
Cálculo y estimación de la varianza en la aplicación del estimador de rastrillo (*Raking*)

Variance Computation and Estimation for the Application of the Raking Estimator

Luis Alfonso Orjuela^a
luis.orjuela@millwardbrown.com

Resumen

La ponderación por calibración es una metodología en la que los pesos muestrales son ajustados de tal forma que cuando se aplican a un determinado conjunto de datos auxiliares, estos reproducen los totales del universo. El presente trabajo muestra el comportamiento que tiene uno de estos estimadores para totales poblacionales: el estimador de rastrillo frente a un estimador clásico como lo es el $\hat{t}_{y\pi}$ en un diseño muestral como el MAS^2 .

Palabras clave: calibración, información auxiliar multivariante, rastrillo, estimador de regresión.

Abstract

Calibration weighting is a methodology under which probability-sample weights are adjusted in such a way that when applied to data they can produce population totals. The present work shows the behavior of one of the estimators for total population: the raking estimator compared to a classical estimator as is the $\hat{t}_{y\pi}$ in a sample design as MAS^2 .

Key words: calibration, multivariate auxiliary information, raking, regression estimators.

1. Introducción

Se pretende estimar los totales para un grupo de variables de estudio a partir de una muestra probabilística m extraída del universo U . Con base en la probabilidad de inclusión, π_k , para cada elemento k de la muestra m , se estima el total

^aMarketing Sciences Assistant. Millward Brown.

poblacional de la variable y , $t_y = \sum_U y_k$, con el estimador de Horvitz-Thompson, $\hat{t}_{y\pi} = \sum_m \frac{y_k}{\pi_k} = \sum_U y_k \frac{I_k}{\pi_k}$ con $I_k = 1$ cuando $k \in m$ y 0 en otro caso. Considerando a I_k como variable aleatoria, se puede demostrar que $\hat{t}_{y\pi}$ es estimador insesgado para t_y . Otra forma de escribirlo es $\hat{t}_{y\pi} = \sum_U d_k y_k = \sum_m d_k y_k$ se define $d_k = \frac{I_k}{\pi_k}$ como el peso o el factor de expansión del elemento k en la muestra.

Deville & Särndal (1992) utilizan el término “estimación de calibración” para describir el estimador $\hat{t}_{yw} = \sum_m w_k y_k$, con $\sum_m w_k \mathbf{x}_k = \sum_U \mathbf{x}_k = \mathbf{t}_x$ para algún vector de variables auxiliares $\mathbf{x}_k = (x_{1k}, x_{2k}, \dots, x_{Pk})^t$ con $\mathbf{t}_x = (t_{x1}, t_{x2}, \dots, t_{xP})^t$ donde t_{xi} es conocida para todo i con $i = 1, \dots, P$. Como consecuencia hay una gran cantidad de conjuntos $\{w_k | k \in m\}$ que satisface la ecuación de calibración:

$$\sum_{k \in m} w_k \mathbf{x}_k = \mathbf{t}_x, \quad (1)$$

Entonces se requiere que la diferencia entre $\{w_k | k \in m\}$ y $\{d_k | k \in m\}$ minimice alguna función de pérdida. Los componentes univariados de la ecuación (1):

$$\sum_{k \in m} w_k x_{pk} = \sum_{k \in U} x_{pk} \quad \text{para } p = 1, \dots, P. \quad (2)$$

son llamadas frecuentemente “ecuaciones de calibración”.

Cuando y_k es una combinación lineal de \mathbf{x}_k , para todo $k \in U$, entonces \hat{t}_{yw} es exactamente igual a t_y .

Esta propiedad es una gran fortaleza de la ponderación por calibración y la razón es porque el estimador de calibración es frecuentemente mucho más eficiente que el estimador de Horvitz-Thompson.

Ya que \hat{t}_{yw} estima perfectamente a t_y cuando $y_k = \mathbf{x}_k^t \beta$, es de esperarse que \hat{t}_{yw} sea un buen estimador cuando $y_k = \mathbf{x}_k^t \beta$ estén altamente correlacionados. Esto puede ser formalizado si se supone que los y_k son variables aleatorias que satisfacen el modelo de predicción lineal.

$$y_k = \mathbf{x}_k^t \beta + \epsilon_k, \quad (3)$$

donde $E(\epsilon_k | \{\mathbf{x}_g | g \in m\}) = 0$ para todo $k \in U$. Bajo este modelo, \hat{t}_{yw} es un estimador asintóticamente insesgado para t_y . La sección 2 muestra algunas definiciones y una forma de estudiar la sensibilidad del método de rastrillo para una correcta simulación de Monte Carlo. Posteriormente se realiza la simulación para diferentes poblaciones creadas artificialmente y por último, en la sección 3 se hacen algunos análisis y conclusiones.

2. Argumentos

2.1. Método de Monte Carlo

Bajo este nombre, se agrupa una serie de procedimientos que analizan distribuciones de variables aleatorias usando simulación de números aleatorios. Éste es un método no determinístico usado para aproximar expresiones matemáticas complejas y costosas de evaluar con exactitud. Debe su nombre al principado de Monte Carlo por ser la capital del juego de azar, al ser la ruleta un generador simple de números aleatorios. El desarrollo sistemático data aproximadamente de 1944 y desde entonces ha sido utilizado como herramienta de investigación.

El método de Monte Carlo proporciona soluciones aproximadas a una gran variedad de problemas matemáticos posibilitando la realización de experimentos con muestreos de números pseudoaleatorios en una computadora. El método es aplicable a cualquier tipo de problema, ya sea estocástico o determinístico. A diferencia de los métodos numéricos que se basan en evaluaciones en N puntos en un espacio M -dimensional para producir una solución aproximada, el método de Monte Carlo tiene un error absoluto de la estimación que decrece como $\frac{1}{\sqrt{N}}$ en virtud del teorema del límite central.

