

COEFICIENTE DE HOMOGAMIA: TAMAÑO MUESTRAL Y NIVEL DE CONFIANZA

Noemí Acreche^{1, 3, 4, 5}

Graciela Caruso^{1, 4, 5}

María V. Albeza^{2, 3, 4, 5}

PALABRAS CLAVE: Homogamia, Selección de parejas, Error de muestreo, Significación

RESUMEN: El coeficiente de homogamia (**H**) mide la correlación entre variantes de los dos sexos (carácter: nativo del lugar). Partiendo de la estructura de parejas de la población de Cachi, se fijaron diferentes valores de error para cada categoría y se calculó el tamaño de la muestra. Asimismo, se definió el tamaño de la muestra y se estimó el error para cada caso. Se exploró la alternativa de fijar el error máximo. Se puso a prueba la propuesta validando los resultados obtenidos a partir de la muestra comparándolos con los de la población. Se generaron 1000 repeticiones sin reposición que permitieron estimar: media e intervalo de confianza para cada categoría de pareja y de **H**. Para evaluar los resultados del procedimiento se simuló la toma de una muestra aleatoria (193 parejas) en la población de Cachi y a partir de ésta se obtuvieron 10.000 muestras del mismo tamaño con reposición (bootstrap). Se estimaron la frecuencia media y los intervalos de confianza de las categorías y de **H**. El coeficiente de homogamia se construyó a partir de la diferencia entre valores observados y esperados de las categorías de parejas. Una prueba de significación adecuada sería la de Chi cuadrado y si la muestra es pequeña, el test exacto de Fisher. Se propone la utilización de software para la evaluación de la significación de **H**. Rev. Arg. Antrop. Biol. 10(2): 71-83, 2008.

1 Cátedra de Evolución. Facultad de Ciencias Naturales. Universidad Nacional de Salta.

2 Cátedra de Bioantropología. Facultad de Ciencias Naturales. Universidad Nacional de Salta.

3 Cátedra de Fundamentos de Antropología Biológica. Facultad de Humanidades. Universidad Nacional de Salta.

4 CIUNSa. Universidad Nacional de Salta.

5 PICTO-UNSa N° 36656.

e-mail: acreche@unsa.edu.ar

Correspondencia a: Dra. Noemí Acreche. JM Leguizamón N° 1966. 4400 Salta. Argentina.

e-mail: acreche@unsa.edu.ar

Recibido 5 Noviembre 2008; aceptado 18 Diciembre 2008.

KEY WORDS: Homogamy, Assortative mating, Sampling error, Significance level

ABSTRACT: Homogamy coefficient (H) measures the correlation of variants from both sexes (native in the village). Basing on the mating structure of the population of Cachi, different errors in each category and sample size were estimated. Results were tested comparing those obtained from the sample and the population. With 1000 repetitions without reposition, means and confidence intervals were estimated for each category and H . From a random sample of 193 couples from the population of Cachi, 10.000 samples of the same size were obtained with reposition (bootstrap). Means and confidence intervals were estimated for each category and H . X^2 and Fisher are suggested as significance tests for H and the use of software to assess H signification is proposed. Rev. Arg. Antrop. Biol. 10(2): 71-83, 2008.

INTRODUCCION

La selección de muestras para el trabajo en biodemografía ha sido soslayada, posiblemente por el hecho de que muchas veces la posibilidad de acceder a los datos limita el diseño de un muestreo aleatorio. En poblaciones pequeñas, se puede optar por relevar la totalidad de los individuos sin que esto represente un gran esfuerzo. Sin embargo, en poblaciones de gran tamaño es necesario trabajar con muestras y por ende fijar criterios y considerar el error estadístico que se suma a los otros errores posibles como los que provienen de la calidad de los datos.

Frecuentemente, en los estudios biodemográficos, se presta atención a la estructura de parejas de la población, ya que esta información permite inferir las preferencias de los individuos en la elección de pareja, la que puede determinar su estructura genética.

Bajo una hipótesis de estabilidad, se espera que las poblaciones se comporten panmícticamente. Una forma de verificar la panmixia con respecto a un carácter determinado, es comparando las frecuencias de las diferentes parejas observadas con las esperadas obtenidas del producto de las frecuencias de los varones y de las mujeres correspondientes.

