

SUPUESTOS Y ESTIMACION DEL TAMAÑO EFECTIVO POBLACIONAL

Noemí Acreche

PALABRAS CLAVES: Deriva génica, Endogamia, Muestreo, Número reproductivo, Razón sexual

RESUMEN: Los efectos de muestreo o endogamia se estudian mediante aproximaciones semejantes siendo sus consecuencias el incremento de la varianza de frecuencias génicas inter-poblacional y la pérdida de variación intra-poblacional, fenómenos vinculados con el incumplimiento del supuesto tamaño infinito de las poblaciones. La variabilidad genética disminuye en estos casos y el Tamaño Efectivo de la Población, N_e , es un parámetro central en estudios cuyo objetivo es la cuantificación de la tasa a la cual la variabilidad genética es erosionada por la deriva génica. En situaciones reales, para medir los efectos de la deriva, es necesario estimar este indicador. Hacerlo sin hacer uso de datos genético-poblacionales tales como cambio en el nivel de heterocigosis o en las frecuencias génicas representa un importante aporte, considerando que esta información no siempre está disponible. Así, se puede contar con datos tales como número censal, distribución de edad y sexo, media y varianza del número de descendientes por padre y con ellos estimar niveles de endogamia y/o de variaciones en heterocigosis y frecuencias génicas. En la mayoría de los modelos, N_e corresponde a una población ideal de tamaño invariable, con igual número de ambos sexos, sin endocruzamiento, en la que cada individuo tiene la misma probabilidad de

Consejo de Investigación de la Universidad Nacional de Salta.

Cátedra de Fundamentos de Antropología Biológica y Cátedra de Evolución. Facultad de Ciencias Naturales. Universidad Nacional de Salta. Av. Bolivia 5150. 4400 Salta. Argentina.

e-mail: acreche@unsa.edu.ar

dejar progenie en la misma cantidad, de modo que el número de hijos por persona se distribuye aproximadamente según Poisson. Se estudiaron 12 poblaciones del NOA que no alcanzan el límite inferior propuesto para valores absolutos de N_e . Se emplearon para las estimaciones, diferentes propuestas que parten de diferentes supuestos, los que se discuten. La mayoría posee N_e efectivos menores que 50, por lo que se puede prever depresión endogámica a corto plazo. Los valores calculados son lo suficientemente bajos como para indicar que se trata de poblaciones en las que los efectos de deriva génica y endogamia pueden tener considerable incidencia. *Rev. Arg. Antrop. Biol.* 6(1): 9-18, 2004.

KEY WORDS: Genetic drift, Endogamy, Reproductive isolation, Sex rate

ABSTRACT: Sampling and endogamy are studied similarly, considering that they increase inter-population variance and decrease intra-population variability. In cases genetic variability is eroded, its rate may be estimated through Effective Population Size, N_e . In order to evaluate the effects of genetic drift, it is necessary to estimate this parameter. To do it without data such as changes in heterocigosity level or changes in allelic frequencies is an important contribution, considering such data are not always available. On the other hand, information such as censal number, age and sex distribution, mean and variance of the number of offspring is easier to obtain. In most models, N_e refers to an ideal population, not varying in size along time, with equal number of males and females, without inbreeding, in which every individual has equal probability of being parent of any one on the next generation. The number of sons per individual has a Poisson distribution. 12 population from North Western Argentina were studied and their Effective Population Sizes were estimated with different models. All have very low values, indicating strong subjection to genetic drift. Assumptions under each model are analysed and discussed. *Rev. Arg. Antrop. Biol.* 6(1): 9-18, 2004.

INTRODUCCION

Los procesos dispersivos son centrales en el estudio de la variabilidad genética de las poblaciones. La evaluación de la magnitud de los cambios que ocasionan puede ser encarada considerando los efectos de muestreo o de endogamia, ya que en ambos casos se incrementan la varianza entre subpoblaciones y la homocigosis como resultado de los tamaños finitos (Caballero, 1994).

El Tamaño Efectivo de la Población, N_e , es un parámetro fundamental en estudios cuyo objetivo es la comprensión del proceso evolutivo, puesto que cuantifica la tasa a la cual la variabilidad genética es erosionada por acción de la deriva génica (Vucetich et al., 1997).