2.1.1. Algoritmo

El algoritmo de simulación Monte Carlo puro, está fundamentado en la generación de números pseudoaleatorios por el método de transformación inversa, el cual se basa en las distribuciones acumuladas de frecuencias:

- Determinar las V.A. (Variable aleatoria) y sus distribuciones acumuladas (F)
- Generar un número aleatorio
- Determinar el valor de la V.A. (Variable aleatoria) para el número aleatorio generado de acuerdo a las clases que tengamos (iterar tantas veces como sea necesario)
- Calcular media, desviación estándar error
- Analizar resultados para distintos tamaños de muestra.

2.2. Desarrollo de la simulación

2.2.1. Estudio de sensibilidad

Para obtener resultados confiables y acertados, se optó por aplicar el método descrito anteriormente (simulación por Monte Carlo).

Para determinar el número de muestras (Q) que se seleccionarán, se probó con diferentes tamaños, 500, 1000, 1500 y 2000 muestras que se replican 15 veces cada una, mostrando el comportamiento de las estimaciones realizadas mediante un diseño de muestreo (MAS^2)¹ y el ajuste por calibración mediante el método de rastrillo, el cual ajusta las celdas de una tabla de dos o más vías mediante los totales conocidos de las celdas marginales. La población que se utiliza tiene una distribución uniforme con respecto a la ocurrencia de las dos variables auxiliares, las cuales tienen tres categorías cada una que se pueden observar con las siguientes tablas de frecuencias, contando con un total de 800.000 hasta 1.200.000 elementos, además cuenta con 2600 conglomerados primarios de muestreo, los cuales no tienen menos de 20 elementos cada uno. Las muestras seleccionadas tienen 20 conglomerados de la primera etapa y un 12% de los registros de cada conglomerado.

Cantidad porcentaje	Variable auxiliar C2				
	1	2	3	Total	
Variable auxiliar C1	1	311096 36.8 %	118706 14.05 %	107511 12.72 %	537313 63.57 %
	2	59360 7.1 %	22488 2.7 %	20665 2.4 %	102513 12.1 %
	3	119305 14.1 %	45010 5.3 %	41036 4.8 %	205351 24.3 %
Total	489761 57.9 %	186204 22.0 %	169212 20.0 %	845177 100 %	

Tabla 1: *Cantidad de elementos por celda*

Suma (en millones)	Variable auxiliar C2			
	1	2	3	
Media				
Varianza				
Variable auxiliar C1	1	12.24 39.3 905.2	4.65 39.2 900.3	4.23 39.4 899.4
	2	2.35 36.6 917	0.89 39.5 910.8	0.81 39.2 907.6
	3	4.7 39.4 909.6	1.76 39.2 908.8	1.6 39.1 916.4

Tabla 2: *Características de y_k por celda*

Al seleccionar las Q muestras y aplicarles estos dos estimadores, se procede a calcular la media y S^2 a todos los Q estimadores resultantes, con lo que se obtiene un estimador de la esperanza del estimador y un estimador de la varianza del estimador respectivamente. A continuación mostramos el comportamiento de estos estimadores con los diferentes tamaños de muestras.

¹Diseño muestral aleatorio simple en dos etapas de muestra.

- Simulación con $Q = 1000$

En la Tabla 3 y Figura 1, se muestran el sesgo relativo y el C.V estimados con base en el promedio y la varianza de mil estimaciones con cada uno de quince experimentos.

<i>Experimento</i>	Sesgo relativo		C.V	
	<i>MAS</i>	<i>IPF</i>	<i>MAS</i>	<i>IPF</i>
1	-2,7 %	2,7 %	31,0 %	2,8 %
2	0,3 %	3,5 %	32,4 %	2,8 %
3	-1,5 %	0,0 %	30,2 %	2,9 %
4	-0,5 %	-5,0 %	31,5 %	2,9 %
5	-1,2 %	-3,1 %	32,7 %	2,8 %
6	1,4 %	-4,4 %	31,5 %	2,9 %
7	-1,5 %	-1,6 %	30,6 %	2,9 %
8	3,1 %	-2,2 %	32,2 %	2,8 %
9	-1,2 %	-3,4 %	31,0 %	2,9 %
10	1,7 %	3,0 %	31,2 %	2,8 %
11	4,9 %	2,3 %	33,0 %	2,7 %
12	-1,0 %	1,9 %	30,7 %	2,7 %
13	-0,6 %	2,1 %	31,6 %	2,9 %
14	-2,2 %	2,7 %	32,6 %	2,7 %
15	0,5 %	-0,1 %	30,6 %	2,9 %

Tabla 3: *Sesgo relativo y coeficiente de variación para 1000 iteraciones*

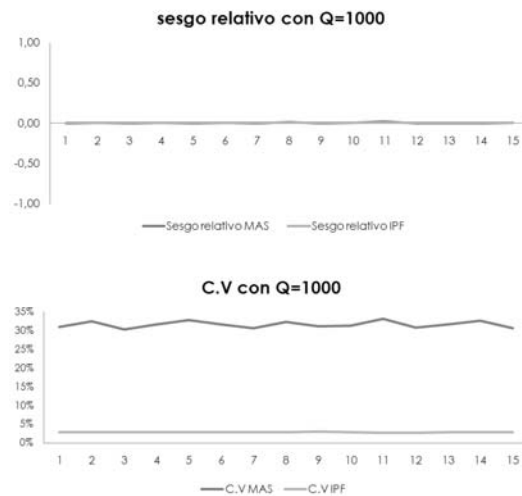


Figura 1: *Sesgo relativo y coeficiente de variación al replicar 1000 veces los dos estimadores*

- para $Q = 1500$ el resultado es:

