Padilla, Alberto

Working Paper

Cotas para la varianza, efecto del diseño y coeficiente de variación de proporciones en el muestreo por conglomerados en dos etapas con tamaños iguales

Working Papers, No. 2014-13

Provided in Cooperation with:
Bank of Mexico, Mexico City

Suggested Citation: Padilla, Alberto (2014) : Cotas para la varianza, efecto del diseño y coeficiente de variación de proporciones en el muestreo por conglomerados en dos etapas con tamaños iguales, Working Papers, No. 2014-13, Banco de México, Ciudad de México

This Version is available at:
https://hdl.handle.net/10419/100132

Banco de México

Documentos de Investigación

Banco de México

Working Papers

N° 2014-13

Cotas para la Varianza, Efecto del Diseño y Coeficiente de Variación de Proporciones en el Muestreo por Conglomerados en Dos Etapas con Tamaños Iguales

Alberto Padilla
Banco de México

Junio 2014

Documento de Investigación
2014-13

Working Paper
2014-13

Cotas para la Varianza, Efecto del Diseño y Coeficiente de Variación de Proporciones en el Muestreo por Conglomerados en Dos Etapas con Tamaños Iguales*

Alberto Padilla†
Banco de México

Resumen

En el problema de estimación de proporciones en el muestreo aleatorio simple, se emplea el valor de la varianza máxima para el cálculo del tamaño de muestra, en caso de no contar con información acerca de la característica por estimar. En este trabajo se extiende dicho resultado a la estimación de proporciones en el muestreo por conglomerados en dos etapas con tamaños iguales, exhibiendo la expresión para la varianza máxima. Como resultado de esto, se construyen cotas para el efecto del diseño y el coeficiente de variación del estimador de proporciones. Se ilustrará con algunos ejemplos el empleo de estas cotas.
Palabras Clave: Varianza máxima; Tamaño de muestra; Efecto del diseño; Coeficiente de variación.

Abstract

In the estimation of proportions using simple random sampling, the maximum value of the variance can be used to compute the sample size when there is no information of the variable of interest. We extend this result to the estimation of proportions under two-stage cluster sampling with equal sizes, showing the expression for the maximum variance. As a by-product it is immediate to obtain bounds for the design effect and the coefficient of variation of the proportion estimator. Some examples are given related to the computation of the bounds.
Keywords: Maximum variance; Sample size; Design effect; Coefficient of variation.
JEL Classification: C80; C83.

*El autor agradece a los participantes del seminario del Banco de México, así como a dos revisores del Banco de México por sus comentarios y sugerencias.
†Dirección General de Investigación Económica. Email: ampadilla@banxico.org.mx.

1. INTRODUCCIÓN

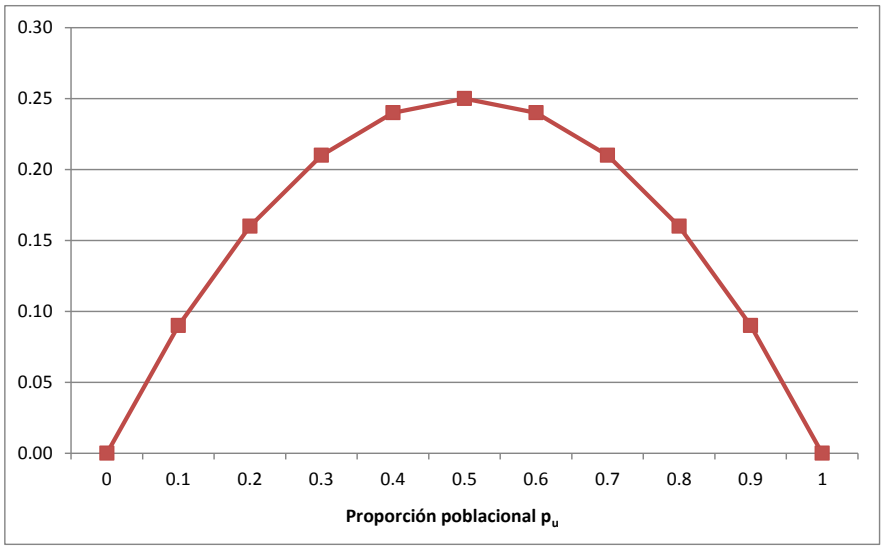
En el cálculo del tamaño de muestra se usa con frecuencia la fórmula asociada al muestreo aleatorio simple, *mas,* y, posteriormente, ésta se ajusta por el efecto del diseño, *efd,* propuesto por Kish (1965). El efecto del diseño se define como el cociente de la varianza de un estimador, bajo un diseño muestral diferente del muestreo aleatorio simple, y la varianza de dicho estimador bajo muestreo aleatorio simple. El cálculo del efecto del diseño requiere del conocimiento de dos varianzas, es decir, de dos cantidades poblacionales. Por otra parte, el efecto del diseño también se emplea como referencia para evaluar la pérdida o ganancia en eficiencia del estimador de un diseño muestral diferente al muestreo aleatorio simple.

En el caso del cálculo del tamaño de muestra para la estimación de proporciones y cuando no se cuenta con información de la característica de interés, puede emplearse el valor máximo de la varianza para el estimador de proporciones bajo *mas*, el cual se alcanza cuando la proporción poblacional adquiere el valor de 0.5, Cochran (1986). Después se aplica un ajuste usando el efecto del diseño, $n = n_{mas} efd(\hat{p})$. En esta expresión $n_{mas} = N' pq/(d^2 + pq/(N-1))$, en la que $N' = N/(N-1)$ y $d^2 = e^2/t^2$, donde N es el número de elementos de la población de interés, e se refiere al error de estimación absoluto, t el valor asentado en tablas de la distribución normal estándar para una confianza prefijada, $q = 1 - p$ y p es una estimación anticipada de p_U, la cual es el valor poblacional que se desea estimar. Esto conduce al tamaño de muestra más grande para una población, error de estimación absoluto y nivel de confianza dados.

El tamaño máximo de muestra para la estimación de una proporción en el *mas* se alcanza por la forma cóncava de la varianza poblacional de la proporción. La fórmula de dicha varianza es $V(\hat{p}) = \frac{N' p_U q_U}{n}(1 - \frac{n}{N})$, donde N', p_U y q_U son como en el párrafo anterior. Esta expresión adquiere el valor máximo cuando $p_U = 1/2$, con n fijo. Al graficar los valores de $p_U\, q_U$ se observa que el valor máximo se tiene con $p_U = 1/2$.

Gráfica 1

Valores de $p_U q_U$



De la gráfica se aprecia que los valores de la varianza son similares en el rango 0.4 a 0.6, por lo cual, en la determinación del tamaño de muestra para *mas*, conviene calcular el tamaño de muestra para diversos valores en dicho rango, cuando no se tenga información de la variable de interés o se conjeture que la proporción poblacional por estimar se encuentra alrededor de los valores mencionados.

Esta propiedad de la varianza para la estimación de proporciones usando *mas* motivó la búsqueda de resultados de este tipo para el muestreo por conglomerados en dos etapas con tamaños iguales, *mc2e,* y se encontró que se han propuesto algunas cotas para los diseños con elección de elementos o conglomerados de primera etapa con probabilidad proporcional al tamaño y con reemplazo, Scott & Smith (1975) y Chaudhuri & Stenger (2005). Estos esquemas han sido estudiados en la literatura; empero, no proporcionan en general resultados que puedan ser empleados con relativa facilidad en la práctica para la etapa de planeación de una encuesta, ya que se basan en el uso de cantidades que no se pueden determinar fácilmente. Como un ejemplo de esto, en el artículo de Scott & Smith (1975) se menciona que el número de conglomerados sea grande o que éstos no varíen mucho en tamaño. Además, se refieren a un diseño distinto al *mc2e*. Con base en la revisión efectuada de la literatura, no se encontraron resultados similares de cotas para el *mc2e*.

Por lo anterior, en este artículo se desarrollan cotas para la varianza de la estimación de proporciones en el muestreo por conglomerados en dos etapas con tamaños iguales y empleando muestreo aleatorio simple en ambas etapas. Una consecuencia de contar con cotas para dicha varianza es que resulta inmediato obtener cotas para el efecto del diseño de dicho esquema muestral, así como para el coeficiente de variación de la proporción estimada. Por otra parte, las fórmulas son sencillas de calcular y únicamente se requieren los elementos de información con los que normalmente se cuenta en la práctica en la etapa de diseño muestral. Las cotas se obtienen al expresar la varianza del estimador de proporciones en el *mc2e* de una manera tal que se aíslan los valores de las proporciones dentro y entre conglomerados de las cantidades relativas al número de conglomerados y

elementos dentro de conglomerados en población y muestra. Esta expresión también permite calcular fácilmente los valores de la varianza del estimador de proporciones en el *mc2e* con diferentes de tamaños de muestra. Se hicieron dos hallazgos importantes, los cuales se mostrarán en los ejemplos, uno de ellos se refiere a la relación entre la varianza y el coeficiente de correlación intraclase. Se encontraron casos en los que la varianza del estimador de proporciones permanece sin cambio o decrece conforme el coeficiente de correlación intraclase aumenta. El otro se refiere a los valores que toma el coeficiente de correlación intraclase, se muestran casos en los que dicha cantidad no siempre alcanza los valores mínimo y máximo.

El artículo se encuentra organizado de la siguiente manera. En la sección 2 se proporciona un breve panorama del muestreo probabilístico, definiciones, notación y la expresión de varianza para el muestreo por conglomerados en dos etapas, también conocido como bietápico. Algunos aspectos de la correlación intraclase y la varianza se encuentran en la sección 3. Las cotas para la varianza, efecto del diseño y coeficiente de variación, junto con varios ejemplos, así como una aplicación de dichas cotas en el cálculo del tamaño de muestra para el muestreo por conglomerados se ilustra en la sección 4.

Es importante hacer notar que los desarrollos que se presentan se refieren a la etapa de planeación en un diseño muestral, en particular, al momento de determinar el tamaño de muestra y no se aborda el tema de la estimación.

2. DEFINICIONES Y NOTACIÓN

En este trabajo se emplea el enfoque del denominado muestreo basado en el diseño para poblaciones finitas, que es otra forma en la que se denomina al muestreo probabilístico. Para una exposición detallada véase Särndal et al. (1992).

2.1 Algunos puntos generales acerca del muestreo probabilístico.

En el muestreo probabilístico, el problema básico consiste en estimar una variable de interés de una población finita, como podría ser estimar el gasto medio en alimentos por hogar en una ciudad. Si se tuviesen recursos suficientes para levantar un censo de todos los hogares de la ciudad en cuestión, se podría calcular dicho gasto y no habría necesidad de recurrir al muestreo. En este ejemplo, el gasto es lo que se conoce como una cantidad poblacional. En muchas situaciones no es factible levantar un censo, entonces se recurre a la extracción de una muestra para estimar la cantidad poblacional. La forma de seleccionar la muestra se conoce como diseño muestral y entre los principales diseños se encuentran los siguientes: el muestreo aleatorio simple, el muestreo aleatorio estratificado, el muestreo por conglomerados, el muestreo sistemático, el muestreo con probabilidades proporcionales a alguna medida de tamaño, entre otros. Para más detalle de estos y otros diseños muestrales usados en la práctica, véase Särndal et al. (1992). Por otra parte, para cada diseño muestral se tiene una expresión matemática particular del estimador de la cantidad poblacional de interés, por ejemplo, en el caso del muestreo aleatorio simple se emplea el promedio aritmético muestral como un estimador del correspondiente promedio poblacional y, como se está empleando un estimador, se construye una fórmula para la varianza de dicho

estimador. La varianza de un estimador es una cantidad poblacional, es decir, depende de cantidades que pueden calcularse al medir todos los elementos de la población de interés y es la que se emplea en la obtención de fórmulas para el cálculo del tamaño de muestra. Por otro lado, al trabajar con datos provenientes de una muestra, para cada diseño muestral, se construye un estimador de la varianza y es el que se emplea para evaluar la precisión del estimador.

Sin pretender abarcar toda la gama de posibilidades que comprende el diseño, ejecución y análisis de una encuesta, a continuación se mencionan las principales etapas en la realización de una encuesta o un plan de muestreo, véase Cochran (1986).

a) Definición de los objetivos de la encuesta o investigación

b) Definición de la población objetivo y población por muestrear

c) Grado de precisión deseado para las variables de interés por estimar

d) Método de medición o método para obtener los datos de la encuesta

e) Marco(s) muestral(es)

f) Diseño muestral con el que se seleccionará la muestra

g) Levantamiento de la información

h) Resumen y análisis de los datos obtenidos

En este artículo nos enfocamos en el inciso f que se refiere al diseño muestral. En el diseño muestral se incluye la construcción del estimador puntual de la característica poblacional de interés, la varianza de dicho estimador y el estimador de varianza. La varianza del estimador es la que se emplea para el cálculo del tamaño de muestra.

Como se mencionó al principio de esta sección, hay varios diseños que se pueden emplear para extraer una muestra de una población. La decisión del tipo de diseño por emplear está sujeta a diversos factores como: características de la población por muestrear, disponibilidad de marcos muestrales de la población de interés, información auxiliar disponible durante la etapa de diseño, costo de extracción de la muestra y medición, tiempo disponible para realizar la encuesta, entre otros. Por ejemplo, si se desea extraer una muestra de personas en una ciudad como el Distrito Federal con el fin de estimar alguna característica de la población finita como el gasto o ahorro mensual por persona mayor de 18 años, no es posible emplear un muestreo aleatorio simple, ya que no se cuenta con un marco muestral de todas las personas mayores de 18 años, que vivan en el Distrito Federal en el lapso en el que se levantará la encuesta. Pero si se cuenta con mapas de manzanas o colonias con número de personas que se haya elaborado previo al levantamiento de la encuesta, como los mapas de áreas geoestádisticas básicas, AGEB, que elabora el INEGI, es posible emplear un muestreo por conglomerados en varias etapas con tamaños desiguales, iguales o proporcionales a alguna medida de tamaño, veáse (INEGI). Las etapas de selección podrían incluir AGEB, manzanas, viviendas y personas.

