

---

## Estrategia de muestreo usando estimadores de regresión generalizada para la estimación de tasas de favoritismo en elecciones presidenciales en Colombia

Sampling Strategy using Generalized Regression Estimators for Presidential Elections Favoritism Rate in Colombia

Linda Johana Torres Celis<sup>a</sup>  
ljtorresc@bt.unal.edu.co

---

### Resumen

Se propone el uso y se evalúa el desempeño de seis estimadores de la tasa de favoritismo en elecciones presidenciales que usen GREG estimadores en la primera etapa de selección. Se utiliza el método de simulación Monte Carlo y se calcula el coeficiente de variación de cada uno de los estimadores para el caso específico de las elecciones presidenciales del periodo 2006, utilizando como información auxiliar los resultados del periodo 2002. Se concluye que los estimadores propuestos no tienen atributos que los hagan preferibles al cociente de pi-estimadores.

**Palabras clave:** GREG-estimadores, linearización de Taylor, simulación Monte Carlo.

### Abstract

Six favoritism rate estimators for presidential elections, which use GREG estimators in the first stage of sampling, are proposed. Its performance is evaluated with Monte Carlo Simulation and by the variation coefficient, calculated for the 2006 Presidential elections case, using the results from 2002 as auxiliary information. It is found that the estimators proposed don't have attributes which makes them better than the totals pi-estimators quotient.

**Key words:** GREG-estimators, Taylor linearization, Monte Carlo simulation.

---

<sup>a</sup>Gestor en la Coordinación de Estudios Económicos. Subdirección de Gestión de Análisis Operacional. Departamento de Impuestos y Aduanas Nacionales (DIAN).

## 1. Introducción

En Bautista (2005) se propone y evalúa una estrategia muestral para la estimación de resultados electorales del 2006. La estrategia propuesta consiste en un diseño en cuatro estratos y tres etapas conformado así :

- Grandes ciudades: inclusión forzosa de 21 municipios; uno de cada 20 sectores cartográficos, mínimo dos por municipio, 60 % de las manzanas por sector y una de cada 25 personas por manzana.
- Ciudades intermedias: 44 de 144 municipios, 10 % de las manzanas por municipio y una de cada 25 personas por manzana.
- Municipios pequeños: 14 de 610 municipios, 60 % de las manzanas por municipio y una de cada 25 personas por manzana.
- Municipios muy pequeños y alejados: 1 de 241 municipios, 60 % de las manzanas por municipio y una de cada 25 personas por manzana.

El estimador es el cociente entre el pi-estimador para el total de votos por el candidato Uribe y el pi-estimador para el total de votos. El coeficiente de variación para la tasa de el candidato Uribe es de 5 %.

Se supone que el uso de una estrategia muestral que utilice GREG estimadores en la primera etapa de selección podría ser más eficiente. Se dispone de información auxiliar valiosa por cuanto se tiene el total de votos y el total por el candidato Uribe por municipio en 2002. Se propone el uso de estimadores de la tasa de favoritismo que utilicen GREG-estimadores, y se evalúa su desempeño. Lo anterior se aplica al caso específico de los resultados de las elecciones del 2006 con el candidato elegido Álvaro Uribe, con el candidato Horacio Serpa y el candidato Carlos Gaviria.

En este artículo se plantea evaluar la conveniencia del uso de GREG estimadores para la estimación de tasas de favoritismo electoral en el caso el candidato Uribe 2006 y también proponer una estrategia muestral utilizando información auxiliar y GREG- estimadores para realizar estimaciones en el caso de tasas de favoritismo en elecciones presidenciales, evaluando el desempeño de la estrategia en términos de precisión y confiabilidad mediante simulación de Monte Carlo.

## 2. GREG Estimadores

Un estimador más sofisticado que el estimador de Horvitz-Thompson para totales es el estimador de regresión generalizada (GREG-estimador). Una estimación obtenida por un GREG-estimador es una suma de valores observados ponderados mediante el producto de un factor de diseño y un factor adicional computado a través del uso de información auxiliar. Es necesario entonces disponer de una o más variables auxiliares que estén, en lo posible, altamente relacionadas con la variable cuyo total se va a estimar (Särndal & Lundström 2005).

Se denota al vector de información auxiliar como  $\mathbf{x}_k$  y su valor para el elemento  $k$  como  $\mathbf{x}_k = (x_{1k}, \dots, x_{Jk})'$  un vector columna con  $J \geq 1$  componentes, construido por una o más variables auxiliares. Así, se tiene el GREG-estimador dado por:

$$\hat{t}_{yGREG} = \hat{t}_{y\pi} + \sum_{j=1}^J \hat{B}_j(t_{x_j} - \hat{t}_{x_j\pi}) \quad (1)$$

donde

$$\mathbf{B} = (\hat{B}_1, \dots, \hat{B}_J)' = \left( \sum_s \mathbf{x}_k \mathbf{x}_k' / \sigma_k^2 \pi_k \right)^{-1} \sum_s \mathbf{x}_k y_k / \sigma_k^2 \pi_k \quad (2)$$

es el vector de coeficientes obtenido al ajustar la regresión de  $y$  sobre  $\mathbf{x}_k$ , usando las parejas  $(y_j, \mathbf{x}_k)$  para los  $k$  elementos de la muestra. El GREG-estimador es aproximadamente insesgado. El sesgo tiende a cero en la medida en que el tamaño de la muestra se incrementa. Cuando hay una relación fuerte entre la variable  $y$  y las variables auxiliares, hay una importante reducción de la varianza al usar Greg-estimadores.