El coeficiente de homogamia (H) mide la correlación entre las variantes de un sexo y las mismas en el otro. En este caso se consideró el carácter nativo del lugar. Este coeficiente toma valores entre -1 y 1, siendo -1 heterogamia completa, 1 homogamia completa y 0 panmixia (Savorgnan, 1949).

Cuando se trabaja con muestras, es necesario adoptar un criterio para evaluar el alejamiento del valor θ que indica panmixia en la población para el carácter considerado.

En este trabajo se considera la elección de una muestra representativa de la estructura de la población para la estimación del coeficiente de homogamia y se propone una prueba de significación para el mismo.

El punto de partida son los datos de una población de la provincia de Salta, Cachi, relevados mediante encuestas domiciliarias en visitas a la totalidad de las viviendas de la localidad (Acreche y Albeza, 2008). Esta información permite poner a prueba el modelo propuesto validando los resultados obtenidos a partir de una muestra y comparándolos con los de la población.

MATERIAL Y METODOS

Para el cálculo del coeficiente de homogamia se utilizó la variable “categoría de pareja”. Se definieron cuatro categorías de parejas (AB , Ab , aB , ab) en función del carácter nativo del lugar/sexo: A , B , a y b , siendo A : varón nativo del lugar, a : varón migrante, B : mujer nativa del lugar y b : mujer migrante. Se registró el número de parejas en cada categoría.

Como la unidad de muestreo es la vivienda, se supone que se registra una pareja por vivienda. Este supuesto surge de analizar la relación número de parejas/número de viviendas en trece localidades del Noroeste Argentino (NOA) relevadas, incluyendo la totalidad de la población (Tabla 1). En Cachi se registraron 385 parejas en 387 viviendas (99,48 %).

Tabla 1
Relación entre número de parejas y número de viviendas en poblaciones del NOA

	Localidad	AB	Ab	aB	ab	Parejas	Viviendas	Proporción
Valle Catchaquí	Cachi	84	51	64	186	385	387	0,994832
	San José	9	6	3	7	25	27	0,925926
	El Barrial	4	2	1	5	12	17	0,705882
	El Pichao	3	10	8	6	27	31	0,870968
	La Alumbraera	0	3	0	4	7	10	0,7
Valle de Lemna	Chicoana	124	66	77	97	364	361	1,00831
	La Isla	7	2	5	26	40	47	0,851064
	San Agustín	5	6	1	13	25	24	1,041667
Puna	Cobres	2	1	0	13	16	26	0,615385
	Chañarcito	1	2	2	12	17	18	0,944444
	Santa Rosa	3	5	4	2	14	22	0,636364
	Olacapato	1	5	1	11	18	24	0,75
	Tolar Grande	0	1	0	19	20	23	0,869565

Elección del tamaño muestral

Para elegir el tamaño de la muestra es necesario conocer la estructura de la población. En el caso aquí analizado, Cachi, la estructura es conocida. Cuando no se cuenta con esta información, es posible tomar una muestra exploratoria que estime esa estructura y luego ajustarla mediante los procedimientos que se detallan a continuación.

a) Partiendo de la estructura de parejas de la población de Cachi y fijando un intervalo de confianza, se tomaron diferentes valores de error y se calculó el tamaño de la muestra para cada categoría de pareja con la ecuación de Cochran (1980):

$$n = \frac{z^2 PQ}{d^2} \Big/ 1 + \frac{1}{N} \left(\frac{z^2 PQ}{d^2} - 1 \right)$$

donde:

z: valor de la distribución normal estandarizada que establece los límites del intervalo de confianza (1- α), para este trabajo: 0,95.

P: Proporción de cada categoría de parejas considerada (AB, Ab, aB, ab) respecto al total de parejas (385).

Q = 1-P.

N: N° parejas en la población.

d: error máximo tolerado: 0,01; 0,025; 0,05; 0,075; 0,10; 0,15 y 0,20.

b) Otra forma de abordar el problema, basado en el modelo anterior (a) y que incluye la corrección para tamaño poblacional finito, es definir el tamaño de la muestra como porcentaje del número de parejas (viviendas). En este caso será necesario estimar el error para los diferentes tamaños y categorías a partir de Cochran (1980):

donde:

$$d = z \sqrt{N - n / N - 1} \sqrt{PQ/n}$$

z: valor de la distribución normal estandarizada que establece los límites del intervalo de confianza (1- α), para este trabajo: 0,95.