2.2 Notación, población y muestreo por conglomerados en 2 etapas.

Sea *U* una población finita de *N* elementos etiquetados como *k=1,…,N, 1<N*. Es usual representar a la población finita por sus etiquetas *k* como *U={1,2,…,k,…,N}*.

Los conglomerados se denotan como *UPM*, unidades primarias de muestreo y a los elementos dentro de conglomerados como *USM*, unidades secundarias de muestreo. *A* y *B* representarán al número de *UPM* en la población y al número de *USM* dentro de cada *UPM* respectivamente; en tanto que *a* y *b* representarán las respectivas cantidades muestrales. Se supone que *A, B, a* y *b* son mayores que uno y *a<A* y *b<B*. El total de elementos en población y muestra se denotan como *N=AB* y *n=ab*, respectivamente. La variable bajo estudio es dicotómica y se representa con y_{ij}, en donde *i* se refiere a la *UPM* y *j* a la *USM*. Dicha variable adquiere el valor de *1* si el *j-ésimo* elemento de la *i-ésima UPM* posee la característica de interés y *0* en otro caso. Se trabaja con las proporciones de la i-ésima *UPM*, $y_i = p_i = \sum_{j=1}^{B} p_{ij} / B$ y la proporción poblacional $y_U = p_U = \sum_{i=1}^{A} p_i / A$. La varianza entre medias de unidades primarias se denota como $s_{1u}^2 = \sum_{i=1}^{A} (y_i - y_U)^2 / (A-1)$, en tanto que la varianza entre elementos dentro de unidades primarias se expresa como $s_{2u}^2 = \sum_{i=1}^{A} \sum_{j=1}^{B} (y_{ij} - y_i)^2 / (A(B-1))$.

Se utiliza la notación *(mas,mas)* para indicar que tanto las *UPM*, como las *USM* en muestra, fueron seleccionadas por muestreo aleatorio simple sin reemplazo de una población conglomerada en la que todas las *UPM* tienen el mismo número de elementos.

2.3 Expresión para la varianza del estimador de proporciones.

La varianza del estimador de la media poblacional en el muestreo por conglomerados en dos etapas con tamaños iguales usando *mas* en ambas etapas, es, véase Cochran (1986):

$$V(\hat{\bar{y}}) = \left(1-\frac{a}{A}\right)\frac{s_{1u}^2}{a} + \left(1-\frac{b}{B}\right)\frac{s_{2u}^2}{ab} = \left(1-\frac{a}{A}\right)\frac{1}{a}\frac{\sum_{i=1}^{A}(\bar{y}_i - \bar{y}_U)}{A-1} + \left(1-\frac{b}{B}\right)\frac{1}{ab}\frac{\sum_{i=1}^{A}\sum_{j=1}^{B}(y_{ij}-\bar{y}_i)^2}{B-1}$$

Bajo *(mas,mas)* un estimador insesgado de la proporción poblacional es $\hat{p} = \sum_{i=1}^{a}\sum_{j=1}^{b} y_{ij} \big/ (ab)$ y la varianza $V(\hat{\bar{y}})$ en términos de la notación en proporciones es:

$$V(\hat{p}) = \alpha\left(\sum_{i=1}^{A} p_i^2 - Ap_U^2\right) + \beta\sum_{i=1}^{A} p_i(1-p_i) \qquad (1)$$

donde, $\alpha = \dfrac{1-a/A}{a(A-1)}$ y $\beta = \dfrac{(1-b/B)B}{abA(B-1)}$.

La varianza $V(\hat{p})$ en (1) puede escribirse como $V(\hat{p}) = \alpha V_1 + \beta V_2$, en la que

$$V_1 = \sum_{i=1}^{A} p_i^2 - Ap_U^2 \quad \text{y} \quad V_2 = \sum_{i=1}^{A} p_i(1-p_i). \qquad (2)$$

Estas fórmulas resultarán útiles en la determinación de las cotas motivo del presente artículo. La expresión para la varianza de un estimador del promedio poblacional en el MC2E, $V(\hat{\bar{y}})$, se encuentra en una gran cantidad de libros de texto y seguramente ha sido bastante usada en la práctica; empero, al menos para la estimación de proporciones y con base en la literatura conocida por el autor, no se había expresado como en (1). La fórmula

(1) es la que permite aislar en la varianza del estimador los efectos de las proporciones de las UPM y USM de los tamaños de población y muestra.

Antes de continuar, es necesario mencionar que en el evento de que todas las proporciones de las *UPM* sean cero o todas sean uno se tiene que $\sum_{i=1}^{A} p_i(1-p_i)=0$ y $\sum_{i=1}^{A} p_i^2 - Ap_U^2 = 0$, respectivamente, por lo cual, tanto V_1 como V_2 definidas en (1), toman el valor cero y será un caso que se excluirá del presente trabajo.

Por otra parte, es importante mencionar que todas las demostraciones se encuentran en el anexo.

2.4 Representación tabular de los valores para poblaciones conglomeradas.

Una forma conveniente de visualizar los datos $y_{ij}=1$ de una población conglomerada en *UPM* como la que nos ocupa es la siguiente:

Tabla 1

Representación de valores $y_{ij}=1$ para poblaciones conglomeradas

	UPM							
USM	**1**	**2**	**3**	·	**i**	·	**A-1**	**A**
1	1	1	1	·	1	·	1	0
2	1	1	0	·	1	·	1	0
3	1	1	0	·	0	·	1	0
·	1	1	0	·	0	·	1	0
j	1	1	0	·	0	·	0	0
·	1	1	0	·	0	·	0	0
B-1	1	1	0	·	0	·	0	0
B	1	1	0	·	0	·	0	0
p_i	p_1	p_2	p_3	·	p_i	·	p_{A-1}	p_A

En este caso, las columnas etiquetadas 1 a *A* representan a los conglomerados o *UPM*, en tanto que los renglones 1 a *B* se refieren a los elementos dentro de *UPM*, es decir, a las *USM*. En la parte inferior de cada columna que representa una *UPM* se encuentra p_i, el promedio de los y_{ij} por *UPM*. Es claro que si todos los y_{ij} dentro de una *UPM* son igual a uno, entonces $p_i=1$ y si todos los los y_{ij} dentro de una *UPM* son cero, $p_i=0$. Este tipo de configuraciones es importante para las cotas, por lo cual se muestra una representación de este tipo de poblaciones conglomeradas en la siguiente tabla.

Tabla2
Configuración $c\rho máx$ de valores $y_{ij}=1$, valor máximo de correlación intraclase

	UPM							
USM	**1**	**2**	**3**	·	**i**	·	**A-1**	**A**
1	1	1	0	·	0	·	0	0
2	1	1	0	·	0	·	0	0
3	1	1	0	·	0	·	0	0
·	1	1	0	·	0	·	0	0
j	1	1	0	·	0	·	0	0
·	1	1	0	·	0	·	0	0
B-1	1	1	0	·	0	·	0	0
B	1	1	0	·	0	·	0	0
p_i	p_1	p_2	p_3	·	p_i	·	p_{A-1}	p_A

En la tabla 2 se tiene una representación de valores $y_{ij}=1$ en la que todos los valores dentro de algunas *UPM* son uno, en este caso el 1 y el 2, y el resto toman el valor de 0. A este tipo de configuración de valores $y_{ij}=1$ se le denominará $c\rho máx$. Esta notación se refiere a que se alcanza el valor máximo del coeficiente de correlación intraclase, véase sección 2.4 y 3.1 del artículo, para la población de interés. Obsérvese que en este tipo de configuraciones el valor de V_2 es cero.

Hay otra configuración de valores $y_{ij}=1$ que también será de utilidad para el análisis de las cotas y se representa en la tabla que se muestra a continuación.

Tabla 3
Configuración *cρmín* de valores $y_{ij}=1$, valor mínimo de correlación intraclase

	UPM							
USM	**1**	**2**	**3**	**·**	**i**	**·**	**A-1**	**A**
1	1	1	1	·	1	·	1	1
2	1	1	1	·	1	·	1	1
3	1	1	1	·	1	·	1	1
·	0	0	0	·	0	·	0	0
j	0	0	0	·	0	·	0	0
·	0	0	0	·	0	·	0	0
B-1	0	0	0	·	0	·	0	0
B	0	0	0	·	0	·	0	0
$\mathbf{p_i}$	$\mathbf{p_1}$	$\mathbf{p_2}$	$\mathbf{p_3}$	**·**	$\mathbf{p_i}$	**·**	$\mathbf{p_{A-1}}$	$\mathbf{p_A}$

En la representación de la tabla 3 se tiene el mismo número de valores $y_{ij}=1$ en todas las *UPM*, por lo cual $p_i=p_u$ para todas las *UPM*. A este tipo de configuración de valores $y_{ij}=1$ se le denominará *cρmín*. Esta notación se refiere a que en este tipo de poblaciones se alcanza el valor mínimo del coeficiente de correlación intraclase, véase sección 2.4 y 3.1 del artículo. Por otra parte, en este tipo de configuraciones el valor de V_1 es cero.

Nótese que no siempre es posible encontrar una configuración *cρmáx* y/o *cρmín* para cualquier población conglomerada en dos etapas con *A*, *B* y p_u dados. Una vez decididos los tamaños *A* y *B* para una población, el valor de p_u nos indica el número de valores $y_{ij}=1$ en la población, siempre que $\sum_{i=1}^{A}\sum_{j=1}^{B} y_{ij} = ABp_U$ sea un entero. Un ejemplo de una población en la que no se alcanza una configuración *cρmáx* se encuentra en la siguiente tabla.

Tabla 4.
Representación en la que no se tiene una configuración *cρmáx*.

	UPM							
USM	**1**	**2**	**3**	**·**	**i**	**·**	**A-1**	**A**
1	1	1	1	·	0	·	0	0
2	1	1	1	·	0	·	0	0
3	1	1	1	·	0	·	0	0
·	1	1	0	·	0	·	0	0
j	1	1	0	·	0	·	0	0
·	1	1	0	·	0	·	0	0
B-1	1	1	0	·	0	·	0	0
B	1	1	0	·	0	·	0	0
$\mathbf{p_i}$	$\mathbf{p_1}$	$\mathbf{p_2}$	$\mathbf{p_3}$	**·**	$\mathbf{p_i}$	**·**	$\mathbf{p_{A-1}}$	$\mathbf{p_A}$

2.5 Expresión para el coeficiente de correlación intraclase.

El coeficiente de correlación intraclase para una población conglomerada en *UPM* y *USM* con tamaños iguales, Cochran (1986), se define como:

$$\rho = \frac{2\sum_{i=1}^{A}\sum_{j=1}^{B-1}\sum_{k>j}^{B}(y_{ij}-\bar{y}_U)(y_{ik}-\bar{y}_U)}{(B-1)(AB-1)s_U^2} \tag{3}$$

En esta fórmula, $\bar{y}_U$ se refiere al promedio poblacional y s_U^2 a la varianza poblacional entre elementos, con $s_U^2 > 0$.

Es importante hacer notar que ρ es una cantidad poblacional y refleja la correlación entre pares de unidades que se encuentran dentro del mismo conglomerado.

3. ALGUNOS ASPECTOS DE LA CORRELACIÓN INTRACLASE

3.1 Expresión para el coeficiente de correlación intraclase.

Debido a que el coeficiente de correlación intraclase se empleará en diversos ejemplos, es necesario contar con una fórmula que facilite su cálculo. Como la fórmula (3) es computacionalmente intensiva para calcular ρ, ya que es necesario evaluar el producto del numerador sobre todos los posibles pares dentro de cada *UPM*, de las expresiones en (2) y la fórmula para el coeficiente de correlación intraclase del capítulo 5.6B de Kish (1965), en términos de la varianza entre y dentro de las *UPM*, se obtiene una expresión sencilla del coeficiente de correlación intraclase ρ en términos de V_1 y V_2 definidas en (1), lo cual se muestra a continuación:

$$\rho = \frac{V_1}{[V_1 + V_2]} - \frac{V_2}{(B-1)[V_1 + V_2]} \tag{4}$$

En esta fórmula se requiere que la población conglomerada sea tal que $A, B \geq 2$ y los valores y_{ij} son tales que $\sum_{i=1}^{A}\sum_{j=1}^{B} y_{ij} / (AB) = p_U \in (0,1)$. Además, con la expresión (4) es inmediato determinar las configuraciones de los valores $y_{ij}=1$ en una población que conducen a los valores mínimo y máximo de dicho coeficiente. Así, los valores mínimo, $-1/(B-1)$, y máximo, 1, de ρ se obtienen con $V_1 = 0$ y $V_2 = 0$, respectivamente. La restricción $A \geq 2$ se debe a que sí $A = 1$, $V_1 = 0$ por construcción y $\rho = -1/(B-1)$. Por otra parte, cabe hacer notar que los valores mínimo y máximo del coeficiente de correlación

intraclase no se alcanzan para cualquier población, ya que esto depende del número de valores $y_{ij}=1$ en la población, la forma en que se encuentran distribuidos en los conglomerados, así como de los valores *A* y *B*. Esto se aprecia con más claridad en el ejemplo 3 de la siguiente sección. El hecho de que el coeficiente de correlación intraclase no alcance los valores mínimo y máximo en todos los casos es algo que no se ha encontrado en la literatura a conocimiento del autor.