Una forma alternativa para expresar el GREG-estimador es:

$$\hat{t}_{yGREG} = \sum_U \hat{y}_k + \sum_s e_{ks} / \pi_k \quad (3)$$

donde

$$\sum_U \hat{y}_k = \left( \sum_U \mathbf{x}_k \right)' \hat{\mathbf{B}}, e_{ks} = y_k - \hat{y}_k y \hat{y}_k = \sum_{j=1}^J \hat{B}_j x_{jk} \quad (4)$$

Otra forma alternativa para el estimador de regresión es:

$$\hat{t}_{yGREG} = \sum_s g_{ks} y_k / \pi_k \quad (5)$$

donde los pesos  $g_{ks}$  dependientes de la muestra están dados por:

$$g_{ks} = 1 + (\mathbf{t}_x - \hat{\mathbf{t}}_{x\pi})' \hat{\mathbf{T}}^{-1} \mathbf{x}_k / \sigma_k^2 \quad (6)$$

con  $\hat{\mathbf{T}} = \sum_s \frac{\mathbf{x}_k \mathbf{x}_k'}{\sigma_k^2 \pi_k}$ ,  $\mathbf{t}_x = (t_{x_1}, \dots, t_{x_J})'$  y  $\hat{\mathbf{t}}_{x\pi} = (\hat{t}_{x_1\pi}, \dots, \hat{t}_{x_J\pi})'$ .

La varianza del estimador es:

$$V(\hat{t}_{yGREG}) = V\left(\sum_s g_{ks} E_k / \pi_k\right) \quad (7)$$

La aproximación a la varianza de un GREG-estimador puede realizarse a través de los residuos  $E_k$  de la regresión, así:

$$AV(\hat{Y}_{GREG}) = \sum \sum_U \Delta_{kl} \frac{E_k}{\pi_k} \frac{E_l}{\pi_l} \quad (8)$$

cuyo estimador está dado por:

$$\hat{V}(\hat{y}_{GREG}) = \sum \sum_s \frac{\Delta_{kl}}{\pi_{kl}} g_{ks} \frac{e_{ks}}{\pi_k} g_{ls} \frac{e_{ls}}{\pi_l} \quad (9)$$

Algunos casos de especial interés son los siguientes: modelo heterocedástico sin intercepto y modelo homocedástico con intercepto para la estimación de totales. En el primer caso se tiene:

$$\hat{t}_{yra} = t_x \frac{\hat{t}_{y\pi}}{\hat{t}_x \pi} \quad (10)$$

En el caso del modelo homocedástico con intercepto se tiene:

$$\hat{t}_{yr} = N[\tilde{y}_s + \hat{B}_2(\bar{x}_U - \tilde{x}_s)] \quad (11)$$

donde  $\hat{B}_2 = \frac{\sum_s (x_k - \tilde{x}_s)(y_k - \tilde{y}_s)/\pi_k}{\sum_s (x_k - \tilde{x}_s)^2/\pi_k}$ ,  $\tilde{x}_s = \frac{\hat{t}_x \pi}{\hat{N}}$ ,  $\tilde{y}_s = \frac{\hat{t}_y \pi}{\hat{N}}$  y  $\hat{N} = \sum_s \frac{1}{\pi_k}$ .

### 3. Estimación de tasas de favoritismo usando GREG-estimadores

El parámetro que se desea estimar es la tasa de favoritismo por un candidato determinado. Se pretende explicar el total de votos a favor mediante una regresión a partir de los votos obtenidos en las elecciones anteriores para el candidato del mismo partido, así como explicar el total de votos por el total de votos obtenidos en las elecciones pasadas. De esta manera, contemplando el uso de los modelos heterocedástico y homocedástico en numerador y denominador, se tienen las combinaciones:

1. Haciendo uso de GREG estimadores tanto en el numerador como en el denominador.
2. Estimando mediante GREG estimadores únicamente al denominador.
3. Estimando mediante GREG estimadores únicamente al numerador.

En cada una de las anteriores estrategias se hace uso de dos modelos. Se contempla el uso de un modelo heterocedástico sin intercepto y el uso de un modelo homocedástico con intercepto, soportados en la información del año 2002 correspondiente a cada variable; esto es, si se desea estimar el total de votos para el año 2006 se utiliza como variable auxiliar el número de votos en cada municipio en el año 2002.

Sean las variables:

$x_k$  := votos por el candidato Uribe en las elecciones del 2002 en el municipio k.

$y_k$  := votos por el candidato Uribe en las elecciones del 2006 en el municipio k.

$w_k$  := votos en las elecciones del 2002 en el municipio k.

$z_k$  := votos en las elecciones del 2006 en el municipio k.

De esta manera, se establecen siete casos de estudio:

$$\hat{C}_1 = \hat{t}_{yra} / \hat{t}_{zra}$$

$$= \left[ \begin{array}{c} \hat{t}_{y\pi} \\ t_x \hat{t}_{x\pi} \end{array} \right] / \left[ \begin{array}{c} \hat{t}_{z\pi} \\ t_w \hat{t}_{w\pi} \end{array} \right]$$

$$\hat{C}_2 = \hat{t}_{yr} / \hat{t}_{zr}$$

$$= \left[ \frac{\hat{t}_{y\pi}}{\hat{N}} + \frac{\hat{N}\hat{t}_{xy\pi} - \hat{t}_{x\pi}\hat{t}_{y\pi}}{\hat{N}\hat{t}_{x^2\pi} - \hat{t}_{x\pi}^2} \left( \frac{t_x}{N} - \frac{\hat{t}_{x\pi}}{\hat{N}} \right) \right] / \left[ \frac{\hat{t}_{z\pi}}{\hat{N}} + \frac{\hat{N}\hat{t}_{wz\pi} - \hat{t}_{w\pi}\hat{t}_{z\pi}}{\hat{N}\hat{t}_{w^2\pi} - \hat{t}_{w\pi}^2} \left( \frac{t_w}{N} - \frac{\hat{t}_{w\pi}}{\hat{N}} \right) \right]$$

$$\hat{C}_3 = \hat{t}_{y\pi} / \hat{t}_{zra}$$

$$= \hat{t}_{y\pi} / \left[ \begin{array}{c} t_w \hat{t}_{z\pi} \\ \hat{t}_{w\pi} \end{array} \right]$$