Por lo menos en poblaciones no estructuradas, la deriva génica es cuantificada por un único parámetro: el Tamaño Efectivo de la Población (Nunney, 1993; Caballero, 1994). Esto lleva, en situaciones reales, a estimar este parámetro para medir los efectos de la deriva. Para ello es necesario considerar una serie de variables poblacionales partiendo de un número de supuestos ya que si bien el efecto dispersivo depende del número de individuos que integran una población, no todos contribuyen con sus gametas a la siguiente generación ni los que lo hacen lo hacen en la misma medida.

Según Hedrick (2000), el concepto de tamaño efectivo poblacional permite la consideración de una población de tamaño N_e en la que todos los adultos con descendencia tienen igual esperanza de ser los padres de cualquier descendiente individual. En otras palabras, las gametas son una muestra al azar de todos los individuos reproductores con probabilidad $1/N_e$ para cada uno producir una gameta en particular.

En este contexto, N_e sería el tamaño de una población ideal que produciría la misma cantidad de endogamia, varianza en frecuencia alélica o pérdida de heterocigosis que la población considerada.

Estimar la población efectiva sin hacer uso de parámetros estrictamente genético-poblacionales tales como el cambio en el nivel de heterocigosis o de frecuencias génicas, representa un importante aporte, considerando que esta información no siempre está disponible. Así, se puede disponer de datos tales como número censal, distribuciones de edad y sexo, media y varianza del número de descendientes por padre y a partir de ellos estimar los niveles de endogamia y/o variaciones en heterocigosis y frecuencias génicas.

Con esta información se podrá abordar el hecho de que no todos los individuos de una población estén en iguales condiciones reproductivas por factores biológicos y socioculturales y la esterilidad de una fracción de la población vinculada a uno o más de estos factores.

Como consecuencia, la estructura reproductiva de una población real generalmente difiere en sus características de las de la población ideal que cumple con ciertos supuestos establecidos y el número efectivo en una población natural puede ser considerablemente menor que el número de adultos en edad reproductiva por no cumplir con uno o más de los siguientes supuestos:

- Reproducción panmíctica,
- No sujeción a selección y correlación entre la fertilidad de los padres y la de sus descendientes,
- Generaciones discretas,

-Variabilidad en el número de descendientes por padres igual que la prevista por la distribución de Poisson,

-Igual número de varones y mujeres,

-Estabilidad temporal del número poblacional.

En la mayoría de los modelos, N_e corresponde a una población ideal de tamaño invariable, con igual número de ambos sexos, sin endocruzamiento, en la que cada individuo tiene idéntica probabilidad de contribuir a la próxima generación, de modo que el número de hijos por persona sigue una distribución binomial.

La distribución del número de progenie por padre se aproxima a una distribución de Poisson cuando N es grande. Si la media del número de hijos por padre, k , es igual a 2, la población se mantiene constante; en caso contrario estaría cambiando de tamaño, lo que incidiría en el tamaño efectivo. Esto no implica necesariamente un número constante de progenie por padre, sino un número variable al azar.

En efecto, la progenie sobreviviente puede no derivar al azar de la generación de sus padres, dependiendo esto del número de gametas aportado por cada individuo y su varianza (Crow y Kimura, 1970). Por otra parte, en poblaciones pequeñas y semiaisladas, la probabilidad de uniones consanguíneas aumenta en las familias en las que la progenie es más numerosa.

Se han propuesto diferentes aproximaciones que dan los mismos valores de N_e o muy similares cuando la población no tiene grandes fluctuaciones (Hedrick, 2000).

Se ha desarrollado una serie de aproximaciones a la estimación del número efectivo poblacional, de acuerdo a los supuestos que se consideran centrales. De este modo, han surgido varios “tipos” de N_e que pueden llevar a diferentes expresiones (Crow y Denniston, 1988).

METODOLOGIA

Si el número de individuos de ambos sexos difiere, se ha propuesto la estimación del número efectivo poblacional mediante la fórmula:

$$N_e = \frac{4N_{ef}N_{em}}{N_{ef} + N_{em}}$$

con N_m y N_f número de varones y de mujeres respectivamente (Crow y Kimura, 1970). En este caso, es el sexo menos frecuente el que más claramente influye en la estimación. Generalmente no entran en consideración, cuando de este modelo se trata, las distribuciones de edad.

El número que representa la población reproductora (N_r) incluye el conjunto de individuos de una población real que contribuye a formar la generación siguiente. Se define como el número de individuos en edad reproductiva con al menos un descendiente vivo entre 0 y 30 años o la fracción de la población constituida por los padres de los individuos que cuentan con entre 0 y 30 años de edad al momento de la recolección de la información.