3.2 Ejemplos del coeficiente de correlación intraclase.

En los dos ejemplos siguientes se muestran sendas representaciones de los valores $y_{ij}=1$ de la población que conducen a los valores máximo y mínimo en una población.

Ejemplo 1: *Valor máximo del coeficiente de correlación intraclase*. Considérese una población con *A=8 UPM*, *B=8 USM* y $p_U=3/8=0.375$, y con los siguientes tamaños de muestra de *UPM* y *USM*, $a=2$, $b=3$. Con estos valores se tiene que $\alpha=0.0536$ y $\beta=0.0149$, α y β fueron definidos en (1). En este ejemplo se tiene una población con 64 elementos en 8 conglomerados o *UPM*, con 8 elementos o *USM*, por conglomerado.

Tabla 5.
Una configuración de los valores $y_{ij}=1$ con la que se alcanza la ρ máxima.

	UPM							
USM	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1	1	1	0	0	0	0	0
2	1	1	1	0	0	0	0	0
3	1	1	1	0	0	0	0	0
4	1	1	1	0	0	0	0	0
5	1	1	1	0	0	0	0	0
6	1	1	1	0	0	0	0	0
7	1	1	1	0	0	0	0	0
8	1	1	1	0	0	0	0	0
$\mathbf{p_i}$	**1**	**1**	**1**	**0**	**0**	**0**	**0**	**0**

En la tabla 5, p_i se refiere a la proporción poblacional para la *i*-ésima *UPM* y, como todos los valores dentro de cada *UPM* son iguales, se tiene que $V_2 = 0$, por lo cual, $V(\hat{p}) = \alpha V_1 = 0.1004$ y $\rho = 1$. Por otra parte, la configuración de valores y_{ij} en la tabla 5 es del tipo *cρmáx*, ya que se alcanzó el valor máximo de la correlación intraclase para esta población y proporción poblacional.

Observación 1: nótese que las condiciones de este ejemplo corresponden a un arreglo de los valores $y_{ij}=1$ en la población tales que el coeficiente de correlación intraclase toma el valor de 1, por lo que se tiene *perfecta homogeneidad* dentro de conglomerados con respecto a la media o proporción poblacional p_U. La *perfecta homogeneidad* se refiere a que todos los valores dentro de cada conglomerado en la población, son mayores que p_U o todos son menores que p_U.

Ejemplo 2: *Valor mínimo del coeficiente de correlación intraclase*. Considérese la misma población del ejemplo 1, solo que las $y_{ij}=1$ se distribuyeron en las 8 *UPM* . Los valores de la población son los mismos que los del ejemplo 2: *A=8 UPM, B=8 USM* y $p_U=3/8=0.37$; en tanto que el número de *UPM* y *USM* en muestra son *a=2*, *b=3*, con el supuesto de selección de *a=2 UPM* y *b=3* elementos o *USM* por *mas*. Por lo anterior, los valores de α y β son los mismos que en el ejemplo 1.

Tabla 6.
Una configuración de los valores $y_{ij}=1$ con los que se alcanza la ρ mínima.

	UPM							
USM	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1
3	1	1	1	1	1	1	1	1
4	0	0	0	0	0	0	0	0
5	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
$\mathbf{p_i}$	**0.375**	**0.375**	**0.375**	**0.375**	**0.375**	**0.375**	**0.375**	**0.375**

En este caso, $V_1(\hat{p})=0$, ya que todas las p_i son iguales a 0.375, $V(\hat{p})=\beta V_2(\hat{p})=0.0279$ y $\rho=-1/(8-1)=-0.1429$. La configuración de valores y_{ij} en la tabla 6 es del tipo $c\rho mín$, ya que se alcanzó el valor mínimo de la correlación intraclase para esta población y proporción poblacional.

Observación 2: las condiciones de este ejemplo se refieren a un arreglo de los valores $y_{ij}=1$ en la población tales que el coeficiente de correlación intraclase toma el valor mínimo, *-1/(B-1)*, es decir, se tiene *perfecta heterogeneidad* dentro de conglomerados con respecto a la media o proporción poblacional p_U. La *perfecta heterogeneidad* se refiere a que dentro de cada conglomerado en la población, hay valores mayores que p_U y menores que p_U.

Observación 3: Es importante mencionar que no siempre se alcanzan los valores mínimo y máximo de la correlación intraclase, lo cual es algo que a conocimiento del autor no se menciona en la literatura del tema. A continuación se muestra un ejemplo de esto.

Ejemplo 3: *Arreglo de valores $y_{ij}=1$ en población para los que no se alcanza el valor máximo del coeficiente de correlación intraclase*. Considérese una población con *A=8 UPM, B=8 USM,* con los tamaños de muestra como los del ejemplo 1, *a=2, b=3,* solo que ahora sea $p_U=20/64=0.313$. Los valores de α y β son los mismos que en el ejemplo 1.

Tabla 7.
Una configuración de los valores $y_{ij}=1$ con los que no se alcanza la ρ máxima.

	UPM							
USM	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1	1	1	0	0	0	0	0
2	1	1	1	0	0	0	0	0
3	1	1	1	0	0	0	0	0
4	1	1	1	0	0	0	0	0
5	1	1	0	0	0	0	0	0
6	1	1	0	0	0	0	0	0
7	1	1	0	0	0	0	0	0
8	1	1	0	0	0	0	0	0
$\mathbf{p_i}$	**1**	**1**	**0.5**	**0**	**0**	**0**	**0**	**0**

En este ejemplo, $V(\hat{p})=\alpha V_1(\hat{p})+\beta V_2(\hat{p})=0.0787+0.0037=0.0824$ y $\rho=0.8338$. Obsérvese que en esta población ρ no alcanza el valor máximo de 1; por lo cual, no es una configuración de valores *cρmáx*. De hecho el valor máximo de ρ en este caso es de 0.8338. Se puede verificar que el valor mínimo posible de ρ para esta población es de -0.1221, por lo cual tampoco es una configuración de valores *cρmín*.

3.3 Valores de la varianza en el caso de correlación intraclase mínima y máxima.

A continuación se muestran dos resultados del muestreo por conglomerados en dos etapas para la varianza de proporciones en los que se aprecia el efecto en la varianza y en el coeficiente de correlación intraclase cuando algunas de las proporciones de las *UPM* son *0*

y el resto toman el valor de 1, condición 1, C1, o cuando todas las proporciones de las *UPM* son iguales a alguna proporción $p=c$, con $0<c<1$, condición 2, C2. Estos valores se emplean en la sección 4.

Condición 1, C1. Las condiciones en las que se tiene un arreglo *cρmáx* para una población conglomerada, véase sección 3.2, son las siguientes: bajo *(mas,mas)*, si $p_i=0$ ó 1 $\forall i \in \{1,\ldots,A\}$ y existen i e j, $i \neq j$, tales que $p_i \neq p_j$, entonces $V(\hat{p}) = \alpha V_1 = \alpha A p_U (1-p_U) = \alpha V_1^{c\rho \max}$ y $\rho = 1$. La varianza que se obtiene con un arreglo *cρmáx* se denota como $V_1^{c\rho\, max}$.

Ejemplo 4: $A=5$, $p_1=p_2=p_3=p_4=0$, $p_5=1, p_U=1/5$ y $V(\hat{p}) = \alpha\, 4/5$.

Si $a=3$, entonces se tiene que $\alpha=0.033$ y $V(\hat{p}) = 0.0267$.

Condición 2, C2. Las condiciones en las que se tiene un arreglo *cρmín* para una población conglomerada, véase sección 3.2, son: bajo *(mas,mas)*, si $p_i=p_U$ con $0<p_U<1$, $\forall i \in \{1,\ldots,A\}$, entonces: $V(\hat{p}) = \beta V_2 = \beta A p_U (1-p_U) = \beta V_2^{c\rho\, min}$ y $\rho = -1/(B-1)$. La varianza que se obtiene con un arreglo *cρmín* se denota como $V_2^{c\rho\, min}$.

Ejemplo 5: $A=5$ y si $p_i = 1/5, \forall i \in \{1,\cdots,5\}$, $a=3$ y $\beta = 0.0667$, entonces se tiene que $V(\hat{p}) = \beta V_2 = \beta\, 4/5 = 0.0533$.

En C1 y C2 se aprecia que tanto $V_1^{c\rho\, max}$ como $V_2^{c\rho\, min}$ son iguales a $A p_U (1-p_U)$. No se hace uso de un solo símbolo para esta última expresión, ya que es importante hacer énfasis en que la varianza del estimador de la proporción proviene de un arreglo de valores $y_{ij}=1$ en la

población que es $c\rho máx$ o $c\rho mín$. Además, en el resultado que se enuncia en la siguiente sección, se observa que el número de *UPM* y *USM* en muestra juegan un papel importante en la determinación de las cotas.

4. COTAS PARA LA VARIANZA

En esta sección se establece el resultado principal de este artículo, el cual es un teorema para las cotas de la varianza del estimador de proporciones para las posibles configuraciones de valores $y_{ij}=1$ en una población conglomerada con p_u dada.

Antes de enunciar el teorema es necesario introducir un par de representaciones de los valores $y_{ij}=1$, así como algunas expresiones para varianzas y sumas de cuadrados, que servirán para comprender mejor la notación usada en el resultado.

4.1 Representaciones y expresiones necesarias para las cotas.

Caso en el que no se alcanza la correlación intraclase mínima posible en una población. En analogía con las poblaciones empleadas en la sección anterior en las que se tenía un arreglo de los valores $y_{ij}=1$ tales que la correlación intraclase era mínima $-1/(B-1)$, considérese la siguiente representación de valores $y_{ij}=1$.

Tabla 8.
Una configuración de los valores $y_{ij}=1$ con los que no se alcanza la ρ mínima

	UPM							
USM	**1**	**2**	**3**	**·**	**i**	**·**	**A-1**	**A**
1	1	1	1	·	1	·	1	1
2	1	1	1	·	1	·	1	1
3	1	1	0	·	0	·	0	0
·	0	0	0	·	0	·	0	0
j	0	0	0	·	0	·	0	0
·	0	0	0	·	0	·	0	0
B-1	0	0	0	·	0	·	0	0
B	0	0	0	·	0	·	0	0
$\mathbf{p_i}$	$\mathbf{p_1}$	$\mathbf{p_2}$	$\mathbf{p_3}$	**·**	$\mathbf{p_i}$	**·**	$\mathbf{p_{A-1}}$	$\mathbf{p_A}$

En esta tabla no se alcanza el valor mínimo de la correlación intraclase *-1/(B-1),* ya que $V_1 > 0$. Por otra parte, obsérvese que $A_{l1} = 2$ *UPM* tienen el mismo valor para las proporciones p_1 y p_2, digamos p_{l1}, en tanto que las $A_{l2} = A - 2$ proporciones restantes p_3 a p_A tienen un valor igual entre sí, digamos p_{l2}, pero distinto a p_{l1}, con $A_{l1} + A_{l2} = A$ y $p_{l1} > p_{l2}$. Si etiquetamos a p_1 y p_2 como $p_{l1,1}$ y $p_{l1,2}$, y hacemos lo mismo para las proporciones p_3 a p_A, pero con p_{l2}, el promedio poblacional p_U puede expresarse como,

$$p_U = \frac{A_{l1}}{A}\bar{p}_{l1} + \frac{A_{l2}}{A}\bar{p}_{l2}, \quad \text{con} \quad \bar{p}_{l1} = \sum_{l=1}^{A_{l1}} \frac{p_{l1,i}}{A_{l1}} \quad \text{y} \quad \bar{p}_{l2} = \sum_{l=A_{l1+1}}^{A_{l1}} \frac{p_{l2,i}}{A_{l2}} \tag{5}$$

y la suma de cuadrados de las proporciones para cada *UPM* se puede descomponer de la siguiente manera:

$$\sum_{i=1}^{A} p_i^2 = \sum_{i=1}^{A_{l1}} p_{l1,i}^2 + \sum_{i=A_{l1+1}}^{A_{l2}} p_{l2,i}^2 = A_{l1}p_{l1}^2 + A_{l2}p_{l2}^2 \tag{6}$$

Con esto, podemos formar varianzas del tipo $V_{1,}$ como se definió en (2), al descomponer la suma de cuadrados de las proporciones asociadas a las *UPM* de la siguiente manera.

$$\sum_{i=1}^{A} p_i^2 - A_{I1}p_{I1}^2 - A_{I2}p_{I2}^2 = \left(\sum_{i=1}^{A_{I1}} p_{I1,i}^2 - A_{I1}p_{I1}^2\right) + \left(\sum_{i=A_{I1}+1}^{A} p_{I2,i}^2 - A_{I2}p_{I2}^2\right) = V_{I1} + V_{I2} \quad (7)$$

Una vez que se cuenta con estos elementos a la mano, expresamos a la varianza dada en (1) para una configuración de valores $y_{ij}=1$ del tipo de la tabla 8. Esto es importante para la demostración del teorema que se encuentra más adelante. Supóngase que $A_{I1} > 0,\ A_{I2} \geq 0$ y $A_{I1} + A_{I2} = A$, los componentes de (1) pueden expresarse como sigue:

$$\beta \sum_{i=}^{A} p_i(1-p_i) = \beta A_{I1}p_{I1}(1-p_{I1}) + \beta A_{I2}p_{I2}(1-p_{I2}) = \beta A_{I1}p_{I1} - \beta A_{I1}p_{I1}^2 + \beta A_{I2}p_{I2} - \beta A_{I2}p_{I2}^2 \quad (8)$$

