UNIVERSIDAD SANTO TOMAS
PRIMER CLAUSTRO UNIVERSITARIO DE COLOMBIA

Comunicaciones en Estadística
Junio 2009, Vol. 2, No. 1

Estimación de la tasa de favoritismo en la elección presidencial mediante el uso del estimador de Regresión Local Polinomial

Estimation of the Popularity Rating for Colombian Presidential Elections Using Local Polynomial Regression Estimators

María Eugenia Aguilera[a]
Maria-Eugenia.Aguilera-Torres@univ-lyon2.fr

Resumen

Se propone el uso y se evalúa el desempeño de varias estrategias muestrales en la estimación de la tasa de favoritismo de cierto candidato en las elecciones presidenciales de 2006, con estimadores basados en Regresión Local Polinomial en la primera etapa de selección. Se utilizan los resultados del comicio 2002 como información auxiliar. Por métodos no analíticos se estima la esperanza y la varianza de cada estimador para luego ser comparados.

***Palabras clave*:** muestreo, estrategia muestral, elecciones presidenciales, estimador asistido por Regresión Local Polinomial, información auxiliar, simulación de Monte Carlo.

Abstract

Several sampling strategies for estimating the popularity rating of a presidential candidate in the colombian elections of 2006, using Local Polynomial Regression Estimators in the first stage of sampling, are proposed and evaluated. The official data records of the 2002 colombian elections are used as the auxiliary information. The performance of these sampling strategies is evaluated by calculating the expectation and the variance of each estimator by non-analytical methods.

***Key words*:** survey sampling, sampling strategy, presidential elections, Local Polynomial Regression estimator, auxiliary information, Monte Carlo simulation.

[a]Estudiante de Maestría en Estadística. Université de Lyon.

1. Introducción

En Bautista (2005) se propone y evalúa un diseño muestral para la estimación de la tasa de favoritismo en elecciones presidenciales en 2006. En éste se considera la aplicación de un diseño MAS^2 (Bautista 1998) con 3 estratos en la primera etapa; que logra una reducción en la varianza del estimador en un 400 % con respecto al diseño con MAS en la primera etapa. En 2007, utilizando el hecho de que el presidente Álvaro Uribe Vélez se presentó como candidato en los comicios del 2002 y 2006, se utilizan los resultados del comicio 2002 como información auxiliar y la estrategia considerada por Bautista en 2005 para evaluar el desempeño de seis estimadores de la tasa de favoritismo asistidos por un modelo de regresión general (GREG-estimadores), y concluye que no hay ventajas importantes al utilizar el cociente de GREG-estimadores frente al cociente de π-estimadores para el caso particular de estimación de la tasa de favoritismo por Álvaro Uribe Vélez en el comicio 2006. Por otro lado, Breidt y Opsomer proponen un estimador de totales asistido por Regresión Local Polinomial, el cual resulta ser asintóticamente insesgado y consistente. Se usa la estrategia muestral, considerada por Bautista para estimar la tasa de favoritismo de Álvaro Uribe Vélez en el comicio 2006, para evaluar el desempeño de los estimadores asistidos por Regresión Local Polinomial propuestos en 2000.

2. Marco teórico

2.1. Tasa de favoritismo

Se llama U al universo de municipios del país, se establecen dos variables para cada uno de éstos (Bautista 2005):

y_k: Número de votos por el candidato Y en el municipio k para el comicio 2006

z_k: Número total de votos en el municipio k para el comicio 2006

La tasa de favoritismo para el candidato Y es el cociente entre la cantidad de votos por el candidato (t_y) sobre la cantidad de votos en el comicio (t_z).

$$C_y = \frac{\sum\limits_{k \in U} y_k}{\sum\limits_{k \in U} z_k} = \frac{t_y}{t_z} \tag{1}$$

La abstención electoral en Colombia es alta y variable entre municipios y sectores poblacionales, lo que convierte a la cantidad total de votantes en cifra aleatoria, haciendo que la tasa de favoritismo del candidato Y sea una tasa y no una proporción.

