab}) + V(\hat{Y}^B_{ab})} \tag{3}$$

El uso de θ_{opt} da la mínima varianza alcanzable:

$$V[\hat{Y}(\theta_{opt})] = V(\hat{Y}^A_a + \hat{Y}^B_b + \hat{Y}^B_{ab}) - \theta^2_{opt}[V(\hat{Y}^A_{ab}) + V(\hat{Y}^B_{ab})].$$

Puede verse en (3) que cuanto más grande sea $V(\hat{Y}^B_{ab})$ respecto a $V(\hat{Y}^A_{ab})$, más grande será θ_{opt}. Nótese que si $Cov(\hat{Y}^B_b, \hat{Y}^B_{ab})$ o $Cov(\hat{Y}^A_a, \hat{Y}^A_{ab})$ es grande en valor absoluto, que posible que θ_{opt} sea más pequeño que 0 o más grande que 1. Cuando los marcos A y B son el mismo, esto es, los dominios a y b son conjuntos vacios, sin embargo, θ_{opt} está entre 0 y 1.

En la práctica, las varianzas y covarianzas en (3) son desconocidas, por lo tanto, el valor óptimo de θ debe ser estimado a partir de los datos. Sea $\hat{theta}_{opt}$ el estimador de θ_{opt} que resulta cuando los estimadores de las vairanzas y las covarianzas se sustituyen en (3). Los pesos ajustados para el método de Hartley se convierten en

$$\tilde{w}^A_{i,H} = \delta_i(a)w^A_i + \hat{\theta}_{opt}\delta_i(ab)w^A_i$$

y

$$\tilde{w}^B_{i,H} = \delta_i(b)w^B_i + (1 - \hat{\theta}_{opt})\delta_i(ab)w^B_i.$$

Dado que $\hat{\theta}$ depende de las varianzas y covarianzas de la respuesta que se está estudiando, los ajustes para los pesos pueden diferir para cada respuesta. Esto puede dar lugar a incoherencias entre las estimaciones. Por ejemplo, supóngase que $\hat{Y}_1(\hat{\theta}_{opt,1})$ estima el total del gasto médico de la población con una edad superior a 65 años, $\hat{Y}_2(\hat{\theta}_{opt,2})$ estima el total en gasto médico en la población con 65 o menos años, y $\hat{Y}_3(\hat{\theta}_{opt,3})$ estima el gasto médico en la toda la población. Si las encuestas tienen un diseño complejo, es muy probable que $\hat{Y}_1(\hat{\theta}_{opt,1}) + \hat{Y}_2(\hat{\theta}_{opt,2}) \neq \hat{Y}_3(\hat{\theta}_{opt,3})$.

0.2.3. El estimador de Fuller - Burmeister

Fuller y Burmeister (1972) propusieron una modificación del estimador de Hartley que incorporaba información adicional relativa a la estimación de N_{ab}. El estimador es

$$\hat{Y}_{FB}(\beta) = \hat{Y}_a^A + \hat{Y}_b^B + \beta_1 \hat{Y}_{ab}^A + (1-\beta_1)\hat{Y}_{ab}^B + \beta_2(\hat{N}_{ab}^A - \hat{N}_{ab}^B) \tag{4}$$

Como con el estimador de Hartley, los parámetros β_1 y β_2 se eligen para minimizar la varianza de $\hat{Y}_{FB}(\beta)$; los valores óptimos son:

$$\begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} = -\begin{bmatrix} V(\hat{Y}_{ab}^A - \hat{Y}_{ab}^B) & Cov(\hat{Y}_{ab}^A - \hat{Y}_{ab}^B, \hat{N}_{ab}^A - \hat{N}_{ab}^B) \\ Cov(\hat{Y}_{ab}^A - \hat{Y}_{ab}^B, \hat{N}_{ab}^A - \hat{N}_{ab}^B) & V(\hat{N}_{ab}^A - \hat{N}_{ab}^B) \end{bmatrix}^{-1} \times \begin{bmatrix} Cov(\hat{Y}_a^A + \hat{Y}_b^B + \hat{Y}_{ab}^B, \hat{Y}_{ab}^A - \\ Cov(\hat{Y}_a^A + \hat{Y}_b^B + \hat{Y}_{ab}^B, \hat{N}_{ab}^A - \end{bmatrix}$$

Como sucedia con el estimador óptimo de Hartley, estas varianzas y covarianzas deben ser estimadas a partir de los datos; esto conlleva el utilizar un conjutno de pesos diferente para cada variable respuesta.

0.2.4. Estimador de marco único

Bankier (1986) y Kalton y Anderson (1986) propusieron métodos de marco único que combinaban la observaciones en un único conjunto de datos y después ajustaban los pesos en los ámbitos de intersección para tratar la multiplicidad. Una unidad i del dominio ab puede ser seleccionada en $\mathcal{S}_A$ y en $\mathcal{S}_B$, por lo que el número esperado de veces que la unidad i de ab es seleccionada es $\pi_i^A + \pi_i^B$. El estimador de un sólo marco de Kalton y Anderson (1986) utiliza $1/(\pi_i^A + \pi_i^B)$ como pesos para las unidades del dominio ab. Así, si $w_i^A = 1/\pi_i^A$ y $w_i^B = 1/\pi_i^B$, el peso ajustado para las unidades muestreadas en el marco A es

$$\tilde{w}_{i,S}^A = \begin{cases} w_i^A & \text{si } i \in a \\ (1/w_i^A + 1/w_i^B)^{-1} & \text{si } i \in ab. \end{cases}$$

Los pesos ajustados para el marco B se definen de forma similar, concretamente

$$\tilde{w}_{i,S}^B = \begin{cases} w_i^B & \text{si } i \in b \\ (1/w_i^A + 1/w_i^B)^{-1} & \text{si } i \in ab. \end{cases}$$

Por lo tanto, el estimador de marco único es

$$\hat{Y}_S = \sum_{i \in \mathcal{S}_A} \tilde{w}_{i,S}^A y_i + \sum_{i \in \mathcal{S}_B} \tilde{w}_{i,S}^B y_i \tag{5}$$

Si los tamaños poblacionales de los marcos N_A y N_B son conocidos, el estimador de marco único puede calibrarse con respecto a estos tamaños usando bien el estimador de razón ajustado (Bankier, 1986; Rao y Skinner, 1996) o bien el estimador de regresión (Lohr y Rao, 2000). Los pesos pueden ser ajustados sustituyendo $\tilde{w}^A_{i,S}$ en el marco A por $\tilde{w}^A_{i,S}(N^A/\sum_j \tilde{w}^A_{j,S})$, sustituyendo el peso $\tilde{w}^B_{i,S}$ en el marco B por $\tilde{w}^B_{i,S}(N^B/\sum_j \tilde{w}^B_{j,S})$ y repitiendo el proceso para los pesos en los marcos A y B hasta que se dé la convergencia. Rao y Skinner (1996) usan un argumento similar al usado por Skinner (1991) y muestran que el proceso de ajuste raking converge demostraron que el proceso de ajuste converge y dieron una expresión explícita para el estimador:

$$\hat{Y}_{Sfrake} = \frac{N_A - \hat{N}^{rake}_{ab}}{\hat{N}^A_a}\hat{Y}^A_a + \frac{N_B - \hat{N}^{rake}_{ab}}{\hat{N}^B_b}\hat{Y}^B_b + \frac{\hat{N}^{rake}_{ab}}{\hat{N}_{abS}}\hat{Y}_{abS} \tag{6}$$

donde

$\hat{Y}_{abS} = \sum_{i\in\mathcal{S}_A} \tilde{w}_{i,S}\delta_i(b)y_i + \sum_{i\in\mathcal{S}_B} \tilde{w}_{i,S}\delta_i(a)y_i,\ \hat{N}_{abS} = \sum_{i\in\mathcal{S}_A} \tilde{w}_{i,S}\delta_i(B) + \sum_{i\in\mathcal{S}_B} \tilde{w}_{i,S}\delta_i(A)$, y $\hat{N}^{rake}_{ab}$ es la menor de las raíces de la ecuación cuadrática:

$$\hat{N}_{abS}x^2 - (\hat{N}_{abS}(N_A + N_B) + \hat{N}^A_{aS} + \hat{N}^B_{bS})x + +\hat{N}_{abS}N_A N_B = 0.$$

Una opción alternativa es ajustar a los totales poblacionales via regresión, que proporciona el estimador:

$$\hat{Y}_{Sfreg} = \hat{Y}_S + \widehat{\beta}'_S(N_A - \hat{N}_{AS}, N_B - \hat{N}_{BS})' \tag{7}$$

donde $\widehat{\beta}'_S = -\widehat{cov}(\hat{N}_{AS}/V(\hat{N}_{AS}), \hat{N}_{BS}/V(\hat{N}_{BS}), \hat{Y}_S$.

Skinner et al. (1994) usaron la estimación de marco único para una muestra agrícola de marcos múltiples. Había varias variables respuesta de interés; cada una correlacionada con una de las posibles variables de estratificación. Se seleccionaron muestras aleatorias estratificadas independientes del marco de muestreo usando una variable de estratificación para cada muestra. Con cuatro muestras independientes $\mathcal{S}_A, \mathcal{S}_B, \mathcal{S}_C$ y $\mathcal{S}_D$, el peso para la unidad j para estar seleccionada, digamos, en la muestra del marco A, se estableció a $\tilde{w}^A_j = (\pi^A_j + \pi^B_j + \pi^C_j + \pi^D_j)^{-1}$. Luego, los pesos fueron ajustados al tamaño de la población N, como describe Skinner (1991).

Si cada muestra es auto-ponderada, esto es, todos los pesos para las unidades de muestreo en $\mathcal{S}_A$ son iguales a w^A y todos los pesos para las unidades de mestreo en $\mathcal{S}_B$ son iguales a w^B, entonces $\hat{Y}_S$ es un caso especial del estimador dado en (1). En este caso, $\hat{Y}_S = \hat{Y}_a^A + \theta_S \hat{Y}_{ab}^A + (1-\theta_S)\hat{Y}_{ab}^B + \hat{Y}_b^B$, con $\theta_S = \pi_i^A/(\pi_i^A + \pi_i^B) = [w^a(1/w^A + 1/w^B)]^{-1}$. El estimador de marco único da mayor peso al estimador de Y_{ab} procedente del marco que tiene unas mayores probabilidades de inclusión. La muestra del marco con las mayores probabilidades de inclusión no siempre tiene la varianza más pequeña, sin embargo, por lo que $V(\hat{Y}_S) \geq V[\hat{Y}(\theta_{opt})]$, siendo ésta última la varianza del estimador de Hartley con el valor óptimo de θ.

A diferencia de los estimadores de Hartley y de Fuller-Burmeister, el estimador de marco único no depende de las covarianzas estimadas para la variable respuesta. Utiliza el mismo conjunto de pesos para cada respuesta que se considere. El cálculo de los pesos en el ámbito ab, sin embargo, requiere del conocimiento de las probabilidades de inclusión de ambos marcos, no únicamente del marco del cual se selecciona la unidad. Si una de las muestras, digamos, por ejemplo, la muestra del marco A, no es autoponderada, entonces π_i^A puede resultar desconocido para una unidad seleccionada en $\mathcal{S}_B$.

0.2.5. Estimador de pseudo máxima verosimilitud

Cuando N_A y N_B son conocidos, Skinner y Rao (1996) propusieron la modificación de un estimador alternativo propuesto por Fuller y Burmeister (1972) para muestras aleatorias simples para obtener un estimador de pseudo máxima verosimilitud (PMV) que puede usarse con diseños complejos. El estimador de PMV usa el mismo conjunto de pesos para todas las variables respuesta y tiene la forma

$$\begin{aligned}\hat{Y}_{PMV}(\theta) = {} & \frac{N_A - \hat{N}_{ab}^{PML}(\theta)}{\hat{N}_a^A}\hat{Y}_a^A + \frac{N_B - \hat{N}_{ab}^{PML}(\theta)}{\hat{N}_b^B}\hat{Y}_b^B \\ & + \frac{\hat{N}_{ab}^{PML}(\theta)}{\theta\hat{N}_{ab}^A + (1-\theta)\hat{N}_{ab}^B}[\theta\hat{Y}_{ab}^A + (1-\theta)\hat{Y}_{ab}^B],\end{aligned} \tag{8}$$

donde $\hat{N}_{ab}^{PML}(\theta)$ es la más pequeñas de las raíces de la ecuación cuadrática $[\theta/N_B + (1-\theta)/N_A]x^2 - [1+\theta\hat{N}_{ab}^A/N_B + (1-\theta)\hat{N}_{ab}^B/N_A]x + \theta\hat{N}_{ab}^A + (1-\theta)\hat{N}_{ab}^B = 0$. Fuller y Burmeister (1972) y Skinner (1991) argumentaron que, cuando una muestra aleatoria simple se toma de cada marco, el estimador dado en (6) puede obtenerse usando los principios de máxima verosimilitud y, por tanto, es asintóticamente eficiente. El estimador de PMV sustituye los estimadores consistentes con el diseño de N_{ab}, Y_{ab}, Y_a, e Y_b por sus estimadores máximo verosímiles correspondientes obtenidos usando muestreo aleatorio simple. El estimador de PMV es, por ello, consistente bajo diseños de muestreo complejos: a diferencia del estimador dado en (4), no depende de las varianzas de $\hat{Y}_{ab}^A$ y $\hat{Y}_{ab}^B$. Aunque para el estimador de PMV no garantiza la optimalidad bajo diseños de muestreo complejos, pero a menudo se comporta bien en la práctica.

Skinner y Rao (1996) sugirieron usar el valor de $\theta = \theta_P$ que minimiza la varianza asintótica de $\hat{N}_{ab}^{PML}(\theta)$:

$$\theta_P = \frac{N_a N_b V(\hat{N}_{ab}^B)}{N_a N_b V(\hat{N}_{ab}^B) + N_b N_a V(\hat{N}_{ab}^A)} \tag{9}$$

El estimador dado en (6) ajusta los estimadores de los tres totales de los dominios Y_a, Y_{ab}, e Y_b mediante el estimador óptimo de N_{ab}. Nótese que el cálculo de θ_P en (7) requiere que los tres dominios a, b, y ab sean no vacíos y requieres que las varianzas de $\hat{N}_{ab}^A$ y $\hat{N}_{ab}^B$ sean positivas. Por tanto, un método diferente debe ser utilizado para determinar θ en (6) si, por ejemplo, la muestra procedente del marco B es post estratificada de modo que $\hat{N}_{ab}^B = N_{ab}$ y $V(\hat{N}_{ab}^B) = 0$. En tales situaciones, un valor para θ puede ser calculado usando los efectos medios del diseño para un subconjunto fijo de variables importantes (véase Lohr y Rao (2006), para un enfoque de esta aproximación y referencias).