N: N° parejas.

n: número de parejas en la muestra, calculado a partir de diferentes porcentajes de N.

P: Proporción de cada categoría de parejas considerada (AB, Ab, aB, ab) respecto al total de parejas.

Q = 1-P

Una alternativa al muestreo exploratorio para obtener la estructura de parejas, es fijar el error máximo. Se estimó el tamaño de la muestra para diferentes tamaños poblacionales (N) y P (0,5). Se consideró este valor de P ya que a esta frecuencia relativa es máxima la probabilidad de error. El cálculo de n se hizo a partir de la ecuación expresada en a), fijando diferentes errores.

Estimación del coeficiente de homogamia H

Se modelizó la estimación de frecuencias de parejas y coeficientes de homogamia (**H**) para Cachi a partir de muestras aleatorias que representan el 5, 10, 15, 20, 25, 30, 40 y 50% del total de parejas.

El coeficiente de homogamia **H** se estimó de acuerdo a Savorgnan (1949) como:

$$H = \frac{(ab * AB) - (Ab * aB)}{\sqrt{a * A * b * B}}$$

Para cada tamaño muestral se generaron 1000 repeticiones sin reposición utilizando el paquete estadístico SYSTAT 11. Estas repeticiones permitieron estimar: frecuencia media e intervalo de confianza (95%) para cada categoría de pareja y **H** medio e intervalo de confianza (95%).

Intervalos de Confianza para una muestra con estructura de población supeuestamente desconocida

Para evaluar los resultados de la aplicación de los procedimientos propuestos, se simuló la toma de una muestra aleatoria en la población de Cachi.

El tamaño muestral se definió de acuerdo a los criterios establecidos en 1): para un error (d) del 5%, el tamaño muestral será máximo si P=0,5. En consecuencia se tomaron 193 parejas aplicando un método de remuestreo sin reposición.

A partir de esta muestra se obtuvieron 10.000 muestras del mismo tamaño con reposición (bootstrap) (SYSTAT 11) y se estimó: frecuencia media e intervalo de confianza de las categorías y coeficiente de homogamia e intervalo de confianza.

Prueba de significación para el coeficiente de homogamia H

Si se trabaja con la población en su totalidad, el coeficiente de homogamia es el parámetro y no un estimador muestral, por lo tanto no será necesaria una prueba de significación para que pruebe el alejamiento de la panmixia.

En caso de trabajar con una muestra, dado que **H** se construye a partir de la diferencia entre los valores observados y esperados de las diferentes categorías de parejas (AB, Ab, aB y ab), una prueba de significación adecuada sería la de Chi cuadrado (con 1 grado de libertad ya que se estiman las frecuencias de cada categoría en cada sexo) y si la muestra es pequeña, el test exacto de Fisher. La hipótesis nula a contrastar será **H**=0.

Si Chi cuadrado es significativo, la población no será panmíctica y el coeficiente es significativamente distinto de cero.

Al ser **H** una expresión de la correlación entre varones y mujeres con idéntico estado para un carácter (en este caso el carácter es la procedencia y los estados son nativo y migrante), se puede considerar análogo al coeficiente de correlación para desequilibrio de ligamiento (D),

$$r = D / \sqrt{ABab}$$

Por lo tanto es posible utilizar programas de genética de poblaciones, por ejemplo ARLEQUÍN (Excoffier et al., 2005), ingresando las frecuencias absolutas de cada tipo de parejas en el lugar de los haplotipos para obtener los valores de Chi cuadrado: $X^2 = nH^2$, su probabilidad asociada p y r^2 .

La equivalencia entre parejas y haplotipos (en el caso de desequilibrio) estaría dada por la analogía expresada en la Tabla 2.