2.2. Estimación de totales

Considere una población finita $U = \{1, \dots i, \dots, N\}$. Para cada $k \in U$, se cuenta con una variable auxiliar x_k. Se extrae una muestra probabilística de U de acuerdo a un diseño muestral de tamaño fijo $p(\cdot)$; donde $p(m)$ es la probabilidad de extraer la muestra m. La probabilidad de inclusión del elemento k es

$$\pi_k = \Pr\{k \in m\} = \sum_{m \ni k} p(m) > 0 \tag{2}$$

Se observa la variable de estudio y_k para todo k en la muestra m. Se desea estimar $t_y = \sum\limits_{k \in U} y_k$

2.3. Estimador de Horvitz - Thompson

Sea

$$I_k = \begin{cases} 1 & si \;\; k \in m \\ 0 & e.\ o.\ c. \end{cases}$$

Entonces, el *pi*-estimador o el estimador de Horvitz-Thompson (1952) está definido como:

$$\widehat{t}_{y\pi} = \sum_{k \in m} \frac{y_k}{\pi_k} = \sum_{k \in U} \frac{y_k}{\pi_k} I_k \tag{3}$$

El estimador de Horvitz-Thompson es insesgado y para el diseño MAS es consistente.

2.4. Estimador asistido por Regresión Local Polinomial

El estimador propuesto en Breidt & Opsomer (2000) está motivado por la idea de modelar la población finita de y_k dada la variable auxiliar x_k, como una realización de la superpoblación infinita ξ en la cual

$$y_k = f(x_k) + \varepsilon_k \tag{4}$$

Donde ε_k son variables aleatorias independientes con media cero y varianza $v(x_k)$. Dado x_k, $f(x_k) = E_\xi[y_k]$ es llamada la función de regresión.

Sea K una función Kernel continua y h su ancho de banda. Sea $\mathbf{y}_U = [y_k]_{k \in U}$ el vector $N-$dimensional de y_k en la población. Entonces, se define la matriz $\mathbf{Z}_{U_k}$ como

$$\mathbf{Z}_{U_k} = \begin{bmatrix} 1 & x_1 - x_k & \cdots & (x_1 - x_k)^q \\ \vdots & \vdots & & \vdots \\ 1 & x_N - x_k & \cdots & (x_N - x_k)^q \end{bmatrix} = \begin{bmatrix} 1 & x_j - x_k & \cdots & (x_j - x_k)^q \end{bmatrix}_{j\in U}, \tag{5}$$

Y se define la matriz $\mathbf{W}_{U_k}$ como

$$\mathbf{W}_{U_k}= diag\left\{\frac{1}{h}K\left(\frac{x_j - x_k}{h}\right)\right\}_{j\in U} \tag{6}$$

Sea e_r, el r-ésimo renglón de $\mathbf{I}_p$ (la matriz identidad de orden p); entonces, el estimador local polinomial Kernel de la función de regresión en x_k basado en la población finita U es:

$$f_k= e_1'\left(\mathbf{Z}_{U_k}'\mathbf{W}_{U_k}\mathbf{Z}_{U_k}\right)^{-1}\mathbf{Z}_{U_k}'\mathbf{W}_{U_k}\mathbf{y}_U \tag{7}$$

En caso de que estos f_k fueran conocidos, un posible estimador insesgado de t_y es el siguiente:

$$t_y^*= \sum_{k\in U} f_k + \sum_{k\in U}\frac{D_k}{\pi_k}I_k \qquad \text{con} \quad D_k = y_k - f_k \tag{8}$$

Nótese que f_k es el estimador de regresión local polinomial para la función desconocida $f(\cdot)$.