En la práctica, $N_a, N_b, V(\hat{N}_{ab}^A)$, y $V(\hat{N}_{ab}^B)$ son estimados a partir de los datos, por lo que un estimador $\hat{\theta}_P$ de θ_P se sustituye en (6). Los pesos ajustados son

$$\tilde{w}_{i,P}^A = \begin{cases} \frac{N_A - \hat{N}_{ab}^{PML}(\hat{\theta}_P)}{\hat{N}_a^A} w_i^A & \text{si } i \in a \\ \frac{\hat{N}_{ab}^{PML}(\hat{\theta}_P)}{\hat{\theta}_P \hat{N}_{ab}^A + (1-\hat{\theta}_P)\hat{N}_{ab}^B} \hat{\theta}_P w_i^A & \text{si } i \in ab. \end{cases}$$

y

$$\tilde{w}_{i,P}^B = \begin{cases} \frac{N_B - \hat{N}_{ab}^{PML}(\hat{\theta}_P)}{\hat{N}_b^B} w_i^B & \text{si } i \in b \\ \frac{\hat{N}_{ab}^{PML}(\hat{\theta}_P)}{\hat{\theta}_P \hat{N}_{ab}^A + (1-\hat{\theta}_P)\hat{N}_{ab}^B} \hat{\theta}_P w_i^B & \text{si } i \in ab. \end{cases}$$

Aunque $\hat{\theta}_P$ depende de las varianzas estimadas del tamaño del dominio de solapamiento, no depende de las covarianzas de las otras variables respuesta. Por tanto, el estimador de PMV usa el mismo conjunto de pesos para cada variable respuesta. Skinner y Rao (1996) y Lohr y Rao (2006) encontraron que el estimador de PMV tiene un error cuadrático medio pequeño y funciona bien con una amplia variedad de diseños de encuestas.

0.2.6. Uso de pesos ajustados para estimar cantidades poblacionales

Cada uno de los métodos de estimación del total de la población que se han planteado en esta sección producen un conjunto de pesos ajustados para las observaciones del marco A y un conjunto de pesos ajustados para las observaciones del marco B. Estos pesos pueden ser usados para estimar los totales poblacionales y otras cantidades. Por ejemplo, para estimar una razón Y/X de dos totales poblacionales Y y X, los pesos ajustados se usan para encontrar $\hat{Y} = \sum_{i \in \mathcal{S}_A} \tilde{w}_{iy}^A y_i + \sum_{i \in \mathcal{S}_B} \tilde{w}_{iy}^B y_i$ y $\hat{X} = \sum_{i \in \mathcal{S}_A} \tilde{w}_{ix}^A x_i + \sum_{i \in \mathcal{S}_B} \tilde{w}_{ix}^B x_i$. Entonces, Y/X se estima por $\hat{Y}/\hat{X}$. Los métodos de Hartley y Fuller-Burmeister, como se dijo anteriormente, permiten que los ajustes para los pesos dependan de la respuesta considerada. Esto puede conducir a anomalías: si X es el número total de ingenieros en la población e Y es el número total de hombres ingenieros, es posible para los pesos de Hartley o de Fuller-Burmeister que se obtenga que $\hat{Y}/\hat{X} > 1$. Los otros estimadores - pormediado, marco único o PMV - no tendrán este problema.

Otras cantidades pueden también ser estimadas usando los pesos ajustados. Para la Encuesta de Finanzas del Consumidor, comentada en la Sección 1, las medianas poblacionales son de interés, tanto como los totales y medias poblacionales. Los pesos ajustados pueden usarse para cualcular medianas, percentiles y otros medidas estadísticas estándar. Para estimar el valor mediano de las acciones para los hogares que poseen algún tipo de acción, podemos usar el valor de m resolviendo

$$\frac{\sum_{i \in \mathcal{S}_A} \tilde{w}_i^A I(y_i \leq m) + \sum_{i \in \mathcal{S}_B} \tilde{w}_i^B I(y_i \leq m)}{\sum_{i \in \mathcal{S}_A} \tilde{w}_i^A + \sum_{i \in \mathcal{S}_B} \tilde{w}_i^B} = \frac{1}{2},$$

donde $I(y_i \leq m) = 1$ si $y_i \leq m$ y 0 en otro caso.

0.2.7. Comparación de los estimadores

Lohr y Rao (2000, 2006) compararon la eficiencia asintótica de los estimadores y demostraron que el estimador de Fuller-Burmeister tiene la mayor eficiencia asintótica de entre todos los estimadores tratados. En la práctica, propiedades como la la facilidad de cálculo, robustez a las hipótesis, y adecuación para la encuesta deberían ser consideradas además de la eficiencia. Aunque el estimador de Fuller-Burmeister tiene la mayor eficiencia asintótica de entre todos los estimadores lineales, tanto el estimador de Fuller-Burmeister como el de Hartley consideran un conjunto de pesos para cada variable respuesta considerada. Además, con más de dos marcos, estos dos estimadores pueden ser inestables ya que dependen de una matriz de covarianzas estimadas de gran dimensión.

En muchas situaciones, un investigador puede querer usar un estimador con pesos fijos para los diferentes dominios, como se describió en la Sección 2.1. En los datos de la SESTAT, por ejemplo, las personas del marco de la Encuesta de Doctorandos representa una pequeña parte del marco de la Encuesta Nacional de Graduados de Bachillerato. La Fundación Nacional de Ciencias usa el estimador

$$\hat{Y} = \hat{Y}_c^C + \hat{Y}_{ac}^A + \hat{Y}_{bc}^B + \hat{Y}_{abc}^A.$$

El total para cada dominio se estima usando exactamente una de las encuestas.

0.3. Estimación de la varianza en encuestas con marcos múltiples

La estimación de la varianza puede ser más complicada para encuestas con marcos múltiples que para encuestas con un único marco. Para el estimador más simple, visto en la Sección 2.1, la estimación de la varianza es sencilla. El estimador del total de la población es

$$\hat{Y}(\theta) = \hat{Y}_a^A + \theta\hat{Y}_{ab}^A + (1-\theta)\hat{Y}_{ab}^B + \hat{Y}_b^B,$$

donde θ es un valor fijo entre 0 y 1. Bajo la hipótesis de que las muestras son tomadas independientemente,

$$V[\hat{Y}(\theta)] = V(\hat{Y}_a^A + \theta\hat{Y}_{ab}^A) + V[(1-\theta)\hat{Y}_{ab}^B + \hat{Y}_b^B].$$

Cada término de la expresión de la varianza puede ser estimado utilizando el diseño de la encuesta, y estas dos partes se suman para obtener un estimador $V[\hat{Y}(\theta)]$. Este mismo método puede ser usado para estimar la varianza del total estimado de la población usando el estimador de marco único.

Las encuestas con marcos múltiples disjuntos fueron tratadas anteriormente como un caso especial de muestreo estratificado. Consecuentemente, se pueden usar métodos estándar para muestras estratificadas para estimar las varianzas.

La estimación de la varianza puede ser más complicada para otros estimadores. Los pesos ajustados para el estimador de Hartley del total poblacional dependen de $\hat{\theta}_{opt}$, el cual es una función de las covarianzas estimadas para ambos marcos. Funciones de los totales, u otros estadísticos como los percentiles, también se basan en un método más complejo basado en estimadores procedentes de ambas muestras.

Pueden usarse varios métodos para estimar las varianzas de las cantidades poblacionales estimadas en encuestas generales de marcos múltiples. Estos métodos incluyen las técnicas de linealización de Taylor, el jackknife, y el bootstrap. Los métodos de linealización de Taylor y el jackknife, tratados en Lohr y Rao (2000), suponen que una caracteristica de interés de la población τ puede expresarse como una función continua doblemente diferenciable de totales poblacionales procedentes de los marcos. Para la linealización de Taylor, las derivadas parciales de esta función se usan conjuntamente con la matriz de covarianzas estimadas de los totales poblacionales estimados del marco A y la matriz de covarianzas estimadas de los totales poblacionales estimados del marco B, para obtener un estimador linealizado de la varianza del estimador $\hat{\tau}$. Por ejemplo, $\tau = Y/X$ podria ser una razón de dos totales poblacionales de una encuestas con dos marcos, con

$$\hat{\tau} = \frac{\hat{Y}(\frac{1}{2})}{\hat{X}(\frac{1}{2})} = \frac{\hat{Y}_a^A + \frac{1}{2}\hat{Y}_{ab}^A + \frac{1}{2}\hat{Y}_{ab}^B + \hat{Y}_b^B}{\hat{X}_a^A + \frac{1}{2}\hat{X}_{ab}^A + \frac{1}{2}\hat{X}_{ab}^B + \hat{X}_b^B}$$

para $\hat{Y}(\frac{1}{2})$ definido en (1). Los totales estimados para el marco A son $(\hat{Y}_a^A, \hat{Y}_{ab}^A, \hat{X}_a^A, \hat{X}_{ab}^A)$ con matriz de covarianza estimada $\mathcal{S}_A$, y los totales estimados para el marco B son $(\hat{Y}_b^B, \hat{Y}_{ab}^B, \hat{X}_b^B, \hat{X}_{ab}^B)$ con matriz de covarianzas estimadas $\mathcal{S}_B$. El estimador linealizado de la varianza es, por tanto,

$$g_A^T S_A g_A + g_B^T S_B g_B,$$

donde, en este caso $g_A = g_B = [\hat{X}(1/2)]^{-1}(1, 1/2, -\hat{\tau}, -\hat{\tau}/2)^T$ es el vector de derivadas empleado para la linealización. Bajo condiciones de regularidad, el estimador linealizado de la varianza es consistente. Sin embargo, requiere, que las derivadas sean calculadas de forma separada para cada estimador que se considere.

El estimador jackknife de la varianza se basa en la propiedad de que las muestras que se toman de los dos marcos son independientes (Lohr y Rao, 2000). Supóngase que se toma una muestra estratificada por conglomerados del marco A y una muestra estratificada por conglomerados independiente de la anterior del marco B. Un estimador jackknife de la varianza lleva a cabo el método jackknife de forma separada en los marcos A y B. Sea $\hat{\tau}_{(hi)}^A$ el estimador de la misma forma que $\hat{\tau}$ cuando las observaciones de la upm i del estrato h de la muestra del marco A no son tenidas en cuenta. De forma similar, sea $\hat{\tau}^B_{(lj)}$ el estimador de la misma forma que $\hat{\tau}$ cuando las observaciones de la upm j del estrato l de la muestra del marco B no son tenidas en cuenta. Luego, si $\tilde{n}^A_h$ es el número de unidades primarias de muestreo en el estrato h de la muestra del marco A y $\tilde{n}^B_l$ es el número de unidades primarias de muestreo en el estrato l de la muestra del marco B, el estimador jackknife de la varianza es

$$v_J(\hat{\tau}) = \sum_{h=1}^{H} \frac{\tilde{n}^A_h - 1}{\tilde{n}^A_h} \sum_{i=1}^{\tilde{n}^A_h} (\tau^A_{(hi)} - \hat{\tau})^2 + \sum_{l=1}^{L} \frac{\tilde{n}^B_l - 1}{\tilde{n}^B_l} \sum_{j=1}^{\tilde{n}^B_l} (\hat{\tau}^B_{(lj)} - \hat{\tau})^2 \tag{10}$$

El estimador jackknife de la varianza es consistente para funciones suaves de las medias poblacionales. Se puede construir un estimador bootstrap de la varianza de forma similar, pero actualmente las propiedades del bootstrap en encuestas con marcos múltiples son objeto de investigación.

0.3.1. Estimación con tres o más marcos

La mayor parte de los estimadores tratados hasta el momento han sido para la situación en la que se tienen dos marcos. Todos ellos pueden extenderse a la situación de Q marcos. El estimador más simple, de nuevo, promedia los estimadores para cada dominio. Para la encuesta con tres marcos representada en la Fig. 4, este estimador es

$$\begin{aligned} \hat{Y}_{ave} =& \hat{Y}^A_a + \hat{Y}^B_b + \hat{Y}^C_c + \frac{1}{2}(\hat{Y}^A_{ab} + \hat{Y}^B_{ab}) + \frac{1}{2}(\hat{Y}^A_{ac} + \hat{Y}^C_{ac}) + \frac{1}{2}(\hat{Y}^B_{bc} + \hat{Y}^C_{bc}) \\ &+ \frac{1}{3}(\hat{Y}^A_{abc} + \hat{Y}^B_{abc} + \hat{Y}^C_{abc}). \end{aligned} \tag{11}$$

Lohr y Rao (2006) desarrollaron y compararon estimadores para encuestas con marcos múltiples cuando $Q \geq 3$. Encontraron que cuando los tamaños muestrales son moderados, los estimadores óptimos de Hartley y Fuller-Burmeister no eran demasiado estables porque requerían la estimación de una matriz de covarianzas muy grande. El estimador de PMV tiene una buena eficiencia y funciona bien bajo un amplio rango de condiciones. El estimador de marco único con estimador de razon ajustado e $\hat{Y}_{ave}$ se comportan bien si los pesos relativos de los estimadores de los dominios están cerca de los valores óptimos.

Una alternativa para trabajar con tres o más marcos son los estimadores de multiplicidad propuestos por Mecatti (2007).

Consideremos $U_1, ...U_Q$ un conjunto de marcos cuya unión cubre la población completa. Los marcos están generalmente solapados. Muestras independientes

$s_1, ...s_Q$ son seleccionadas de los marcos de forma independiente y posiblemente con diferentes diseños.

El parámetro Y puede escribirse como la suma sobre los marcos:

$$Y = \sum_{q=1}^{Q} \sum_{i \in U_q} y_i \alpha_{q(i)}$$

donde $\alpha_{q(i)}$ son los factores de ajuste de multiplicidad correspondientes al marco q para cada unidad i y $\sum_q \alpha_{q(i)} = 1$ nos asegura que la uidad no es contada más de una vez. La elección más simple de $\alpha_{q(i)}$ es la inversa de la multiplicidad de i, m_i, (el número de marcos en la que está incluida).

Mecatti introduce el siguiente estimador:

$$\widehat{Y}_{SM} = \sum_{q=1}^{Q} \sum_{i \in U_q} \frac{y_i 1_{i \in U_q}}{m_i \pi_{i(q)}} \tag{12}$$

con $\pi_{i(q)}$ la probabilidad de inclusión de primer orden de la unidad i en el marco q Este estimador requiere poca información, pero puede ser inestable cuando alguna unidad tiene probabilidad de inclusión pequeña.

Recientemente Singh y Mecatti (2011) proponen una generalización de este estimador que denominan estimadores de Horvitz-Thompson ajustados por la multiplicidad (estimadores GMHT), que están definidos mediante la expresión:

$$\widehat{Y}_{GMHT} = \sum_{q=1}^{Q} \sum_{i \in U_q} y_i \alpha_{q(i)} \frac{\delta_{i(q)}}{E(\delta_{i(q)})} \tag{13}$$

para distintos valores de $\alpha_{q(i)}$ y de $\delta_{i(q)}$.

Cada estimador en esta clase es diseño insesgado por construcción y muy simple analíticamente a pesar del número Q de marcos. El estimador de Horvitz Thompson es un caso especial para $Q = 1$ y $\alpha_{i(q)} = 1$.