Tabla 2

Analogía entre parejas y haplotipos

		Mujeres (locus II)		
		Alelos	B (nativa)	b (migrante)
Varones (locus I)	A (nativo)	AB	Ab	
	a (migrante)	aB	ab	

sexo = locus
 carácter nativo/migrante = alelos
 parejas = haplotipos: AB, Ab, aB, ab

Dado que r^2 es siempre un valor positivo, para determinar si **H** indica homogamia o heterogamia será necesario comparar el total de homógamos vs heterógamos o bien se puede considerar el signo de D en la matriz de salida para homógamos y heterógamos. El signo (+) indica exceso y el (-) defecto.

Es importante tener en cuenta que en las poblaciones grandes, el mayor riesgo de error se dará cuando una categoría, en este caso parejas, se encuentre cercana a 0,5 mientras que en las poblaciones pequeñas el mayor riesgo estará asociado a frecuencias cercanas a 0, con la posibilidad de no inclusión de la categoría en la muestra.

RESULTADOS

En una primera instancia y partiendo de la estructura de parejas de la población de Cachi (Tabla 3) se estimaron tamaños muestrales para diferentes errores máximos tolerables (Figura 1).

Tabla 3
Estructura de la población de Cachi por parejas

Población	N	N Parejas	Categoría	Proporción
CACHI	1897	385	P(AB)	0,21818
			P(Ab)	0,13247
			P(aB)	0,16623
			P(ab)	0,48312

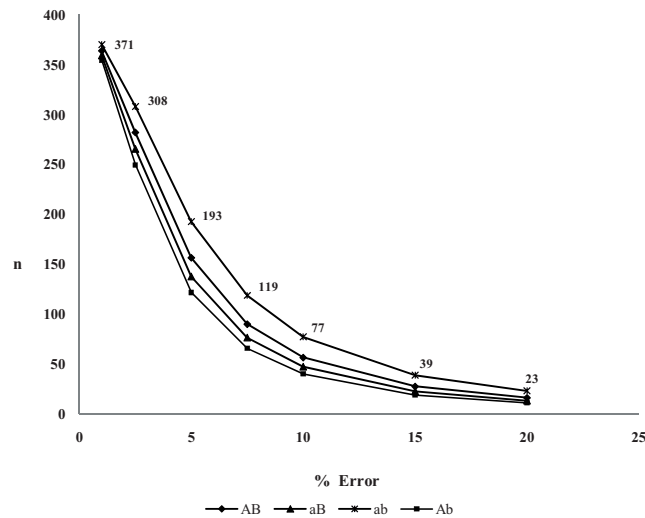


Figura 1
Tamaño muestral en función del error.

En la Figura 1 se observa que para cada error corresponden cuatro tamaños muestrales. Se desprende que dada una estructura de parejas, el tamaño muestral considerando un error determinado, varía según la categoría de pareja como consecuencia de que cada categoría presenta diferente frecuencia. Así, para un error del 5%, los valores extremos de n estimado son 121,28 y 192,43. Para cualquier valor de error, los mayores tamaños poblacionales son los calculados con base en la combinación ab (varón y mujer migrantes), cuya frecuencia es la más cercana a 0,5, valor para el que se estimará el mayor tamaño muestral ya que a esta frecuencia la desviación típica es mayor.

La Figura 2 muestra los diferentes errores cuando el tamaño muestral es una proporción fija del número de parejas de la población.

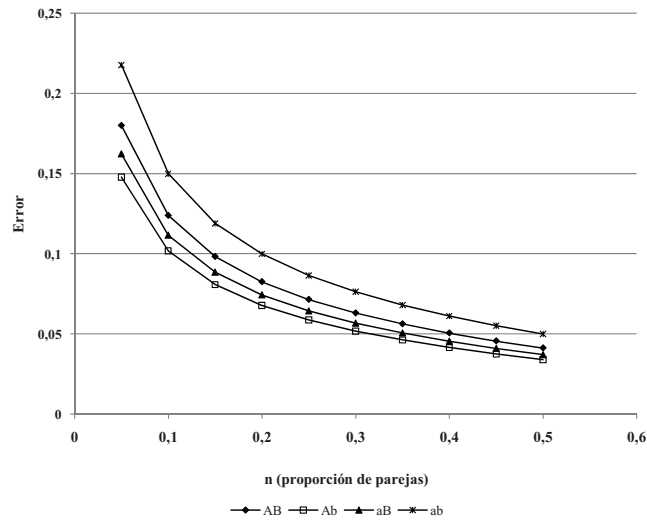


Figura 2

Diferentes errores para tamaño muestral como proporción fija de N parejas.