<i>Experimento</i>	Sesgo relativo		C.V	
	<i>MAS</i>	<i>IPF</i>	<i>MAS</i>	<i>IPF</i>
1	-2,5 %	5,3 %	31,0 %	2,8 %
2	-0,5 %	0,0 %	31,7 %	2,9 %
3	4,6 %	-2,0 %	32,5 %	2,8 %
4	0,9 %	5,3 %	31,3 %	2,9 %
5	2,4 %	0,3 %	31,6 %	2,8 %
6	-0,7 %	-2,6 %	32,7 %	2,8 %
7	-3,8 %	4,0 %	31,1 %	2,8 %
8	2,4 %	-1,2 %	31,6 %	2,9 %
9	3,5 %	6,6 %	31,1 %	2,9 %
10	1,4 %	-0,2 %	31,9 %	2,8 %
11	-0,3 %	2,1 %	31,9 %	2,9 %
12	-2,8 %	4,0 %	31,9 %	2,8 %
13	-0,3 %	-0,5 %	31,1 %	2,8 %
14	5,2 %	2,6 %	31,6 %	2,8 %
15	-1,5 %	-0,8 %	32,6 %	2,8 %

Tabla 4: *Sesgo relativo y coeficiente de variación para 1500 iteraciones*

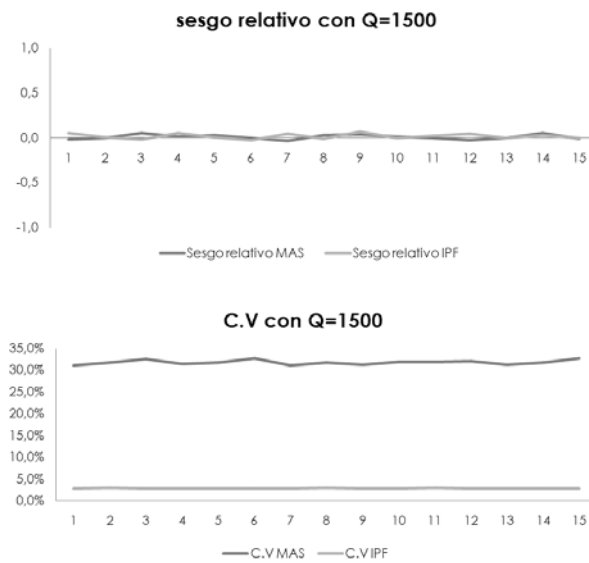


Figura 2: *Sesgo relativo y coeficiente de variación al replicar 1500 veces los dos estimadores*

- para $Q = 2000$ es:

<i>Experimento</i>	Sesgo relativo		C.V	
	<i>MAS</i>	<i>IPF</i>	<i>MAS</i>	<i>IPF</i>
1	-1,0 %	-0,1 %	30,8 %	2,8 %
2	-0,3 %	0,0 %	31,5 %	2,8 %
3	2,3 %	-0,1 %	32,9 %	2,9 %
4	0,5 %	0,1 %	30,5 %	2,9 %
5	0,5 %	0,1 %	32,2 %	2,9 %
6	1,4 %	0,0 %	31,7 %	2,8 %
7	-1,5 %	0,0 %	31,6 %	2,8 %
8	-1,2 %	0,0 %	31,5 %	2,8 %
9	0,8 %	0,1 %	31,6 %	2,8 %
10	0,5 %	0,0 %	31,9 %	2,9 %
11	-1,3 %	0,0 %	31,8 %	2,8 %
12	-0,2 %	0,0 %	31,7 %	2,9 %
13	0,0 %	0,0 %	30,9 %	2,9 %
14	0,2 %	-0,1 %	31,3 %	2,9 %
15	0,7 %	0,1 %	31,8 %	2,8 %

Tabla 5: *Sesgo relativo y coeficiente de variación para 2000 iteraciones*

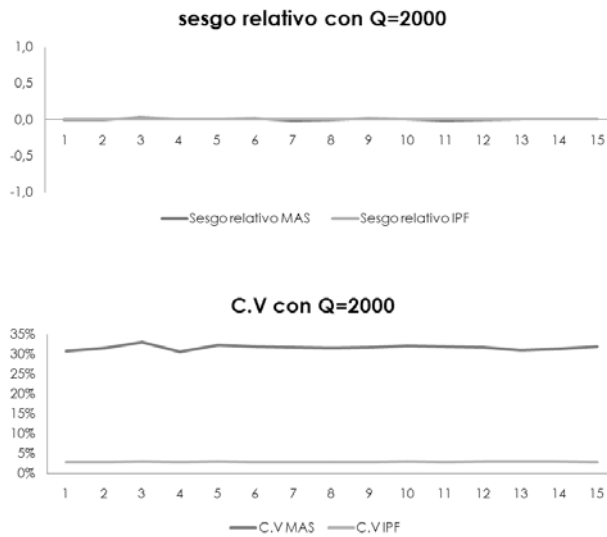


Figura 3: *Sesgo relativo y coeficiente de variación al replicar 2000 veces los dos estimadores*

Como se observa, el C.V. es estable desde las 1000 muestras, pero se opta por desarrollar el resto de simulaciones con un tamaño de 1500 muestras, ya que aunque el sesgo no es estable, los valores en los que ronda no son importantes y puesto que

el $\hat{t}_{y\pi}$ es insesgado, el sesgo registrado es causado por el error de la simulación. Por otro lado, el proceso de calcular las 1500 réplicas, tarda entre 3 a 4 horas, y realizar un número mayor de réplicas no aportaría mucho con respecto al incremento de tiempo requerido.