En el caso de inferencia en poblaciones finitas, f_k no puede ser calculado debido a que solo se conocen los y_k en la muestra $m \subseteq U$. Por tanto, se reemplazará cada f_k por un estimador consistente basado en la muestra

$$\widehat{f}_k^0 = e_1'\left(\mathbf{Z}_{m_k}'\mathbf{W}_{m_k}\mathbf{Z}_{m_k}\right)^{-1}\mathbf{Z}_{m_k}'\mathbf{W}_{m_k}\mathbf{y}_m \tag{9}$$

con $\mathbf{y}_m = [y_k]_{k\in m}$ el vector n-dimensional de y_k en la muestra $m \subseteq U$ y

$$\mathbf{Z}_{mk} = \begin{bmatrix} 1 & x_1 - x_k & \cdots & (x_1 - x_k)^q \\ \vdots & \vdots & & \vdots \\ 1 & x_n - x_k & \cdots & (x_n - x_k)^q \end{bmatrix} = \begin{bmatrix} 1 & x_j - x_k & \cdots & (x_j - x_k)^q \end{bmatrix}_{j\in m},$$

$$\mathbf{W}_{mk} = diag\left\{\frac{1}{h\pi_k}K\left(\frac{x_j - x_k}{h}\right)\right\}_{j\in m}$$

El estimador de regresión local polinomial para la población total, propuesto por Breidt & Opsomer (2000) es

$$\widetilde{t}_{0y} = \sum_{k \in U} \widehat{f}_k^0 + \sum_{k \in U} \frac{y_k - \widehat{f}_k^0}{\pi_k} I_k \tag{10}$$

2.5. Estimación de un cociente

Se desea estimar $C_y = \frac{t_y}{t_z}$. Un posible estimador sería $\widehat{C}_y = \frac{\widehat{t}_y}{\widehat{t}_z}$. Se usa el método de linearización de Taylor para aproximar la varianza de $\widehat{C}_y$:

$$\begin{aligned}
\widehat{C}_y &= C_y + \frac{\Im \widehat{C}_y}{\Im \widehat{t}_y}|_{\widehat{t}_y = t_y}(\widehat{t}_y - t_y) + \frac{\Im \widehat{C}_z}{\Im \widehat{t}_z}|_{\widehat{t}_z = t_z}(\widehat{t}_z - t_z) + R \\
\widehat{C}_y &= C_y + \frac{\Im \widehat{C}_y}{\Im \widehat{t}_y}|_{\widehat{t}_y = t_y}\widehat{t}_y + \frac{\Im \widehat{C}_z}{\Im \widehat{t}_z}|_{\widehat{t}_z = t_z}\widehat{t}_z - \frac{\Im \widehat{C}_y}{\Im \widehat{t}_y}|_{\widehat{t}_y = t_y} t_y - \frac{\Im \widehat{C}_z}{\Im \widehat{t}_z}|_{\widehat{t}_z = t_z} t_z + R
\end{aligned}$$

$$\begin{aligned}
Var(\widehat{C}_y) &= Var(C_y + \frac{\Im \widehat{C}_y}{\Im \widehat{t}_y}|_{\widehat{t}_y = t_y}\widehat{t}_y + \frac{\Im \widehat{C}_z}{\Im \widehat{t}_z}|_{\widehat{t}_z = t_z}\widehat{t}_z - \frac{\Im \widehat{C}_y}{\Im \widehat{t}_y}|_{\widehat{t}_y = t_y} t_y - \frac{\Im \widehat{C}_z}{\Im \widehat{t}_z}|_{\widehat{t}_z = t_z} t_z \\
&\quad +R) \\
&= Var(\frac{\Im \widehat{C}_y}{\Im \widehat{t}_y}|_{\widehat{t}_y = t_y}\widehat{t}_y + \frac{\Im \widehat{C}_z}{\Im \widehat{t}_z}|_{\widehat{t}_z = t_z}\widehat{t}_z + R) \\
&\approx Var(\frac{\Im \widehat{C}_y}{\Im \widehat{t}_y}|_{\widehat{t}_y = t_y}\widehat{t}_y + \frac{\Im \widehat{C}_z}{\Im \widehat{t}_z}|_{\widehat{t}_z = t_z}\widehat{t}_z)
\end{aligned}$$