$$\hat{C}_4 = \hat{t}_{y\pi} / \hat{t}_{zr}$$

$$= \hat{t}_{y\pi} / \left[ \frac{\hat{t}_{z\pi}}{\hat{N}} + \frac{\hat{N}\hat{t}_{wz\pi} - \hat{t}_{w\pi}\hat{t}_{z\pi}}{\hat{N}\hat{t}_{w^2\pi} - \hat{t}_{w\pi}^2} \left( \frac{t_w}{N} - \frac{\hat{t}_{w\pi}}{\hat{N}} \right) \right]$$

$$\hat{C}_5 = \hat{t}_{yra} / \hat{t}_{z\pi}$$

$$= \left[ \begin{array}{c} \hat{t}_{y\pi} \\ t_x \hat{t}_{x\pi} \end{array} \right] / \hat{t}_{z\pi}$$

$$\hat{C}_6 = \hat{t}_{yr} / \hat{t}_{z\pi}$$

$$= \left[ \frac{\hat{t}_{y\pi}}{\hat{N}} + \frac{\hat{N}\hat{t}_{xy\pi} - \hat{t}_{x\pi}\hat{t}_{y\pi}}{\hat{N}\hat{t}_{x^2\pi} - \hat{t}_{x\pi}^2} \left( \frac{t_x}{N} - \frac{\hat{t}_{x\pi}}{\hat{N}} \right) \right] / \hat{t}_{z\pi}$$

$$\hat{C}_7 = \hat{t}_{y\pi} / \hat{t}_{z\pi}$$

Cada uno de los cocientes anteriores es una función de totales; así, el análisis de su precisión es necesario hacerlo mediante la aproximación de Taylor (Särndal et al. 1993).

Calculando las derivadas parciales de los cocientes dados anteriormente, respecto a cada uno de los pi-estimadores para totales y evaluando cuando el vector de los estimadores de totales es igual al vector de totales, se obtienen las siguientes transformaciones  $u_k$  para los conglomerados (municipios):

- $u_{k1} = u_{k\pi} - \frac{t_y}{t_z} \left( \frac{x_k}{t_x} - \frac{w_k}{t_w} \right)$
- $u_{k2} = u_{k\pi} - \frac{1}{t_z} \left[ \left( x_k - \frac{t_x}{N} \right) \left( \frac{Nt_{xy} - t_x t_y}{Nt_{x2} - t_x^2} \right) - \left( \frac{t_y}{t_z} \right) \left( w_k - \frac{t_w}{N} \right) \left( \frac{Nt_{zw} - t_z t_w}{Nt_{w2} - t_w^2} \right) \right]$
- $u_{k3} = u_{k\pi} + \frac{t_y}{t_z} \frac{w_k}{t_w}$
- $u_{k4} = u_{k\pi} + \frac{t_y}{t_z} \left[ \frac{1}{N} - \left( \frac{1}{t_z} \right) \left( w_k - \frac{t_w}{N} \right) \left( \frac{Nt_{zw} - t_z t_w}{Nt_{w2} - t_w^2} \right) \right]$
- $u_{k5} = u_{k\pi} - \frac{t_y}{t_z} \frac{x_k}{t_x}$
- $u_{k6} = u_{k\pi} + \frac{t_y}{Nt_z} - \left( \frac{1}{t_z} \right) \left( x_k - \frac{t_x}{N} \right) \left( \frac{Nt_{xy} - t_x t_y}{Nt_{x2} - t_x^2} \right)$
- $u_{k\pi} = \frac{1}{t_z} \left( y_k - z_k \frac{t_y}{t_z} \right)$

De forma análoga se obtienen las fórmulas necesarias para evaluar el desempeño de los estimadores propuestos para el caso del favoritismo por el candidato Serpa o por el candidato el candidato Gaviria del partido Polo Democrático.

## 4. Aplicación a datos electorales 2006

Se realiza una partición de los municipios según los valores absolutos de las transformaciones obtenidas mediante el método de aproximación de Taylor para el cociente de pi-estimadores:

- Grandes ciudades: inclusión forzosa de 21 municipios en la primera etapa.
- Municipios intermedios: M.A.S. sobre 758 municipios.
- Municipios muy pequeños: M.A.S. sobre 241 municipios.

El universo de estudio es el estrato intermedio conformado por las ciudades intermedias y los municipios pequeños (758 municipios). En primera medida, se analiza la conveniencia del uso de los resultados de las elecciones en el 2002 (RNEC 2007a) como variables auxiliares que expliquen el comportamiento de los resultados en el 2006 (RNEC 2007b). En los gráficos de dispersión para el cruce de cada par de variables se puede observar la relación que existe.