Si se parte del hecho de que lo frecuente es desconocer la estructura de pareja en la población, tal como fuera señalado en Material y Métodos, se puede estimar el máximo tamaño de la muestra considerando un error aceptable, por ejemplo del 5%, lo cual tendrá lugar cuando alguno de los tipos de pareja se encuentre a una frecuencia igual a 0,5 (Figura 3). Así, para el caso de Cachi, con 385 parejas y un error máximo del 5%, se deberían muestrear 193 parejas. En la Tabla 4 se representan los tamaños muestrales, expresados como porcentaje de N , correspondientes a diferentes errores.

TAMAÑO MUESTRAL Y NIVEL DE CONFIANZA

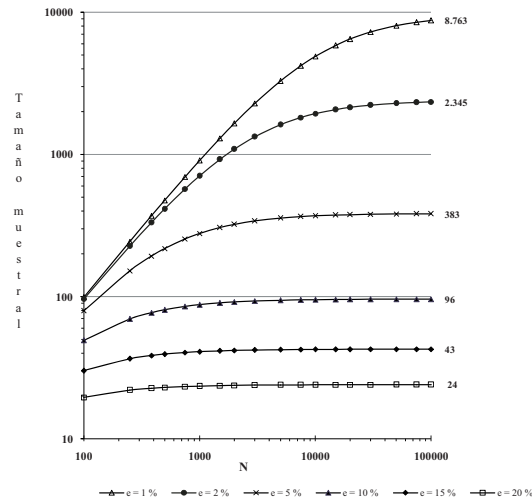


Figura 3
Estimación del máximo tamaño muestral para cada error, en función del tamaño poblacional (escala logarítmica).

Tabla 4
Tamaño de la muestra en % N para diferentes tamaños poblacionales y errores

N parejas	Error (%)					
	1	2	5	10	15	20
100	98,98	96,04	79,51	49,24	30,13	19,52
250	97,47	90,60	60,67	27,84	14,63	8,80
385	96,16	86,21	50,01	20,01	10,00	5,89
500	95,06	82,79	43,50	16,14	7,88	4,59
750	92,77	76,22	33,90	11,37	5,39	3,11
1.000	90,58	70,62	27,77	8,77	4,10	2,35
1.500	86,50	61,56	20,40	6,02	2,77	1,58
2.000	82,77	54,57	16,12	4,58	2,09	1,19
3.000	76,20	44,46	11,36	3,10	1,40	0,79
5.000	65,77	32,45	7,14	1,89	0,85	0,48
7.500	56,15	24,25	4,87	1,26	0,57	0,32
10.000	48,99	19,36	3,70	0,95	0,43	0,24
15.000	39,04	13,80	2,50	0,64	0,28	0,16
20.000	32,44	10,72	1,88	0,48	0,21	0,12
30.000	24,25	7,41	1,26	0,32	0,14	0,08
50.000	16,11	4,58	0,76	0,19	0,09	0,05
75.000	11,35	3,10	0,51	0,13	0,06	0,03
100.000	8,76	2,34	0,38	0,10	0,04	0,02

A fin de obtener una distribución de las estimaciones de las frecuencias relativas de las categorías de pareja y del coeficiente de homogamia en muestras cuyos tamaños fueron obtenidos como proporción de N, se realizó una simulación partiendo de una población con estructura de parejas conocida (Cachi).

La simulación obtenida a partir de 1000 muestras independientes sin reposición, permitió construir intervalos de confianza para las estimaciones (Tabla 5) y analizar el aumento de precisión con el incremento del tamaño de la muestra y la aproximación a los valores reales de los parámetros incluidos.