Para el caso del π-estimador, es sencillo encontrar una expresión para la aproximación de la varianza del estimador; pero en el caso del estimador de Regresión Local Polinomial, cada valor a predecir es función del kernel a escoger, del grado del polinomio y de la variable auxiliar; por tanto, resulta difícil y poco práctico (cada caso particular tendría una expresión distinta) deducirla por métodos exclusivamente analíticos. Se usará un método alterno no analítico para estimar la varianza del estimador de Regresión Local Polinomial.

2.6. Método de Monte Carlo

El método Monte Carlo es un método numérico que permite resolver problemas físicos, matemáticos y estadísticos mediante la simulación de variables aleatorias. Este procedimiento resulta a veces más simple que la tentativa de construir un modelo analítico para las operaciones complejas en las cuales participa una gran cantidad de elementos cuyos factores causales están entrelazados en forma compleja. Cualquier problema probabilístico puede ser resuelto por medio del método de Monte Carlo, pero resulta justificado sólo cuando el procedimiento de sorteo es más simple que el cálculo analítico.Generalmente se utiliza el método de Monte Carlo en los siguientes casos:

- Para simular operaciones complicadas y complejas que tienen muchos factores aleatorios que accionan recíprocamente.
- Para comprobar la aplicabilidad de métodos analíticos más simples y aclarar las condiciones de su aplicación.
- Para elaborar las correcciones de las fórmulas analíticas del tipo fórmulas empíricas de la técnica.

Algoritmo

1. Se elabora un programa para la realización de una prueba aleatoria. En este caso habría que tomar una muestra aleatoria del universo de municipios.
2. Se mide la variable de estudio o se toma una decisión de acuerdo a la prueba aleatoria realizada. Para el caso en cuestión, se estima el total de las variables de estudio y su cociente.
3. Se repite un número determinado de veces (por decir, Q) de modo que cada experimento sea independiente de los restantes. Para este contexto, se toman Q muestras aleatorias independientes entre sí y se les aplica el estimador.
4. Al final se toma la media de los resultados de todos los experimentos.
5. Además el error es como regla, proporcional a la magnitud $\sqrt{\frac{D}{Q}}$ donde D es una constante y Q el número de pruebas. Esta fórmula permite ver que para disminuir el error 10 veces (es decir que para obtener en el resultado otra cifra decimal exacta) es preciso aumentar Q (es decir, el número de muestras) en 100 veces.

3. Descripción de los datos

Para la estimación de la tasa de favoritismo del candidato Y, se cuenta con el registro electoral (RNEC 2006) de cada municipio del país para los comicios 2002 y 2006 organizado en la Tabla 1. Sin embargo, en el cálculo del total se incluyó la votación por otros candidatos.

Las variables de estudio son: la votación por Álvaro Uribe Vélez (candidato Y) y la votación total; ambos casos para cada municipio del país y para el período electoral 2006-2010. Las variables auxiliares son: la votación por Álvaro Uribe Vélez y la votación total; ambos casos para cada municipio del país y para el período electoral 2002-2006. Se propone extraer 80 municipios mediante un muestreo estratificado del tipo IF (estrato de inclusión forzosa)- ESTMAS (estratificado MAS)- UNO (Bautista 2005), es decir, se investigan todos los municipios del primer estrato, se extraen muestras MAS en el estrato intermedio y en el último estrato se extraen de uno a tres municipios.