Usando la igualdad (6) para αV_1 se tiene que:

$$\alpha\left(\sum_{i=1}^{A} p_i^2 - Ap_U^2\right) = \alpha\left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2\right) - \alpha A p_U^2 \quad (9)$$

Por lo cual, usando (8) y (9), la varianza (1), $V(\hat{p}) = \alpha\left(\sum_{i=1}^{A} p_i^2 - Ap_U^2\right) + \beta\sum_{i=1}^{A} p_i(1-p_i)$, puede expresarse como:

$$V(\hat{p}) = \alpha\left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2\right) - \alpha A p_U^2 - \beta A_{I1}p_{I1}^2 - \beta A_{I2}p_{I2}^2 + \beta A_{I1}p_{I1} + \beta A_{I2}p_{I2} \quad (10)$$

Sustituyendo (5) en (10) y sumando y restando $\beta A p_U^2$ en (10), la varianza queda como,

$$V(\hat{p}) = \beta A p_U(1-p_U) + (\alpha - \beta)\left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2 - Ap_U^2\right) \quad (11)$$

En esta fórmula, sí $\alpha = \beta$ y/o sí $A_{I2} = 0$, entonces $A_{I1} = A$ y $p_{I1} = p_U$, por lo que el segundo término del lado derecho de (11) es igual a cero y la varianza toma la forma de C2:

$$V(\hat{p}) = \beta A p_U(1-p_U) \quad (12)$$

Caso en el que no se alcanza la correlación intraclase máxima de una población. En analogía con las poblaciones empleadas en la sección anterior en las que se tenía un arreglo de los valores $y_{ij}=1$ tales que la correlación intraclase era máxima, es decir, con un valor de 1, considérese la siguiente representación de valores $y_{ij}=1$:

Tabla 9.
Una configuración de los valores $y_{ij}=1$ con los que no se alcanza la ρ máxima

	UPM							
USM	**1**	**2**	**3**	**·**	**i**	**·**	**A-1**	**A**
1	1	1	1	·	0	·	0	0
2	1	1	1	·	0	·	0	0
3	1	1	1	·	0	·	0	0
·	1	1	0	·	0	·	0	0
j	1	1	0	·	0	·	0	0
·	1	1	0	·	0	·	0	0
B-1	1	1	0	·	0	·	0	0
B	1	1	0	·	0	·	0	0
$\mathbf{p_i}$	$\mathbf{p_1}$	$\mathbf{p_2}$	$\mathbf{p_s}$	**·**	$\mathbf{p_i}$	**·**	$\mathbf{p_{A-1}}$	$\mathbf{p_A}$

En este caso, supóngase que $A \geq 2$ y sean A_{S1} aquellas *UPM* cuyas $p_i = 1$; $A_{S2} = 1$ una *UPM* con $p_i = p_s$, $p_s \in (0,1)$, A_{S3} la(s) *UPM* con $p_i = 0$ y $A_{S1}+A_{S2}+A_{S3} = A$. Para ser congruente con la restricción mencionada antes de iniciar la sección 2.3, es decir, $\sum_{i=1}^{A} p_i^2 - Ap_U^2 > 0$, en la tabla que se muestra a continuación se encuentran las posibles combinaciones de casos admisibles para los valores de A_{S1}, A_{S2} y A_{S3}.

Tabla 10.
Combinación de valores admisibles para A_{S1}, A_{S2} y A_{S3}.

Número de Combinación	A_{S1}	A_{S2}	A_{S3}	Admisible
1	=0	=0	=0	No
2	=0	=0	>0	No
3	=0	=1	=0	No
4	=0	=1	>0	Sí
5	>0	=0	=0	No
6	>0	=0	>0	Sí
7	>0	=1	=0	Sí
8	>0	=1	>0	Sí

La combinación número 6 corresponde a las configuraciones de valores $y_{ij}=1$ en las que el coeficiente de correlación intraclase toma el valor 1 y se tiene cuando $A_{S2}=0$, por lo que $A_{S1}+A_{S3}=A$, con $A_{S1}\in\{1,2,\cdots,A-1\}$. Por otra parte, la combinación número 8 corresponde a la de la tabla 9. El que una combinación sea no admisible se refiere a arreglos de valores y_{ij} en la población para los cuales no aplican las cotas del teorema, no a configuraciones que no se encuentren en la práctica. A continuación se construye la expresión para la varianza (1) en términos de las configuraciones admisibles de la tabla 10, por lo cual escribimos los componentes de (1) como sigue:

$$\beta\sum_{i=1}^{A}p_i(1-p_i)=\beta p_{S2}(1-p_{S2})=\beta p_{S2}-\beta p_{S2}^2 \tag{13}$$

$$\alpha\left(\sum_{i=1}^{A}p_i^2-Ap_U^2\right)=\alpha\left(A_{S1}+p_{S2}^2-Ap_U^2\right) \tag{14}$$

Antes de continuar, es importante notar que $Ap_U = \sum_{i=1}^{A} p_i = \sum_{i=1}^{A_{S1}} p_i + p_{S2} = A_{S1} + p_{S2}$, por lo que $A_{S1} = Ap_U - p_{S2}$ y sustituyendo este término en (14) se tiene que:

$$\alpha\left(\sum_{i=1}^{A} p_i^2 - Ap_U^2\right) = \alpha\left(Ap_U - p_{S2} + p_{S2}^2 - Ap_U^2\right) = \alpha Ap_U(1-p_U) - \alpha p_{S2}(1-p_{S2}) \tag{15}$$

La varianza (1), $V(\hat{p}) = \alpha\left(\sum_{i=1}^{A} p_i^2 - Ap_U^2\right) + \beta\sum_{i=1}^{A} p_i(1-p_i)$, adquiere la siguiente forma, usando (13) y (15):

$$V(\hat{p}) = \alpha Ap_U(1-p_U) - (\alpha-\beta)p_{S2}(1-p_{S2}) \tag{16}$$

En esta fórmula, sí $\alpha = \beta$ y/ó sí $A_{S2} = 0$, entonces el segundo término del lado derecho de (16) es igual a cero y la varianza toma la forma de C1:

$$V(\hat{p}) = \alpha Ap_U(1-p_U) \tag{17}$$

En la fórmula (16), el número de combinación 4 de la tabla 15, corresponde a una población en la que solo una de las *UPM* tiene una $p_i \in (0,1)$ y $A_{S1} = 0$, por lo que la varianza en (16) adquiere la siguiente forma:

$$V(\hat{p}) = \alpha\left(1 - \frac{1}{A}\right)p_{S2}^2 - \beta p_{S2}(1-p_{S2}) \tag{18}$$

Antes de continuar, recordemos que tanto $V_1^{c\rho\, max}$ como $V_2^{c\rho\, min}$ son iguales a $Ap_U(1-p_U)$.

4.2 Cotas para la varianza, coeficiente de variación y efecto del diseño.

Teorema: bajo *(mas,mas)*, α, β, V_1 y V_2 definidas en (1) y (2), α *y* β fijos, $A,B \geq 2$ y para cualquier permutación de los valores y_{ij} de la población tal que $\sum_{i=1}^{A}\sum_{j=1}^{B} y_{ij}/(AB) = p_U^* \in (0,1)$, con p_U^* fijo, el valor de $V(\hat{p})$ satisface alguna de las siguientes desigualdades:

(a) si $\alpha>\beta$, $\beta V_2^{c\rho\, min} + (\alpha - \beta)\left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2 - Ap_U^2\right) \leq V(\hat{p}) \leq \alpha V_1^{c\rho\, max} - (\alpha - \beta)p_{S2}(1-p_{S2})$,

(b) si $\alpha<\beta$, $\alpha V_1^{c\rho\, max} - (\alpha - \beta)p_{S2}(1-p_{S2}) \leq V(\hat{p}) \leq \beta V_2^{c\rho\, min} + (\alpha - \beta)\left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2 - Ap_U^2\right)$,

(c) si $\alpha=\beta=\gamma$, $V(\hat{p}) = \gamma A p_U(1-p_U)$.

Demostración: véase el Anexo 2.

Corolario 1: bajo las condiciones del teorema, $A = B$ y para cualquier población conglomerada que admita las configuraciones *cρmín* y *cρmáx* y cualquier permutación de los valores y_{ij} de la población tal que $\sum_{i=1}^{A}\sum_{j=1}^{B} y_{ij}/(AB) = p_U^* \in (0,1)$, con p_U^* fijo, el valor de $V(\hat{p})$ satisface alguna de las siguientes desigualdades:

(a) si $\alpha>\beta$, $\beta V_2^{c\rho\, min} \leq V(\hat{p}) \leq \alpha V_1^{c\rho\, max}$,

(b) si $\alpha<\beta$, $\alpha V_1^{c\rho\, max} \leq V(\hat{p}) \leq \beta V_2^{c\rho\, min}$,

(c) si $\alpha=\beta=\gamma$, $V(\hat{p}) = \gamma A p_U(1-p_U)$.

Corolario 2: Si $A = B$, $A > 2$, A par, $a = A/2 + 1$ y $b = 2$, entonces se tiene que:

(a) $\alpha=\beta$,

(b) $efd=1$.

Es importante mencionar que en el corolario 1 se exhiben las cotas mínima y máxima para la varianza del muestreo por conglomerados en dos etapas para poblaciones cuyos valores $y_{ij}=1$ tienen las dos configuraciones $c\rho mín$ y $c\rho máx$.

En el inciso (b) del corolario 2, al ser el efecto del diseño igual a uno, se tiene que la varianza del estimador de proporciones bajo *(mas,mas)* es igual a la del muestreo aleatorio simple, por lo cual no hay efecto de conglomeración al permutar los valores $y_{ij}=1$ de la población. Por otra parte, puede parecer poco factible tener un tamaño de muestra que sea un poco más grande que la mitad de las *UPM*; empero, esto podría usarse en poblaciones que tienen pocos conglomerados.

Observación 4: es importante recalcar que las cotas son válidas para una población en la cual solo se permutan los valores de las $y_{ij}=1$, pero se mantiene fijo el valor de la proporción poblacional p_U, así como las constantes α y β.

Cabe hacer notar que dado un arreglo de los valores $y_{ij}=1$ en la población, las cotas dependen de los valores α y β. Cuando $\alpha<\beta$, la configuración $c\rho mín$ en la población, la cual corresponde al valor mínimo de la correlación intraclase, se asocia con la cota superior para las tres cantidades, la varianza del estimador de proporciones, el coeficiente de variación y el efecto del diseño; en tanto que la configuración $c\rho máx$ en la población, la cual corresponde al valor máximo de la correlación intraclase, se asocia con la cota inferior para las tres cantidades mencionadas. Cuando $\alpha=\beta$, la varianza del estimador de proporciones

permanece sin cambio al permutar los valores de las $y_{ij}=1$, manteniendo fijo por supuesto el valor de la proporción poblacional p_U. En este último caso, lo que se modifican son los valores de V_1 y V_2.

El resultado obtenido para las cotas cuando $\alpha<\beta$ ó $\alpha=\beta$ es algo nuevo en opinión del autor, ya que en la literatura del muestreo por conglomerados en dos etapas, generalmente se menciona que la varianza del estimador se incrementa conforme el coeficiente de correlación intraclase crece. A la luz de estos resultados, es necesario aclarar que la relación entre la varianza y la correlación intraclase depende del signo de α-β.

A continuación se muestran las fórmulas de las cotas para el efecto del diseño y el coeficiente de variación en el caso de que satisfagan las condiciones del teorema y el corolario 1. Se enuncian como resultados ya que se trata de hechos que pueden verificarse fácilmente a partir del teorema y el corolario 1, usando las definiciones del efecto del diseño y el coeficiente de variación; sin embargo, son cantidades útiles en la práctica y conviene presentarlas de forma resumida.

Resultado 1: bajo las condiciones del teorema, el valor del efecto del diseño, $V(\hat{p})/V_{mas}(\hat{p})$, satisface alguna de las siguientes desigualdades:

(a) si $\alpha>\beta$, $efd_2^{c\rho mín} \le efd(\hat{p}) \le efd_1^{c\rho max}$

(b) si $\alpha<\beta$, $efd_1^{c\rho max} \le efd(\hat{p}) \le efd_2^{c\rho mín}$

(c) si $\gamma=\alpha=\beta$, $efd(\hat{p}) = \gamma An(N-1)/(1-f)N$,

donde,

$$efd_1^{c\rho max} = \alpha An(N-1)/(1-f)N - (\alpha-\beta)p_{S2}(1-p_{S2})/V_{mas}(\hat{p}),$$

$$efd_2^{c\rho min} = \beta An(N-1)/(1-f)N + (\alpha-\beta)\left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2 - Ap_U^2\right)/V_{mas}(\hat{p}) \text{ y}$$

$$V_{mas}(\hat{p}) = (1-f)N' p_U(1-p_U)/n, \text{ con } f = n/N \text{ y } N' = N/(N-1).$$

En este resultado, *f=n/N* se refiere a la fracción de muestreo de elementos, como si la muestra de tamaño *n=ab* hubiese sido extraída por *mas* de la población de *N=AB* elementos y $N' = N/(N-1)$. Como se mencionó en la introducción el efecto del diseño, *efd*, fue propuesto por Kish (1965) como una medida de eficiencia de diseños muestrales distintos al muestreo aleatorio simple. Por otra parte, cuando se tiene que $A = B$ y la población tiene las configuraciones *cρmín* y *cρmáx*, las cotas para el efecto del diseño adquieren una forma simple, lo cual se encuentra en el siguiente resultado.