2.2.2. Descripción de las poblaciones

Para el ejercicio de comparación se consideraron 27 universos construidos así: todos tienen dos variables auxiliares de tres categorías cada una. Tres poblaciones diferentes según la distribución como se generan variables auxiliares y la variable de interés, Uniforme, Normal y Gamma; el segundo factor es la frecuencia en la cantidad de individuos pertenecientes a las diferentes categorías de las variables auxiliares, estas frecuencias están distribuidas de una manera uniforme, con tendencia no definida y con tendencia marcada hacia un extremo de ellas (Figura 4). El tercer factor se refiere a los valores de la variable de interés y_k , donde al igual que en las frecuencias de individuos, en el primer caso se encuentran valores muy similares en todas las celdas, en el segundo, se encuentran con una tendencia no definida, y en el tercero con tendencia marcada hacia unas celdas específicas, como se muestra en la Figura 5.

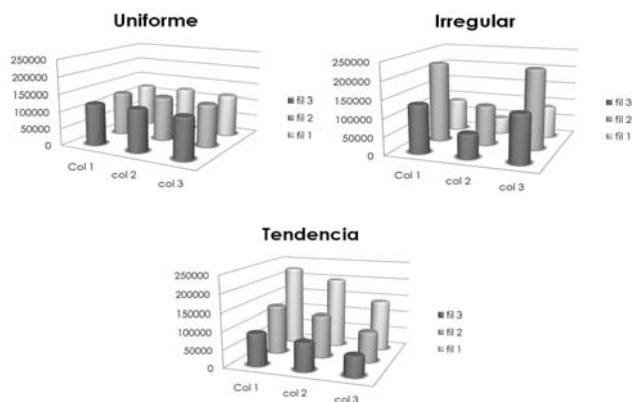


Figura 4: Cantidad de individuos por celda (clasificación de acuerdo a las dos variables auxiliares)

A cada población se le aplican cuatro formas de selección de muestras. Todas por medio de un diseño MAS^2 pero cambiando los parámetros de selección, en la primera se toman 20 conglomerados primarios de muestreo (CPM) con una selección del 12% de los elementos dentro de cada conglomerado, la segunda, 20 CPM y una selección del 25% de los elementos dentro de cada conglomerado, la tercera 70 CPM y 12% dentro de cada conglomerado, y por último 70 CPM y 25% dentro de cada conglomerado.

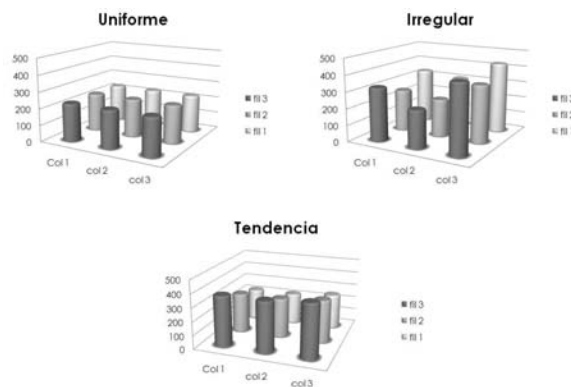


Figura 5: Promedio del valor de la variable de interés y_k clasificada por medio de las dos variables auxiliares

3. Resultados

Para los tres grupos de escenarios discriminados por el primer factor (distribuciones uniforme, normal y gamma) se obtienen prácticamente los mismos resultados, con lo que se procede a describir el grupo generado por la distribución uniforme y para los restantes se presentan los resultados en los anexos.

Para los 36 escenarios generados por medio de la distribución uniforme se nota que el estimador de rastrillo (IPF) tiene una reducción “abismal” en el coeficiente de variación con respecto al $\hat{t}_{y\pi}$ ya que el CV que se registra al emplear el estimador de Horvitz-Thompson, ronda los valores de 32 % cuando la selección de conglomerados es menor (20CPM) y se reduce al 16 % cuando en la muestra se escogen 70 CPM.

Lo anterior era de esperarse, pero los valores que registra el C.V al aplicar el estimador de rastrillo están alrededor del 1 % (ver Figura 7); en cuanto al sesgo relativo, los valores registrados por la simulación arrojan valores entre el $\pm 5\%$ para ambos estimadores que puede ocasionarse por errores que puede generar la simulación y teniendo en cuenta que el $\hat{t}_{y\pi}$ es un estimador insesgado.

Para observar mejor el comportamiento del coeficiente de variación para los dos estimadores, se puede apreciar que tanto las frecuencias de registros en las celdas como las tres diferentes formas de distribuir los valores de los y_k no afectan en lo mínimo a este valor, sólo se registra una caída cuando se cambian los parámetros de la muestra (mayor número de CPM).

20 CPM			
Frecuencia	y_k	Sesgo <i>MAS</i>	Sesgo <i>IPF</i>
Uniforme	Uniforme	0.8 %	-0.01 %
	Irregular	-0.2 %	-0.01 %
	Tendencia	-0.7 %	-0.01 %
Irregular	Uniforme	-0.8 %	0.02 %
	Irregular	1.6 %	0.01 %
	Tendencia	-0.8 %	0.00 %
Tendencia	Uniforme	0.2 %	0.01 %
	Irregular	0.6 %	-0.03 %
	Tendencia	1.3 %	-0.01 %

Tabla 6: *Estimación de SESGO RELATIVO para poblaciones seleccionadas con 20 CPM*

70 CPM			
Frecuencia	y_k	Sesgo <i>MAS</i>	Sesgo <i>IPF</i>
Uniforme	Uniforme	0.0 %	0.02 %
	Irregular	-0.6 %	-0.01 %
	Tendencia	-0.2 %	0.02 %
Irregular	Uniforme	0.0 %	0.01 %
	Irregular	-0.3 %	-0.01 %
	Tendencia	-0.4 %	0.00 %
Tendencia	Uniforme	-0.8 %	0.00 %
	Irregular	0.3 %	0.01 %
	Tendencia	0.6 %	0.00 %

Tabla 7: *Estimación de SESGO RELATIVO para poblaciones seleccionadas con 70 CPM*