Tabla 5

Número promedio de parejas por categoría, media de **H** e intervalos de confianza (95%) para diferente tamaños muestrales

n	AB	Ab	aB	ab	coeficiente de homogamia <i>H</i>		
	Frecuencias Promedio (Intervalo 95 %)				Media	Intervalo (95 %)	
20	4,302 (4,189; 4,415)	2,73 (2,626; 2,824)	3,318 (3,216; 3,419)	9,65 (9,515; 9,785)	0,3483	0,3346	0,3620
39	8,576 (8,427; 8,725)	5,074 (4,952; 5,196)	6,493 (6,357;6,629)	18,857 (18,670; 19,043)	0,3644	0,3555	0,3732
58	12,782 (12,598;12,9669)	7,645 (7,503; 7,787)	9,673 (9,515; 9,831)	27,9 (27,680; 28,119)	0,3596	0,3520	0,3671
77	16,807 (16,611; 10,061)	10,224 (10,062; 10,386)	12,795 (12,615; 12,974)	37,174 (36,933;37,415)	0,3587	0,3527	0,3647
97	21,267 (21,052; 21,482)	12,856 (12,680; 3,032)	16,074 (15,879; 16,269)	46,803 (46,536;47,069)	0,3606	0,3554	0,3658
116	25,369 (25,129; 25,599)	15,357 (15,173; 15,541)	19,189 (18,968; 19,410)	56,085 (55,818; 56,351)	0,3613	0,3566	0,3660
185	40,482 (40,234;40,729)	24,524 (24,319; 24,729)	30,6 (30,366; 30,833)	89,394 (89,080; 89,708)	0,3609	0,3577	0,3642
193	42,08 (41,824; 42,336)	25,33 (25,126; 25,534)	32,444 (32,223; 32,665)	93,146 (92,844; 93,448)	0,3581	0,3551	0,3612

Por otra parte, se simuló la toma de una muestra aleatoria en la población de Cachi bajo el supuesto de desconocimiento de la estructura de parejas. El tamaño de la muestra obtenido es de 193 parejas considerando un error d del 5% y una proporción P de 0,5 mediante el método de remuestreo sin reposición.

Para construir el intervalo de confianza a partir de esta muestra, se obtuvieron 10.0000 réplicas mediante bootstrap (Tabla 6).

Tabla 6

Estimación del coeficiente de homogamia para una muestra al azar

Categorías de Parejas	Frecuencia (Intervalo 95 %)
AB	41 (29,871; 52,123)
Ab	23 (14,142; 31,856)
aB	34 (23,627; 44,373)
ab	95 (81,366; 108,634)
Coficiente de Homogamia	0,3642 (0,2284; 0,4964)

Se obtuvieron las distribuciones de las cuatro categorías de pareja y coeficientes de homogamia verificándose la inclusión de los parámetros reales de la población en los intervalos contruidos a partir de una muestra aleatoria de estructura desconocida.

Tanto el coeficiente de homogamia como la prueba de Chi cuadrado ($H=0,3641$, $X^2_{1\text{ gr}}=25,60$, $p<0,01$) calculados en base a la muestra de 193 parejas de Cachi y proporcionado por el programa ARLEQUÍN, coincidieron en su totalidad con los valores obtenidos mediante procedimientos de cálculo regulares (ver Material y Métodos).

DISCUSION

Considerando que para cada valor de error se calculan cuatro tamaños muestrales (uno para cada categoría), se puede optar por uno de los siguientes criterios para elegir el tamaño de la muestra, previa definición del error: a) tomar el valor máximo, asegurando de ese modo errores iguales o inferiores al elegido en todas las categorías; b) tomar un valor de n promedio entre los estimados para un mismo error en las cuatro categorías de pareja.

Otra observación es que las diferencias en las estimaciones del tamaño entre las categorías de pareja es menor cuando los errores son muy pequeños. En la medida en que aumenta el error, aumenta también la dispersión en los tamaños muestrales estimados, debiendo considerarse este efecto al momento de elegir el tamaño de la muestra en base al error.

Se observa que a partir de una muestra del 20% de las parejas de la población la caída del error es menos abrupta (Figura 2). Por otra parte, a medida que aumenta el tamaño de la muestra también hay menos diferencias entre los errores estimados para cada tipo de pareja. En una población como Cachi, muestreando entre el 40 y 50% de la población, el error de estimación de la frecuencia relativa de cada tipo de pareja estará entre 4 y 5%. A pesar de que parezca demasiado considerar un muestreo del 50% de la población, en términos de esfuerzo de muestreo relevar la mitad de las viviendas es una opción aceptable si consideramos que el error promedio entre las frecuencias de las cuatro categorías de parejas rondará el 5%.