	Serpa		Uribe		Total	
Depto-Mpio	2002	2006	2002	2006	2002	2006
Antioquia-Medellín	56178	20751	387510	445907	520400	604941
Antioquia-Abejorral	1100	223	2724	4546	4606	5349
Antioquia-Abriaquí	76	20	382	572	655	719
Antioquia-Alejandría	81	44	604	1302	1021	1520
⋮	⋮	⋮	⋮	⋮	⋮	⋮
Vaupés-Mitú	1860	506	877	1818	3074	3468
Vaupés-Taraira	3	77	5	630	14	810
Vichada-Puerto Carreño	2173	125	915	571	3299	1066
Vichada-La Primavera	581	80	659	1700	1342	1926

Tabla 1: *Organización de los datos del registro nacional electoral para el año 2006*

De acuerdo al tamaño de muestra propuesto, este diseño alcanza un coeficiente de variación de 3.92 (Bautista 2005) en la primera etapa usando π-estimador; se espera con el estimador por Regresión Local Polinomial, reducirlo o, al menos, igualarlo. Los estratos quedan definidos de la siguiente forma (Bautista 2005):

- Ciudades de inclusión forzosa: los 13 municipios con mayor votación en el país. Éstos son Bogotá, Medellín, Cali, Barranquilla, Bucaramanga, Cúcuta, Cartagena, Manizales, Ibagué, Pereira, Villavicencio, Bello y Armenia. Se incluyen todas en la muestra. Contempla el 44.2 % de los votantes del país.
- Municipios intermedios: son 758 municipios, de los cuales se tomará una muestra tamaño 65. Contempla aproximadamente el 53.7 % de los votantes del país.
- Un último estrato que totaliza el 2 % de la población. Se tomará una muestra de tamaño 2 y se usará únicamente el π-estimador.

Es de especial interés el estrato intermedio, ya que concentra la mitad de los votantes del país, tiene la mayor cantidad de municipios y es al cual se le aplicará el estimador de regresión local polinomial.

3.1. Estrato Intermedio

Se observa que la distancia entre la votación en cada municipio es directamente proporcional a su tamaño, por lo que hay muchos municipios con votaciones muy parecidas, y unos pocos, muy alejados del grupo, con votaciones muy altas.

La separación proporcional al tamaño de la variable auxiliar puede generar problemas a la hora de elegir el ancho de banda. Hay que tener en cuenta que el ancho de banda es el encargado de definir cuantos vecinos tiene cierto registro.

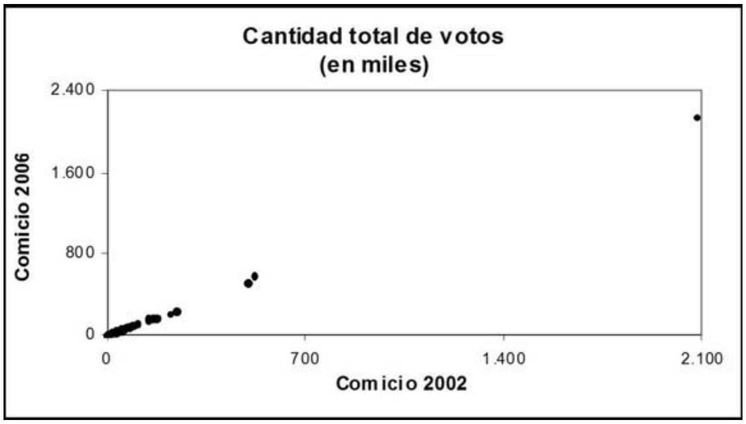

Figura 1: *Diagrama de dispersión de la votación total por municipio de 2006 con respecto a la votación total por municipio de 2002*

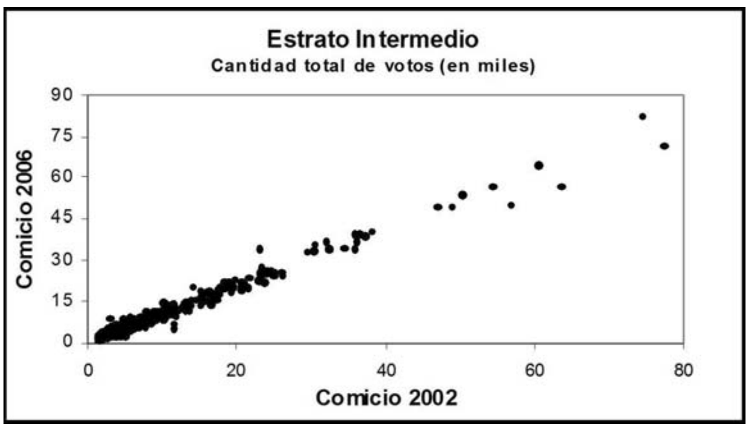

Figura 2: *Diagrama de dispersión de la votación total por municipio con respecto a la votación en el período anterior*