20 CPM			
Frecuencia	y_k	C.V <i>MAS</i>	C.V <i>IPF</i>
Uniforme	Uniforme	33 %	1.1 %
	Irregular	30 %	0.8 %
	Tendencia	31 %	0.9 %
Irregular	Uniforme	32 %	1.2 %
	Irregular	33 %	0.8 %
	Tendencia	31 %	0.9 %
Tendencia	Uniforme	31 %	1.1 %
	Irregular	32 %	0.8 %
	Tendencia	31 %	0.9 %

Tabla 8: *Estimación de C.V para poblaciones seleccionadas con 20 CPM*

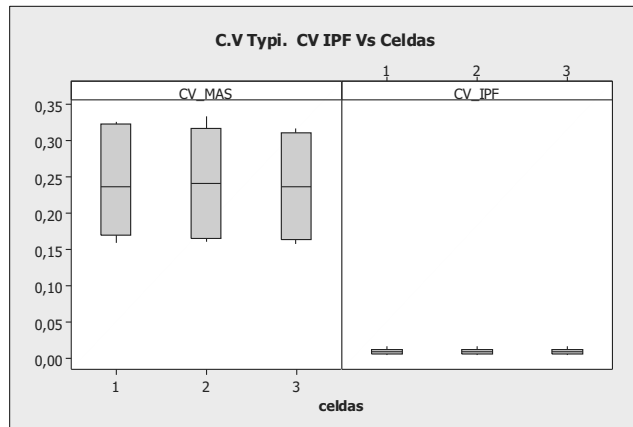


Figura 6: Gráfico de caja del coeficiente de variación con respecto a las tres formas de distribuir las cantidades de elementos por celda

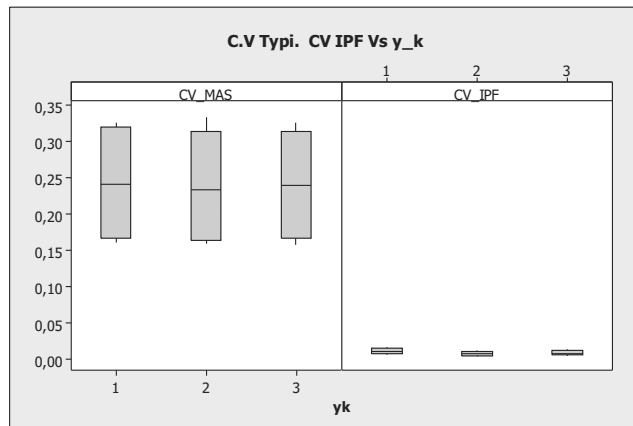


Figura 7: Gráfico de caja del coeficiente de variación con respecto a las tres configuraciones del total de y_k por celda

3.1. Caso extremo

Como para todos los escenarios el método IPF es una buena opción para aplicar, se realiza el ejercicio con un caso “extremo” donde se configura el universo de la siguiente forma:

Para esta configuración se obtienen resultados más favorables para el π -estimador que para el IPF en términos del C.V.

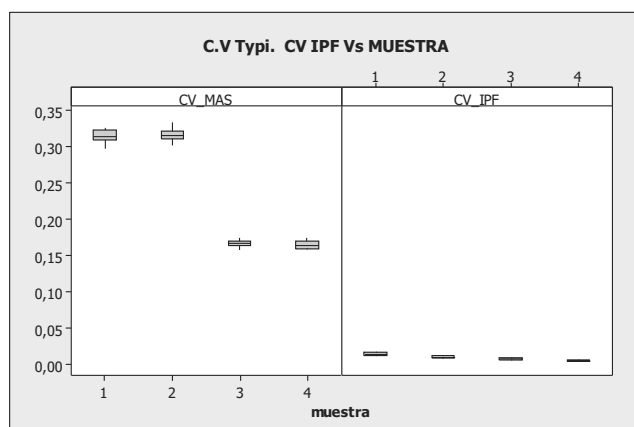


Figura 8: Gráfico de caja del coeficiente de variación con respecto a los cuatro diseños de muestra

70 CPM			
Frecuencia	y_k	C.V MAS	C.V IPF
Uniforme	Uniforme	17 %	0.6 %
	Irregular	16 %	0.4 %
	Tendencia	17 %	0.5 %
Irregular	Uniforme	16 %	0.6 %
	Irregular	17 %	0.4 %
	Tendencia	17 %	0.5 %
Tendencia	Uniforme	16 %	0.6 %
	Irregular	16 %	0.4 %
	Tendencia	16 %	0.5 %

Tabla 9: Estimación de C.V para poblaciones seleccionadas con 20 CPM

Cantidad porcentaje	Variable auxiliar C2			
	1	2	3	Total
1	108286	26121	20353	154760
	11 %	3 %	2 %	15 %
Variable auxiliar	2	7995	220545	14060
	1 %	22 %	1 %	24 %
C1	3	9125	20736	596176
	1 %	2 %	58 %	61 %
Total	125406	267402	630589	1023397
	12 %	26 %	62 %	100 %

Tabla 10: Cantidad de elementos por celda

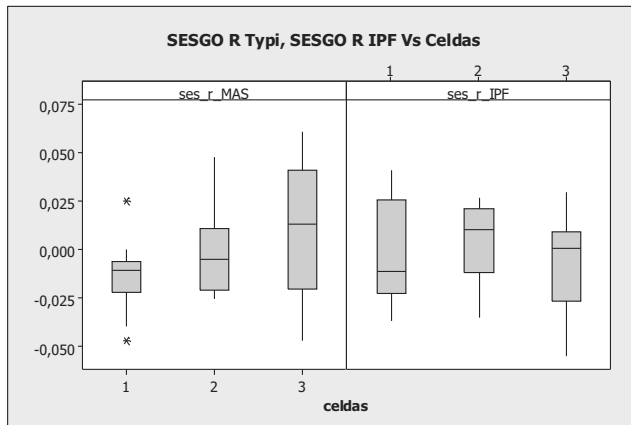


Figura 9: Gráfico de caja del sesgo relativo con respecto a las tres formas de distribuir las cantidades de elementos por celda