La Figura 3 puede ser utilizada como referencia para estimar los tamaños muestrales, fijado el error para poblaciones de diferente N . En el caso de la población cosmopolita de la ciudad de Salta, en la que se estima que el número de viviendas es aproximadamente 100.000 y teniendo en cuenta el supuesto de que la relación pareja: vivienda es 1:1, el número de viviendas/parejas que se debería relevar es 383.

Si bien a priori, este tamaño de muestra parece insuficiente para incluir aquellas categorías de pareja cuya frecuencia sea muy baja en la población, utilizando la dis-

tribución hipergeométrica se calculó la probabilidad de que la muestra contenga al menos 1 (una) pareja de una categoría cuya frecuencia sea de 0,01 en la población, siendo esta probabilidad igual a 0,979. Por otra parte, la probabilidad de que la muestra contenga exactamente la frecuencia de la pareja en la población es de 0,196. Duplicando el tamaño de la muestra, la probabilidad de contener al menos 1 pareja es de 0,9995 y de contener exactamente la frecuencia de la categoría de pareja en la población es de 0,14.

De los porcentajes observados en la Tabla 4, se desprende que cuando las poblaciones son pequeñas, el esfuerzo de muestreo deberá ser mayor en términos relativos, considerando un mismo error. En las poblaciones pequeñas el principal error va a estar asociado a no incluir en la muestra aquellas categorías con frecuencias muy bajas, mientras que en las poblaciones grandes, el mayor error corresponderá a las categorías cuyas frecuencias se encuentren próximas a 0,5.

Con una muestra cercana al 20%, en una población de 100 parejas el error es de 0,2, en una de 1500 es de 0,05 y en una de 10.000 es de 0,02 (Tabla 4).

En la Tabla 5 se observa que a medida que se incrementa el tamaño muestral, se reduce la amplitud del intervalo de confianza y la media se aproxima a los parámetros conocidos consignados en la Tabla 2. En el caso de \mathbf{H} , siendo el valor poblacional 0,359, la muestra de 193 individuos seleccionada presenta un valor muy próximo.

Partiendo de una muestra aleatoria en una población de estructura desconocida y replicándola mediante bootstrap, se puede brindar una estimación del parámetro \mathbf{H} con su correspondiente intervalo de confianza.

En el caso de \mathbf{H} muestral es importante evaluar su alejamiento del valor 0 y el intervalo de confianza es una forma de hacerlo. Otras formas de evaluar la significación de \mathbf{H} son mediante la prueba de Chi cuadrado o el test exacto de Fisher.

El uso del programa ARLEQUÍN simplifica la obtención de \mathbf{H} y las pruebas de significación mediante el simple ingreso de las frecuencias de las categorías de pareja estimadas a partir del muestreo, constituyendo una valiosa herramienta cuando se trabaja con un volumen importante de datos.

CONCLUSIONES

Es necesario incluir las estimaciones de error cuando los parámetros de interés se expresan como proporciones.

Dado que en las poblaciones pequeñas las estimaciones de parámetros están sujetas a grandes errores, sería conveniente incluirlas en su totalidad.

En poblaciones pequeñas, el principal error está asociado a las categorías infrecuentes y en las poblaciones grandes, a aquellas cuyas frecuencias se aproximan a 0,5.

La analogía entre los modelos de desequilibrio de ligamiento y homogamia permite la utilización de programas genéticos en el cálculo de H y su nivel de significación.

BIBLIOGRAFIA CITADA

Acreche N y Albeza MV (2008) Selección de parejas y migración en la provincia de Salta. San Salvador de Jujuy, IX Jornadas Regionales de Investigación en Humanidades y Ciencias Sociales.

Cochran WG (1980) Técnicas de Muestreo. México, Cia. Editorial Continental.

Excoffier L, Laval G y Schneider S (2005) ARLEQUÍN ver.3.0: An integrated software package for population genetics data analysis.

<http://arlequin3http://cmpg.unibe.ch/software/arlequin3>.

Savorgnan F (1949) Matrimonial Selection and the Amalgamation for Heterogeneous Groups. Cultural Assimilation of Immigrants. Cambridge, Cambridge University Press, pp.59-67.

SYSTAT software versión 11.