Si el ancho de banda es muy pequeño, el estimador pierde robustez (se deja influenciar por atípicos), sin tener en cuenta que algunos municipios podrían no tener vecinos en la muestra. Si el ancho de banda es muy grande, el ajuste aparte de ser muy suave, podría carecer de sentido, ya que se puede predecir la votación de municipios grandes a partir de la de unos pequeños y viceversa. La cuestión entonces es si en la muestra sólo se cuenta con 64 posibles vecinos con valores entre 16 y 100000, ¿cuál sería el ancho de banda adecuado?

Para tal fin, se proponen varias soluciones:

- Hallar un ancho de banda adecuado para ajustar los valores más separados de la variable auxiliar, ya que en últimas, éstos son más importantes para la estimación de la tasa de favoritismo. Inconveniente: el ajuste para los municipios pequeños será muy suave.
- Después de extraída la muestra, se les asigna a los datos dos posibles anchos de banda de acuerdo a la separación entre éstos. Es decir, hasta cierto punto límite de la variable auxiliar, el ancho de banda será h y a partir de éste el ancho de banda será $2h$.
- En lugar de mantener un ancho de banda fijo, se mantiene un vecindario fijo. Es decir, el ancho de banda cambia de acuerdo al vecindario deseado; de esta forma se garantiza la existencia de vecindario para todos los registros.
- Aplicarle transformaciones a la variable auxiliar, de tal forma que la separación entre registros no sea muy ámplia o se haga regular. Aquí se analiza el comportamiento del logaritmo de ésta y de su rango.

Las Figuras 3 y 4 muestran el comportamiento de la variable de estudio con respecto a ciertas transformaciones de la variable auxiliar. Para el primer caso (logaritmo de ésta), a pesar de que no se logra total regularidad en las distancias, se observan pocos municipios alejados del resto y a una menor distancia.

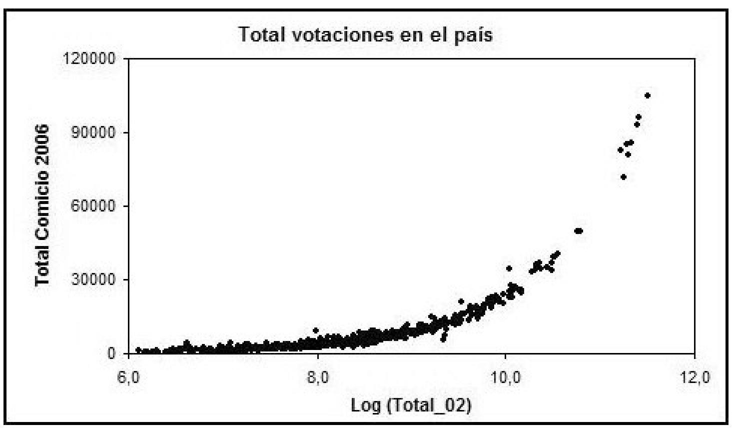

Figura 3: *Diagrama de dispersión de la votación total por municipio con respecto al logaritmo de su votación en el período anterior*

La sugerencia es cambiar la votación obtenida en 2002 por su rango como se muestra en la figura 4. En este caso, la separación entre los registros de la variable auxiliar es regular; se encuentran cambios bruscos en la variable de estudio. Se observa que en este caso, la correlación entre la variable de estudio y la variable auxiliar a usar es prácticamente cero.