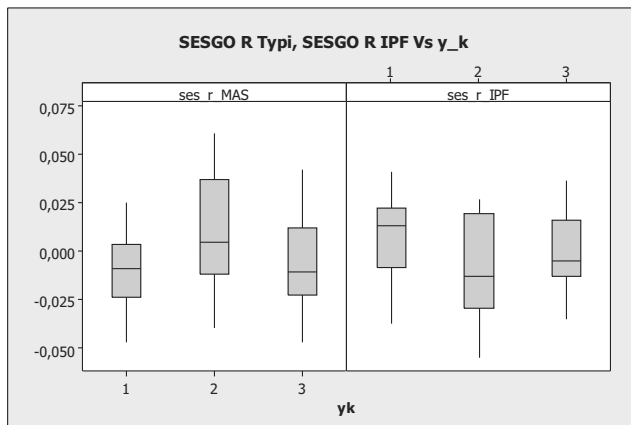


Figura 10: Gráfico de caja del sesgo relativo con respecto a las tres configuraciones del total de y_k por celda

4. Conclusiones

- El sesgo producido por IPF crece respecto al π -estimador en una proporción que es de poca importancia.
- El coeficiente de variación de IPF se pasa de 30% con el π -estimador al 2% con IPF estimador para 20 CPM en la muestra de la primera etapa.
- El coeficiente de variación se reduce del 17% al 1% cuando se pasa del π -estimador al IPF-estimador con tamaño de muestra de 70 CPM.

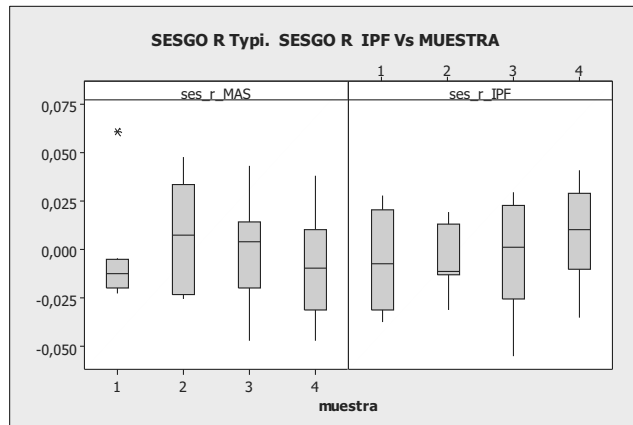


Figura 11: Gráfico de caja del sesgo relativo con respecto a los cuatro diseños de muestra

Suma (en millones)		Variable auxiliar C2		
Media		1	2	3
Varianza				
Variable auxiliar C1	1	3.24	15.01	10.68
		29.9	574.8	524.8
		6.89	107317.4	99192.79
Variable auxiliar C2	2	5.1	2.2	7.6
		638.5	10	540.5
		117824.9	2.7	101653.8
Variable auxiliar C3	3	4.43	9.43	1.19
		485.5	454.7	2
		99691.4	98022.9	0.45

Tabla 11: Características y_k por celda

	π – estimador	IPF – estimador
C.V	21.4%	28.1%
Sesgo relativo	0.01%	0.05%

Tabla 12: Valores de y_k

- La ganancia por aplicar IPF-estimador es entonces siempre muy grande frente al π -estimador.
- Para todos los universos generados, se obtiene la misma ganancia, es decir, la reducción del C.V no depende ni de la cantidad de y_k por celda ni del total.
- También se obtienen resultados muy parecidos cuando se varía el tamaño y la forma de la tabla

- Las pruebas realizadas con casos extremos de distribución de elementos por celda, fila o columna, son las únicas que arrojan resultados diferentes, en donde resulta que el π -estimador es mejor que el IPF-estimador. Estos resultados son válidos para la estimación de totales, aunque basados en simulaciones particulares, no pueden extrapolarse para la estimación de cocientes.

Agradecimientos

Gracias al maestro que no sólo me enseñó técnicas estadísticas y de muestreo, sino que se encargó de mostrarme cómo la disciplina y la pasión con que se hace cada labor, entre otras cosas, son herramientas para lograr la felicidad... Ojalá pudiera hacer entender entre éstas pocas líneas la admiración que siento al gran hombre que dedicó su vida a formar excelentes profesionales, Leonardo Bautista².

Recibido: 10 de febrero de 2008

Aceptado: 24 de abril de 2009

Referencias

- Bautista, J. (1998), *Diseños de muestreo estadístico*, Universidad Nacional de Colombia.
- Deville, J. & Särndal, C.-E. (1992), 'Calibration estimators in survey sampling', *JASA* pp. 376-382.
- Särndal, C.-E. & Lundström, S. (2005), *Estimation in Surveys with nonresponse*, Jhon wiley & Sons, Ltd, Sweden.
- Särndal, C.-E., Swensson, B. & Wretman, J. (1993), *Model Assisted Survey Sampling*, Springer, New York.
- SAS (2004), Sas onlinedoc 9.1.3., SAS Institute Inc, Cary, NC.

²Este artículo es resultado del trabajo de grado para optar por el título de estadístico titulado *Cálculo y estimación de la varianza en la aplicación del estimador de rastrillo*, dirigido por el maestro Leonardo Bautista y presentado en la Universidad Nacional de Colombia en 2007.

A. Resultados Población Normal

- Resultados para las poblaciones creadas por medio de la distribución NORMAL.