El estimador $\widehat{Y}_{GMHT}$ es una combinación lineal de estimadores del tipo Horvitz-Thompson independientes, por lo que su varianza exacta bajo el diseño se obtiene fácilmente como

$$Var(\widehat{Y}_{GMHT}) = \sum_{q=1}^{Q} \sum_{i \in U_q} z_{i(q)}^2 Var(\delta_{i(q)}) + \sum_{i \neq j} \sum_{\in U_q} z_{i(q)} z_{j(q)} Cov(\delta_{i(q)}, \delta_{j(q)}) \quad (14)$$

la cual, para diseños como tamaño muestral fijo se reduce a

$$Var(\widehat{Y}_{GMHT}) = \sum_{q=1}^{Q} \sum_{i<j} \sum_{\in U_q} -Cov(\delta_{i(q)}, \delta_{j(q)})(z_{i(q)} - z_{j(q)})^2 \quad (15)$$

donde

$$z_{i(q)} = \frac{y_i \alpha_{q(i)}}{E(\delta_{i(q)})}$$

recoge todos los componentes no aleatorios de $\widehat{Y}_{GMHT}$. La forma de Sen-Yates-Grundy para los estimadores insesgados de la varianza para diseños con tamaño muestral fijo en todos los marcos, considerando $E(\delta_{i(q)}\delta_{j(q)}) > 0$ para cualquier par de unidades $i \neq j \in U_q, q = 1, ..., Q$, es la dada por

$$v(\widehat{Y}_{GMHT}) = \sum_{q=1}^{Q} \sum_{i<j} \sum_{\in U_q} \frac{-Cov(\delta_{i(q)}, \delta_{j(q)})}{E(\delta_{i(q)}\delta_{j(q)})} (z_{i(q)} - z_{j(q)})^2 \delta_{i(q)} \delta_{j(q)} \quad (16)$$

Nótese que $Cov\delta_{i(q)}, \delta_{j(q)}) = 0$ para muestras de marcos independientes. Para $\delta_{i(q)} = 1_{i \in s_q}$ las ecuaciones (15) y (16) se reducen a la fórmula usual de la varianza pra la suma de estimadores independientes de tipo Horvitz-Thompson, es decir

$$Var(\widehat{Y}_{GMHT}) = \sum_{q=1}^{Q} \sum_{i<j} \sum_{\in U_q} (\pi_{i(q)}\pi_{j(q)} - \pi_{ij(q)}) \left(\frac{y_i \alpha_{q(i)}}{\pi_{i(q)}} - \frac{y_j \alpha_{q(j)}}{\pi_{j(q)}} \right)^2 \quad (17)$$

y

$$v(\widehat{Y}_{GMHT}) = \sum_{q=1}^{Q} \sum_{i<j} \sum_{\in s_q} \frac{(\pi_{i(q)}\pi_{j(q)} - \pi_{ij(q)})}{\pi_{ij(q)}} \left(\frac{y_i \alpha_{q(i)}}{\pi_{i(q)}} - \frac{y_j \alpha_{q(j)}}{\pi_{j(q)}} \right)^2 \quad (18)$$

Por último, en el caso especial en que $\delta_{i(q)} = 1_{i \in s_q}$ y se emplee un muestreo aleatorio simple en cada marco, se tiene que $E(\delta_[i(q)) = n_q/N_q$ (f_q, la fracción de muestreo para el marco q) para todo $i \in U_q$, $E(\delta_{i(q)} \cdot \delta_{j(q)}) = f_q(n_q - 1)/(N_q - 1)$, para todo $i \neq j \in U_q$, y las ecuaciones (14) y (16) se simplifican, de manera que se elimina la doble suma

$$Var(\widehat{Y}_{GMHT}) = \sum_{q=1}^{Q} N_q^2 \frac{1 - f_q}{n_q} \frac{1}{N_q - 1} \left[\sum_{i \in U_q} y_i^2 \alpha_{q(i)}^2 - N_q \left(\frac{1}{N_q} \sum_{i \in U_q} y_i \alpha_{q(i)} \right)^2 \right] \tag{19}$$

y

$$v(\widehat{Y}_{GMHT}) = \sum_{q=1}^{Q} N_q^2 \frac{1 - f_q}{n_q} \frac{1}{n_q - 1} \left[\sum_{i \in s_q} y_i^2 \alpha_{q(i)}^2 - n_q \left(\frac{1}{n_q} \sum_{i \in s_q} y_i \alpha_{q(i)} \right)^2 \right] \tag{20}$$

Todos los estimadores de marco múltiple disponibles en la literatura pueden considerarse como un caso especial de $\widehat{Y}_{GMHT}$ tomando adecuadamente los valores de $\alpha_{q(i)}$ y de $\delta_{i(q)}$. En efecto, y a modo de ejemplo, se puede demostrar que el estimador con ajuste proporcional de la multiplicidad propuesto por Kalton y Anderson en 1986, puede expresarse tal y como sigue:

$$\widehat{Y}_{PM} = \sum_{q=1}^{Q} \sum_{i \in U_q} y_i \alpha_{q(i)} \frac{1_{i \in s_q}}{\pi_{i(q)}} = \sum_{q=1}^{Q} \sum_{i \in U_q} \frac{y_i 1_{i \in s_q}}{m_i \bar{\pi}_i}, \bar{\pi}_i = m_i^{-1} \sum_{q' \in K(i)} \pi_i(q') \tag{21}$$

siendo $\alpha_{q(i)} = \pi_{i(q)} / \sum_{q' \in K(i)} \pi_{i(q')}$ y $\delta_{i(q)} = 1_{i \in s_q}$.

El estimador que Mecatti propuso en 2007, al cual se ha hecho referencia algunos párrafos atrás, también forma parte de la clase de estimadores GMHT, ya que puede expresarse del siguiente modo:

$$\widehat{Y}_{SM} = \sum_{q=1}^{Q} \sum_{i \in U_q} \frac{y_i 1_{i \in s_q}}{m_i \pi_{i(q)}} \tag{22}$$

por lo que $\alpha_{q(i)} = m_i^{-1}$ y $\delta_{i(q)} = 1_{i \in s_q}$

0.4. Nuevas aplicaciones y retos para las encuestas con marcos múltiples

Históricamente, la mayor parte del desarrollo de la teoría de las encuestas con marcos múltiples ha sido motivada por la situación en la cual un marco es un marco de área y el otro marco es un marco de lista. Como las poblaciones se diversifican demográfica y tecnológicamente, un buen número de investigadores han estudiado el uso de encuestas con marcos múltiples para obtener marcos y métodos de recolección de datos alternativos que puedan proporcionar reducciones en los costes.

Internet ofrece muchas posibilidades para el uso de encuestas con marcos múltiples. Internet puede proporcionar un método barato para la recolección de datos, pero un marco de usuarios de internet raramente incluye la totalidad de la población de interés. Blair y Blair (2006) consideraron el uso de encuestas con dos marcos para muestrear poblaciones raras en las que la web se usaba para muestrear a los miembros de la población rara que tenían acceso a internet, y un método general como la encuesta telefónica se usaba para las otras personas. Ellos consideraron utilizar un grupo ”online” de personas que habían accedido a participar en el estudio como el marco B. A menudo, un grupo ”online” tiene centenares de miles de miembros, e información demográfica y de otros tipos puede ser conocida para los miembros del grupo. En algunos caso, por tanto, los miembros del grupo que pertenecen a poblaciones raras puede ser fácilmente identificados y se puede contactar con ellos, proporcionando una muestra relativamente grande con un coste pequeño. En esta aproximación, es importante que se tome una probabilística del marco muestral del grupo; una muestra de conveniencia de personas que por casualidad visitan una página web y hacen una encuesta no puede ser usada para hacer inferencia sobre la población (Couper, 2000).

Desafortunadamente, en la práctica un grupo de este tipo a menudo está formado por voluntarios; los estimadores obtenidos únicamente a partir de la muestra del grupo online, por tanto, no representan cantidades poblacionales porque los miembros del grupo son auto-seleccionados. Una aproximación con dos marcos, combinando información del grupo online (marco B en la Fig. 3) con información de una encuesta sobre la población general (marco A), puede reducir el sesgo al cubrir las partes de la población que no estan en el grupo online. El grupo online a menudo tiene únicamente una pequeña fracción de la población. Escribiendo el total de la población como $Y = Y_a + Y_{ab}$, a menudo se dará el caso de que Y_{ab}, la parte del total procedente de las unidades presentes en ambos marcos, es despreciable. Algunos investigadores han sugerido que la muestra del grupo online puede considerarse como representativa de la población con acceso a internet, pero esta es una hipótesis muy fuerte, y usualmente injustificada. El diseño con dos marcos, sin embargo, permite a los investiadores detectar diferencias entre las respuestas de los encuestados de cada uno de los marcos. En algunos casos, puede ser posible usar estimadores basados en el modelo o post-estratificación para ajustar posibles sesgos en la muestra del panel online, pero se necesita de más investigación en este tema.

Estas nuevas aplicaciones plantean retos para el uso de encuestas con marcos múltiples. Cuando las encuestas de diferentes marcos se toman usando diferentes modelos - por ejemplo, cuando las personas con acceso a internet rellenan un cuestionario online mientras que a las personas de la encuesta telefónica un entrevistador les realiza una serie de preguntas - es posible que las diferencias en los estimadores se deban a efectos del modo de realización de la encuesta más que a la variabilidad del muestreo. En este caso, $\hat{Y}_{ab}^{A}$ y $\hat{Y}_{ab}^{B}$ no están ni siquiera estimando la misma cantidad.

También es importante asegurar que la misma cantidad está siendo estimada en el dominio de solapamiento cuando se usan métodos de marcos múltiples para combinar la información recolectada de diferentes encuestas, como sugirieron Elliot y Davis (2005). Ellos combinaron la información de la Encuesta Nacional de Salud de Estados Unidos (ENS) y del Sistema de Vigilancia del Comportamiento de Factores de Riesgo de Estados Unidos (SVCFR) para estimar la prevalencia del tabaquismo en los condados de Estados Unidos. Ellos asumieron que los estimadores de la ENS eran menos sesgados y ajustaron los pesos para las observaciones en el SVCFR. Esta aproximción puede dar estimadores más precisos de la prevalencia del tabaquismo, pero se debe de tener cuidado y comprobar si la "prevalencia del tabaquismo" medida en una encuesta es la misma que la "prevalencia del tabaquismo" medida en la otra encuesta: orden de las preguntas, efectos de la redacción, o características de los entrevistadores o de los administradores de la encuesta que pueden causar que los supuestos teóricos subyacentes varíen en las dos encuestas.

Los estimadores de los totales poblacionales dados en la Sección 2 suponen que los estimadores de todos los dominios son, aproximadamente, insesgados. Esto no es necesariamente asi cuando hay falta de respuesta o errores de medida.Para la Encuesta de Finanzas del Consumidor de Estados Unidos tratada en la Sección 1 (Bucks et al., 2006), el marco A es un marco de área que contiene la totalidad de la población de Estados Unidos, mientras que el marco B es un marco de lista de los hogares más ricos construídos a partir de la información sobre los impuestos. Las tasas de respuesta difieren ampliamente en las dos muestras. La tasa de respuesta en la muestra del marco A usando el marco de área es aproximadamente del 70 %, mientras que la tasa de respuesta en la muestra de lista está cerca del 30 %. En los hogares con mayores ingresos, es más probable que rechacen participar en la encuesta. Los estimadores con dos marcos para los totales poblacionales necesitan ser ajustados para la falta de respuesta. Esto puede hacer, por ejemplo, realizando ajustes de post-estratificación en los pesos modificados construídos usando uno de los métodos de la Sección 2. Se pueden realizar ajustes adicionales para los pesos de forma separada dentro de cada marco si se conoce información adicional. La estimación por calibración (Deville y Särndal, 1992) puede ser usada para dar un conjunto final de pesos que satisfaga las restricciones de post-estratificación para los marcos separados y para su unión.

Brick et al. (2006) probaron a usar encuestas con dos marcos para reducir el sesgo de la cobertura para encuestas telefónicas. En las condiciones de la Fig. 1, ellos usaron un marco de números de teléfono fijo como marco A y un marco de números de teléfono móvil como marco B. Se dieron cuenta de que en la muestra del marco B había más hogares que sólo disponían de teléfono móvil (dominio b) de los esperados. Tucker et al. (2007) encontraron que el 31 % de los hogares en el dominio de solapamiento ab, aún disponiendo tanto de teléfono fijo como móvil, informaron que raramente recibían llamadas en el teléfono móvil. Aunque estos hogares estan oficialmente en el dominio de solapamiento, es muy probable que no puedan ser localizados en el caso de ser seleccionados en la muestra de hogares con teléfono móvil. Consecuentemente, una parte desproporcionada de la falta de respuesta en la encuesta de hogares con teléfono móvil procede del dominio ab. Las probabilidades de que un hogar no conteste difieren para cada marco y también difieren para los dominios, lo cual dificulta el ajuste de la no respuesta.

Todos los estimadores para las encuestas con marcos múltiples requieren que los miembros de los dominios de las unidades muestreadas sea conocido. En una encuesta con marcos disjuntos, se debe ser capaz de reconocer si una unidad del marco de área está en el marco de lista, por lo que cada unidad poblacional pertenece exactamente a un marco. Esto no es siempre fácil de hacer. In la Encuesta Nacional de Familias de América descrita en la Sección 1.1, los hogares del marco A que tenían teléfono no eran entrevistados. Para evitar la multiplicidad, esta separación debe ser efectiva: si una proporción sustancial de hogares del marco A entrevistados tienen teléfono realmente, este tipo de hogares estará sobrerrepresentado en la encuesta. En encuestas con marcos múltiples con solapamientro, la clasificación errónea de unidades muestreadas en los dominios equivocados puede conllevar a estimadores sesgados de cantidades poblacionales. Si los miembros de un dominio deben ser determinados a través de la encuesta, por ejemplo, cuando a una persona localizada a través de un marco de teléfonos fijos se le pregunta si dispone de teléfono móvil, son necesarias unas pruebas previas para asegurar que la asignación al dominio es efectiva. Meccati (2007) apuntó que estimadores específicos pueden reducir el efecto de ciertos tipos de errores de clasificación errónea. Si $\hat{Y}_{ave}$ en (8) se usa para estimar un total poblacional, el ajuste para el peso en marcos múltiples para una persona muestreada depende únicamente del número de marcos que contengan al individuo; la clasificación errónea de una persona del dominio *ac* en el dominio *bc* no cambia el estimador. En general, sin embargo, se pueden construir escenarios particulares en los que cada uno de los estimadores es sensible a los errores de clasificación.

Las encuestas con marcos múltiples tienen un gran potencial para mejorar la precisión de los estimadores en regiones donde una encuesta tiene un tamaño muestral demasiado pequeño para proporcionar suficiente precisión para los estimadores. Estas regiones se denominan "áreas pequeñas". Rao (2003a, p. 23) sugirió usar diseños con dos marcos para mejores los estimadores para áreas pequeñas para las subpoblaciones de interés. Como ejemplo, Rao trató la Encuesta Holandesa de Demanda de Vivienda, en la cual la encuesta principal de entrevistas personales se complementa con encuestas telefónicas en algunos municipios. Entonces, la estimación con dos marcos puede usarse para encontrar estimadores de cantidades poblacionales dentro de estos municipios. Hay otro uso potencial para las encuestas con dos marcos: complementar una encuesta general cuando se puede querer más precisión para subgrupos de población en el futuro. La ENS de Estados Unidos ha sido diseñada para que en el futuro, los datos de los estados de la ENS puedan ser integrados con datos complementarios procedentes de una encuesta telefónica con marcado aleatorio. (Botman et al., 2000, p.4).

Las encuestas con marcos múltiples están comenzando a usarse mucho más, y en numerosas situaciones, son el mejor método para obtener buena cobertura controlando los costes de la recolección de datos. Deben ser diseñadas y analizadas cuidadosamente, sin embargo, para tener en cuenta la multiplicidad de las unidades poblacionales entre los marcos. El diseño necesita tener en cuenta errores ajenos al muestreo y posibles errores de clasificación, así como los errores propios del muestreo.

Bibliografía

[1] Bankier, M.D. (1986). *Estimators Based on Several Stratified Samples With Applications to Multiple Frame Surveys*. Journal of the American Statistical Association, Vol. 81, 1074 - 1079.

[2] Blair, E. and Blair, J. (2006). *Dual Frame Web-Telephone Sampling for Rare Groups*. Journal of Official Statistics, Vol. 22, No. 2, 211 - 220.

[3] Botman et al. (2000). *Design and Estimation for the National Health Interview Survey, 1995 - 2004*. National Center for Health Statistics, Vital Health Stat, 2(130).

[4] Brick, J.M. et al. (2002). *Two-Phase List-assisted RDD Sampling*. Journal of Official Statistics, Vol. 18, No. 2, 203 - 215.