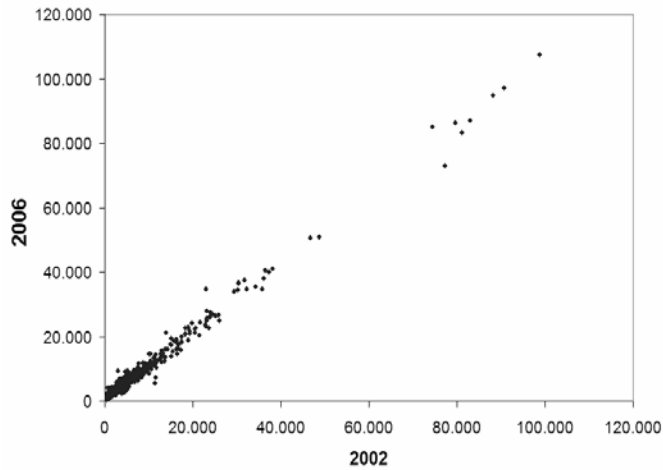


Figura 1: *Dispersión votos municipales por el candidato Uribe*

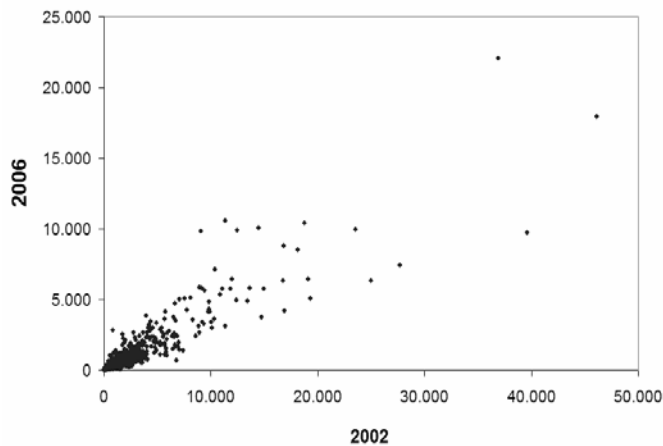


Figura 2: *Dispersión votos municipales por el candidato Serpa*

Para realizar la estimación de los totales para el 2006, se utilizan GREG-estimadores para totales usando los modelos heterocedástico sin intercepto y homocedástico con intercepto. En todo caso, sin importar el total que se desee estimar, ni el estimador que se utilice (de los seis propuestos), se encuentra que la ganancia en precisión para la estimación de totales, al usar GREG-estimadores, es considerable frente a la precisión del pi-estimador.

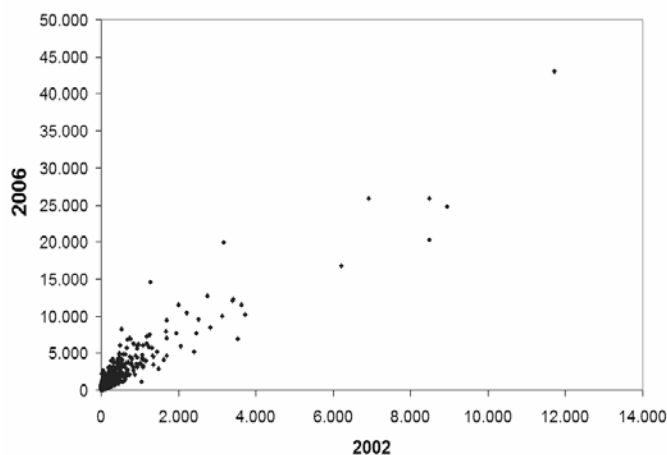


Figura 3: *Dispersión votos municipales por candidato del Polo*

Al tomar un tamaño muestral de 20 municipios, se pasa de un coeficiente de variación de 34% a un 4% al usar Greg-estimadores para estimar el total de votos; de 34% a un 12% usando el modelo sin intercepto y a un 10% usando el modelo con intercepto, cuando se estima el total de votos por el candidato Uribe. En la estimación del total de votos por el candidato Uribe el estimador más preciso es el que usa el modelo homocedástico con intercepto.

Para el candidato Serpa, la precisión de los Greg-estimadores considerados para estimar el total de votos por el candidato Serpa coincide, permitiéndose en ambos casos pasar de un coeficiente de variación de 38% para una muestra de 20 municipios a un coeficiente de 16%.

En la estimación del número total de votos por el candidato Gaviria el estimador más preciso resulta ser el que utiliza el modelo homocedástico con intercepto. Así, se logra pasar de un coeficiente de variación de 44% a un 15% con un tamaño muestral de 20 municipios.

En general, se observa que los GREG-estimadores son considerablemente más precisos que el pi-estimador en la estimación de totales. En particular, se observa que con el modelo homocedástico con intercepto se obtienen mejores resultados en lo que se refiere a precisión del estimador.

Se espera que si el uso de la información auxiliar mejora la precisión para la estimación de los totales, se asegure un comportamiento similar para la estimación de cocientes.

Al hacer un análisis sobre el coeficiente de variación de cada uno de los estimadores propuestos, probando las fórmulas obtenidas con los datos censales del estrato in-



termedio y para cada uno de los candidatos mencionados, se obtienen los resultados de las Figuras 4, 5 y 6.

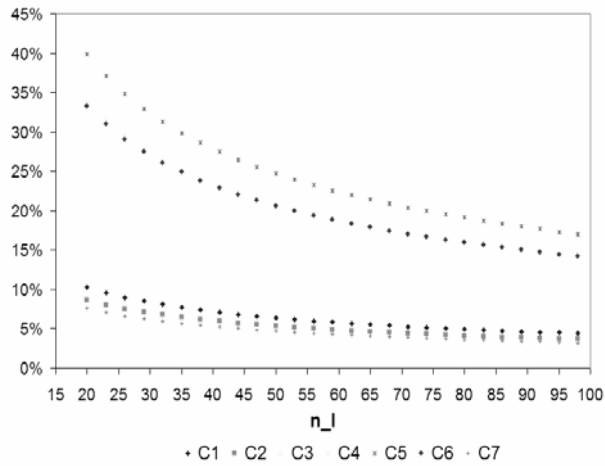


Figura 4: *Coefficiente de variación por estimador para la tasa de favoritismo por el candidato Uribe*

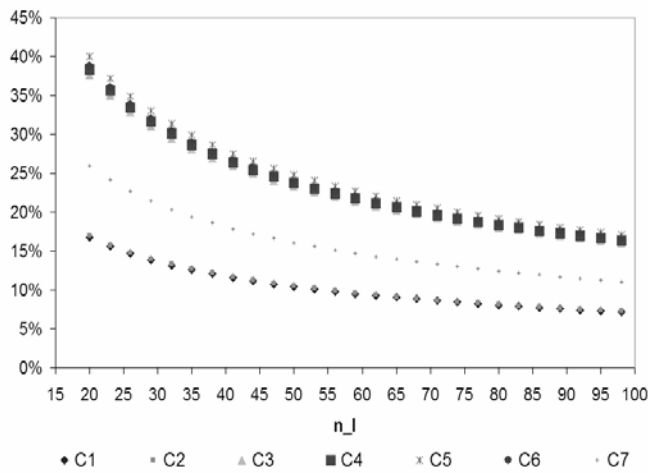


Figura 5: *Coefficiente de variación por estimador para la tasa de favoritismo por el candidato Serpa*