Frecuencia	y_k	20CPM		70 CPM	
		Sesgo <i>MAS</i>	Sesgo <i>IPF</i>	Sesgo <i>MAS</i>	Sesgo <i>IPF</i>
Uniforme	Uniforme	2.0 %	1.7 %	0.8 %	3.6 %
	Irregular	-2.2 %	1.3 %	0.6 %	4.7 %
	Tendencia	-0.4 %	1.3 %	0.7 %	2.2 %
Irregular	Uniforme	1.7 %	1.7 %	0.8 %	3.0 %
	Irregular	-1.2 %	1.3 %	0.7 %	2.1 %
	Tendencia	-0.4 %	1.4 %	0.7 %	6.6 %
Tendencia	Uniforme	1.0 %	1.8 %	0.9 %	-1.3 %
	Irregular	-3.1 %	1.3 %	0.7 %	1.0 %
	Tendencia	-5.0 %	1.5 %	0.8 %	0.2 %

Tabla 13: *Sesgo relativo para poblaciones generadas a partir de una distribución Normal*

Frecuencia	y_k	20 CPM		70 CPM	
		C.V <i>MAS</i>	C.V <i>IPF</i>	C.V <i>MAS</i>	C.V <i>IPF</i>
Uniforme	Uniforme	31.1 %	1.7 %	16.9 %	-1.9 %
	Irregular	30.3 %	1.3 %	16.1 %	8.8 %
	Tendencia	32.4 %	1.3 %	17.5 %	2.4 %
Irregular	Uniforme	30.5 %	1.7 %	15.9 %	2.4 %
	Irregular	31.6 %	1.3 %	16.3 %	3.5 %
	Tendencia	30.6 %	1.4 %	16.3 %	5.4 %
Tendencia	Uniforme	31.7 %	1.8 %	17.2 %	0.3 %
	Irregular	29.7 %	1.3 %	16.6 %	2.8 %
	Tendencia	31.4 %	1.5 %	16.9 %	-1.1 %

Tabla 14: *C.V para poblaciones generadas a partir de una distribución Normal*

B. Resultados Población Gamma

- Resultados para las poblaciones creadas por medio de la distribución GAMMA.

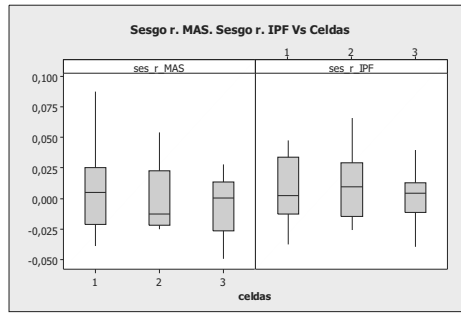


Figura 12: Gráfico de caja del sesgo relativo con respecto a las tres formas de distribuir las cantidades de elementos por celda

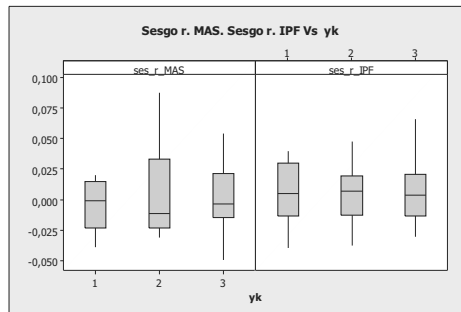


Figura 13: Gráfico de caja del sesgo relativo con respecto a las tres formas de distribuir el valor de y_k

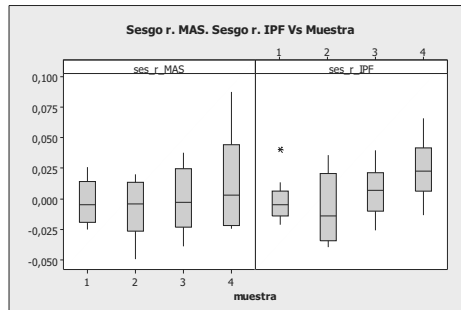


Figura 14: Gráfico de caja del sesgo relativo con respecto a los cuatro parámetros de selección de muestra

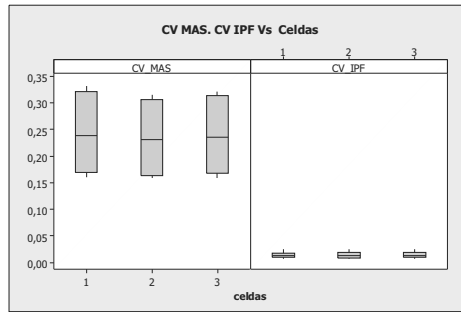


Figura 15: Gráfico de caja del sesgo relativo con respecto a las tres formas de distribuir las cantidades de elementos por celda

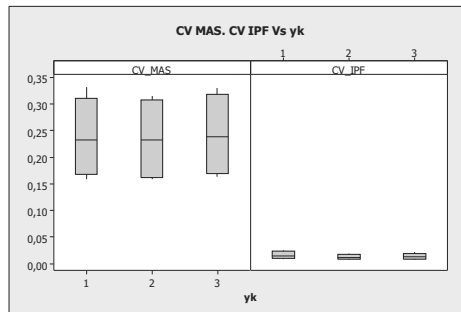


Figura 16: Gráfico de caja del sesgo relativo con respecto a las tres formas de distribuir el valor de y_k

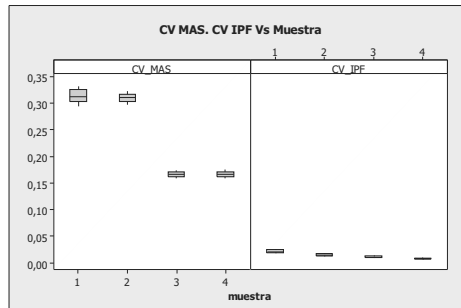


Figura 17: Gráfico de caja del sesgo relativo con respecto a los cuatro parámetros de selección de muestra