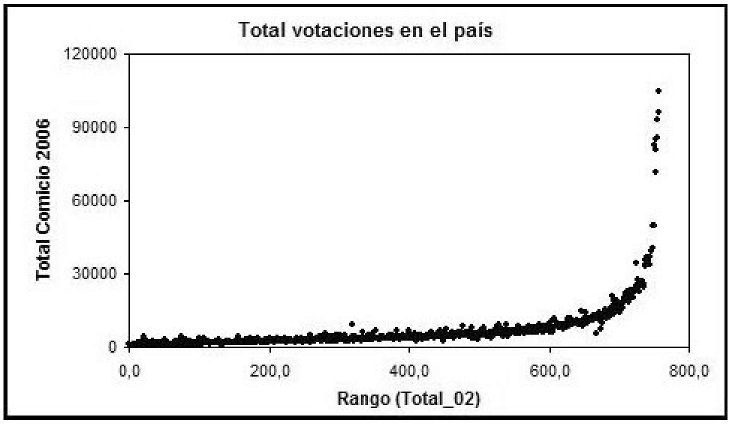

Figura 4: *Diagrama de dispersión de la votación total por municipio con respecto al rango de su votación en el período anterior*

4. Resultados

Se compara el desempeño de los estimadores de Horvitz-Thompson (HT), Regresión Local Polinomial, con los registros electorales de 2002 como variable auxiliar (RLP), Regresión Local Polinomial, con el logaritmo de los registros electorales de 2002 como variable auxiliar (RLP2) y Regresión Local Polinomial con el rango de los registros electorales de 2002 como variable auxiliar (RLP3); los tres últimos con ajuste cuadrático ($q = 2$). Se usa el kernel Epanechnikov,

$$K(t) = \frac{3}{4}(1 - t^2)I_{\{|t| \leq 1\}} \tag{11}$$

para los tres estimadores de regresión local polinomial. A modo de prueba, se simularon 200 posibles muestras del estrato intermedio y se halló el ancho de banda promedio requerido para que cada elemento tenga seis vecinos en la muestra. Los resultados son los siguientes:

	Candidato Y		**Total**	
	h fijo	h variable	h fijo	h variable
RLP	9000	6000	11000	8000
RLP1	1.2	1.0	1.4	1.2
RLP2	40	25	40	25

Tabla 2: *Anchos de banda fijos empleados para la estimación local polinomial*

Los anteriores anchos de banda propuestos, son comparados con los anchos de banda hallados por el criterio empírico de la cuarta parte del rango de la variable auxiliar. Esto, únicamente para ancho de banda fijo.

	Candidato Y	**Total**
RLP	14914	24606
RLP1	2.06	1.35
RLP2	189	189

Tabla 3: *Anchos de banda fijos empleados para la estimación local polinomial. Criterio empírico*

Luego de haber fijado los anchos de banda, se extraen 5000 muestras aleatorias simples del estrato intermedio de tamaño $n = 65$. A cada muestra se le aplican los cuatro estimadores y se evalúa el comportamiento promedio de éstos. Específicamente, se estiman el sesgo relativo, la probabilidad de cobertura, el coeficiente de variación de la estrategia y error cuadrático medio de la estrategia. A continuación se encuentran los totales de votación y la tasa de favoritismo del candidato Y.

Valores reales en el estrato intermedio	
Total votos candidato Y	3284647
Total votos presidenciales	5322380
Tasa de Favoritismo(%)	61.7

Tabla 4: *Anchos de banda fijos empleados para la estimación local polinomial. Criterio empírico*

A modo de resumen de resultados obtenidos, se muestra únicamente el error cuadrático medio de la estrategia con respecto al error cuadrático medio obtenido por el π-estimador.