[5] Brick, J.M. et al. (2006). *Nonresponse Bias in a Dual Frame Sample of Cell and Landline Number*. Public Opinion Quarterly 70(5), 780 - 793.

[6] Bucks et al. (2006). *Recent Changes in U.S. Family Finances: Evidence from the 2001 and 2004 Survey of Consumer Finances*. Federal Reserve Bulletin, A1 - A38

[7] Couper, M.P. (2000). *Web Surveys: A Review of Issues and Approaches*. Public Opinion Quarterly, Vol. 64, 464 - 494.

[8] Deville, J.C. and Sarndal, C.E. (1992). *Calibration estimators in survey Sampling*. Journal of the American Statistical Association, Vol. 87, 376 - 382.

[9] Elliot, M.R. and Davis, W.W. (2005). *Obtaining cancer risk factor prevalence estimates in small areas: combining data from two surveys*. Journal of the Royal Statistical Society, Vol. 54, No. 3, 595 - 609.

[10] Fuller, W.A. and Burmeister, L.F. (1972). *Estimation for Samples Selected From Two Overlapping Frames* ASA Proceedings of the Social Statistics Sections, 245 - 249.

[11] González-Villalobos, A. and Wallace, M. A. (1996). *Multiple Frame Agriculture Surveys, Vols. 1 and 2. Rome: Food and Agriculture Organization of the United Nations.*

[12] Haines, D.E. and Pollock, K.H. (1998). *Estimating the number of active and successful bald eagle nests: An application of the dual frame method.* Journal of Environmental and Ecological Statistics, Vol. 5, 245 - 256.

[13] Hájek, J. (1981). *Sampling from a Finite Population.* New York: Marcel Dekker.

[14] Hansen, M.H. et al. (1953). *Sample Survey Methods and Theory, Vols. I and II.* New Your: John Wiley and Sons.

[15] Hartley, H.O. (1962). *Multiple Frame Surveys.* Proceedings of the American Statistical Association, Social Statistics Sections, 203 - 206.

[16] Hartley, H.O. (1974). *Multiple frame methodology and selected applications.* Sankhya C., Vol. 36, 99 - 118.

[17] Iachan, R. and Dennis, M.L. (1993). *A multiple frame approach to sampling the homeless and transient population.* Journal of Official Statistics, Vol. 9, 747 - 764.

[18] Kalton, G. and Anderson, D.W. (1986) *Sampling Rare Populations.* Journal of the Royal Statistical Society, Ser. A, Vol. 149, 65 - 82.

[19] Lohr, S. and Rao, J.N.K. (2000). *Inference in Dual Frame Surveys.* Journal of the American Statistical Association, Vol. 95, 271 - 280.

[20] Lohr, S. and Rao, J.N.K. (2006). *Estimation in Multiple-frame Surveys.* Journal of the American Statistical Association, Vol. 101, 1019 - 1030.

[21] Mecatti, F. (2007). *A Single Frame Multiplicity Stimator for Multiple Frame Surveys.* Survey Methodology, Vol. 33, 151 - 158.

[22] Rao, J.N.K. and Skinner, C.J. (1996). *Estimation in Dual Frame Surveys with Complex Designs.* Proceedings of the Survey Method Section, Statistical Society of Canada, 63 - 68.

[23] Rao, J.N.K. (2003a). *Small Area Estimation.* Hoboken: New York: John Wiley and Sons Inc., 23.

[24] Singh A.C. and Mecatti, F. (2011). *Generalized Multiplicity-Adjusted Horvitz-Thompson Estimation as a Unified Approach to Multiple Frame Surveys*. Journal of Official Statistics, Vol. 27, No. 4, 633 - 650.

[25] Skinner, C.J. (1991). *On the Efficiency of Raking Ratio Estimation for Multiple Frame Surveys*. Journal of the American Statistical Association, Vol. 86, 779 - 784.

[26] Skinner, C.J. et al. (1994). *Multiple frame sampling for multivariate stratification*. International Statistical Review, Vol. 62, 333 - 347.

[27] Tucker, C. et al. (2007). *Household Telephone service and Usage Patterns in the United States in 2004: Implications for Telephone Samples*. Public Opinion Quarterly, Vol. 71, 3 - 22.