Frecuencia	y_k	20CPM		70 CPM	
		Sesgo <i>MAS</i>	Sesgo <i>IPF</i>	Sesgo <i>MAS</i>	Sesgo <i>IPF</i>
Uniforme	Uniforme	1.2 %	1.3 %	4.4 %	-3.0 %
	Irregular	2.7 %	1.0 %	-0.3 %	-4.0 %
	Tendencia	-7.1 %	4.9 %	-0.9 %	1.9 %
Irregular	Uniforme	-0.2 %	0.3 %	1.4 %	1.9 %
	Irregular	2.4 %	-5.5 %	-1.9 %	-9.7 %
	Tendencia	0.7 %	0.1 %	4.2 %	1.4 %
Tendencia	Uniforme	2.2 %	2.6 %	2.2 %	-3.4 %
	Irregular	0.6 %	-0.8 %	-1.7 %	-2.7 %
	Tendencia	-1.0 %	1.7 %	-1.4 %	1.3 %

Tabla 15: *Sesgo relativo para poblaciones generadas a partir de una distribución Normal*

Frecuencia	y_k	20 CPM		70 CPM	
		C.V <i>MAS</i>	C.V <i>IPF</i>	C.V <i>MAS</i>	C.V <i>IPF</i>
Uniforme	Uniforme	30.8 %	1.8 %	16.2 %	0.9 %
	Irregular	32.2 %	0.2 %	16.7 %	0.1 %
	Tendencia	30.9 %	0.2 %	16.5 %	0.1 %
Irregular	Uniforme	31.2 %	0.8 %	16.4 %	0.4 %
	Irregular	31.5 %	0.4 %	15.8 %	0.2 %
	Tendencia	31.5 %	0.2 %	16.6 %	0.1 %
Tendencia	Uniforme	31.8 %	0.8 %	17.1 %	0.4 %
	Irregular	31.0 %	0.4 %	16.1 %	0.2 %
	Tendencia	31.0 %	0.2 %	17.0 %	0.1 %

Tabla 16: *Sesgo relativo para poblaciones generadas a partir de una distribución Normal*

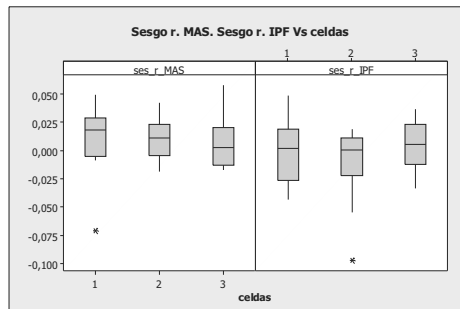


Figura 18: *Gráfico de caja del sesgo relativo con respecto a las tres formas de distribuir las cantidades de elementos por celda*

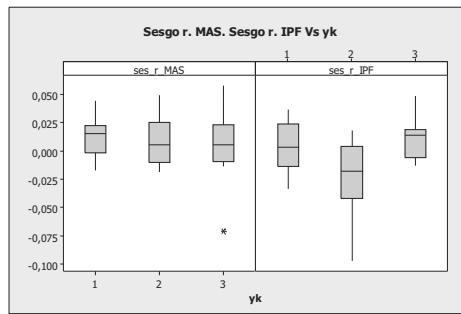


Figura 19: Gráfico de caja del sesgo relativo con respecto a las tres formas de distribuir el valor de y_k

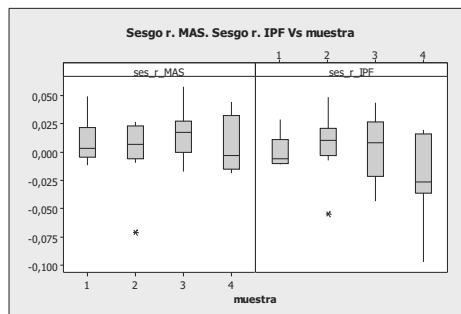


Figura 20: Gráfico de caja del sesgo relativo con respecto a los cuatro parámetros de selección de muestra

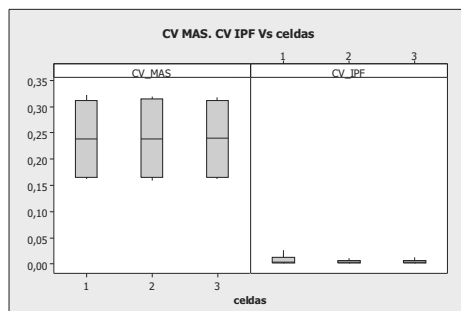


Figura 21: Gráfico de caja del sesgo relativo con respecto a las tres formas de distribuir las cantidades de elementos por celda

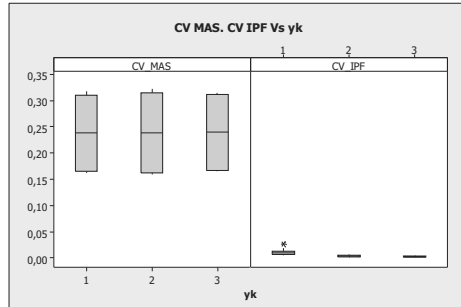


Figura 22: Gráfico de caja del sesgo relativo con respecto a las tres formas de distribuir el valor de y_k

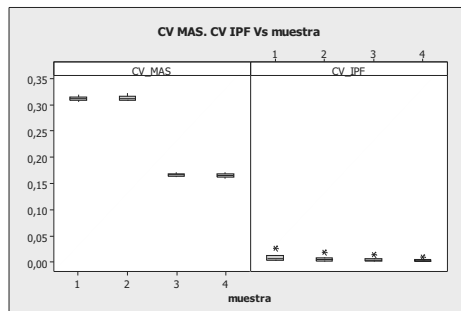


Figura 23: Gráfico de caja del sesgo relativo con respecto a los cuatro parámetros de selección de muestra