Resultado 2: bajo las condiciones del corolario 1, el efecto del diseño, $V(\hat{p})/V_{mas}(\hat{p})$, $V_{mas}(\hat{p}) > 0$, satisface alguna de las siguientes desigualdades:

(a) si $\alpha>\beta$, $\beta\, efd_2^{c\rho min} \le efd(\hat{p}) \le \alpha\, efd_1^{c\rho max}$

(b) si $\alpha<\beta$, $\alpha\, efd_1^{c\rho max} \le efd(\hat{p}) \le \beta\, efd_2^{c\rho min}$

(c) si $\gamma=\alpha=\beta$, $efd(\hat{p}) = \gamma An(N-1)/(1-f)N$,

donde,

$$efd_1^{c\rho max} = An(N-1)/(1-f)N = efd_2^{c\rho min} \text{ y}$$

$V_{mas}(\hat{p}) = (1-n/N)N' p_U(1-p_U)/n$, con $f=n/N$.

Recordemos que el coeficiente de variación para el estimador de una proporción se define como $cv(\hat{p}) = \sqrt{V(\hat{p})}/p$, con $p > 0$. Debido a la importancia de esta cantidad en el ámbito estadístico, en el resultado 3 se encuentran las cotas para dicha cantidad. Cuando se tiene que $A = B$ y la población admite las configuraciones *cρmín* y *cρmáx*, las cotas para el coeficiente de variación adquieren una forma simple, lo cual se encuentra en el resultado 4.

Resultado 3: bajo las condiciones del teorema, el coeficiente de variación, $\sqrt{V(\hat{p})}/p$, satisface alguna de las siguientes desigualdades:

(a) si $\alpha > \beta$, $cv_2^{c\rho\, min} \le cv(\hat{p}) \le cv_1^{c\rho\, max}$

(b) si $\alpha < \beta$, $cv_1^{c\rho\, max} \le cv(\hat{p}) \le cv_2^{c\rho\, min}$

(c) si $\gamma = \alpha = \beta$, $cv(\hat{p}) = \sqrt{\gamma A(1-p_U)/p_U}$,

donde, $cv_1^{c\rho\, max} = \sqrt{\alpha A p_U(1-p_U) - (\alpha-\beta)p_{s2}(1-p_{s2})}/p_U$ y

$$cv_2^{c\rho mín} = \sqrt{\beta A p_U(1-p_U) + (\alpha-\beta)\left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2 - Ap_U^2\right)}/p_U$$

Resultado 4: bajo las condiciones del corolario 1, el coeficiente de variación, $\sqrt{V(\hat{p})}/p$, satisface alguna de las siguientes desigualdades:

(a) si $\alpha > \beta$, $\sqrt{\beta}\, cv_2^{c\rho\, min} \le cv(\hat{p}) \le \sqrt{\alpha}\, cv_1^{c\rho\, max}$

(b) si $\alpha < \beta$, $\sqrt{\alpha}\, cv_1^{c\rho\, max} \le cv(\hat{p}) \le \sqrt{\beta} cv_2^{c\rho\, min}$

(c) si $\gamma=\alpha=\beta$, $cv(\hat{p})=\sqrt{\gamma A(1-p_U)/p_U}$,

donde, $cv_1^{c\rho max}=\sqrt{A(1-p_U)/p_U}=cv_2^{c\rho min}$.

En la siguiente sección se mostrarán diversos ejemplos para ilustrar los valores de las cotas.

5 EJEMPLOS DE COTAS

Ejemplo 6: *α>β, varianza entre cotas máxima y mínima*. Considérese la misma población del ejemplo 2, con las $y_{ij}=1$ acomodadas de manera diferente a los casos de los ejemplos 2 y 4 y los valores de *a*, *b*, *α* y *β* son los mismos, $\alpha=0.0536$ y $\beta=0.0149$, así como la selección de *UPM* y *USM* por *mas*. En este caso, $V(\hat{p})=\alpha V_1(\hat{p})+\beta V_2(\hat{p})=0.0285+.02=0.0485$ y $\rho=0.1810$. En este ejemplo, el valor de ρ se encuentra entre el ρ mínimo, -0.1429, y el ρ máximo que es 1. Una representación de la población en términos de los valores y_{ij} es como sigue:

Tabla 11.
Configuración de los valores y_{ij} con los que la varianza se encuentra entra la cota mínima y máxima.

	UPM							
USM	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1	1	1	1	1	1	0	0
2	1	1	1	1	1	1	0	0
3	1	1	1	1	1	0	0	0
4	1	1	1	1	0	0	0	0
5	1	1	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0
7	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0
$\mathbf{p_i}$	**0.750**	**0.625**	**0.500**	**0.500**	**0.375**	**0.250**	**0**	**0**

Ejemplo 7: *$\alpha>\beta$, varios arreglos de los valores $y_{ij}=1$ para mostrar las cotas mínima y máxima*. Se usa la población del ejemplo 1 con $p_U=0.375$, solo que ahora $a=3$, $b=3$, con estos valores se tiene que $\alpha=0.0298$, $\beta=0.0099$. Los valores de *A*, *B*, *a* y *b*, satisfacen las condiciones del inciso a del corolario 1. En la tabla 12 se muestra el valor de la varianza del estimador de la proporción, los valores αV_1 y βV_2, las contribuciones relativas a la varianza *V* de αV_1 y βV_2, el efecto del diseño, *efd*, el coeficiente de variación, *cv*, así como el coeficiente de correlación intraclase para siete configuraciones de valores $y_{ij}=1$. Dos de las siete configuraciones corresponden al mínimo y máximo del coeficiente de correlación intraclase.

Tabla 12.
Valores de αV_1 y βV_2 para diversas configuraciones de los valores y_{ij}, $\alpha>\beta$.

ρ	*V*	αV_1	βV_2	$\alpha V_1/V$	$\beta V_2/V$	*efd*	*cv*
-0.1429	0.0186	0.0000	0.0186	0%	100%	0.818	36%
-0.0095	0.0229	0.0065	0.0164	28%	72%	1.009	40%
0.1429	0.0279	0.0140	0.0140	50%	50%	1.227	45%
0.3143	0.0335	0.0223	0.0112	67%	33%	1.473	49%
0.6000	0.0428	0.0363	0.0065	85%	15%	1.882	55%
0.8667	0.0515	0.0493	0.0022	96%	4%	2.264	60%
1.0000	0.0558	0.0558	0.0000	100%	0%	2.455	63%

Al final de la sección 3 se mostró $V_1^{c\rho\,max}=V_2^{c\rho\,min}=Ap_U(1-p_U)$, por lo cual, las cotas superior, $\beta V_2^{c\rho\,min}$, e inferior, $\alpha V_1^{c\rho\,max}$, según el inciso a del corolario 1 toman los valores *0.0558* y *0.0186*. En la tabla 12 se aprecia que tanto los valores de la varianza, *V*, como el del componente αV_1, crecen conforme el coeficiente de correlación intraclase se incrementa, lo cual está de acuerdo con el inciso a del corolario 1. Para el caso del *efd* se presenta el caso conocido de que esta cantidad crece al incrementarse la correlación intraclase.

Ejemplo 8: *α=β, misma varianza independientemente del arreglo de valores* $y_{ij}=1$. Considérese la población del ejemplo 1, solo que ahora sean *a=5*, *b=2*, con estos valores se tiene que *α=0.0107*, *β=0.0107* y *α=β*. Los valores de *A*, *a* y *b*, satisfacen las condiciones del corolario 2. Por otra parte, el valor de la proporcional poblacional sigue siendo $p_U=0.375$. Los títulos de las columnas son como en el ejemplo 7 y dos de las siete configuraciones corresponden al mínimo y máximo del coeficiente de correlación intraclase.

Tabla 13.
Valores de αV_1 y βV_2 para diversas configuraciones de los valores y_{ij}, *α=β*.

ρ	*V*	αV_1	βV_2	$\alpha V_1/V$	$\beta V_2/V$
-0.1429	0.0201	0.0000	0.0201	0%	100%
-0.0095	0.0201	0.0023	0.0177	12%	88%
0.1429	0.0201	0.0050	0.0151	25%	75%
0.3143	0.0201	0.0080	0.0121	40%	60%
0.6000	0.0201	0.0131	0.0070	65%	35%
0.8667	0.0201	0.0177	0.0023	88%	12%
1.0000	0.0201	0.0201	0.0000	100%	0%

En la tabla 13 se observa que los valores de la varianza, *V*, son iguales para las permutaciones de los valores $y_{ij}=1$ que se hicieron en esta población, lo cual está de acuerdo con el corolario 1; sin embargo, lo que cambia para cada configuración que se hizo son los valores de V_1 y V_2, así como el coeficiente de correlación intraclase. En las dos últimas columnas de la tabla 13 se tienen los valores del tamaño relativo de los componentes de varianza y se aprecia en este caso que, conforme la correlación intraclase crece, así lo hace el componente de variación entre *UPM*. Se puede comprobar que el *efd*

toma el valor 1 para todos los valores de la varianza *V* de la tabla 13 y que el coeficiente de variación permanece sin cambio con un valor de 38%.

Ejemplo 9: *α<β, relación inversa entre la varianza y la correlación intraclase*. Considérese la población del ejemplo 1, solo que ahora sean *a=6, b=2*, con estos valores se tiene que *α=0.0060, β=0.0089* y *α<β*. Los valores de *α* y *β* satisfacen el inciso b del corolario 1. Por otra parte, el valor de la proporcional poblacional sigue siendo $p_U=0.375$. Los títulos de las columnas son como en el ejemplo 7 y dos de las siete configuraciones corresponden al mínimo y máximo del coeficiente de correlación intraclase.

Tabla 14.
Valores de αV_1 y βV_2 para diversas configuraciones de los valores y_{ij}, $\alpha=\beta$.

ρ	*V*	αV_1	βV_2	$\alpha V_1/V$	$\beta V_2/V$	*efd*	*cv*
-0.1429	0.0167	0.0000	0.0167	0%	100%	1.038	35%
-0.0095	0.0161	0.0013	0.0148	8%	92%	0.998	34%
0.1429	0.0153	0.0028	0.0126	18%	82%	0.952	33%
0.3143	0.0145	0.0045	0.0100	31%	69%	0.900	32%
0.6000	0.0131	0.0073	0.0059	55%	45%	0.813	31%
0.8667	0.0118	0.0099	0.0020	83%	17%	0.733	29%
1.0000	0.0112	0.0112	0.0000	100%	0%	0.692	28%

Recordemos que $V_1^{c\rho\,max}=V_2^{c\rho mín}=Ap_U(1-p_U)$, por lo cual, las cotas superior, $\beta V_2^{c\rho\,min}$, e inferior, $\alpha V_1^{c\rho\,max}$, según el inciso b del corolario 1 toman los valores *0.0167* y *0.0112*. En la tabla 14 se observa que los valores de la varianza, *V*, son más grandes conforme el valor de la correlación intraclase es más pequeño, lo cual está de acuerdo con el inciso b del corolario 1. Este ejemplo hace evidente lo que se mencionó en la subsección 4.2: un valor creciente del coeficiente de correlación intraclase no necesariamente implica un valor mayor de la varianza. También cambian para cada configuración realizada los valores de V_1

y V_2, así como el coeficiente de correlación intraclase. En las dos últimas columnas se encuentran los valores del tamaño relativo de los componentes de varianza y se aprecia en este caso que un incremento en la correlación intraclase, va asociado a un crecimiento en el componente de variación entre las *UPM*, αV_1.

Ejemplo 10: A continuación se calculan las cotas para el coeficiente de variación (lím inf cv y lím sup cv) de la proporción estimada, desviación estándar (lím inf desv y lím sup desv) y efecto del diseño (lím inf efd y lím sup efd) para una población con *A=8*, a=*2*, *B=10*, tamaños de submuestreo, *b*, de *2* a *4 USM* y p_U=*0.5*. Como $\alpha > \beta$, para todos los valores de *b* en este ejercicio, los límites inferior y superior para la varianza coincide con los arreglos *cpmín* y *cpmáx* en la población. Por este motivo, las cotas inferior y superior que se aplican para la varianza, el coeficiente de variación y el efecto del diseño, son las que se encuentran en el inciso a del corolario 1, del resultado 4 y del resultado 2.