Un modelo más utilizado considera el Número Reproductivo (N_r), la media y la varianza de los hijos sobrevivientes. En este caso:

$$N_e = \frac{2N_r - 2}{(k - 1) + \frac{V_k}{k}} \quad (\text{Magalhães y Arce Gómez, 1987})$$

siendo k el número promedio de hijos actualmente vivos por progenitor y V_k su varianza.

Se estimó la población reproductora, considerando en el grupo reproductivo masculino los mayores de 15 años y femenino mayores de 15 y menores de 49 años. Sin embargo, este número reproductivo puede no ser indicativo del tamaño real de la población que debe ser usado en estudios evolutivos.

Si bien se han propuesto otras aproximaciones, éstas son las más frecuentemente utilizadas. Una tercera aproximación a considerar en este trabajo integraría ambas propuestas mediante la estimación de N_c a partir del modelo que considera varianza y media reproductiva incluyendo en el cálculo, en lugar de N_r o simplemente N , N_c con la consiguiente consideración de la razón sexual. A tales efectos, se computarían los números efectivos de los respectivos sexos siendo:

$$N_e = \frac{4N_{ef}N_{em}}{N_{ef} + N_{em}}$$

Se estimaron los valores de N_r y N_c tomando en cuenta estas consideraciones para poblaciones de la Puna (Santa Rosa de los Pastos Grandes, Cobres, Tolar Grande, Olacapato y Chañarcito), del Valle Calchaquí (San José, El Barrial, El Pichao y La Alumbreira) y del Valle de Lerma (La Isla, San Agustín y Chicoana).

Con el objeto de relativizar los valores de N_c , se consideró su relación con el N censal. Esto tiene sentido para facilitar su valor comparativo. Sin embargo, la interpretación de la razón depende en gran medida del significado operativo del denominador. Según Vucetich et al. (1997), idealmente N debería representar siempre el número de individuos potencialmente reproductivos. Sin embargo, se puede considerar la inclusión de N_r en la razón, representando los individuos realmente reproductivos más apropiada.

Se estudió la distribución del número de hijos actualmente vivos para el total de los individuos por zona y por localidad. Asimismo se comparó la distribución de hijos actualmente vivos por individuo y su ajuste a una distribución de Poisson con media igual a dos.

Por otra parte, se puso a prueba la desviación de la razón sexual de la proporción ideal mediante el ajuste a una distribución Binomial en cada uno de los casos.

RESULTADOS

El número de hijos actualmente vivos, para todas las poblaciones estudiadas, incluyendo en el análisis los individuos mayores de 14 años, no se distribuye según Poisson ($p=0,00$). Resultados similares se obtuvieron para la población puneña en su conjunto ($p=0,038$), la del Valle Calchaquí ($p=0,002$) y la del Valle de Lerma ($p=0,00$).

En el análisis por localidad, se obtuvieron diferencias significativas para Chicoana ($p=0,00$), San Agustín ($p=0,043$) y San José ($p=0,013$).

Si se incluyen en el análisis sólo aquellos individuos que pertenecen a la población reproductora (N_r), para el total se obtuvieron valores de no correspondencia a la distribución de Poisson ($p=0,00$), lo mismo que para las tres zonas analizadas separadamente (Puna, $p=0,029$; Valle Calchaquí, $p=0,009$; Valle de Lerma, $p=0,00$). En el mismo análisis, por localidad, sólo Chicoana ($p=0,00$) presenta valores significativos.

En la Tabla 1 se consignan los valores de N_c para las 12 poblaciones estudiadas, calculadas según los tres modelos mencionados, a partir de datos biodemográficos. Asimismo, en la misma tabla se da cuenta de la información referente a N y las proporciones que los diferentes N efectivos representan de este total. En términos generales, se observa que al calcular N_c a partir de N reproductivo, Media y Varianza de Hijos Vivos y al calcularlo ajustando estos mismos valores a la razón sexual no se presentan mayores diferencias, mientras que teniendo en cuenta sólo la razón sexual, los valores son verdaderamente dispares.

En estas poblaciones, la varianza reproductiva es alta, sugiriendo una distribución no aleatoria de gametas por padre, que puede deberse a factores genéticos, ambientales o accidentales. La razón V_k/k es un índice de la dirección y cantidad en que el número efectivo se desvía del real.