Apéndice: Programas

0.4.1. Muestreo estratificado en Marco A y con probabilidades proporcionales al tamaño en el Marco B. Estimación Horvitz-Thompson de la varianza

```
STR + PPS

################################################################################################
     ############################ SIMULATED POPULATIONS ##################################
################################################################################################
#######################
##### Varianza HT #####
#######################
varHT<-function (Ys, pikls)
{
    if (any(is.na(pikls)))
        stop("there are missing values in pikls")
    if (nrow(pikls) != ncol(pikls))
        stop("pikls is not a square matrix")
    if (any(is.na(Ys)))
        stop("there are missing values in y")
    if (length(Ys) != nrow(pikls))
        stop("y and pikls have different sizes")
    piks = diag(pikls)
    ds<-1/piks
    ns = length(piks)
    ss <- 0
    for (k in 1:ns) {
        ss2 <- 0
        for (l in 1:ns)
                ss2 <- ss2 + (1 - piks[k] * piks[l]/pikls[k, l]) * ds[k] * Ys[k] * ds[l] * Ys[l]
        ss <- ss + ss2
}
    evar = as.numeric(ss)
}


#######################
################################################################################################

#########################################################################################
#PEL
##########################################
#  R function for solving g_1(lambda)=0  #
```

```
########################################
Lag2<-function(u,ds,mu)
{
   n<-length(ds)
   u<-u-rep(1,n)%*%t(mu)
   M<-0*mu
   dif<-1
   tol<-1e-08
   while(dif>tol){
      D1<-0*mu
      DD<-D1%*%t(D1)
      for(i in 1:n){
         aa<-as.numeric(1+t(M)%*%u[i,])
         D1<-D1+ds[i]*u[i,]/aa
         DD<-DD-ds[i]*(u[i,]%*%t(u[i,]))/aa^2
                   }
      D2<-solve(DD,D1,tol=1e-12)
      dif<-max(abs(D2))
      rule<-1
      while(rule>0){
         rule<-0
         if(min(1+t(M-D2)%*%t(u))<=0) rule<-rule+1
         if(rule>0) D2<-D2/2
                   }
      M<-M-D2
   }
   return(M)
}
library(sampling)
########################################
#diferentes tama\~{n}os de muestra, escenarios y estimadores
ntm<-4;nesc<-3;nr<-10; ne<- 4 + 4

for (sce in 1:nesc) {
set.seed(123456789)
#population data 6 strata = Comu x Compa = (Anda,Resto)x(1,2,3) main y-variable=Ingresos~N(2000,500)
datos=rbind(matrix(rep("Anda",1650),1650,1,byrow=TRUE),matrix(rep("Resto",700),700,1,byrow=TRUE))
c(rep(1,1000), rep(2,500), rep(3,150), rep(1,250),rep(2,150),rep(3,300))->c1
datos=cbind.data.frame(matrix(rep(1,2350),2350,1,byrow=TRUE),datos,c1,rnorm(2350, mean = 5000, sd = 500))
names(datos)=c("Uno","Comu","Compa","Ingresos")

# auxiliary variables
equis1<- (datos$Ingresos-rnorm(2350, mean = 500, sd = 300))/0.5
equis2<- (datos$Ingresos-rep(1,2350)-rnorm(2350, mean = 700, sd = 500))/1.2

# variables for Midzuno sampling
tamA<-as.integer(runif(2350,18,90))
tamB<-abs(datos$Ingresos-rnorm(2350, mean = 300, sd = 200))

datos<- data.frame(datos, equis1,equis2,tamA,tamB)
names(datos)=c("Uno","Comu","Compa","Ingresos", "AuxiliarA", "AuxiliarB","tamA","tamB")

if (sce==1) {#OVERLAP DOMAIN SMALL
rbinom(2350,2,0.3)->dom #con 0=a, 1=b y 2=ab
rep(0,2350)->MarcoA;rep(0,2350)->a;rep(0,2350)->MarcoB;rep(0,2350)->b;rep(0,2350)->ab
for (i in 1:2350) {if (dom[i]==0) {dom[i]<-"a" ;MarcoA[i]<-1;MarcoB[i]<-0;a[i]<-1;ab[i]<-0;b[i]<-0}
   if (dom[i]==2) {dom[i]<-"ab" ;MarcoA[i]<-1;MarcoB[i]<-1;a[i]<-0;ab[i]<-1;b[i]<-0}
   if (dom[i]==1) {dom[i]<-"b"  ;MarcoA[i]<-0;MarcoB[i]<-1;a[i]<-0;ab[i]<-0;b[i]<-1}}
    }
if (sce==2) {#OVERLAP DOMAIN LARGE
rbinom(2350,2,0.5)->dom
rep(0,2350)->MarcoA;rep(0,2350)->a;rep(0,2350)->MarcoB;rep(0,2350)->b;rep(0,2350)->ab
for (i in 1:2350) {if (dom[i]==0) {dom[i]<-"a" ;MarcoA[i]<-1;MarcoB[i]<-0;a[i]<-1;ab[i]<-0;b[i]<-0}#
   if (dom[i]==1) {dom[i]<-"ab";MarcoA[i]<-1;MarcoB[i]<-1;a[i]<-0;ab[i]<-1;b[i]<-0}
   if (dom[i]==2) {dom[i]<-"b" ;MarcoA[i]<-0;MarcoB[i]<-1;a[i]<-0;ab[i]<-0;b[i]<-1}}
    }
if (sce==3) {#OVERLAP DOMAIN MEDIUM
rbinom(2350,2,0.5)->dom
rep(0,2350)->MarcoA;rep(0,2350)->a;rep(0,2350)->MarcoB;rep(0,2350)->b;rep(0,2350)->ab
for (i in 1:2350) {if (dom[i]==1) {dom[i]<-"a" ;MarcoA[i]<-1;MarcoB[i]<-0;a[i]<-1;ab[i]<-0;b[i]<-0}
   if (dom[i]==2) {dom[i]<-"ab";MarcoA[i]<-1;MarcoB[i]<-1;a[i]<-0;ab[i]<-1;b[i]<-0}
   if (dom[i]==0) {dom[i]<-"b" ;MarcoA[i]<-0;MarcoB[i]<-1;a[i]<-0;ab[i]<-0;b[i]<-1}}
    }

dato<- data.frame(MarcoA, MarcoB,datos, dom,a,b,ab)

#sample and population size
subset(dato,dato$MarcoA==1)->datoA
subset(dato,dato$MarcoB==1)->datoB
as.vector(table(datoA$Compa,datoA$Comu))->tmEA
as.vector(table(datoB$Compa,datoB$Comu))->tmEB

for (tm in 1:ntm){if (tm==1) {c(15,20,15,20,15,20)->tmeA; c(25,20,25,20,25,20)->tmeB}
  if (tm==2) {c(30,40,30,40,30,40)->tmeA; c(25,20,25,20,25,20)->tmeB}
```

```
                if (tm==3) {c(15,20,15,20,15,20)->tmeA; c(50,40,50,40,50,40)->tmeB}
  if (tm==4) {c(30,40,30,40,30,40)->tmeA; c(50,40,50,40,50,40)->tmeB}

nA<-sum(tmeA);nB<-sum(tmeB)

# we compute the inclusion probabilities and add them to the data
# stratified random sampling
tmeA/tmEA->probas
tmeB/tmEB->probas2
(tmeA-1)/(tmEA-1)->pidosA
(tmeB-1)/(tmEB-1)->pidosB
NMA<-nrow(datoA)
NMB<-nrow(datoB)
ProbA<-rep(0,2350)
EA<-rep(0,2350); EB<-rep(0,2350); qA<-rep(0,2350) ; qB<-rep(0,2350)
for (i in 1:2350) if (dato$Comu[i]=="Anda" & dato$Compa[i]==1) {ProbA[i]<- probas[1]; EA[i]<-1;qA[i]<-pidosA[1]}
for (i in 1:2350) if (dato$Comu[i]=="Anda" & dato$Compa[i]==2) {ProbA[i]<- probas[2]; EA[i]<-2;qA[i]<-pidosA[2]}
for (i in 1:2350) if (dato$Comu[i]=="Anda" & dato$Compa[i]==3) {ProbA[i]<- probas[3]; EA[i]<-3;qA[i]<-pidosA[3]}
for (i in 1:2350) if (dato$Comu[i]=="Resto"& dato$Compa[i]==1) {ProbA[i]<- probas[4]; EA[i]<-4;qA[i]<-pidosA[4]}
for (i in 1:2350) if (dato$Comu[i]=="Resto"& dato$Compa[i]==2) {ProbA[i]<- probas[5]; EA[i]<-5;qA[i]<-pidosA[5]}
for (i in 1:2350) if (dato$Comu[i]=="Resto"& dato$Compa[i]==3) {ProbA[i]<- probas[6]; EA[i]<-6;qA[i]<-pidosA[6]}
ProbB<-rep(0,2350)
for (i in 1:2350) if (dato$Comu[i]=="Anda" & dato$Compa[i]==1) {ProbB[i]<- probas2[1]; EB[i]<-1;qB[i]<-pidosB[1]}
for (i in 1:2350) if (dato$Comu[i]=="Anda" & dato$Compa[i]==2) {ProbB[i]<- probas2[2]; EB[i]<-2;qB[i]<-pidosB[2]}
for (i in 1:2350) if (dato$Comu[i]=="Anda" & dato$Compa[i]==3) {ProbB[i]<- probas2[3]; EB[i]<-3;qB[i]<-pidosB[3]}
for (i in 1:2350) if (dato$Comu[i]=="Resto"& dato$Compa[i]==1) {ProbB[i]<- probas2[4]; EB[i]<-4;qB[i]<-pidosB[4]}
for (i in 1:2350) if (dato$Comu[i]=="Resto"& dato$Compa[i]==2) {ProbB[i]<- probas2[5]; EB[i]<-5;qB[i]<-pidosB[5]}
for (i in 1:2350) if (dato$Comu[i]=="Resto"& dato$Compa[i]==3) {ProbB[i]<- probas2[6]; EB[i]<-6;qB[i]<-pidosB[6]}
data<- data.frame(dato,ProbA,ProbB,EA,EB,qA,qB)

#  Midzuno sampling
subset(data,data$MarcoA==1)->daA
subset(data,data$MarcoB==1)->daB
inclusionprobabilities(daA$tamA,sum(tmeA))->pA
inclusionprobabilities(daB$tamB,sum(tmeB))->pB


# Probabilities and auxiliary variables in frames
for (i in 1:2350) if (data$dom[i]=="a") {data$ProbB[i]<- 0;data$AuxiliarB[i]<-0}
for (i in 1:2350) if (data$dom[i]=="b") {data$ProbA[i]<- 0;data$AuxiliarA[i]<-0}
ppsA<-rep(0,2350); ppsB<-rep(0,2350)
j<-1; for (i in 1:2350) if ((data$MarcoA[i]==1) & (j<=length(pA)) ) {ppsA[i]<-pA[j];j<-j+1}
j<-1; for (i in 1:2350) if ((data$MarcoB[i]==1) & (j<=length(pB)) ) {ppsB[i]<-pB[j];j<-j+1}

# finally, the data and the frames
data<- data.frame(data,ppsA,ppsB)
subset(data,data$MarcoA==1)->datA
subset(data,data$MarcoB==1)->datB

#en el ambiente estaran SpiklA, SpiklB, MpiklA, MpiklB en toda la poblacion
#second order inclusion probabilities STR
 SpiklA=matrix(0,NMA,NMA)
 for(i in 1:NMA)
 {for(j in 1:NMA){
  if ((i!=j) & (datA$EA[i]!=datA$EA[j])) SpiklA[i,j]<-datA$ProbA[i]*datA$ProbA[j]
if ((i!=j) & (datA$EA[i]==datA$EA[j])) SpiklA[i,j]<-datA$qA[i]
SpiklA[i,i]<-datA$ProbA[i]
}
 }
SpiklB=matrix(0,NMB,NMB)
 for(i in 1:NMB)
 {for(j in 1:NMB){
  if ((i!=j) & (datB$EB[i]!=datB$EB[j])) SpiklB[i,j]<-datB$ProbB[i]*datB$ProbB[j]
if ((i!=j) & (datB$EB[i]==datB$EB[j])) SpiklB[i,j]<-datB$qB[i]
SpiklB[i,i]<-datB$ProbB[i]
}
 }

# second order inclusion probabilities Midzuno sampling
MpiklA=matrix(0,NMA,NMA)
MpiklB=matrix(0,NMB,NMB)
UPmidzunopi2(pA)->MpiklA
UPmidzunopi2(pB)->MpiklB

# if we need save the data write.table(data,"s4M.txt")

###############################################################################################

res<- matrix(rep(0,nr),nr,ne+1,byrow=TRUE)
#repeticiones
for (k in 1:nr){

################################# TO DRAW THE SAMPLE ###############################
s1=strata(datA,c("Compa","Comu"),size=tmeA, method="srswor")
```

```
getdata(datA,s1)->sA
SpiklsA=SpiklA[sA[,23],sA[,23]]

#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
UPmidzuno(datB$ppsB,eps=1e-6)->s2
getdata(datB,s2)->sBB
Strato<- rep("NA",nB) #en el otro he tenido que poner rep(NA,nB)
datB$ppsB[s2==1]->pep
sB<-data.frame( sBB$MarcoA, sBB$MarcoB, sBB$Uno, sBB$Ingresos, sBB$AuxiliarA,sBB$AuxiliarB, sBB$tamA,sBB$tamB,sBB$dom,sBB$a,sBB$b,sBB$ab, sBB$ProbA, sBB$ProbB,
names(sB)<-c("MarcoA", "MarcoB", "Uno", "Ingresos", "AuxiliarA","AuxiliarB", "tamA","tamB","dom","a","b","ab", "ProbA", "ProbB", "EA", "EB", "qA", "qB","ppsA",
MpiklsB=MpiklB[s2==1,s2==1]

rep(1,nA)->ma;rep(0,nA)->mb;data.frame(ma,mb,sA)->sA1
rep(0,nB)->ma;rep(1,nB)->mb;data.frame(ma,mb,sB)->sB1
rbind(sA1,sB1)->s

s$ProbB<-s$ppsB

##################################################################

#estimacion tama\~{n}os dominios
val<- s$Uno[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
pikl<-SpiklsA[sA1$a==1,sA1$a==1]
HTestimator(val,pro)->Nhat.a.A  #esto da igual que sum(s$ma*s$a*miA*diA) despues de calcular diA
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A
varHT(val,pikl)->Vhat.Nhat.a.A


val<- s$Uno[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
pikl<-SpiklsA[sA1$ab==1,sA1$ab==1]
HTestimator(val,pro)->Nhat.ab.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A
varHT(val,pikl)->Vhat.Nhat.ab.A

val<- s$Uno[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
pikl<-MpiklsB[sB1$ab==1,sB1$ab==1]
HTestimator(val,pro)->Nhat.ab.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B
varHT(val,pikl)-> Vhat.Nhat.ab.B

val<- s$Uno[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
pikl<-MpiklsB[sB1$b==1,sB1$b==1]
HTestimator(val,pro)->Nhat.b.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B
varHT(val,pikl)-> Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Tama\~{n}os
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

tam<-data.frame(Tama\~{n}os, Dominio,Estimacion,Error.estandar)

#coincide con #sum(s$ma*s$a*(1/s$ProbA));sum(s$ma*s$ab*(1/s$ProbA));sum(s$mb*s$ab*(1/s$ProbB));sum(s$mb*s$b*(1/s$ProbB))

##########################################################################################


#estimacion ingresos
val<- s$Ingresos[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
pikl<-SpiklsA[sA1$a==1,sA1$a==1]
HTestimator(val,pro)->Nhat.a.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A
varHT(val,pikl)->Vhat.Nhat.a.A

val<- s$Ingresos[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
pikl<-SpiklsA[sA1$ab==1,sA1$ab==1]
HTestimator(val,pro)->Nhat.ab.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A
varHT(val,pikl)->Vhat.Nhat.ab.A

val<- s$Ingresos[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
pikl<-MpiklsB[sB1$ab==1,sB1$ab==1]
HTestimator(val,pro)->Nhat.ab.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B
```

```
varHT(val,pikl)->Vhat.Nhat.ab.B

val<- s$Ingresos[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
pikl<-MpiklsB[sB1$b==1,sB1$b==1]
HTestimator(val,pro)->Nhat.b.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B
varHT(val,pikl)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Ingresos
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

dir<-data.frame(Ingresos, Dominio,Estimacion,Error.estandar)

#########################################################################################

#estimador FWA Fixed Weight Ajusted Lohr December 2011 p200 (3) Brick et al. 2006 toman 1/2

theta1_2<-1/2
FWA1_2<- dir[1,3]+theta1_2*dir[2,3]+(1-theta1_2)*dir[3,3]+dir[4,3]

#alternativa modificando los pesos del dise\~{n}o
#aqui se definen los pesos del dise\~{n}o

miA<- s$ma
for (i in 1:length(s$ma)) if (s$ma[i]==1 & s$ab[i]==1) miA[i]<-theta1_2
rep(0,length(s$ma))-> diA
for (i in 1:length(s$ma)) if (s$ProbA[i]!=0) diA[i]<-1/s$ProbA[i]

miB<-s$mb
for (i in 1:length(s$mb)) if (s$mb[i]==1 & s$ab[i]==1) miB[i]<-theta1_2
rep(0,length(s$mb))-> diB
for (i in 1:length(s$mb)) if (s$ProbB[i]!=0) diB[i]<-1/s$ProbB[i]


#########################################################################################

#otra alternativa desde Lohr y Rao pagina 272 expresion (2)

theta_p<-(tam[1,3]*NMB*(tam[3,4]^2))/(tam[1,3]*NMB*(tam[3,4]^2)+tam[4,3]*NMA*(tam[2,4]^2))
#ecuacion segundo grado ca2*x*x - cb2*x + cc2=0
ca2<- theta_p/NMB + (1-theta_p)/NMA
cb2<- 1 + theta_p*(tam[2,3]/NMB) + (1-theta_p)*(tam[3,3]/NMA)
cc2<- theta_p*tam[2,3] + (1-theta_p)*tam[3,3]

raiz2 <- (cb2 - sqrt(cb2*cb2 - 4*ca2*cc2))/(2*ca2)
hatNab<- cc2


uno2<- (NMA - raiz2)/tam[1,3]
dos2<- (NMB - raiz2)/tam[4,3]
tres2<- raiz2/hatNab
hatYab<- theta_p*dir[2,3] + (1-theta_p)*dir[3,3]

PMLA2<-uno2*dir[1,3]+dos2*dir[4,3]+tres2*hatYab

#########################################################################################
#Single frame estimator desde Lohr pag 201 Bankier (1986) Kalton y Anderson (1986)

miAS<-s$ma
for (i in 1:length(s$ma)) if (s$ma[i]==1 & s$ab[i]==1) miAS[i]<-diB[i]/(diA[i]+diB[i])

miBS<-s$mb
for (i in 1:length(s$mb)) if (s$mb[i]==1 & s$ab[i]==1) miBS[i]<-diA[i]/(diA[i]+diB[i])


SF<-sum(s$ma*miAS*diA*s$Ingresos) + sum(s$mb*miBS*diB*s$Ingresos)
#########################################################################################
#Single frame estimator raking ratio desde Lohr y Rao pag 273 Rao and Skinner (1996)

rep(0,length(s$ma))->wi
for (i in 1:length(s$ma)) if (s$a[i]==1)  wi[i]<-diA[i]
for (i in 1:length(s$mb)) if (s$b[i]==1)  wi[i]<-diB[i]
for (i in 1:length(s$ma)) if (s$ab[i]==1) wi[i]<-1/(s$ProbA[i]+s$ProbB[i])

#este debe dar lo mismo que el anterior
YS<-sum(s$ma*wi*s$Ingresos)+sum(s$mb*wi*s$Ingresos)

NabS<- sum(s$ma*s$ab*wi)+sum(s$mb*s$ab*wi)
#esto de abajo sacado del articulo original de Rao and Skinner (1996)
```

```
NaS <- sum(s$ma*s$a*wi)
NbS <- sum(s$mb*s$b*wi)

#ecuacion segundo grado ca1*x*x - cb1*x + cc1=0
ca1<- NabS
cb1<- NabS*(NMA + NMB) + NaS*NbS
cc1<- NabS*NMA*NMB

raiz1 <- (cb1 - sqrt(cb1*cb1 - 4*ca1*cc1))/(2*ca1)

uno1<- (NMA - raiz1)/tam[1,3]
dos1<- (NMB - raiz1)/tam[4,3]
tres1<- raiz1/NabS


YabS<-  sum(s$ma*s$ab*wi*s$Ingresos)+sum(s$mb*s$ab*wi*s$Ingresos)
PMLRR<-uno1*dir[1,3]+dos1*dir[4,3]+tres1*YabS

#estimadores FWA1_2;PMLA2;SF;PMLRR


###########################################
#Estimadores PEL

N<- sum(data$a)+sum(data$b)+sum(data$ab)
die<-wi/sum(wi)

#no auxiliary information
die->pi
PEL<-sum(pi*s$Ingresos)*N

media1<-as.matrix(c(sum(data$MarcoA),sum(data$MarcoB))/N)
Xs1<-as.matrix(data.frame(s$MarcoA,s$MarcoB))
Lag2(Xs1,as.matrix(wi),media1)->landa1
rep(0,nrow(s))->pi1
for (i in 1:nrow(s)) pi1[i]<-die[i]/(1 + t(landa1)%*%(Xs1[i,]-media1))
PEL1<-sum(pi1*s$Ingresos)*N


media3<-as.matrix(c(sum(data$MarcoA),sum(data$MarcoB),sum(data$AuxiliarA),sum(data$AuxiliarB))/N)
Xs3<-as.matrix(data.frame(s$MarcoA,s$MarcoB,s$AuxiliarA,s$AuxiliarB))
condi<- ((sce==2) & (tm==1) & ((k==605) |(k==679) | (k==725))) |
        ((sce==2) & (tm==3) & (k==725)) |
        ((sce==1) & (tm==3) & (k==725))
if ( condi ) {PEL3<-sum(data$Ingresos)} else { Lag2(Xs3,as.matrix(wi),media3)->landa3
rep(0,nrow(s))->pi3
for (i in 1:nrow(s)) pi3[i]<-die[i]/(1 + t(landa3)%*%(Xs3[i,]-media3))
PEL3<-sum(pi3*s$Ingresos)*N }

fi<-((tam[3,4]^2))/((tam[3,4]^2)+(tam[2,4]^2))
NabP<-fi*tam[2,3]+(1-fi)*tam[3,3]
hatN<-NMA+NMB-NabP

media4<-as.matrix(c(sum(data$MarcoA),sum(data$MarcoB))/hatN)
Xs4<-as.matrix(data.frame(s$MarcoA,s$MarcoB))
Lag2(Xs4,as.matrix(wi),media4)->landa4
rep(0,nrow(s))->pi4
for (i in 1:nrow(s)) pi4[i]<-die[i]/(1 + t(landa4)%*%(Xs4[i,]-media4))
PEL4<-sum(pi4*s$Ingresos)*hatN


media6<-as.matrix(c(sum(data$MarcoA),sum(data$MarcoB),sum(data$AuxiliarA),sum(data$AuxiliarB))/hatN)
Xs6<-as.matrix(data.frame(s$MarcoA,s$MarcoB,s$AuxiliarA,s$AuxiliarB))
condi1<-(sce==1) & (tm==1) & (k==393)
if (condi1) {PEL6<-sum(data$Ingresos)} else { Lag2(Xs6,as.matrix(wi),media6)->landa6
rep(0,nrow(s))->pi6
for (i in 1:nrow(s)) pi6[i]<-die[i]/(1 + t(landa6)%*%(Xs6[i,]-media6))
PEL6<-sum(pi6*s$Ingresos)*hatN}


#estimadores PEL; PEL1; PEL2;PEL3;PEL4;PEL5;PEL6
#########################################################################################

#########################################################################################
#########################################################################################
#resumen hasta ahora
#c(FWA1_2,PMLA2,SF,PMLRR)-> e1
#c(CAL.gNAB,CAL.gN0,CAL.gN1,CAL.gN2,CAL.gxAzB, CAL.g1, CAL.g2, CAL.g_2,CAL.g_90)->ecl
#c(CAL.grNAB,CAL.grN0,CAL.grN1,CAL.grN2,CAL.grxAzB, CAL.gr1, CAL.gr2, CAL.gr_2,CAL.gr_90)->ecr
#c(CAL.gtNAB,CAL.gtN0,CAL.gtN1,CAL.gtN2,CAL.gtxAzB, CAL.gt1, CAL.gt2, CAL.gt_2,CAL.gt_90)->ect
#c(PEL, PEL1, PEL2,PEL3,PEL4,PEL5,PEL6)-> epel
#e1;ecl;ecr;ect;epel
#
#NMA;NMB
```

```
#tam
#dir
 PEL1; PEL3;PEL4;PEL6
res[k,]<-c(k,   FWA1_2,  PMLA2,  SF,  PMLRR,   PEL1, PEL3,PEL4,PEL6)


}#fin for
colnames(res)=c("k","FWA1_2","PMLA2","SF","PMLRR","PEL1", "PEL3","PEL4","PEL6")

#
TY<-sum(data$Ingresos)
ecm<-rep(0,ncol(res)-1)
bias<-rep(0,ncol(res)-1)
for (i in 2:ncol(res)) mean((res[,i]-TY)^2)/TY^2->ecm[i-1]
for (i in 2:ncol(res)) mean((res[,i]-TY))/TY->bias[i-1]
ge<-(1-(ecm/ecm[3]))
rbind(bias,ecm,ge)->qq
round(qq*100,3)->qq #ESTAN EN %
colnames(qq)<-colnames(res)[-1]


if (tm==1) t(qq)->ver1; if (tm==2) t(qq)->ver2;if (tm==3) t(qq)->ver3;if (tm==4) t(qq)->ver4

} #fin for tm

li1<-c(NMA,NMB,tmEA,tmEB)
li2<-c(cor(datA$Ingresos,datA$AuxiliarA),cor(datB$Ingresos,datB$AuxiliarB),cor(datA$Ingresos,datA$tamA),cor(datB$Ingresos,datB$tamB))
print (li1, useSource = TRUE);print(li2, useSource = TRUE);print(table(data$dom), useSource = TRUE)
print (cbind(ver1,ver2,ver3,ver4), useSource = TRUE)


} #fin for sce
```