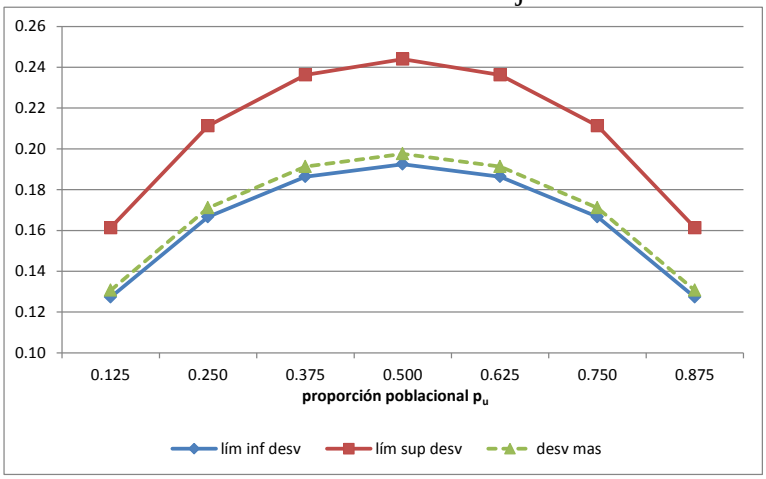
Tabla 15.
Cotas mínima y máxima para la varianza, coeficiente de variación y efecto del diseño

b=	2	3	4
α =	0.054	0.054	0.054
β =	0.028	0.016	0.010
α - β =	0.026	0.037	0.043
lím inf desv=	0.236	0.180	0.144
lím sup desv=	0.327	0.327	0.327
lím inf cv=	0.471	0.360	0.289
lím sup cv=	0.655	0.655	0.655
lím inf efd=	0.924	0.830	0.731
lím sup efd=	1.782	2.745	3.762

De la Tabla 15 se aprecia que el límite inferior para la desviación estándar disminuye conforme *b* se incrementa, lo cual es una propiedad del *mas* dentro de cada *UPM*; empero, la diferencia entre la cota mínima y máxima crece ya que el número de *UPM* en muestra está fijo. Para las cotas del coeficiente de variación se observa un comportamiento similar; sin embargo, para el efecto del diseño, la cota superior crece al incrementarse $n=ab$.

Ejemplo 11: *$\alpha>\beta$, cotas para la varianza con diferentes valores de p_U.* En la siguiente gráfica se encuentran los valores mínimos, línea azul, y máximos, línea roja, de la desviación estándar para una población conglomerada de 80 elementos con $A=8$, $a=3$, $B=10$, $b=2$, $\alpha=0.0298$, $\beta=0.0185$ y p_U tomando los siguientes valores: 0.125, 0.250, 0.375, 0.5, 0.625, 0.750 y 0.875. Como $\alpha>\beta$, se emplea el inciso a del corolario 1 para las cotas inferior y superior de la varianza del estimador de proporciones. Para la misma población y valores de p_U se calculó la desviación estándar del estimador de la proporción poblacional bajo *mas*, línea verde punteada. En este caso, se considera a la población sin conglomerar y se tienen $N=AB=8x10=80$ elementos y una muestra de $n=ab=3x2=6$ elementos.

Gráfica 2
Valores mínimo y máximo de la desviación estándar para el diseño (*mas*,*mas*), así como los de la desviación estándar bajo *mas*.



En la Gráfica 2 se observa que el valor máximo de la desviación estándar en el esquema (*mas*,*mas*), tanto para el límite inferior como para el superior, se tiene con p_U=0.5, lo cual es similar al caso de la varianza máxima para el estimador de proporciones bajo *mas*, la cual se alcanza cuando la proporción poblacional adquiere el valor de 0.5, como se mostró en la introducción del presente artículo.

En esta gráfica se aprecia el efecto de la conglomeración, obsérvese que la desviación estándar del estimador de la proporción poblacional bajo *mas*, línea verde punteada, tiene un valor apenas mayor que el límite inferior de la desviación estándar para el diseño *(mas,mas)*. Esto implica que casi siempre se trabajará con un *efd* mayor que uno en esta población al usar el muestreo por conglomerados en dos etapas.

Ejemplo 12: *Efecto de un diseño muestral en el error de estimación absoluto y selección del tamaño de muestra*. Suponga que se tiene una unidad habitacional con 175 edificios de departamentos y cada edificio tiene 8 departamentos. Se desea calcular el tamaño de muestra para estimar la proporción de departamentos que sufrieron algún robo en el último mes y como estimación anticipada de p_U usamos $p_U^* = 0.15$. Para seleccionar el número de departamentos por edificio en muestra, de la Tabla 15 se desprende que el rango de la varianza disminuye conforme los valores de submuestreo de *b* son cercanos a *2*, con *A* y *a* fijos, por lo cual evaluaremos el error de estimación absoluto con *a* entre *15* y *35* edificios y $b \in \{3,4\}$. Con estos datos, *N=AB=1400*. Como la representación tabular de esta población requiere 175 columnas, no es conveniente mostrarla; sin embargo, como se supone que $p_U^* = 0.15$, esto implica que hay 210 valores $y_{ij}=1$ en la población. Debido a que no parece

razonable que en todos los edificios haya habido robos, supóngase que aproximadamente un 26% de los edificios no ha tenido este tipo de eventos y que en una cantidad similar de edificios solo se ha dado un evento de este tipo por edificio. De esta información se tiene que aproximadamente 46 edificios no han tenido robos, 48 han tenido robos en un departamento por edificio y los restantes 81 edificios han tenido robos a dos departamentos por edificio. La suma de estas 3 cantidades resulta en 175 edificios. Esto es un supuesto de una configuración de valores $y_{ij}=1$ en la población que en la práctica del muestreo por conglomerados se hace con el coeficiente de correlación intraclase. Con esta información ya se está en condiciones de calcular V_1 y V_2 usando (2), así $V_1 = 5.8125$ y $V_2 = 20.4375$, por lo que, de la fórmula (4), $\rho = 0.11$. Nótese que esta configuración de valores $y_{ij}=1$ no corresponde a una en la que se pueda alcanzar el mínimo o máximo valor de la varianza.

En las tablas A1 y A2 del Anexo 1, se encuentran los efectos en la varianza y el error de estimación, entre otras cantidades, cuando el número de *UPM* en muestra crece para dos valores de *b*, *3* y *4*. El error de estimación se construyó usando un desvío normal, $t_{\alpha/2}=1.645$. De las tablas A1 y A2 se observa que dicho error disminuye conforme el número de *UPM* crece de *15* a *35*, para los dos valores de *b*. Para el caso de *b=3*, Tabla A1, el error de estimación varía entre *0.0824* y *00531*, y cuando *b=4*, Tabla A2, dicho error va de *0.0692* y *0.0442*. Con estos datos, ya se está en condiciones de seleccionar los valores de *a* y *b*, dependiendo del número de edificios que se puedan visitar en el lapso de levantamiento de la encuesta. Por ejemplo, un error de estimación de *0.08* alrededor de $p_U^* =0.15$ podría no proporcionar la información requerida. Se requeriría que el error estuviera más concentrado alrededor de $p_U^* =0.15$, por lo cual las cantidades cercanas a *0.05* parecen

adecuadas, en caso de que sean costeables. Supóngase que se cuenta con recursos para visitar a lo más el 20% de los edificios, es decir, $A=175x0.20=35$ edificios, así podrían elegirse valores de $a=35$ y $b=4$, con lo cual el error de estimación esperado es 0.0442.

Las implicaciones de este plan en términos de los errores de estimación mínimo y máximo, e_a, se pueden evaluar con las cotas del teorema. Con estos datos, $n=140$ y usando un desvío normal, $t_{\alpha/2}=1.645$, se tiene que $\alpha=0.000131$, $\beta=0.0000233$, $\alpha>\beta$, por lo que aplicando el inciso (a) del teorema, las cotas inferior y superior de la varianza son, 0.000568 y 0.002234 respectivamente. De esta manera, el error de estimación absoluto se encuentra entre 0.039 y 0.094.

Con este ejemplo solo se pretende ilustrar un posible uso de una de las cotas y es una simplificación del proceso de determinación de diversas cantidades en el cálculo del tamaño de muestra. Por ejemplo, la determinación del número de *USM* en muestra puede estar influida por el costo asociado al submuestreo en las *UPM* y no se consideró un ajuste al tamaño de muestra por no respuesta.

6. CONCLUSIONES

Se propusieron cotas para la varianza, el efecto del diseño y el coeficiente de variación en el caso de la estimación de proporciones para el muestreo por conglomerados en dos etapas con tamaños iguales, suponiendo muestreo aleatorio simple en las dos etapas de selección. Estas cotas facilitan el cálculo del tamaño de muestra y también permiten evaluar los valores mínimo y máximo posibles de la varianza del estimador de la proporción. También

se construyó una expresión para el coeficiente de correlación intraclase poblacional en términos de varianzas entre y dentro de conglomerados. A través de varios ejemplos se observó que los tamaños de muestra para unidades primarias y secundarias de muestreo para una población conglomerada tienen efecto en la determinación de las cotas inferior y superior. Se mostraron situaciones en las cuales, dependiendo de los tamaños de muestra de unidades primarias y secundarias, las cotas inferior y superior son iguales o se tienen casos de relación inversa entre el valor del coeficiente de correlación intraclase y la varianza. Ciertamente este tipo de casos no corresponden a situaciones que se den con frecuencia en la práctica, pero podrían ser de interés en situaciones particulares, como poblaciones con pocos conglomerados y unidades secundarias de muestreo. Las cotas para el efecto del diseño también permiten evaluar los tamaños mínimo, máximo de muestra que se tendrían en un diseño muestral con conglomerados de tamaños iguales, al usar muestreo aleatorio en ambas etapas con la metodología mencionada en el primer párrafo de la introducción. Por supuesto, también pueden calcularse varios valores del efecto del diseño y evaluar su impacto en el tamaño de muestra para diversas configuraciones de los valores de la variable de interés en la población.

Por otra parte, las fórmulas son sencillas de calcular y únicamente se requieren los elementos de información con los que normalmente se cuenta en la práctica en la etapa de diseño muestral. Además, como se aprecia en los ejemplos, no solo se encuentran las cotas superior e inferior, sino que se pueden calcular para diversas configuraciones de los valores de la variable de interés en la población, las cantidades poblacionales como: el coeficiente

de correlación intraclase, varianza, efecto del diseño, coeficiente de variación y contribución a la variabilidad de los componentes entre y dentro de varianza.

Como temas de investigación a futuro se evaluará la posible extensión de las cotas a variables que tengan más de dos categorías y se estudiará el tema de la estimación y la relación con las cotas. Asimismo se estudiará la posible extensión al muestreo de conglomerados en dos etapas con probabilidades proporcionales a alguna medida de tamaño sin reemplazo.

Bibliografía

Chaudhuri, A. & Stenger, H. *Survey Sampling: theory and methods*, 2nd ed., Chapman & Hall/CRC, 2005.

Cochran, W., *Técnicas de Muestreo*, Ed. CECSA, México, 1986.

INEGI, *Encuesta Nacional de Ingresos y Gastos de los Hogares 2008*. *DiseñoMuestral*.

Kish, L., *Survey Sampling*, New York: Wiley & Sons, 1965.

Padilla, Terán, A. M. "*Cotas para la varianza, efecto del diseño y coeficiente de variación de proporciones en el muestreo por conglomerados en dos etapas con tamaños iguales*". Memorias electrónicas en extenso de la 3ª Semana Internacional de la Estadística y la Probabilidad, Puebla de Zaragoza, Puebla, México. Junio 2010, CD ISBN: 978-607-487-162-3.

Särndal, C.E., Swensson, B. & Wretman, J.H., *Model Assisted Survey Sampling*, Springer-Verlag, New York, 1992.

Scott, A.J. & Smith, T.M.F., *Minimax designs for sample surveys*, Biometrika, Vol. 62, No. 2, pp. 353-357, Aug. 1975.

Anexo 1

Cantidades para evaluar el tamaño de muestra del ejemplo 12

Tabla A1

Número de elementos por muestrear en cada UPM, b=3

a	*n*	α	β	αV_1	βV_2	$v(\hat{p})$	error absoluto	$\alpha V_1 / V(\hat{p})$	$\beta V_2 / V(\hat{p})$	V_{mas}	*efd*
15	45	0.00035	0.00009	0.00066	0.00185	0.0025	0.0824	26.2%	73.8%	0.0027	0.915
16	48	0.00033	0.00009	0.00061	0.00174	0.0023	0.0797	26.0%	74.0%	0.0026	0.915
17	51	0.00031	0.00008	0.00057	0.00164	0.0022	0.0773	25.9%	74.1%	0.0024	0.916
18	54	0.00029	0.00008	0.00054	0.00154	0.0021	0.0751	25.8%	74.2%	0.0023	0.916
19	57	0.00027	0.00007	0.00051	0.00146	0.0020	0.0730	25.7%	74.3%	0.0021	0.917
20	60	0.00025	0.00007	0.00048	0.00139	0.0019	0.0711	25.6%	74.4%	0.0020	0.918
21	63	0.00024	0.00006	0.00045	0.00132	0.0018	0.0693	25.4%	74.6%	0.0019	0.918
22	66	0.00023	0.00006	0.00043	0.00126	0.0017	0.0677	25.3%	74.7%	0.0018	0.919
23	69	0.00022	0.00006	0.00041	0.00121	0.0016	0.0661	25.2%	74.8%	0.0018	0.919
24	72	0.00021	0.00006	0.00039	0.00116	0.0015	0.0647	25.1%	74.9%	0.0017	0.920
25	75	0.00020	0.00005	0.00037	0.00111	0.0015	0.0633	24.9%	75.1%	0.0016	0.920
26	78	0.00019	0.00005	0.00035	0.00107	0.0014	0.0620	24.8%	75.2%	0.0015	0.921
27	81	0.00018	0.00005	0.00034	0.00103	0.0014	0.0608	24.7%	75.3%	0.0015	0.921
28	84	0.00017	0.00005	0.00032	0.00099	0.0013	0.0597	24.6%	75.4%	0.0014	0.922
29	87	0.00017	0.00005	0.00031	0.00096	0.0013	0.0586	24.4%	75.6%	0.0014	0.923
30	90	0.00016	0.00005	0.00030	0.00093	0.0012	0.0576	24.3%	75.7%	0.0013	0.923
31	93	0.00015	0.00004	0.00029	0.00090	0.0012	0.0566	24.2%	75.8%	0.0013	0.924
32	96	0.00015	0.00004	0.00028	0.00087	0.0011	0.0556	24.1%	75.9%	0.0012	0.924
33	99	0.00014	0.00004	0.00026	0.00084	0.0011	0.0547	23.9%	76.1%	0.0012	0.925
34	102	0.00014	0.00004	0.00026	0.00082	0.0011	0.0539	23.8%	76.2%	0.0012	0.925
35	105	0.00013	0.00004	0.00025	0.00079	0.0010	0.0531	23.7%	76.3%	0.0011	0.926