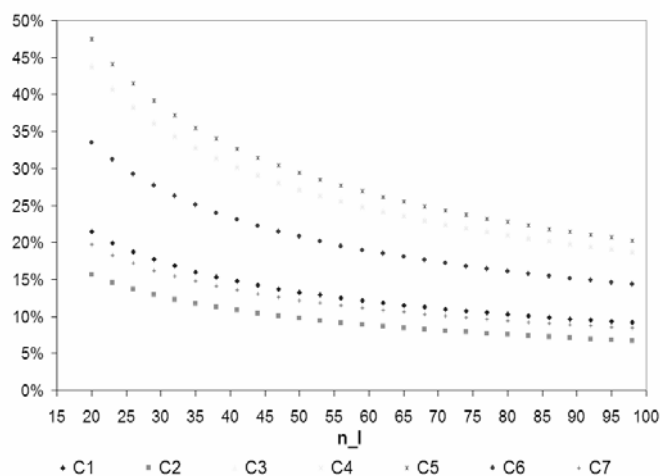


Figura 6: *Coefficiente de variación por estimador para la tasa de favoritismo por el candidato Gaviria*

El coeficiente de variación muestra que, en el caso de el candidato Uribe, los estimadores asistidos por regresión son menos precisos que el estimador que utiliza únicamente los pi-estimadores para totales. Para dicho estimador se obtienen resultados aceptables con un tamaño de muestra de alrededor de 55 municipios.

En el caso del candidato Serpa se observa que hay una mejora utilizando GREG-estimadores en el numerador y en el denominador, sin importar cuál sea el modelo de regresión ajustado; mientras que en el caso de el candidato Gaviria sólo se alcanza una mejora considerable del coeficiente de variación para tamaños de muestra demasiado pequeños. Según éstos resultados, en general, el uso de GREG estimadores no mejoró la precisión en comparación al estimador tradicional (sin usar regresión). Sin embargo, dado que todos estos estimadores son sesgados, es necesario realizar la comparación no sólo en términos de precisión, sino además en términos de confiabilidad.

La confiabilidad de cada estimador debe ser calculada a partir del universo de todas las posibles muestras que arroja el diseño, por lo cual se utiliza el método de simulación de Monte Carlo (Särndal et al. 1993), seleccionando así 5000 muestras de tamaño 55. Para cada una de las 5000 muestras se calcula la estimación y la variación estimada, logrando con esto un intervalo de confianza estimado para el parámetro. Así, se puede determinar el promedio de las estimaciones como una aproximación a la esperanza del estimador, la varianza de las estimaciones y el promedio de las varianzas, como estimadores de la varianza del estimador. Adicionalmente se mira en qué porcentaje los intervalos de confianza cubren al parámetro; esto es una aproximación a la confiabilidad del estimador.

Además, se calcula el sesgo absoluto del estimador como la diferencia entre el promedio de los estimativos y el parámetro, y el sesgo relativo como la comparación del sesgo relativo respecto al parámetro.

El procedimiento anterior se realiza para cada uno de los estimadores propuestos y para cada uno de los candidatos. Los resultados de la consulta electoral del 2006 en los municipios del estrato en que se concentra la estimación fueron 60.59 % para el candidato Uribe, 14.74 % para el candidato Serpa y 21.43 % para el candidato Gaviria.

Los resultados obtenidos con la simulación de Monte Carlo son los mostrados en las siguientes tablas:

$\hat{C}$	$\tilde{c}$ (%)	$S_c^2$	Tasa de cobertura (%)	Sesgo Absoluto (%)	$\frac{100\tilde{c}-C}{C}$ (%)
sin int/sin int	60.95	0.0013	91.60	0.36	0.6
con int/con int	60.47	0.0009	87.74	0.13	0.2
pi-est/sin int	59.69	0.0158	88.92	0.91	1.5
pi-est/con int	59.71	0.0143	88.78	0.88	1.5
sin int/pi-est	64.85	0.0229	84.00	4.26	7.0
con int/pi-est	63.69	0.0153	84.20	3.09	5.1
pi-est/pi-est	60.67	0.0007	92.24	0.07	0.1

Tabla 1: *Precisión y Confiabilidad. Favoritismo por el candidato Uribe*

El estimador que no utiliza regresión es el que mejor desempeño tiene para la estimación de la tasa de favoritismo por el candidato Uribe, tanto en términos de precisión como en términos de confiabilidad (cobertura).

El sesgo obtenido al utilizar Greg-estimadores no es importante en la mayoría de los casos, excepto cuando se utiliza sólo en el numerador.

Utilizando greg-estimadores, aunque no se obtienen resultados mejores a los obtenidos con la razón de pi-estimadores, se observa que el que mejor desempeño lo tiene el estimador que utiliza el modelo de regresión homocedástico tanto en el numerador como en el denominador.

$\hat{C}$	$\tilde{c}$ (%)	$S_c^2$	Tasa de cobertura (%)	Sesgo Absoluto (%)	$\frac{100\tilde{c}-C}{C}$ (%)
sin int/sin int	14.75	0.0002	91.36	0.01	0.1
con int/con int	14.87	0.0002	87.10	0.13	0.8
pi-est/sin int	14.48	0.0012	90.54	0.26	1.8
pi-est/con int	14.49	0.0011	90.50	0.25	1.7
sin int/pi-est	15.74	0.0013	85.40	1.00	6.8
con int/pi-est	15.81	0.0014	85.40	1.07	7.2
pi-est/pi-est	14.86	0.0005	91.46	0.12	0.8

Tabla 2: *Precisión y Confiabilidad. Favoritismo por el candidato Serpa*

En este caso, el mejor estimador resulta ser el estimador asistido por un modelo de regresión heterocedástico sin intercepto en el numerador y en el denominador, ya que su sesgo y su varianza son los menores. Sin embargo, se observa que el estimador que no utiliza regresión tiene sesgo y varianza ligeramente mayores; de esta manera, al tener mayor varianza aumenta el rango de cobertura de los intervalos de confianza, haciendo que la confiabilidad del estimador sea mayor.