0.4.2. Muestreo con probabilidades proporcionales en ambos Marcos A y B. Estimación Horvitz-Thompson de la varianza

```
modificaciones para calibev PPS + PPS

################################# TO DRAW THE SAMPLE ################################
UPmidzuno(datA$ppsA,eps=1e-6)->s1
getdata(datA,s1)->sAA
Strato<- rep("NA",nA)
datA$ppsA[s1==1]->pep
sA<-data.frame( sAA$MarcoA, sAA$MarcoB, sAA$Uno, sAA$Ingresos, sAA$AuxiliarA,sAA$AuxiliarB, sAA$tamA,sAA$tamB,sAA$dom,sAA$a,sAA$b,sAA$ab, sAA$ProbA, sAA$ProbB,
names(sA)<-c("MarcoA", "MarcoB", "Uno", "Ingresos", "AuxiliarA","AuxiliarB", "tamA","tamB","dom","a","b","ab", "ProbA", "ProbB", "EA", "EB", "qA", "qB","ppsA",
MpiklsA=MpiklA[s1==1,s1==1]


#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
UPmidzuno(datB$ppsB,eps=1e-6)->s2
getdata(datB,s2)->sBB
Strato<- rep("NA",nB)
datB$ppsB[s2==1]->pep
sB<-data.frame( sBB$MarcoA, sBB$MarcoB, sBB$Uno, sBB$Ingresos, sBB$AuxiliarA,sBB$AuxiliarB, sBB$tamA,sBB$tamB,sBB$dom,sBB$a,sBB$b,sBB$ab, sBB$ProbA, sBB$ProbB,
names(sB)<-c("MarcoA", "MarcoB", "Uno", "Ingresos", "AuxiliarA","AuxiliarB", "tamA","tamB","dom","a","b","ab", "ProbA", "ProbB", "EA", "EB", "qA", "qB","ppsA",
MpiklsB=MpiklB[s2==1,s2==1]

rep(1,nA)->ma;rep(0,nA)->mb;data.frame(ma,mb,sA)->sA1
rep(0,nB)->ma;rep(1,nB)->mb;data.frame(ma,mb,sB)->sB1
rbind(sA1,sB1)->s


s$ProbA<-s$ppsA;s$ProbB<-s$ppsB

#################################################################

#estimacion tama\~{n}os dominios
val<- s$Uno[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
pikl<-MpiklsA[sA1$a==1,sA1$a==1]
HTestimator(val,pro)->Nhat.a.A  #esto da igual que sum(s$ma*s$a*miA*diA) despues de calcular diA
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A
varHT(val,pikl)->Vhat.Nhat.a.A
```

```
val<- s$Uno[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
pikl<-MpiklsA[sA1$ab==1,sA1$ab==1]
HTestimator(val,pro)->Nhat.ab.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A
varHT(val,pikl)->Vhat.Nhat.ab.A

val<- s$Uno[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
pikl<-MpiklsB[sB1$ab==1,sB1$ab==1]
HTestimator(val,pro)->Nhat.ab.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B
varHT(val,pikl)-> Vhat.Nhat.ab.B

val<- s$Uno[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
pikl<-MpiklsB[sB1$b==1,sB1$b==1]
HTestimator(val,pro)->Nhat.b.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B
varHT(val,pikl)-> Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Tama\~{n}os
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

tam<-data.frame(Tama\~{n}os, Dominio,Estimacion,Error.estandar)

#coincide con #sum(s$ma*s$a*(1/s$ProbA));sum(s$ma*s$ab*(1/s$ProbA));sum(s$mb*s$ab*(1/s$ProbB));sum(s$mb*s$b*(1/s$ProbB))

##########################################################################################


#estimacion ingresos
val<- s$Ingresos[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
pikl<-MpiklsA[sA1$a==1,sA1$a==1]
HTestimator(val,pro)->Nhat.a.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A
varHT(val,pikl)->Vhat.Nhat.a.A

val<- s$Ingresos[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
pikl<-MpiklsA[sA1$ab==1,sA1$ab==1]
HTestimator(val,pro)->Nhat.ab.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A
varHT(val,pikl)->Vhat.Nhat.ab.A

val<- s$Ingresos[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
pikl<-MpiklsB[sB1$ab==1,sB1$ab==1]
HTestimator(val,pro)->Nhat.ab.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B
varHT(val,pikl)->Vhat.Nhat.ab.B

val<- s$Ingresos[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
pikl<-MpiklsB[sB1$b==1,sB1$b==1]
HTestimator(val,pro)->Nhat.b.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B
varHT(val,pikl)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Ingresos
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

dir<-data.frame(Ingresos, Dominio,Estimacion,Error.estandar)
```

0.4.3. Muestreo con probabilidades proporcionales al tamaño en el Marco A y Muestreo estratificado en Marco B. Estimación Horvitz-Thompson de la varianza

```
modificaciones para calibev PPS + STR


################################# TO DRAW THE SAMPLE ################################

#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
UPmidzuno(datA$ppsA,eps=1e-6)->s1
getdata(datA,s1)->sAA
Strato<- rep(NA,nA)
datA$ppsA[s1==1]->pep
sA<-data.frame( sAA$MarcoA, sAA$MarcoB, sAA$Uno, sAA$Ingresos, sAA$AuxiliarA,sAA$AuxiliarB, sAA$tamA,sAA$tamB,sAA$dom,sAA$a,sAA$b,sAA$ab, sAA$ProbA, sAA$ProbB,
names(sA)<-c("MarcoA", "MarcoB", "Uno", "Ingresos", "AuxiliarA","AuxiliarB", "tamA","tamB","dom","a","b","ab", "ProbA", "ProbB", "EA", "EB", "qA", "qB","ppsA",
MpiklsA=MpiklA[s1==1,s1==1]

s2=strata(datB,c("Compa","Comu"),size=tmeB, method="srswor")
getdata(datB,s2)->sB
SpiklsB=SpiklB[sB[,23],sB[,23]]

rep(1,nA)->ma;rep(0,nA)->mb;data.frame(ma,mb,sA)->sA1
rep(0,nB)->ma;rep(1,nB)->mb;data.frame(ma,mb,sB)->sB1
rbind(sA1,sB1)->s

s$ProbA<-s$ppsA

##################################################################

#estimacion tama\~{n}os dominios
val<- s$Uno[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
pikl<-MpiklsA[sA1$a==1,sA1$a==1]
HTestimator(val,pro)->Nhat.a.A  #esto da igual que sum(s$ma*s$a*miA*diA) despues de calcular diA
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A
varHT(val,pikl)->Vhat.Nhat.a.A


val<- s$Uno[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
pikl<-MpiklsA[sA1$ab==1,sA1$ab==1]
HTestimator(val,pro)->Nhat.ab.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A
varHT(val,pikl)->Vhat.Nhat.ab.A

val<- s$Uno[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
pikl<-SpiklsB[sB1$ab==1,sB1$ab==1]
HTestimator(val,pro)->Nhat.ab.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B
varHT(val,pikl)-> Vhat.Nhat.ab.B

val<- s$Uno[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
pikl<-SpiklsB[sB1$b==1,sB1$b==1]
HTestimator(val,pro)->Nhat.b.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B
varHT(val,pikl)-> Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Tama\~{n}os
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

tam<-data.frame(Tama\~{n}os, Dominio,Estimacion,Error.estandar)

#coincide con #sum(s$ma*s$a*(1/s$ProbA));sum(s$ma*s$ab*(1/s$ProbA));sum(s$mb*s$ab*(1/s$ProbB));sum(s$mb*s$b*(1/s$ProbB))

#########################################################################################


#estimacion ingresos
val<- s$Ingresos[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
pikl<-MpiklsA[sA1$a==1,sA1$a==1]
HTestimator(val,pro)->Nhat.a.A
```

```
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A
varHT(val,pikl)->Vhat.Nhat.a.A

val<- s$Ingresos[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
pikl<-MpiklsA[sA1$ab==1,sA1$ab==1]
HTestimator(val,pro)->Nhat.ab.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A
varHT(val,pikl)->Vhat.Nhat.ab.A

val<- s$Ingresos[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
pikl<-SpiklsB[sB1$ab==1,sB1$ab==1]
HTestimator(val,pro)->Nhat.ab.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B
varHT(val,pikl)->Vhat.Nhat.ab.B

val<- s$Ingresos[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
pikl<-SpiklsB[sB1$b==1,sB1$b==1]
HTestimator(val,pro)->Nhat.b.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B
varHT(val,pikl)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Ingresos
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

dir<-data.frame(Ingresos, Dominio,Estimacion,Error.estandar)
```

0.4.4. Muestreo estratificado en ambos Marcos A y B. Estimación Horvitz-Thompson de la varianza

```
modificaciones para calibev STR + STR
```

```
################################ TO DRAW THE SAMPLE ###############################
s1=strata(datA,c("Compa","Comu"),size=tmeA, method="srswor")
getdata(datA,s1)->sA
SpiklsA=SpiklA[sA[,23],sA[,23]]

s2=strata(datB,c("Compa","Comu"),size=tmeB, method="srswor")
getdata(datB,s2)->sB
SpiklsB=SpiklB[sB[,23],sB[,23]]

rep(1,nA)->ma;rep(0,nA)->mb;data.frame(ma,mb,sA)->sA1
rep(0,nB)->ma;rep(1,nB)->mb;data.frame(ma,mb,sB)->sB1
rbind(sA1,sB1)->s

##################################################################

#estimacion tama\~{n}os dominios
val<- s$Uno[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
pikl<-SpiklsA[sA1$a==1,sA1$a==1]
HTestimator(val,pro)->Nhat.a.A  #esto da igual que sum(s$ma*s$a*miA*diA) despues de calcular diA
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A
varHT(val,pikl)->Vhat.Nhat.a.A


val<- s$Uno[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
pikl<-SpiklsA[sA1$ab==1,sA1$ab==1]
HTestimator(val,pro)->Nhat.ab.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A
varHT(val,pikl)->Vhat.Nhat.ab.A
```

```
val<- s$Uno[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
pikl<-SpiklsB[sB1$ab==1,sB1$ab==1]
HTestimator(val,pro)->Nhat.ab.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B
varHT(val,pikl)-> Vhat.Nhat.ab.B

val<- s$Uno[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
pikl<-SpiklsB[sB1$b==1,sB1$b==1]
HTestimator(val,pro)->Nhat.b.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B
varHT(val,pikl)-> Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Tama\~{n}os
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

tam<-data.frame(Tama\~{n}os, Dominio,Estimacion,Error.estandar)

#coincide con #sum(s$ma*s$a*(1/s$ProbA));sum(s$ma*s$ab*(1/s$ProbA));sum(s$mb*s$ab*(1/s$ProbB));sum(s$mb*s$b*(1/s$ProbB))

##########################################################################################


#estimacion ingresos
val<- s$Ingresos[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
pikl<-SpiklsA[sA1$a==1,sA1$a==1]
HTestimator(val,pro)->Nhat.a.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A
varHT(val,pikl)->Vhat.Nhat.a.A

val<- s$Ingresos[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
pikl<-SpiklsA[sA1$ab==1,sA1$ab==1]
HTestimator(val,pro)->Nhat.ab.A
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A
varHT(val,pikl)->Vhat.Nhat.ab.A

val<- s$Ingresos[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
pikl<-SpiklsB[sB1$ab==1,sB1$ab==1]
HTestimator(val,pro)->Nhat.ab.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B
varHT(val,pikl)->Vhat.Nhat.ab.B

val<- s$Ingresos[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
pikl<-SpiklsB[sB1$b==1,sB1$b==1]
HTestimator(val,pro)->Nhat.b.B
#varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B
varHT(val,pikl)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Ingresos
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

dir<-data.frame(Ingresos, Dominio,Estimacion,Error.estandar)
```

0.4.5. Muestreo con probabilidades proporcionales en ambos Marcos A y B. Estimación de la varianza por el método de Deville

```
modificaciones para varest PPS + PPS

# MpiklA, MpiklB, SpiklA y SpiklB no son necesarias

################################ TO DRAW THE SAMPLE ###############################
```

```
#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
UPmidzuno(datA$ppsA,eps=1e-6)->s1
getdata(datA,s1)->sA
#MpiklsA=MpiklA[s1==1,s1==1]

#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
UPmidzuno(datB$ppsB,eps=1e-6)->s2
getdata(datB,s2)->sB
#piklsB=piklB[s2==1,s2==1]

rep(1,nA)->ma; rep(0,nA)->mb; data.frame(ma,mb,sA)->sA1
rep(0,nB)->ma; rep(1,nB)->mb; data.frame(ma,mb,sB)->sB1
rbind(sA1,sB1)->s

s$ProbA<-s$ppsA; s$ProbB<-s$ppsB

################################################################

#estimacion tama\~{n}os dominios
val<- s$Uno[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
HTestimator(val,pro)->Nhat.a.A  #esto da igual que sum(s$ma*s$a*miA*diA) despues de calcular diA
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A

val<- s$Uno[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A

val<- s$Uno[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B

val<- s$Uno[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
HTestimator(val,pro)->Nhat.b.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Tama\~{n}os
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

tam<-data.frame(Tama\~{n}os, Dominio,Estimacion,Error.estandar)

#coincide con #sum(s$ma*s$a*(1/s$ProbA));sum(s$ma*s$ab*(1/s$ProbA));sum(s$mb*s$ab*(1/s$ProbB));sum(s$mb*s$b*(1/s$ProbB))

#########################################################################################


#estimacion ingresos
val<- s$Ingresos[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
HTestimator(val,pro)->Nhat.a.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A


val<- s$Ingresos[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A

val<- s$Ingresos[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B

val<- s$Ingresos[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
HTestimator(val,pro)->Nhat.b.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Ingresos
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

dir<-data.frame(Ingresos, Dominio,Estimacion,Error.estandar)
```

```
modificaciones para varest PPS + STR

# MpiklA, MpiklB, SpiklA y SpiklB no son necesarias

################################ TO DRAW THE SAMPLE ###############################
#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
UPmidzuno(datA$ppsA,eps=1e-6)->s1
getdata(datA,s1)->sAA

# the results of UPmidzuno and strata are different
Strato<- rep(NA,nA)
datA$ppsA[s1==1]->pep
sA<-data.frame( sAA$MarcoA, sAA$MarcoB, sAA$Uno, sAA$Ingresos, sAA$AuxiliarA,sAA$AuxiliarB, sAA$tamA,sAA$tamB,sAA$dom,sAA$a,sAA$b,sAA$ab, sAA$ProbA, sAA$ProbB,
names(sA)<-c("MarcoA", "MarcoB", "Uno", "Ingresos", "AuxiliarA","AuxiliarB", "tamA","tamB","dom","a","b","ab", "ProbA", "ProbB", "ppsA", "ppsB", "Compa","Comu'

#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
s2=strata(datB,c("Compa","Comu"),size=tmeB, method="srswor")
getdata(datB,s2)->sB

rep(1,nA)->ma; rep(0,nA)->mb; data.frame(ma,mb,sA)->sA1
rep(0,nB)->ma; rep(1,nB)->mb; data.frame(ma,mb,sB)->sB1
rbind(sA1,sB1)->s

# los estimadores estan implementados con s$ProbA y s$ProbB
s$ProbA<-s$ppsA
################################################################

#estimacion tama\~{n}os dominios
val<- s$Uno[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
HTestimator(val,pro)->Nhat.a.A  #esto da igual que sum(s$ma*s$a*miA*diA) despues de calcular diA
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A

val<- s$Uno[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A

val<- s$Uno[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B

val<- s$Uno[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
HTestimator(val,pro)->Nhat.b.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Tama\~{n}os
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

tam<-data.frame(Tama\~{n}os, Dominio,Estimacion,Error.estandar)

#coincide con #sum(s$ma*s$a*(1/s$ProbA));sum(s$ma*s$ab*(1/s$ProbA));sum(s$mb*s$ab*(1/s$ProbB));sum(s$mb*s$b*(1/s$ProbB))

#########################################################################################


#estimacion ingresos
val<- s$Ingresos[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
HTestimator(val,pro)->Nhat.a.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A


val<- s$Ingresos[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A

val<- s$Ingresos[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B

val<- s$Ingresos[s$mb==1 & s$b==1]
```

```
pro<- s$ProbB[s$mb==1 & s$b==1]
HTestimator(val,pro)->Nhat.b.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Ingresos
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

dir<-data.frame(Ingresos, Dominio,Estimacion,Error.estandar)
```