Tabla A2

Número de elementos por muestrear en cada UPM, b=4

a	*n*	alfa	beta	alfa V1	beta V2	$v(\hat{p})$	error absoluto	$\alpha V_1 / V(\hat{p})$	$\beta V_2 / V(\hat{p})$	V mas	*efd*
15	60	0.00035	0.00005	0.00066	0.00111	0.0018	0.0692	37.1%	62.9%	0.0020	0.869
16	64	0.00033	0.00005	0.00061	0.00104	0.0017	0.0669	37.0%	63.0%	0.0019	0.870
17	68	0.00031	0.00005	0.00057	0.00098	0.0016	0.0648	36.8%	63.2%	0.0018	0.870
18	72	0.00029	0.00005	0.00054	0.00093	0.0015	0.0629	36.7%	63.3%	0.0017	0.871
19	76	0.00027	0.00004	0.00051	0.00088	0.0014	0.0612	36.5%	63.5%	0.0016	0.871
20	80	0.00025	0.00004	0.00048	0.00083	0.0013	0.0596	36.4%	63.6%	0.0015	0.872
21	84	0.00024	0.00004	0.00045	0.00079	0.0012	0.0581	36.2%	63.8%	0.0014	0.873
22	88	0.00023	0.00004	0.00043	0.00076	0.0012	0.0567	36.1%	63.9%	0.0014	0.873
23	92	0.00022	0.00004	0.00041	0.00073	0.0011	0.0554	35.9%	64.1%	0.0013	0.874
24	96	0.00021	0.00003	0.00039	0.00070	0.0011	0.0541	35.8%	64.2%	0.0012	0.874
25	100	0.00020	0.00003	0.00037	0.00067	0.0010	0.0530	35.6%	64.4%	0.0012	0.875
26	104	0.00019	0.00003	0.00035	0.00064	0.0010	0.0519	35.5%	64.5%	0.0011	0.876
27	108	0.00018	0.00003	0.00034	0.00062	0.0010	0.0508	35.3%	64.7%	0.0011	0.876
28	112	0.00017	0.00003	0.00032	0.00060	0.0009	0.0499	35.2%	64.8%	0.0010	0.877
29	116	0.00017	0.00003	0.00031	0.00058	0.0009	0.0489	35.0%	65.0%	0.0010	0.878
30	120	0.00016	0.00003	0.00030	0.00056	0.0009	0.0481	34.9%	65.1%	0.0010	0.878
31	124	0.00015	0.00003	0.00029	0.00054	0.0008	0.0472	34.7%	65.3%	0.0009	0.879
32	128	0.00015	0.00003	0.00028	0.00052	0.0008	0.0464	34.5%	65.5%	0.0009	0.879
33	132	0.00014	0.00002	0.00026	0.00051	0.0008	0.0457	34.4%	65.6%	0.0009	0.880
34	136	0.00014	0.00002	0.00026	0.00049	0.0007	0.0449	34.2%	65.8%	0.0008	0.881
35	140	0.00013	0.00002	0.00025	0.00048	0.0007	0.0442	34.1%	65.9%	0.0008	0.881

Anexo 2
Demostraciones

A continuación se encuentran las demostraciones del teorema y corolarios, así como la verificación de las condiciones C1 y C2. El fin de la demostración se denota por □.

Verificación de la expresión (4).

La expresión para el coeficiente de correlación intraclase de la sección 5.6B de Kish (1965), en términos de varianzas entre y dentro de *UPM* es:

$$\rho = \frac{\frac{A-1}{A}s_{U1}^2 - \frac{1}{B}s_{U2}^2}{\frac{N-1}{N}s_U^2}$$, en la que $s_{U1}^2 = \frac{V_1}{A-1}$, $s_{U2}^2 = \frac{B}{A(B-1)}V_2$ y

$$\frac{N-1}{N}s_U^2 = \frac{A-1}{A}s_{U1}^2 + \frac{B-1}{B}s_{U2}^2 .$$

Sustitúyanse estas tres cantidades en la expresión para ρ y se tiene que:

$$\rho = \frac{V_1}{V_1+V_2} - \frac{1}{(B-1)(V_1+V_2)} .$$ □

Verificación de la condición C1.

Si todas las proporciones de las *UPM* tienen valor cero o uno y recordando que $p_U \in (0,1)$, la varianza en (1) queda como $V(\hat{p}) = \alpha\left(\sum_{i=1}^{A} p_i^2 - Ap_U^2\right)$, es decir, $V_2 = 0$. Como las p_i

toman el valor cero o uno, $\sum_{i=1}^{A} p_i^2 = \sum_{i=1}^{A} p_i = Ap_U$ y $V(\hat{p}) = \alpha\left(\sum_{i=1}^{A} p_i^2 - Ap_U^2\right) = \alpha Ap_U(1-p_U)$. Como $V_2 = 0$, al sustituir este valor en (4) el coeficiente de correlación intraclase toma el valor 1. □

Verificación de la condición C2.

Si todas las proporciones de las *UPM* tienen el mismo valor $p_i = p_U$ y recordando que $p_U \in (0,1)$, la varianza en (1) queda como $V(\hat{p}) = \beta\sum_{i=1}^{A} p_i(1-p_i)$, es decir, $V_1 = 0$. Como las p_i tienen el mismo valor, $V(\hat{p}) = \beta\sum_{i=1}^{A} p_i(1-p_i) = \beta Ap_U(1-p_U)$. Como $V_1 = 0$, al sustituir este valor en (4) el coeficiente de correlación intraclase toma el valor $-1/(B-1)$. □

<u>Demostración del teorema.</u>

Debido a que la demostración de este teorema es larga ya que se hace por incisos, tipo de cota, inferior o superior, y cuando es necesario por casos, se empleará la notación en negritas e itálica FinPru Teo-inciso-tipo de cota para indicar el fin de prueba para cada inciso, tipo de cota y, si es aplicable el número de combinación según la tabla 10. Por ejemplo, ***FinPru Teo-(a)-ci*** se refiere al fin de la demostración del inciso a, cota inferior del teorema.

Teorema, inciso a, cota inferior, Teo-(a)-ci. Sean,

$ci = \beta Ap_U(1-p_U) + (\alpha - \beta)\left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2 - Ap_U^2\right)$ y

$v = V(\hat{p}) = \alpha(\sum_{i=1}^{A} p_{vi}^2 - Ap_U^2) + \beta\sum_{i=1}^{A} p_{vi}(1-p_{vi})$, el subíndice *v* en p_{vi} se refiere al valor de la proporción correspondiente a la *i*-ésima *UPM* usado en la varianza $V(\hat{p})$. Es necesario demostrar que *v-ci≥0*, para lo cual tomaremos la diferencia:

$$v - ci = \alpha\sum_{i=1}^{A} p_{vi}^2 - \alpha Ap_U^2 + \beta\sum_{i=1}^{A} p_{vi} - \beta\sum_{i=1}^{A} p_{vi}^2 - \beta Ap_U + \beta Ap_U^2$$

$$-(\alpha-\beta)\left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2\right) + \alpha Ap_U^2 - \beta Ap_U^2$$

Como $\beta\sum_{i=}^{A} p_i = \beta Ap_U$ y los términos αAp_U^2 y βAp_U^2 se cancelan, se tiene que,

$v - ci = (\alpha-\beta)\left(\sum_{i=1}^{A} p_{vi}^2 - \left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2\right)\right)$, hágase $\sum_{i=1}^{A} p_{vi}^2 = \sum_{i=1}^{A_{I1}} p_{vi}^2 + \sum_{i=1}^{A_{I2}} p_{vi}^2$

Y recordando que $\sum_{i=1}^{A_{I1}} p_{vi}^2 - A_{I1}p_{I1}^2$ y $\sum_{i=1}^{A_{I2}} p_{vi}^2 - A_{I2}p_{I2}^2$ son varianzas tipo V_I como en (7), entonces:

$$v - ci = (\alpha-\beta)\left(\sum_{i=1}^{A} p_{vi}^2 - \left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2\right)\right) = (\alpha-\beta)(V_{I1} + V_{I2}) \geq 0 \qquad \text{(D1)}$$

Como $\alpha - \beta > 0$ y si $\sum_{i=1}^{A} p_{vi}^2 - \left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2\right) > 0$, entonces *ci* es la cota inferior del teorema, inciso a. ***FinPru Teo-(a)-ci.***

Teorema, inciso b, cota superior, Teo-(b)-cs.

Si en (D1), $\alpha - \beta < 0$, $v - ci = (\alpha-\beta)\left(\sum_{i=1}^{A} p_{vi}^2 - \left(A_{I1}p_{I1}^2 + A_{I2}p_{I2}^2\right)\right) = (\alpha-\beta)(V_{I1} + V_{I2}) < 0$ y se tiene que *v-ci<0*, por lo cual *ci* es una cota superior *cs*. ***FinPru Teo-(b)-cs***

Antes de continuar en la demostración del teorema, es necesario efectuar algunas manipulaciones algebraicas para expresar cantidades que serán útiles en la prueba. Sean,

$cs = \alpha A p_U (1-p_U) - (\alpha - \beta) p_{S2} (1-p_{S2})$ y $v = V(\hat{p}) = \alpha(\sum_{i=1}^{A} p_{vi}^2 - A p_U^2) + \beta \sum_{i=1}^{A} p_{vi}(1-p_{vi})$; es necesario mostrar que *cs-v≥0*, para lo cual tomaremos la diferencia:

$$cs - v = \alpha A p_U - \alpha A p_U^2 - (\alpha - \beta) p_{S2}(1-p_{S2}) - \alpha \sum_{i=1}^{A} p_{vi}^2 + \alpha A p_U^2 - \beta \sum_{i=1}^{A} p_{vi} + \beta \sum_{i=1}^{A} p_{vi}^2$$

Como $\beta \sum_{i=}^{A} p_{vi} = \beta A p_U$ y los términos $-\alpha A p_U^2$ y $\alpha A p_U^2$ se cancelan, se tiene que,

$$cs - v = (\alpha - \beta)\left(A p_U - p_{S2}(1-p_{S2}) - \sum_{i=1}^{A} p_{vi}^2\right) \qquad \text{(D2)}$$

Mostraremos por casos que *cs-v≥0* haciendo referencia a la tabla 10, para ello se usará la notación (A_{S1}, A_{S2}, A_{S3}), en la que cada componente se relaciona con el valor de A_{Si}. Así, una tripleta $(>0, =1, >0)$ hace referencia al número de combinación 8 de la tabla 15.

Teorema, inciso a, cota superior, Teo-(b)-cs-comb6.

Caso 1, número de combinación 6 $(>0, =0, >0)$.

Obsérvese que $A_{S2} = 0$ implica que $p_{S2} = 0$ ó $p_{S2} = 1$, por lo cual:

$$cs - v = (\alpha - \beta)\left(A p_U - \sum_{i=1}^{A} p_{vi}^2\right) \qquad \text{(D3)}$$

En esta fórmula, los valores de p_i asociados a $A p_U = \sum_{i=1}^{A} p_i$ son 1, por lo que se cumple que:

$$p_{vi}^2 \le p_{vi} \text{ y } \sum_{i=1}^{A} p_{vi}^2 \le \sum_{i=1}^{A} p_{vi} = Ap_U = \sum_{i=1}^{A} p_i \qquad \text{(D4)}$$

Usando esta última desigualdad en (D3) y el hecho de que $\alpha - \beta > 0$ se tiene que:

$$cs - v = (\alpha - \beta)\left(Ap_U - \sum_{i=1}^{A} p_{vi}^2\right) \ge 0 \qquad \text{(D5)}$$

Como $\alpha - \beta > 0$ y $p_{S2} = 0$ ó $p_{S2} = 1$, entonces *cs* es la cota superior en el inciso a del teorema. ***FinPru Teo-(b)-cs-comb6.***

Teorema, inciso b, cota inferior, Teo-(b)-ci.

Si $\alpha < \beta$ en (D5), entonces *cs* es una cota inferior *ci* y se obtiene el resultado para la cota inferior del inciso b del teorema. ***FinPru Teo-(b)-ci-comb6.***

Teorema, inciso a, cota superior, Teo-(a)-cs-comb4.

Caso 2, número de combinación 4 $(=0,=1,>0)$.