El sesgo aumenta en los estimadores que usan GREG-estimadores únicamente en el numerador.

$\hat{C}$	$\hat{c}$ (%)	$S_{\hat{c}}^2$	Tasa de cobertura (%)	Sesgo Absoluto (%)	$\frac{100\hat{c}-C}{\hat{C}}$ (%)
sin int/sin int	21.95	0.0008	84.68	0.52	2.4
con int/con int	21.98	0.0007	82.40	0.55	2.5
pi-est/sin int	21.13	0.0033	87.80	0.30	1.4
pi-est/con int	21.13	0.0031	87.82	0.30	1.4
sin int/pi-est	23.47	0.0042	79.62	2.04	9.5
con int/pi-est	23.18	0.0029	80.88	1.75	8.2
pi-est/pi-est	21.30	0.0006	90.60	0.13	0.6

Tabla 3: *Precisión y Confiabilidad. Favoritismo por el candidato Gaviria*

Entre los estimadores asistidos por regresión, el estimador que utiliza el modelo homocedástico tanto en numerador como en denominador es el que mejor se comporta en cuanto a precisión. Sin embargo, al compararlo con el estimador  $\hat{C}_7$ , se observa que su sesgo es mayor, por lo que finalmente la confiabilidad hace que se prefiera a éste último sobre el primero. Entre los estimadores que usan GREG-estimadores, los menos sesgados son los que lo hacen en el denominador. En todos los casos se observa que el desempeño de los GREG-estimadores que sólo utilizan regresión en el numerador o en el denominador hace que éstos sean obsoletos para la estimación de las tasas de favoritismo; esto es, en caso de utilizar regresión para la estimación de la tasa de favoritismo, los resultados dan evidencia de que sería más conveniente utilizar la regresión tanto en el numerador como en el denominador. Se cree que el hecho de que el estimador que no utiliza regresión resulte mejor que los asistidos por regresión puede deberse a la presencia de cambios bruscos en las tasas de favoritismo para cada municipio. Aunque hay un buen ajuste en la explicación de la cantidad de votos en el 2006 por la del 2002 (tanto para cada partido como en el total), en el proceso construcción de los estimadores se ignoran las relaciones entre las tasas de favoritismo en cada periodo y municipio.

Lo anterior se puede observar en la Figura 7, en donde se muestra que una gran cantidad de municipios en los que en el 2002 hubo una baja tasa de favoritismo por el candidato Uribe, en el 2006 presentaron una tasa mayor. En algunos casos como en el Municipio de Chicas en Boyacá, se pasa de un 5 % (99 votos) a un 91 % (1493 votos) de favoritismo por el candidato Uribe (casos similares son Chorrito y Silos en Santander, Chitagá en Norte de Santander, Topaipi en Cundinamarca), o como el caso de El Carmen de Bolívar, donde pasa de un 24 % (2386) a un 78 % (11488

votos) en 2006 (casos similares son Agustín Codazzi y Bosconia en el Cesar, Paz de Ariporo en Casanare, María la baja, Achí y Aracataca en Bolívar, y Ayapel en Córdoba).

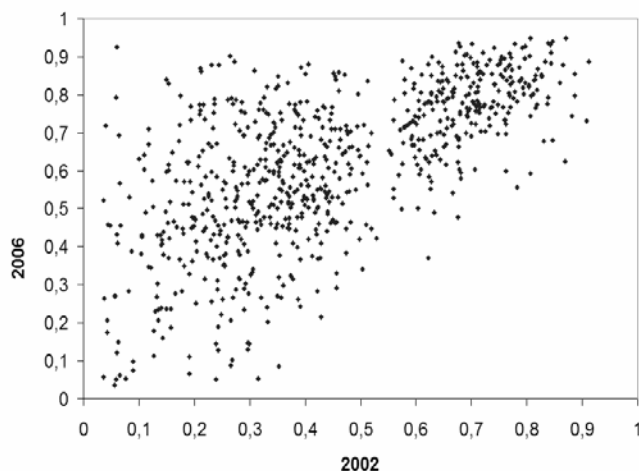


Figura 7: *Tasa de favoritismo municipal por el candidato Uribe*

Por otra parte, alrededor del 83% de los municipios de este estrato presenta un aumento en la tasa de favoritismo por el candidato Uribe en el 2006 con respecto al 2002. Los departamentos con mayor número de municipios en los que el aumento es mayor al 40% son: 10 en Bolívar, 7 en Boyacá, 6 en Casanare, 11 en el Cesar, 6 en el Chocó y 13 en Cundinamarca.

En cuanto a el candidato Serpa, se tiene que en alrededor del 98% de los municipios el favoritismo por el candidato Serpa decrece, como el caso del municipio de Chiscas, donde se pasa de un 90% (1493 votos) a un 4% (78) en el 2006. Un comportamiento similar presentan los municipios de Cerrito, María la baja, Chitagá, Silos y Coloso. En otros municipios como Montería se pasa de un 51% (46.126 votos) a un 18% (17.923 votos), de manera similar esto se presenta en los municipios de Valledupar, Soacha, Floridablanca, Neiva, Palmira, Maicao, Pasto, Riohacha, Tunja y Sogamoso, donde las diferencias en votos son mayores a 10.000.

En alrededor de un 97% de los municipios se presenta un aumento en el favoritismo por el candidato del Polo; aunque dichos aumentos en la mayoría de los casos, no son contundentes. Los cambios más notorios se concentran en el Departamento de Nariño (Cumbitara, Policarpa, Santacruz, Mallama, Linares, Sapuyes, Samaniego y Guaitarilla), donde el aumento es superior al 50% (superior a 1060 votos y menor a 4300), en Riohacha donde la diferencia es de 51% (13324 votos), y en ciudades intermedias como Pasto, Palmira, Popayán, Valledupar, Neiva y Tunja, donde el aumento está entre 12 y 28% con una diferencia de votos entre 10000 y 32000.