0.4.6. Muestreo estratificado en Marco A y con probabilidades proporcionales al tamaño en el Marco B. Estimación de la varianza por el método de Deville

```
modificaciones para varest STR + PPS

# MpiklA, MpiklB, SpiklA y SpiklB no son necesarias

################################# TO DRAW THE SAMPLE ###############################
#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
s1=strata(datA,c("Compa","Comu"),size=tmeA, method="srswor")
getdata(datA,s1)->sA

#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
UPmidzuno(datB$ppsB,eps=1e-6)->s2
getdata(datB,s2)->sBB
Strato<- rep("NA",nB)
datB$ppsB[s2==1]->pep
sB<-data.frame( sBB$MarcoA, sBB$MarcoB, sBB$Uno, sBB$Ingresos, sBB$AuxiliarA,sBB$AuxiliarB, sBB$tamA,sBB$tamB,sBB$dom,sBB$a,sBB$b,sBB$ab, sBB$ProbA, sBB$ProbB,
names(sB)<-c("MarcoA", "MarcoB", "Uno", "Ingresos", "AuxiliarA","AuxiliarB", "tamA","tamB","dom","a","b","ab", "ProbA", "ProbB", "ppsA", "ppsB", "Compa","Comu"
#piklsB=piklB[s2==1,s2==1]

rep(1,nA)->ma; rep(0,nA)->mb; data.frame(ma,mb,sA)->sA1
rep(0,nB)->ma; rep(1,nB)->mb; data.frame(ma,mb,sB)->sB1
rbind(sA1,sB1)->s

s$ProbB<-s$ppsB
##################################################################

#estimacion tama\~{n}os dominios
val<- s$Uno[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
HTestimator(val,pro)->Nhat.a.A  #esto da igual que sum(s$ma*s$a*miA*diA) despues de calcular diA
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A

val<- s$Uno[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A

val<- s$Uno[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B

val<- s$Uno[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
HTestimator(val,pro)->Nhat.b.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Tama\~{n}os
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

tam<-data.frame(Tama\~{n}os, Dominio,Estimacion,Error.estandar)
```

```
#coincide con #sum(s$ma*s$a*(1/s$ProbA));sum(s$ma*s$ab*(1/s$ProbA));sum(s$mb*s$ab*(1/s$ProbB));sum(s$mb*s$b*(1/s$ProbB))

##########################################################################################


#estimacion ingresos
val<- s$Ingresos[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
HTestimator(val,pro)->Nhat.a.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A


val<- s$Ingresos[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A

val<- s$Ingresos[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B

val<- s$Ingresos[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
HTestimator(val,pro)->Nhat.b.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Ingresos
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

dir<-data.frame(Ingresos, Dominio,Estimacion,Error.estandar)
```

0.4.7. Muestreo estratificado en ambos Marcos A y B. Estimación de la varianza por el método de Deville

```
modificaciones para varest STR + STR

# MpiklA, MpiklB, SpiklA y SpiklB no son necesarias

################################# TO DRAW THE SAMPLE ###############################
s1=strata(datA,c("Compa","Comu"),size=tmeA, method="srswor")
getdata(datA,s1)->sA

#SpiklsA=SpiklA[s1==1,s1==1]

s2=strata(datB,c("Compa","Comu"),size=tmeB, method="srswor")
getdata(datB,s2)->sB

#SpiklsB=SpiklB[s2==1,s2==1]

rep(1,nA)->ma;rep(0,nA)->mb;data.frame(ma,mb,sA)->sA1
rep(0,nB)->ma;rep(1,nB)->mb;data.frame(ma,mb,sB)->sB1
rbind(sA1,sB1)->s

#################################################################

#estimacion tama\~{n}os dominios
val<- s$Uno[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
HTestimator(val,pro)->Nhat.a.A  #esto da igual que sum(s$ma*s$a*miA*diA) despues de calcular diA
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A

val<- s$Uno[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A

val<- s$Uno[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.B
```

```
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B

val<- s$Uno[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
HTestimator(val,pro)->Nhat.b.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Tama\~{n}os
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

tam<-data.frame(Tama\~{n}os, Dominio,Estimacion,Error.estandar)

#coincide con #sum(s$ma*s$a*(1/s$ProbA));sum(s$ma*s$ab*(1/s$ProbA));sum(s$mb*s$ab*(1/s$ProbB));sum(s$mb*s$b*(1/s$ProbB))

#########################################################################################


#estimacion ingresos
val<- s$Ingresos[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
HTestimator(val,pro)->Nhat.a.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A


val<- s$Ingresos[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A

val<- s$Ingresos[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B

val<- s$Ingresos[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
HTestimator(val,pro)->Nhat.b.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Ingresos
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

dir<-data.frame(Ingresos, Dominio,Estimacion,Error.estandar)
```

0.4.8. Muestreo con probabilidades proporcionales al tamaño en el Marcos A y Muestreo estratificado en el Marco B. Estimación de la varianza por el método de Deville

```
# MpiklA, MpiklB, SpiklA y SpiklB no son necesarias

################################ TO DRAW THE SAMPLE ##############################
#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
UPmidzuno(datA$ppsA,eps=1e-6)->s1
getdata(datA,s1)->sAA

# the results of UPmidzuno and strata are different
Strato<- rep(NA,nA)
datA$ppsA[s1==1]->pep
sA<-data.frame( sAA$MarcoA, sAA$MarcoB, sAA$Uno, sAA$Ingresos, sAA$AuxiliarA,sAA$AuxiliarB, sAA$tamA,sAA$tamB,sAA$dom,sAA$a,sAA$b,sAA$ab, sAA$ProbA, sAA$ProbB,
names(sA)<-c("MarcoA", "MarcoB", "Uno", "Ingresos", "AuxiliarA","AuxiliarB", "tamA","tamB","dom","a","b","ab", "ProbA", "ProbB", "ppsA", "ppsB", "Compa","Comu"

#set.seed( as.integer((as.double(Sys.time())*1000+Sys.getpid()) %% 2^31) )
s2=strata(datB,c("Compa","Comu"),size=tmeB, method="srswor")
getdata(datB,s2)->sB
```

```
rep(1,nA)->ma; rep(0,nA)->mb; data.frame(ma,mb,sA)->sA1
rep(0,nB)->ma; rep(1,nB)->mb; data.frame(ma,mb,sB)->sB1
rbind(sA1,sB1)->s

# los estimadores estan implementados con s$ProbA y s$ProbB
s$ProbA<-s$ppsA
################################################################

#estimacion tama\~{n}os dominios
val<- s$Uno[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
HTestimator(val,pro)->Nhat.a.A  #esto da igual que sum(s$ma*s$a*miA*diA) despues de calcular diA
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A

val<- s$Uno[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A

val<- s$Uno[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B

val<- s$Uno[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
HTestimator(val,pro)->Nhat.b.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Tama\~{n}os
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

tam<-data.frame(Tama\~{n}os, Dominio,Estimacion,Error.estandar)

#coincide con #sum(s$ma*s$a*(1/s$ProbA));sum(s$ma*s$ab*(1/s$ProbA));sum(s$mb*s$ab*(1/s$ProbB));sum(s$mb*s$b*(1/s$ProbB))

##########################################################################################


#estimacion ingresos
val<- s$Ingresos[s$ma==1 & s$a==1]
pro<- s$ProbA[s$ma==1 & s$a==1]
HTestimator(val,pro)->Nhat.a.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.a.A


val<- s$Ingresos[s$ma==1 & s$ab==1]
pro<- s$ProbA[s$ma==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.A
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.A

val<- s$Ingresos[s$mb==1 & s$ab==1]
pro<- s$ProbB[s$mb==1 & s$ab==1]
HTestimator(val,pro)->Nhat.ab.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.ab.B

val<- s$Ingresos[s$mb==1 & s$b==1]
pro<- s$ProbB[s$mb==1 & s$b==1]
HTestimator(val,pro)->Nhat.b.B
varest(val,Xs=NULL,pro,w=NULL)->Vhat.Nhat.b.B

#resumen
c(" ", " ", " ", " ")->Ingresos
c(" a "," ab "," ba ", " b ")->Dominio
c(Nhat.a.A,Nhat.ab.A,Nhat.ab.B,Nhat.b.B)->Estimacion
c(Vhat.Nhat.a.A,Vhat.Nhat.ab.A,Vhat.Nhat.ab.B,Vhat.Nhat.b.B)->est.var.est.tam.dom.dir
sqrt(est.var.est.tam.dom.dir)->Error.estandar

dir<-data.frame(Ingresos, Dominio,Estimacion,Error.estandar)
```

Resultados de las simulaciones

Muestreo estratificado en Marco A y con probabilidades proporcionales al tamaño en el Marco B. Estimación Horvitz-Thompson de la varianza

N_{A}=1309 N_{B}=1251 ρ_{XY}^{A}=0.858750892 ρ_{ZY}^{B}=0.709358009 N_{h}^{A}=535 279 78 148 101 168 ρ_{YT}^{B}=0.928502839
N_{a}=1099 N_{ab}=210 N_{b}=1041

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.081	0.029	36.534	-0.029	0.017	22.381	0.005	0.023	52.438	0.038	0.012	41.112
{\sc pml}	0.184	0.030	35.703	0.225	0.027	-25.224	0.106	0.017	64.342	0.155	0.013	38.743
{\sc sf}	-0.084	0.046	0.000	-0.034	0.022	0.000	0.024	0.048	0.000	0.021	0.021	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.083	0.025	46.476	-0.025	0.016	25.627	-0.016	0.015	68.081	0.054	0.010	53.235

N_{A}=1746 N_{B}=1790 ρ_{XY}^{A}=0.859823075 ρ_{ZY}^{B}=0.708688942 N_{h}^{A}=734 377 116 187 115 217 ρ_{YT}^{B}=0.927119945
N_{a}=560 N_{ab}=1186 N_{b}=604

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.121	0.076	38.956	0.149	0.049	28.687	-0.065	0.060	55.651	0.013	0.034	41.196
{\sc pml}	-0.063	0.103	16.450	0.092	0.102	-48.758	-0.007	0.049	63.488	-0.024	0.052	9.555
{\sc sf}	-0.074	0.124	0.000	0.132	0.068	0.000	-0.126	0.135	0.000	0.023	0.057	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.229	0.073	40.964	0.071	0.047	31.122	-0.186	0.055	59.014	-0.044	0.033	42.570

N_{A}=1790 N_{B}=1164 ρ_{XY}^{A}=0.85956256 ρ_{ZY}^{B}=0.70922481 N_{h}^{A}=781 375 114 186 111 223 ρ_{YT}^{B}=0.93095981
N_{a}=1186 N_{ab}=604 N_{b}=560

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge

{\sc fwa(1/2)}	-0.023	0.072	53.967	0.050	0.038	36.976	0.004	0.061	62.473	0.045	0.033	53.892
{\sc pml}	0.139	0.048	69.088	0.148	0.048	20.558	0.110	0.026	83.873	0.003	0.022	69.798
{\sc sf}	-0.108	0.155	0.000	0.039	0.061	0.000	-0.016	0.163	0.000	0.082	0.072	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.023	0.043	72.174	0.054	0.031	48.897	0.007	0.029	82.123	0.001	0.020	72.834

Muestreo con probabilidades proporcionales en ambos Marcos A y B. Estimación Horvitz-Thompson de la varianza

N_{A}=1309 N_{B}=1251 ρ_{XY}^{A}=0.858750892 ρ_{ZY}^{B}=0.709358009 ρ_{YT}^{A}=0.004916936 ρ_{YT}^{B}=0.928502839
N_{a}=1099 N_{ab}=210 N_{b}=1041

	$n_{A}=105$ $n_{B}=135$			$n_{A}=210$ $n_{B}=135$			$n_{A}=105$ $n_{B}=270$			$n_{A}=210$ $n_{B}=270$		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	0.036	0.063	0.882	0.015	0.033	3.615	-0.027	0.059	8.965	0.050	0.028	1.569
{\sc pml}	0.218	0.023	63.710	0.189	0.017	49.007	0.081	0.014	78.380	0.174	0.010	64.396
{\sc sf}	0.035	0.064	0.000	0.017	0.034	0.000	-0.026	0.064	0.000	0.052	0.029	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.020	0.021	66.858	0.017	0.015	55.732	-0.066	0.015	76.967	0.049	0.010	66.487

N_{A}=1746 N_{B}=1790 ρ_{XY}^{A}=0.859823075 ρ_{ZY}^{B}=0.708688942 ρ_{YT}^{A}=0.008601539 ρ_{YT}^{B}=0.927119945
N_{a}=560 N_{ab}=1186 N_{b}=604

	$n_{A}=105$ $n_{B}=135$			$n_{A}=210$ $n_{B}=135$			$n_{A}=105$ $n_{B}=270$			$n_{A}=210$ $n_{B}=270$		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.043	0.104	-7.553	0.089	0.064	6.268	0.007	0.087	12.785	0.025	0.047	-6.273
{\sc pml}	0.169	0.073	24.799	0.067	0.052	23.839	0.112	0.038	61.458	0.040	0.034	23.297
{\sc sf}	-0.017	0.097	0.000	0.093	0.068	0.000	-0.027	0.100	0.000	0.021	0.044	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.128	0.065	32.777	-0.077	0.045	34.073	-0.112	0.045	54.476	-0.074	0.028	35.215

N_{A}=1790 N_{B}=1164 ρ_{XY}^{A}=0.85956256 ρ_{ZY}^{B}=0.70922481 ρ_{YT}^{A}=0.01601011 ρ_{YT}^{B}=0.93095981
N_{a}=1186 N_{ab}=604 N_{b}=560

	$n_{A}=105$ $n_{B}=135$			$n_{A}=210$ $n_{B}=135$			$n_{A}=105$ $n_{B}=270$			$n_{A}=210$ $n_{B}=270$		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	0.080	0.112	12.828	0.007	0.064	-1.268	0.060	0.120	24.464	-0.028	0.050	11.054
{\sc pml}	0.278	0.043	66.687	0.269	0.033	48.917	0.190	0.024	85.182	0.074	0.018	68.064

{\sc sf}	0.061	0.129	0.000	0.019	0.064	0.000	0.058	0.159	0.000	-0.040	0.056	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	0.041	0.040	68.740	0.060	0.030	53.549	0.040	0.028	82.288	-0.027	0.017	69.024

Muestreo con probabilidades proporcionales al tamaño en el Marco A y Muestreo estratificado en Marco B. Estimación Horvitz-Thompson de la varianza

N_{A}=1309 N_{B}=1251 ρ_{XY}^{A}=0.858750892 ρ_{ZY}^{B}=0.709358009 ρ_{YT}^{A}=0.004916936 N_{h}^{B}=557 257 84 127 67 159
N_{a}=1099 N_{ab}=210 N_{b}=1041

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	0.106	0.069	8.450	-0.022	0.041	20.018	-0.086	0.057	9.546	0.008	0.030	8.661
{\sc pml}	0.349	0.041	45.474	0.145	0.025	52.647	0.110	0.033	48.342	0.097	0.020	39.379
{\sc sf}	0.097	0.075	0.000	-0.018	0.052	0.000	-0.074	0.063	0.000	0.005	0.033	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	0.052	0.025	66.333	-0.028	0.019	64.078	-0.137	0.016	74.501	-0.004	0.011	66.279

N_{A}=1746 N_{B}=1790 ρ_{XY}^{A}=0.859823075 ρ_{ZY}^{B}=0.708688942 ρ_{YT}^{A}=0.008601539 N_{h}^{B}=781 375 114 186 111 223
N_{a}=560 N_{ab}=1186 N_{b}=604