5. Discusión y conclusiones

Se presentaron estimativos inadmisibles tanto para totales como para la tasa de favoritismo; por ejemplo, en la simulación 1861 el total estimado para el candidato es de -193845321, el total de votaciones estimado es 4945037; y la tasa de favoritismo estimada fue $-3920\,\%$. El estimador propuesto por Breidt y Opsomer (2000), a pesar de tener resultados levemente mejores en totales, no resulta práctico a la hora de hacer estimaciones por muestreo para el caso particular de elecciones presidenciales en Colombia.

		$\text{ECM}(\widehat{t}_y)/\text{ECM}(\widehat{t}_{y\pi})$			
			h fijo	h	vecinda.
		Arbitrario	$\frac{1}{4}(x_{\text{máx}} - x_{\text{mín}})$	variable	fijo
	RLP	2.82	1.47	17.18	0.77
Total Candidato Y	RLP1	1.99	0.95	2.65	0.34
	RLP2	6.24	3.30	13.90	0.61
	RLP	108.8	2.93	0.72	0.82
Total Votación	RLP1	18.2	150.3	46.8	0.39
	RLP2	8.15	0.31	16.9	0.40
	RLP	65.40	30.15	404.9	27.6
Total de Favoritismo	RLP1	123.6	130.1	122.0	15.3
	RLP2	195.4	5.15	318.3	24.4

Tabla 5: *Cociente entre el ECM de los estimadores de regresión local polinomial y el pi-estimador*

El mejor desempeño para totales lo obtuvo la combinación ancho de banda fijo empírico y la transformación rango, aunque el vecindario fijo tuvo estimaciones eficientes para todos los casos y todas las transformaciones, demostrando que el problema radicaba en el ancho de banda a escoger y el vecindario del punto en la muestra. Por otro lado, no hay estimador con mejor desempeño (en términos de eficiencia) que el de Horvitz-Thompson para el caso de elecciones presidenciales en Colombia.

Al igual que para totales, hubo casos en los cuales la estimación de la tasa de favoritismo dió mayor a uno (en valor absoluto) o negativa. Nótese que la combinación de mayor eficiencia fue ancho de banda fijo para la transformación rango de la variable auxiliar, aunque no fue mayor que la del π-estimador.

Agradecimientos

Agradezco al profesor Leonardo Bautista por sus múltiples y valiosos comentarios[1].

Recibido: 1 de febrero de 2009
Aceptado: 15 de abril de 2009

[1]Este artículo es resultado del trabajo de grado para optar por el titulo de estadístico titulado *Estimación de la tasa de favoritismo en la elección presidencial mediante el uso del estimador de Regresión local polinomial*, dirigido por el maestro Leonardo Bautista y presentado en la Universidad Nacional de Colombia en 2007

Referencias

Bautista, L. (1998), *Diseños de muestreo estadístico*, Universidad Nacional de Colombia, Bogotá, Colombia.

Bautista, L. (2000), *Diseño y desarrollo de encuestas*, Simposio Colombiano de Estadística. Universidad Nacional de Colombia, San Andrés, Colombia.

Bautista, L. (2005), 'Estrategia de muestreo para la estimación de la tasa de favoritismo en la elección presidencial', *Revista Colombiana de Estadistica* **28**(1), 33–62.

Breidt, F. & Opsomer, J. (2000), 'Local polynomial regression estimators in survey sampling', *Annals of Statistics* **28**(4), 1026–1053.

Guacaneme, F. (2007), Intervalos de confianza para pronósticos no paramétricos de la inflación colombiana., *in* 'Memorias Simposio de Estadística', Vol. 11.

RNEC (2002), *Elecciones presidenciales de 2002 en Colombia*, Registraduría Nacional del Estado Civil, Bogotá, Colombia.

RNEC (2006), *Elecciones presidenciales de 2006 en Colombia*, Registraduría Nacional del Estado Civil, Bogotá, Colombia.

Särndal, C., Swensson, B. & Wretman, J. (1995), *Model Assisted Survey Sampling*, Springer, New York.

Wand, M. & Jones, M. (1995), *Kernel Smoothing*, Chapman and Hall, London.