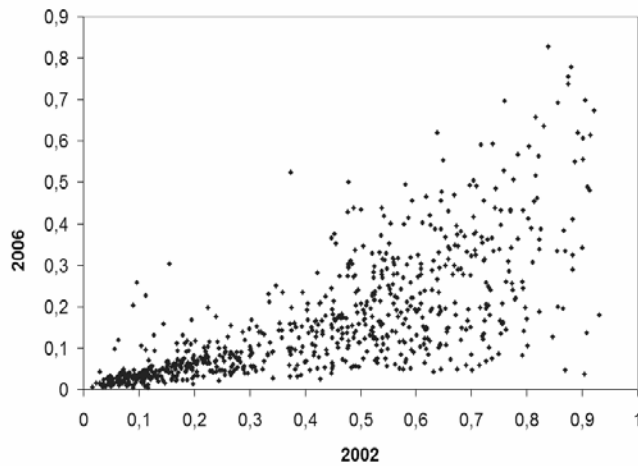


Figura 8: *Tasa de favoritismo municipal por el candidato Serpa*

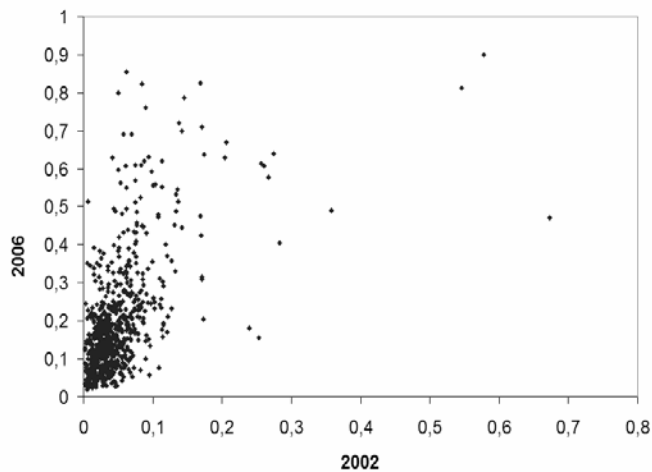


Figura 9: *Tasa de favoritismo municipal por el candidato del Polo*

Por lo anterior, se decide conformar dos estratos en donde los cambios en las tasas de favoritismo por el candidato Uribe del 2002 al 2006 tengan una relación similar.

Los municipios en los que se propone el uso de GREG-estimadores quedan particionados en tres estratos. Al estrato superior pertenecen los 342 municipios que se encuentran por encima de la línea de tendencia del gráfico de dispersión de las tasas de favoritismo por el candidato Uribe en 2002 y 2006 y por debajo del límite

que considera 40 veces la desviación estándar del intercepto (donde el favoritismo por el candidato Uribe resulta ser de 70.73 %). De manera análoga, se conforma el estrato inferior con 320 municipios. En este estrato el favoritismo por el candidato Uribe alcanza el 55.76 %. La conformación de los estratos es la que se muestra en la Figura 10.

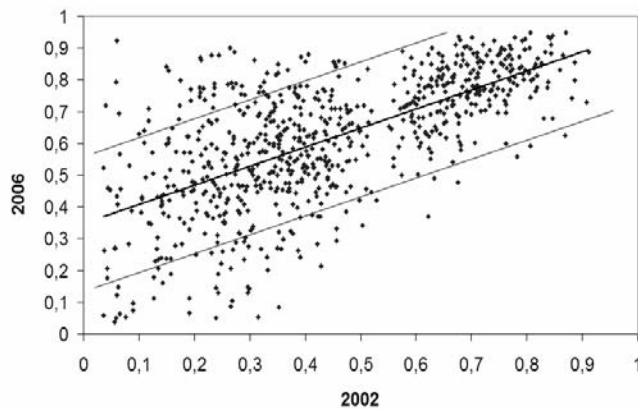


Figura 10: *Estratificación. Tasa de favoritismo municipal por el candidato Uribe*

Con esta estratificación se logra un comportamiento similar en las tasas de favoritismo por el candidato Uribe. Los resultados de la precisión de los estimadores en cada uno de dichos estratos son los que se muestran en las Figuras 11 y 12.

Se encuentra que en la estratificación no se presentan mejoras en el coeficiente de variación, excepto en el estrato inferior donde se mejora la precisión, aunque no de forma considerable. Para el estimador de modelo homocedástico, en el numerador y el denominador se necesita una muestra de tamaño 48 en el estrato inferior para alcanzar un cv de 3% y una de 22 para un 5%; mientras que el estimador de cociente entre pi-estimadores necesita muestras de 60 y 28 respectivamente, lo que puede hacer la diferencia en términos de costos. El desempeño que los estimadores muestran en términos de sesgo y precisión, determinado por medio de la simulación de Monte Carlo seleccionando muestras de tamaño 25 en el estrato superior y de tamaño 35 en el inferior se encuentra a continuación.

En este estrato, el estimador más preciso y con menor sesgo es el que no recurre al uso de regresión. El estimador  $\hat{C}_1$  es el único que lo supera en términos de confiabilidad debido a que su sesgo relativo es menor y sus intervalos de confianza. Aunque en este estrato hay un comportamiento relativamente homogéneo con respecto a los cambios en la tasa de favoritismo por el candidato Uribe, se muestra que dicha homogeneidad no mejora el uso de los estimadores asistidos por regresión.