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	0.052	0.122	10.582	-0.027	0.074	30.276	-0.057	0.102	9.489	0.052	0.055	10.527
{\sc pml}	0.100	0.180	-31.308	-0.194	0.086	19.274	-0.035	0.177	-56.621	-0.071	0.080	-30.282
{\sc sf}	0.071	0.137	0.000	-0.086	0.107	0.000	-0.072	0.113	0.000	0.085	0.062	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.085	0.077	43.529	-0.185	0.052	51.641	-0.167	0.058	48.323	-0.052	0.035	43.112

N_{A}=1790 N_{B}=1164 ρ_{XY}^{A}=0.85956256 ρ_{ZY}^{B}=0.70922481 ρ_{YT}^{A}=0.01601011 N_{h}^{B}=485 248 70 127 74 160
N_{a}=1186 N_{ab}=604 N_{b}=560

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	0.021	0.116	9.958	-0.032	0.068	10.542	-0.023	0.118	21.245	-0.063	0.054	7.364

{\sc pml}	0.148	0.138	-7.296	0.140	0.074	3.231	0.167	0.139	7.760	0.013	0.073	-25.992
{\sc sf}	0.003	0.128	0.000	-0.040	0.076	0.000	0.019	0.150	0.000	-0.077	0.058	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.048	0.048	62.588	-0.017	0.036	52.700	-0.089	0.031	79.075	-0.068	0.023	60.933

Muestreo estratificado en ambos Marcos A y B. Estimación Horvitz-Thompson de la varianza

N_{A}=1309 N_{B}=1251 ρ_{XY}^{A}=0.858750892 ρ_{ZY}^{B}=0.709358009 N_{h}^{A}=535 279 78 148 101 168 N_{h}^{B}=557 257 84 127 67
N_{a}=1099 N_{ab}=210 N_{b}=1041

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.011	0.033	11.468	-0.066	0.023	7.093	-0.054	0.025	37.461	-0.004	0.016	9.098
{\sc pml}	0.681	0.043	-15.961	0.461	0.029	-18.393	0.494	0.032	20.081	0.354	0.019	-10.503
{\sc sf}	-0.018	0.037	0.000	-0.053	0.024	0.000	-0.016	0.041	0.000	-0.003	0.017	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.015	0.031	15.407	-0.081	0.022	9.037	-0.095	0.020	49.746	-0.008	0.015	12.000

N_{A}=1746 N_{B}=1790 ρ_{XY}^{A}=0.859823075 ρ_{ZY}^{B}=0.708688942 N_{h}^{A}=734 377 116 187 115 217 N_{h}^{B}=781 375 114 186 1
N_{a}=560 N_{ab}=1186 N_{b}=604

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.027	0.088	7.538	0.034	0.062	9.222	-0.129	0.066	39.557	0.040	0.042	6.842
{\sc pml}	0.689	0.104	-9.149	0.560	0.072	-5.756	0.470	0.074	32.755	0.380	0.045	-1.594
{\sc sf}	-0.011	0.095	0.000	0.018	0.069	0.000	-0.137	0.109	0.000	0.049	0.045	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.164	0.089	6.370	-0.047	0.061	10.410	-0.253	0.064	41.577	-0.024	0.042	6.498

N_{A}=1790 N_{B}=1164 ρ_{XY}^{A}=0.85956256 ρ_{ZY}^{B}=0.70922481 N_{h}^{A}=781 375 114 186 111 223 N_{h}^{B}=485 248 70 127 74
N_{a}=1186 N_{ab}=604 N_{b}=560

	$n_{A}=105$	$n_{A}=210$	$n_{A}=105$	$n_{A}=210$
	$n_{B}=135$	$n_{B}=135$	$n_{B}=270$	$n_{B}=270$
n_{h}^{A}	15,20,15,20,15,20	30,40,30,40,30,40	15,20,15,20,15,20	30,40,30,40,30,40
n_{h}^{B}	25,20,25,20,25,20	25,20,25,20,25,20	50,40,50,40,50,40	50,40,50,40,50,40

	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.082	0.078	35.979	0.011	0.044	7.875	-0.079	0.062	55.965	0.010	0.037	35.755
{\sc pml}	0.748	0.092	24.505	0.538	0.052	-9.600	0.624	0.070	49.791	0.432	0.042	26.799
{\sc sf}	-0.120	0.122	0.000	0.012	0.047	0.000	-0.058	0.140	0.000	0.050	0.058	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.115	0.059	52.001	-0.028	0.041	13.378	-0.122	0.035	74.974	-0.043	0.027	52.743

Muestreo estratificado en Marco A y con probabilidades proporcionales al tamaño en el Marco B. Estimación de la varianza por el método de Deville

N_{A}=1309 N_{B}=1251 ρ_{XY}^{A}=0.858750892 ρ_{ZY}^{B}=0.709358009 N_{h}^{A}=535 279 78 148 101 168 ρ_{YT}^{B}=0.928502839
N_{a}=1099 N_{ab}=210 N_{b}=1041

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.081	0.029	36.534	-0.029	0.017	22.381	0.005	0.023	52.438	0.038	0.012	41.112
{\sc pml}	-0.009	0.033	28.468	0.130	0.032	-44.724	0.009	0.018	63.896	0.101	0.013	35.435
{\sc sf}	-0.084	0.046	0.000	-0.034	0.022	0.000	0.024	0.048	0.000	0.021	0.021	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.083	0.025	46.476	-0.025	0.016	25.627	-0.016	0.015	68.081	0.054	0.010	53.235

N_{A}=1746 N_{B}=1790 ρ_{XY}^{A}=0.859823075 ρ_{ZY}^{B}=0.708688942 N_{h}^{A}=734 377 116 187 115 217 ρ_{YT}^{B}=0.927119945
N_{a}=560 N_{ab}=1186 N_{b}=604

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.121	0.076	38.956	0.149	0.049	28.687	-0.065	0.060	55.651	0.013	0.034	41.196
{\sc pml}	-0.062	0.103	16.907	0.117	0.098	-43.140	-0.009	0.049	63.541	-0.017	0.051	11.069
{\sc sf}	-0.074	0.124	0.000	0.132	0.068	0.000	-0.126	0.135	0.000	0.023	0.057	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.229	0.073	40.964	0.071	0.047	31.122	-0.186	0.055	59.014	-0.044	0.033	42.570

N_{A}=1790 N_{B}=1164 ρ_{XY}^{A}=0.85956256 ρ_{ZY}^{B}=0.70922481 N_{h}^{A}=781 375 114 186 111 223 ρ_{YT}^{B}=0.93095981
N_{a}=1186 N_{ab}=604 N_{b}=560

	$n_{A}=105$	$n_{A}=210$	$n_{A}=105$	$n_{A}=210$

	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.023	0.072	53.967	0.050	0.038	36.976	0.004	0.061	62.473	0.045	0.033	53.892
{\sc pml}	0.094	0.049	68.417	0.129	0.049	19.302	0.089	0.026	83.781	-0.006	0.022	69.653
{\sc sf}	-0.108	0.155	0.000	0.039	0.061	0.000	-0.016	0.163	0.000	0.082	0.072	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.023	0.043	72.174	0.054	0.031	48.897	0.007	0.029	82.123	0.001	0.020	72.834

Muestreo estratificado en ambos Marcos A y B. Estimación de la varianza por el método de Deville

N_{A}=1309 N_{B}=1251 ρ_{XY}^{A}=0.858750892 ρ_{ZY}^{B}=0.709358009 N_{h}^{A}=535 279 78 148 101 168 N_{h}^{B}=557 257 84 127 67
N_{a}=1099 N_{ab}=210 N_{b}=1041

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.011	0.033	11.468	-0.066	0.023	7.093	-0.054	0.025	37.461	-0.004	0.016	9.098
{\sc pml}	0.462	0.039	-6.088	0.230	0.025	-3.058	0.221	0.027	33.207	0.184	0.017	1.334
{\sc sf}	-0.018	0.037	0.000	-0.053	0.024	0.000	-0.016	0.041	0.000	-0.003	0.017	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.015	0.031	15.407	-0.081	0.022	9.037	-0.095	0.020	49.746	-0.008	0.015	12.000

N_{A}=1746 N_{B}=1790 ρ_{XY}^{A}=0.859823075 ρ_{ZY}^{B}=0.708688942 N_{h}^{A}=734 377 116 187 115 217 N_{h}^{B}=781 375 114 186 1
N_{a}=560 N_{ab}=1186 N_{b}=604

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.027	0.088	7.538	0.034	0.062	9.222	-0.129	0.066	39.557	0.040	0.042	6.842
{\sc pml}	0.489	0.096	-0.516	0.414	0.066	3.368	0.187	0.056	48.561	0.254	0.044	1.275
{\sc sf}	-0.011	0.095	0.000	0.018	0.069	0.000	-0.137	0.109	0.000	0.049	0.045	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.164	0.089	6.370	-0.047	0.061	10.410	-0.253	0.064	41.577	-0.024	0.042	6.498

N_{A}=1790 N_{B}=1164 ρ_{XY}^{A}=0.85956256 ρ_{ZY}^{B}=0.70922481 N_{h}^{A}=781 375 114 186 111 223 N_{h}^{B}=485 248 70 127 74
N_{a}=1186 N_{ab}=604 N_{b}=560

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{A}	15,20,15,20,15,20			30,40,30,40,30,40			15,20,15,20,15,20			30,40,30,40,30,40		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.082	0.078	35.979	0.011	0.044	7.875	-0.079	0.062	55.965	0.010	0.037	35.755
{\sc pml}	0.356	0.066	45.750	0.311	0.045	5.400	0.087	0.033	76.752	0.113	0.027	53.866
{\sc sf}	-0.120	0.122	0.000	0.012	0.047	0.000	-0.058	0.140	0.000	0.050	0.058	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.115	0.059	52.001	-0.028	0.041	13.378	-0.122	0.035	74.974	-0.043	0.027	52.743

Muestreo con probabilidades proporcionales al tamaño en el Marcos A y Muestreo estratificado en el Marco B. Estimación de la varianza por el método de Deville

N_{A}=1309 N_{B}=1251 ρ_{XY}^{A}=0.858750892 ρ_{ZY}^{B}=0.709358009 ρ_{YT}^{A}=0.004916936 N_{h}^{B}=557 257 84 127 67 159
N_{a}=1099 N_{ab}=210 N_{b}=1041

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	0.106	0.069	8.450	-0.022	0.041	20.018	-0.086	0.057	9.546	0.008	0.030	8.661
{\sc pml}	0.372	0.036	51.635	0.153	0.023	56.194	0.167	0.022	65.568	0.156	0.015	55.755
{\sc sf}	0.097	0.075	0.000	-0.018	0.052	0.000	-0.074	0.063	0.000	0.005	0.033	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	0.052	0.025	66.333	-0.028	0.019	64.078	-0.137	0.016	74.501	-0.004	0.011	66.279

N_{A}=1746 N_{B}=1790 ρ_{XY}^{A}=0.859823075 ρ_{ZY}^{B}=0.708688942 ρ_{YT}^{A}=0.008601539 N_{h}^{B}=781 375 114 186 111 223
N_{a}=560 N_{ab}=1186 N_{b}=604

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	0.052	0.122	10.582	-0.027	0.074	30.276	-0.057	0.102	9.489	0.052	0.055	10.527
{\sc pml}	0.543	0.107	21.588	0.046	0.062	41.525	0.324	0.066	41.703	0.171	0.047	24.344

{\sc sf}	0.071	0.137	0.000	-0.086	0.107	0.000	-0.072	0.113	0.000	0.085	0.062	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.085	0.077	43.529	-0.185	0.052	51.641	-0.167	0.058	48.323	-0.052	0.035	43.112

N_{A}=1790 N_{B}=1164 ρ_{XY}^{A}=0.85956256 ρ_{ZY}^{B}=0.70922481 ρ_{YT}^{A}=0.01601011 N_{h}^{B}=485 248 70 127 74 160
N_{a}=1186 N_{ab}=604 N_{b}=560

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
n_{h}^{B}	25,20,25,20,25,20			25,20,25,20,25,20			50,40,50,40,50,40			50,40,50,40,50,40		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	0.021	0.116	9.958	-0.032	0.068	10.542	-0.023	0.118	21.245	-0.063	0.054	7.364
{\sc pml}	0.477	0.071	44.720	0.301	0.049	35.946	0.321	0.040	73.525	0.170	0.029	50.589
{\sc sf}	0.003	0.128	0.000	-0.040	0.076	0.000	0.019	0.150	0.000	-0.077	0.058	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.048	0.048	62.588	-0.017	0.036	52.700	-0.089	0.031	79.075	-0.068	0.023	60.933

Muestreo con probabilidades proporcionales en ambos marcos A y B. Estimación de la varianza por el método de Deville

N_{A}=1309 N_{B}=1251 ρ_{XY}^{A}=0.858750892 ρ_{ZY}^{B}=0.709358009 ρ_{YT}^{A}=0.004916936 ρ_{YT}^{B}=0.928502839
N_{a}=1099 N_{ab}=210 N_{b}=1041

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	0.036	0.063	0.882	0.015	0.033	3.615	-0.027	0.059	8.965	0.050	0.028	1.569
{\sc pml}	0.064	0.029	54.764	0.170	0.028	18.157	-0.008	0.015	76.590	0.127	0.013	54.807
{\sc sf}	0.035	0.064	0.000	0.017	0.034	0.000	-0.026	0.064	0.000	0.052	0.029	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.020	0.021	66.858	0.017	0.015	55.732	-0.066	0.015	76.967	0.049	0.010	66.487

N_{A}=1746 N_{B}=1790 ρ_{XY}^{A}=0.859823075 ρ_{ZY}^{B}=0.708688942 ρ_{YT}^{A}=0.008601539 ρ_{YT}^{B}=0.927119945
N_{a}=560 N_{ab}=1186 N_{b}=604

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	-0.043	0.104	-7.553	0.089	0.064	6.268	0.007	0.087	12.785	0.025	0.047	-6.273
{\sc pml}	0.017	0.095	2.430	0.153	0.088	-29.208	0.031	0.045	54.598	0.007	0.048	-8.647

{\sc sf}	-0.017	0.097	0.000	0.093	0.068	0.000	-0.027	0.100	0.000	0.021	0.044	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	-0.128	0.065	32.777	-0.077	0.045	34.073	-0.112	0.045	54.476	-0.074	0.028	35.215

N_{A}=1790 N_{B}=1164 ρ_{XY}^{A}=0.85956256 ρ_{ZY}^{B}=0.70922481 ρ_{YT}^{A}=0.01601011 ρ_{YT}^{B}=0.93095981
N_{a}=1186 N_{ab}=604 N_{b}=560

	$n_{A}=105$			$n_{A}=210$			$n_{A}=105$			$n_{A}=210$		
	$n_{B}=135$			$n_{B}=135$			$n_{B}=270$			$n_{B}=270$		
	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge	bias	ecm	ge
{\sc fwa(1/2)}	0.080	0.112	12.828	0.007	0.064	-1.268	0.060	0.120	24.464	-0.028	0.050	11.054
{\sc pml}	0.171	0.047	63.158	0.196	0.043	31.707	0.123	0.025	84.303	0.021	0.020	63.696
{\sc sf}	0.061	0.129	0.000	0.019	0.064	0.000	0.058	0.159	0.000	-0.040	0.056	0.000
$\hbox{{\sc sf}}_{\hbox{\tiny rake}}$	0.041	0.040	68.740	0.060	0.030	53.549	0.040	0.028	82.288	-0.027	0.017	69.024