Obsérvese que $A_{S2} = 1$ implica que $p_{S2} \in (0,1)$ y como $A_{S1} = 0$, por lo cual, usando (18):

$$cs - v = \alpha\left(1 - \frac{1}{A}\right)p_{S2}^2 + \beta p_{S2}(1 - p_{S2}) - \alpha\sum_{i=1}^{A} p_{vi}^2 + \alpha A p_U^2 - \beta\sum_{i=1}^{A} p_{vi} + \beta\sum_{i=1}^{A} p_{vi}^2 \qquad \text{(D6)}$$

Sustituyendo $Ap_U = p_{S2} = \sum_{i=1}^{A} p_{vi}$ en (D6) se tiene que:

$$cs - v = (\alpha - \beta)p_{S2}^2 - \alpha\frac{p_{S2}^2}{A} + \beta p_{S2} - (\alpha - \beta)\sum_{i=1}^{A} p_{vi}^2 + \alpha A p_U^2 - \beta A p_U$$

y cancelando en la ecuación anterior los términos $\alpha A p_U^2$ y $\beta A p_U$, se obtiene:

$$cs - v = (\alpha - \beta)\left(p_{S2}^2 - \sum_{i=1}^{A} p_{vi}^2\right) \quad \text{(D7)}$$

Como $p_{S2} = \sum_{i=1}^{A} p_{vi}$ al elevar al cuadrado se satisface lo siguiente,

$$p_{S2}^2 = \left(\sum_{i=1}^{A} p_{vi}\right)^2 = \sum_{i=1}^{A} p_{vi}^2 + 2\sum_{i=1}^{A-1}\sum_{j>i}^{A} p_{vi}p_{vj} \geq \sum_{i=1}^{A} p_{vi}^2 \quad \text{(D8)}$$

De la desigualdad en (D8) se concluye que *cs-v≥0* y la igualdad se cumple si $p_{S2} = p_{vi}$. Este caso no tiene equivalencia en el corolario 1. ***FinPru Teo-(a)-cs-comb4*.**

Teorema, inciso b, cota inferior, Teo-(b)-ci-comb4.

Si $\alpha < \beta$ en (D7), entonces *cs* es una cota inferior *ci* y se obtiene el resultado para la cota inferior del inciso b del teorema. ***FinPru Teo-(b)-ci-comb4*.**

Teorema, inciso a, cota superior, Teo-(a)-cs-comb7.

Caso 3, número de combinación 7 $(>0,=1,=0)$.

Aquí se tiene que $A_{S1} > 0$, $A_{S3} = 0$, $A_{S2} = 1$, $p_{S2} \in (0,1)$ y $A = A_{S1} + A_{S2} = A_{S1} + 1 \geq 2$, por lo cual:

$cs = \alpha A p_U (1 - p_U) - (\alpha - \beta) p_{S2} (1 - p_{S2})$ y

$$v = \alpha\left(\sum_{i=1}^{A} p_{vi}^2 - A p_U^2\right) + \beta\sum_{i=1}^{A} p_{vi}(1 - p_{vi}) = (\alpha - \beta)\sum_{i=1}^{A} p_{vi}^2 - \alpha A p_U^2 + \beta\sum_{i=1}^{A} p_{vi}$$

Ahora se demostrará que *cs-v≥0*.

$$cs - v = \alpha A p_U (1 - p_U) - (\alpha - \beta) p_{S2} (1 - p_{S2}) - (\alpha - \beta)\sum_{i=1}^{A} p_{vi}^2 - \alpha A p_U^2 + \beta\sum_{i=1}^{A} p_{vi}$$

En esta expresión se cancelan los términos $\alpha A p_U^2$ y se tiene que.

$$cs - v = (\alpha - \beta)\left[A p_U + p_{S2}^2 - p_{S2} - \sum_{i=1}^{A} p_{vi}^2\right] \quad \text{(D9)}$$

Como $A p_U = \sum_{i=1}^{A} p_i = A_{S1} + p_{S2}$, la fórmula en (D9) se puede expresar de la siguiente forma:

$$cs - v = (\alpha - \beta)\left[A_{S1} + p_{S2}^2 - \sum_{i=1}^{A} p_{vi}^2\right] \quad \text{(D10)}$$

Antes de continuar, es necesario notar que en el caso de la cota superior en esta combinación no hay *UPM* con $p_i = 0$ y que $A_{S2} = 1$, por lo que las A_{S1} *UPM* tienen valor $p_i = 1$ y la *UPM* asociada a A_{S2} tiene una proporción $p_i = p_{S2}$ y $p_{S2} \in (0,1)$. De aquí se desprenden 2 posibilidades para la configuración de valores $y_{ij}=1$ en esta población (recordemos que toda configuración debe ser tal que su promedio sea p_U):

a) La configuración de valores $y_{ij}=1$ es igual a la de los valores de la cota superior, en cuyo caso $cs=v$.

b) La configuración de valores $y_{ij}=1$ es distinta a la de los valores de la cota superior, y se cumple que $0 < p_{S2} < p_{v,s2} < 1$ y una(s) p_{vi} son tales que: $0 < p_{vi} < 1$ y otras $p_{vi} = 1$.

Con estos elementos a la mano, se requiere demostrar que la expresión del lado derecho en (D10) es mayor que cero. Denótese a $A_{S1,d}$ como el número de columnas A_{S1} en las que $0 < p_{vi} < p_i = 1$ y como $0 < p_{S2} < p_{v,s2} < 1$, la diferencia $0 < p_{v,s2} - p_{S2}$ tiene que ser igual a

$A_{S1,d} - \sum_{i=1}^{A_{S1,d}} p_{vi}$. De esta manera se cumple lo siguiente $A_{S1,d} - \sum_{i=1}^{A_{S1,d}} p_{vi} = p_{v,s2} - p_{S2}$ y al despejar A_{S1} se obtiene:

$$A_{S1,d} = \sum_{i=1}^{A_{S1,d}} p_{vi} + p_{v,s2} - p_{S2} \qquad \text{(D11)}$$

Como $\sum_{i=1}^{A_{S1,d}} p_{vi} > \sum_{i=1}^{A_{S1,d}} p_{vi}^2$ y $p_{v,s2}^2 > p_{S2}^2$, de (D11) se tiene que:

$\sum_{i=1}^{A_{S1,d}} p_{vi} + p_{v,s2} - p_{S2} > \sum_{i=1}^{A_{S1,d}} p_{vi}^2 + p_{v,s2}^2 - p_{S2}^2$ y $A_{S1,d} + p_{S2}^2 - \sum_{i=1}^{A_{S1,d}} p_{vi}^2 - p_{v,s2}^2 > 0$. Esta última desigualdad es equivalente a la del lado derecho de (D10), ya que en (D10) se encuentran las $p_i = 1$ asociadas a la cota superior y las $p_{vi} = 1$ de la configuración de valores $y_{ij}=1$ según el inciso (b). Por lo anterior y dado que $\alpha - \beta > 0$, se cumple la desigualdad:

$$(\alpha - \beta)\left[A_{S1} + p_{S2}^2 - \sum_{i=1}^{A} p_{vi}^2\right] > 0 \qquad \text{(D12)}$$

Esta combinación no tiene un resultado para el corolario 1. ***FinPru Teo-(a)-cs-comb7.***

Teorema, inciso b, cota inferior.

Si $\alpha < \beta$ en (D12), entonces *cs* es una cota inferior *ci* y se obtiene el resultado para la cota inferior del inciso b del teorema. ***FinPru Teo-(b)-ci-comb7.***

Teorema, inciso a, cota superior, Teo-(a)-cs-comb8.

Caso 4, número de combinación 8 $(>0,=1,>0)$.

Para esta combinación, $A_{S1} > 0$, $A_{S3} > 0$, $A_{S2} = 1$, $p_{S2} \in (0,1)$, por lo cual:

$cs = \alpha A p_U (1-p_U) - (\alpha - \beta) p_{S2} (1-p_{S2})$ y

$$v = \alpha\left(\sum_{i=1}^{A} p_{vi}^2 - A p_U^2\right) + \beta \sum_{i=1}^{A} p_{vi}(1-p_{vi}) = (\alpha-\beta)\sum_{i=1}^{A} p_{vi}^2 - \alpha A p_U^2 + \beta \sum_{i=1}^{A} p_{vi}$$

Antes de comenzar la demostración, obsérvese que si la configuración de valores $y_{ij}=1$ para evaluar la varianza *v*, es tal que se acumulan todos en las *UPM* asociadas a las A_{S1} y en la *UPM* asociada a A_{S2}, entonces se tiene el caso 3.

Ahora se demostrará que $cs\text{-}v \geq 0$.

$$cs - v = \alpha A p_U (1-p_U) - (\alpha - \beta) p_{S2} (1-p_{S2}) - (\alpha-\beta)\sum_{i=1}^{A} p_{vi}^2 - \alpha A p_U^2 + \beta \sum_{i=1}^{A} p_{vi}$$

En esta expresión se cancelan los términos $\alpha A p_U^2$ y se tiene que.

$$cs - v = (\alpha-\beta)\left[A p_U + p_{S2}(1-p_{S2}) - \sum_{i=1}^{A} p_{vi}^2\right] \qquad \text{(D13)}$$

Como $\sum_{i=1}^{A} p_{vi}^2 < \sum_{i=1}^{A} p_{vi} = A p_U$ y $\alpha - \beta > 0$, se tiene que $A p_U - \sum_{i=1}^{A} p_{vi}^2 > 0$ y

$$cs - v = (\alpha-\beta)\left[A p_U + p_{S2}(1-p_{S2}) - \sum_{i=1}^{A} p_{vi}^2\right] > 0. \qquad \text{(D14)}$$

Esta combinación no tiene un resultado equivalente para el corolario 1. ***FinPru Teo-(a)-cs-comb8*.**

Teorema, inciso b, cota inferior.

Si $\alpha < \beta$ en (D14), entonces *cs* es una cota inferior *ci* y se obtiene el resultado para la cota inferior del inciso b del teorema. ***FinPru Teo-(a)-cs-comb8*.**

Teorema, inciso c, igualdad entre las cotas superior e inferior.

Si $\alpha = \beta$ en las cotas superior e inferior de los incisos a y b del teorema, la varianza se reduce a $V(\hat{p}) = \gamma A p_U (1 - p_U)$, con $\alpha = \beta = \gamma$. □

<u>Demostración del corolario 1.</u>

Se emplea una notación similar al teorema para indicar el fin de la prueba por inciso y tipo de cota.

Corolario 1, inciso a, cota inferior, Cor1-(a)-ci.

Como $\alpha - \beta > 0$ y si $A_{I2} = 0$, de (D1) se tiene que $\sum_{i=1}^{A} p_{vi}^2 - \left(A_{I1} p_{I1}^2 + A_{I2} p_{I2}^2\right) = \sum_{i=1}^{A} p_{vi}^2 - A p_U^2$, la cual es una varianza tipo V_1 como en (2), entonces *ci* es la cota inferior en el inciso a del corolario 1. ***FinPru Cor1-(a)-ci.***

Corolario 1, inciso b, cota superior, Cor1-(b)-cs.

Si $\alpha - \beta < 0$ y $A_{I2} = 0$, de (D1) se tiene que $\sum_{i=1}^{A} p_{vi}^2 - \left(A_{I1} p_{I1}^2 + A_{I2} p_{I2}^2\right) = \sum_{i=1}^{A} p_{vi}^2 - A p_U^2$, la cual es una varianza tipo V_1 como en (2), entonces *ci* es la cota superior en el inciso b del corolario 1. ***FinPru Cor1-(b)-cs.***

Corolario 1, inciso a, cota superior, Cor1-(a)-cs.

Esta combinación de valores $y_{ij}=1$ es del tipo *cρmáx* para la cual se aplica el corolario 1, por lo que *cs* es la cota superior en el inciso a del corolario 1. ***FinPru Cor1-(a)-cs.***

Corolario 1, inciso b, cota inferior, Cor1-(b)-ci.

Si $\alpha - \beta < 0$, *cs* se convierte en una cota inferior *ci* y se tiene el resultado del inciso b para el corolario 1. ***FinPru Cor1-(b)-ci.***

Corolario 1, inciso c, igualdad entre las cotas superior e inferior.

Si $\alpha = \beta$ en las cotas superior e inferior de los incisos a y b del corolario 1, la varianza se reduce a $V(\hat{p}) = \gamma A p_U (1 - p_U)$, con $\alpha = \beta = \gamma$. □

Demostración del corolario 2.

Se emplea una notación similar al teorema para indicar el fin de la prueba por inciso y tipo de cota.

Corolario 2, inciso a, Cor2-(a).

Como $A = B$, $a = \frac{A}{2} + 1$, $b = 2$, de (1) se tiene lo siguiente:

$$\alpha = \frac{\left(1 - \frac{a}{A}\right)}{a(A-1)} = \frac{1 - \frac{1}{A}\left(\frac{A}{2} + 1\right)}{\left(\frac{A}{2} + 1\right)(A-1)} = \frac{A-2}{A(A-1)(A+2)} \quad \text{y}$$

$$\beta = \frac{\left(1 - \frac{2}{A}\right)}{2a(A-1)} = \frac{A-2}{2\left(\frac{A}{2} + 1\right)(A-1)} = \frac{A-2}{A(A+2)(A-1)} = \alpha \text{ . } \textbf{\textit{FinPru Cor2-(a).}}$$

Corolario 2, inciso b, Cor2-(b).

Cuando $\alpha = \beta = \gamma$, la varianza $V(\hat{p}) = \gamma A p_U (1 - p_U)$ según el inciso c del teorema y el corolario 1. Por otra parte, nótese que $n = ab = \left(\frac{A}{2} + 1\right)2 = A + 2$, $N = AB = A^2$ y $\alpha = \frac{A-2}{A(A+2)(A-1)}$. El efecto del diseño es:

$$efd = \frac{V(\hat{p})}{V_{mas}(\hat{p})}, \text{ con } V_{mas}(\hat{p}) = \left(1 - \frac{n}{N}\right)\frac{N}{N-1}\frac{p_u(1-p_U)}{n}$$

$$efd = \frac{\alpha A p_U (1-p_U)}{\frac{(N-n)p_U(1-p_U)}{(N-1)n}} = \frac{\alpha A(A+2)(A^2-1)}{A^2 - A - 2} = \frac{(A-2)(A+1)}{(A-2)(A+1)} = 1.$$ ***FinPru Cor2-(b).*** □