Tabla 1

Localidades y zonas: K_r media de hijos actualmente vivos, población reproductora; Vk_r varianza de hijos actualmente vivos, población reproductora. Valores de N_c estimados considerando Razón sexual (N_{c1}); N_r , media y varianza de hijos (N_{e2}) y todas las variables (N_{e3}) y sus relaciones con N reproductivo

Zona	Lugar	N	K_r	Vk_r	Total	N_r		N_c			N_{e1}/N_r	N_{e2}/N_r	N_{e3}/N_r
						Masculino	Femenino	N_{c1}	N_{c2}	N_{c3}			
Valle Catchaqui	La Alumbraera	45	5,21	5,57	14	7	7	14,00	4,92	4,51	1,00	0,35	0,32
	El Pichao	133	3,43	5,62	42	23	19	41,62	20,15	20,23	0,99	0,48	0,48
	El Barrial	62	4,29	10,01	21	12	9	20,57	7,11	9,02	0,98	0,34	0,43
	San José	154	4,04	6,83	50	24	26	49,92	20,72	20,43	1,00	0,41	0,41
Valle de Lerma	Chicoana	2097	3,87	5,98	656	328	328	656,00	296,70	298,57	1,00	0,45	0,45
	La Isla	242	4,03	4,90	75	41	34	70,65	35,39	35,10	0,94	0,47	0,47
	San Agustín	129	3,79	8,51	42	23	19	41,62	16,28	15,64	0,99	0,39	0,37
Puna	Cobres	141	4,29	8,35	42	16	26	39,62	15,66	14,19	0,94	0,37	0,34
	Chañarcito	137	5,19	7,02	36	17	19	35,89	12,63	12,13	1,00	0,35	0,34
	Olacapato	150	5,34	7,48	41	22	19	40,78	13,93	13,36	0,99	0,34	0,33
	Santa Rosa	114	4,74	7,46	31	16	15	30,97	11,29	10,80	1,00	0,36	0,35
	Tolar Grande	99	3,28	5,58	36	18	18	36,00	17,58	16,96	1,00	0,49	0,47

DISCUSION

Para Wright (1978), el concepto de Número Efectivo Poblacional es útil para propósitos comparativos y se transforma en una necesidad práctica cuando se trata con poblaciones naturales. Sin embargo, al no existir uniformidad en los enfoques para la estimación de este parámetro, la comparación se dificulta. Como se ha visto, diversas variables entran en el cálculo: distribución, media y varianza del número de hijos actualmente vivos, razón sexual, estructura de edad, solapamiento generacional, apareamientos no panmícticos y variaciones en el tamaño de la población, entre los más relevantes.

A partir de los primeros trabajos de Wright (1931) en los que se define el problema del N efectivo, se ha propuesto una serie de modelos para aproximar la estimación (Crow y Kimura, 1970). Como se ha mencionado, entre los más utilizados se encuentra el que considera las desviaciones en la razón sexual. De los supuestos establecidos, este modelo sólo considera uno, por lo que los valores que se alcanzan podrían estar alejados de los reales.

Nunney (1993) precisa que una razón sexual sesgada puede ser entendida en términos de éxito reproductivo ya que los individuos del sexo raro tienen un éxito mayor que los del más común, de modo que la varianza general aumenta en la medida en que aumenta el sesgo en la razón sexual. Se puede esperar que N_c sea máximo cuando el número de varones es igual al de mujeres, igual a $N/2$. Por otra parte, según Fontdevilla y Moya (2000), la desviación de la razón sexual provoca una dramática disminución de N_c , razón por la cual tampoco sería recomendable ignorar este parámetro.

La razón sexual puede variar ampliamente no sólo en el espacio, sino también en el tiempo, por lo que sería importante realizar estudios que abarquen varias generaciones. Salzano y Callegari-Jacques (1988) reportan datos demográficos de grupos Toba que en 1942 presentaban una razón sexual de 136 y en 1970 de 88.

Para el segundo modelo utilizado, se considera el valor de N_r . En el cálculo de este parámetro, la restricción etaria se refiere al reemplazo generacional en nuestra especie (Freire-Maia, 1974). A partir de la definición, quedan excluidos todos aquellos individuos que no han alcanzado aún la edad reproductora, los que la han dejado atrás y los que, por razones biológicas o sociales, no participan de la reproducción.

En cuanto al tiempo de generación, se ha fijado en 30 años. Sin embargo, es un valor que puede variar entre poblaciones y en el tiempo. En este caso sería más adecuado considerar en N_r a los padres de los hijos cuya edad máxima fuera la estimación indirecta del tiempo de generación (ya que no se puede construir una tabla de vida): la edad media al nacimiento del primer hijo, dato con el que se cuenta para todas las poblaciones aquí consideradas. Se mantuvo no obstante el límite de 30 años, a fin de mantener la posibilidad de comparar resultados con los de la bibliografía, con la consecuencia de que los valores de N_c quedarían sobreestimados.