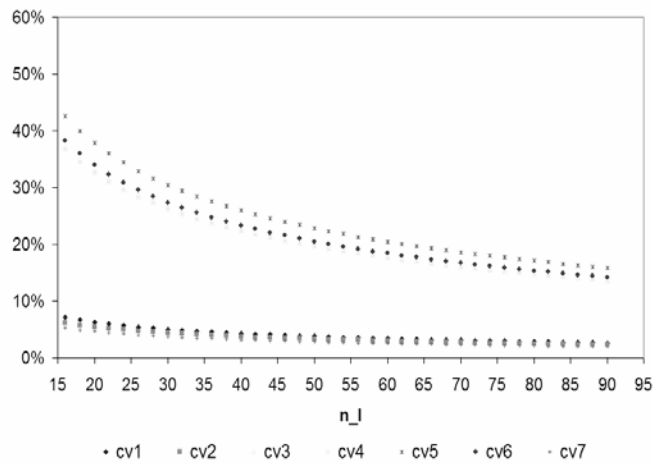


Figura 11: *Coeficiente de variación. Estrato Superior*

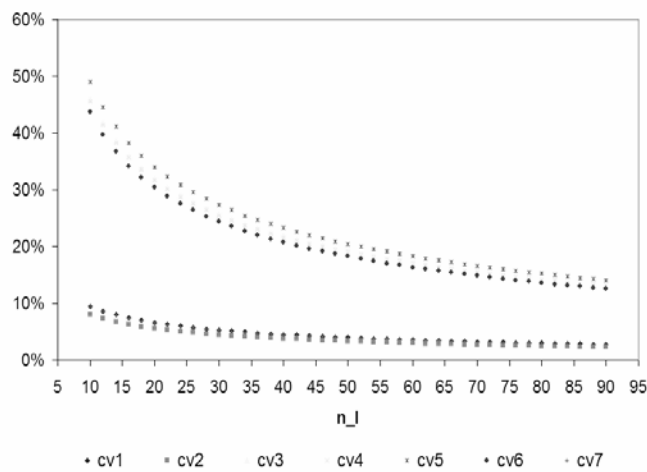


Figura 12: *Coeficiente de variación. Estrato Inferior*

Los estimadores  $\hat{C}_1$  y  $\hat{C}_2$  son al menos tan precisos como  $\hat{C}_7$ ; sin embargo, su sesgo es mayor y su confiabilidad es menor; lo que los hace inadmisibles.

Al parecer, los cambios presentes en la tasa de favoritismo para el candidato el candidato Uribe inciden en el comportamiento de los GREG-estimadores, pero la estratificación realizada buscando homogeneidad en los cambios, no genera ventajas considerables sobre el desempeño de los estimadores.



$\hat{C}$	$\bar{\hat{c}}$ (%)	$S_{\hat{c}}^2$	Tasa de cobertura (%)	Sesgo Absoluto (%)	$\frac{100\bar{\hat{c}}-C}{\bar{C}}$ (%)
$\hat{C}_1$	71.08	0.0019	92.58	0.35	0.5
$\hat{C}_2$	69.46	0.0017	82.84	1.27	1.8
$\hat{C}_3$	67.88	0.0465	80.50	2.85	4.0
$\hat{C}_4$	67.76	0.0412	79.98	2.97	4.2
$\hat{C}_5$	81.44	0.0581	72.28	10.71	15.1
$\hat{C}_6$	78.41	0.0392	71.80	7.68	10.9
$\hat{C}_7$	71.12	0.0009	88.20	0.39	0.6

Tabla 4: *Precisión y Confiabilidad. Favoritismo por el candidato Uribe. Estrato Superior*

$\hat{C}$	$\bar{\hat{c}}$ (%)	$S_{\hat{c}}^2$	Tasa de cobertura (%)	Sesgo Absoluto (%)	$\frac{100\bar{\hat{c}}-C}{\bar{C}}$ (%)
$\hat{C}_1$	56.10	0.0007	88.84	0.34	0.6
$\hat{C}_2$	55.88	0.0006	85.76	0.11	0.2
$\hat{C}_3$	54.18	0.0171	88.96	1.59	2.9
$\hat{C}_4$	54.20	0.0167	88.80	1.57	2.8
$\hat{C}_5$	61.18	0.0242	81.20	5.41	9.7
$\hat{C}_6$	60.51	0.0192	81.16	4.74	8.5
$\hat{C}_7$	55.73	0.0008	90.96	0.03	0.1

Tabla 5: *Precisión y Confiabilidad. Favoritismo por el candidato Uribe. Estrato Inferior*

## 5. Conclusiones

No hay ventajas importantes al utilizar el cociente de GREG-estimadores frente al cociente de pi-estimadores, para el caso específico de estimación de la tasa de favoritismo por el candidato el candidato Uribe en el periodo de elecciones 2006.

Si se trata de utilizar el cociente de GREG-estimadores, entonces éstos deben estar tanto en el numerador como en el denominador.

El uso de GREG-estimadores no mejora la eficiencia de cocientes de totales aunque la bondad de ajuste de los modelos sea muy buena para la estimación de totales.

Este resultado con base en datos reales pero muy particulares, muestra que para cocientes y quizá para otra función de totales, el comportamiento de estimadores asistidos por modelos no tiene las ventajas que se obtienen cuando se estiman totales.

## Agradecimientos

A quien una vez dijo:

...Suponiendo que Dios existe, que cuando me muera vaya al cielo y pueda hablar con Él, y que adicionalmente Él me conteste lo que le pregunte, le preguntaría: ¿existe un estimador insesgado para cocientes?, ¿cuál es? y ¿cómo es su varianza?

Gracias profesor Bautista<sup>1</sup>.

**Recibido: 5 de febrero de 2008**  
**Aceptado: 25 de marzo de 2009**

## Referencias

- Bautista, L. (2005), 'Estrategia de muestreo para la estimación de la tasa de favoritismo en la elección presidencial', *Revista Colombiana de Estadística* **28**(1), 39–62.
- RNEC (2007a), *Elecciones presidenciales de 2002 en Colombia*, Registraduría Nacional del Estado Civil.
- RNEC (2007b), *Elecciones presidenciales de 2006 en Colombia*, Registraduría Nacional del Estado Civil.
- 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, Sweden.

---

<sup>1</sup>Este artículo es resultado del trabajo de grado para optar por el título de estadístico titulado *Estrategia de muestreo usando estimadores de regresión generalizada para la estimación de tasas de favoritismo en elecciones presidenciales en Colombia*, dirigido por el maestro Leonardo Bautista y presentado en la Universidad Nacional de Colombia en 2007.