En este caso se considera la varianza reproductiva y es apropiada para poblaciones en las que la misma alcanza valores considerables. Esta varianza sugiere una distribución no aleatoria de las gametas por padre. Asimismo, al entrar en consideración N_r en lugar de N o simplemente el número de adultos, entra en consideración la corrección para el solapamiento generacional.

En una población ideal, N y N_e son iguales. Si el número de hijos tiene mayor varianza que la esperada, N_e es menor que N , mientras que si la varianza es menor, N_e será mayor (Crow y Kimura, 1970). De este modo, la razón V_k/k es un índice de la dirección y cantidad por la cual el número efectivo se desvía del real (Crow y Denniston, 1988).

Hedrick (2000) menciona las poblaciones humanas de Japón y Suecia en las que el tamaño de la familia es uniforme, lo que lleva N_e a valores mayores que los de N_r .

En la tercera aproximación se ha considerado un número de variables poblacionales que inciden sobre la determinación y estimación del número efectivo poblacional. Se ha calculado N_e para cada sexo separadamente, tomando en consideración el respectivo N_r (con lo que entra en consideración el solapamiento generacional), la media y la varianza de hijos actualmente vivos (con lo que se atiende al problema de las desviaciones de distribución teórica, a las fluctuaciones en el tamaño y en alguna medida al supuesto de no acción de la selección natural) y finalmente se han utilizado los N_e de cada sexo en la estimación final haciendo uso de la fórmula que considera las desviaciones de la razón sexual.

Es esta última aproximación, por lo tanto, la que se considera más apropiada para la estimación del Número Efectivo Poblacional por incluir correcciones para el mayor número de las desviaciones de los supuestos que subyacen en estos modelos.

Nunney y Campbell (1993) señalan que bajo un amplio rango de condiciones, la razón N_e/N estaría en el intervalo 0,25-1,00. En un trabajo más reciente Vucetich et al. (1997) llegaron a la conclusión de que N_e/N está generalmente cerca de 0,5 y sólo raramente fuera del rango 0,25-0,75. Tomando estos valores como referencia, en ninguna de las localidades estudiadas se alcanza este piso.

CONCLUSIONES

Las poblaciones estudiadas se encuentran sujetas al efecto de la deriva génica, con tamaños efectivos muy por debajo de los considerados como mínimos necesarios para el mantenimiento de la variabilidad.

A los efectos de comparar resultados, es necesario explicitar los supuestos y se sugiere la utilización del modelo 3 propuesto en este trabajo para la estimación de N_e .

BIBLIOGRAFIA CITADA

Caballero A (1994) Developments in the prediction of effective population size. *Heredity* 73:657-679.

Crow JF y Denniston C (1988) Inbreeding and variance effective population numbers. *Evolution* 42(3):482-495.

Crow JF y Kimura M (1970) *An Introduction to Population Genetics Theory*. New York, Harper and Row.

Fontdevilla A y Moya A (2000) *Introducción a la Genética de Poblaciones*. España, Editorial Síntesis.

Freire-Maia N (1974) *Genética de Populações Humanas*. Universidade de São Paulo, HUCITEC Editora.

Hedrick PW (2000) *Genetics of Populations*. Sudbury, Massachusetts. Jones and Bartlett Publishers, Inc.

Magalhães JCM y Arce-Gómez B (1987) Study on a brazilian isolate I. Population structure and random genetic drift. *Hum. Hered.* 37:278-284.

Nunney L (1993) The influence of mating system and overlapping generations on effective population size. *Evolution* 47(5):1329-1341.

Nunney L y Campbell K (1993) Assessing minimum viable population size: demography meets population genetics. *Tree* 8(7):234-239.

Salzano F y Callegari-Jacques SM (1988) *South American Indians. A Case Study in Evolution*. Oxford, Clarendon Press.

Vucetich JA, Waite TA y Nunney L (1997) Fluctuating population size and the ratio of effective to census population size. *Evolution* 51(6):2017-2021.

Wright S (1978) *Evolution and the Genetics of Populations*. Vol. 4: Variability within and among natural populations. Chicago University Press.

Wright S (1931) Evolution in mendelian populations. *Genetics* 16:97-159.