ESTADÍSTICA ESPAÑOLA
Vol. 38, Núm. 141, 1996, págs. 237 a 244

Acerca de la finitud en las poblaciones

por
VICENTE MANZANO ARRONDO
Departamento de Psicología Experimental
Universidad de Sevilla

RESUMEN

Las expresiones «población finita» y «población infinita» son frecuentemente utilizadas. Pero su uso no viene acompañado de una comprensión fundamentada de qué es lo que significan cada uno de estos términos y en qué circunstancias puede considerarse a una población finita o infinita. En este trabajo se pretende facilitar recursos prácticos para definirse acerca de la finitud de una población, en el contexto del muestreo aleatorio simple.

Palabras clave: poblaciones finitas, poblaciones infinitas, muestreo.

Clasificación AMS: 62D05.

1. INTRODUCCIÓN

En *teoría* del muestreo, al igual que en cualquier otro área de conocimiento teórico-matemático, la preocupación básica no es la adecuación de los modelos matemáticos a las situaciones prácticas, sino la generación de estos modelos bajo determinados supuestos. Debido a esta circunstancia, es cada vez más frecuente encontrar la expresión *la práctica del muestreo* (Salgado, 1990) como conjunto de conocimientos generados a raíz de adecuar *la teoría del muestreo* a los imperativos fácticos. Una de las puntas de este iceberg es el hecho de que si bien la teoría del muestreo contempla poblaciones finitas e infinitas (el espacio muestral compuesto por infinitas tiradas de un dado, por ejemplo), en el caso de las investigaciones en contextos sociales, lo más que nos encontramos son grandes colectivos humanos, pero estrictamente hablando no son poblaciones infinitas.

Si no existiera ningún otro tipo de información, se debería observar que ningún investigador social llega a utilizar la expresión «población infinita», ni los algoritmos algebraicos que genera la teoría del muestreo para el cálculo de errores típicos y tamaños de muestra en el contexto de estas poblaciones. Pero no es así. Tanto para el cálculo del tamaño de la muestra como para el proceso de estimación de las funciones poblacionales, los investigadores en ciencias sociales utilizan frecuentemente el recurso de las poblaciones de tamaño infinito. Considerando la circunstancia de que, estrictamente hablando, no existen las poblaciones infinitas en ciencias sociales, el obviar este hecho puede llevar a consecuencias negativas. ¿Qué consecuencias?: estimar tamaños de muestra, varianzas y errores incorrectos por sobreestimados, incrementando el coste global del estudio. ¿En qué medida el error es despreciable?: conforme el tamaño de la población se acerque más al infinito. No obstante, lo cerca o lejos que el tamaño de una población se encuentre del considerado *prácticamente infinito* no es constante en cualquier situación, sino que depende del contexto concreto: estimador, función poblacional a estimar, riesgos I y II, error de precisión o tamaño del efecto o el modelo de muestreo.

Se puede considerar que la utilización de expresiones de cálculo bajo el supuesto de poblaciones de tamaños infinitos, constituye una actitud conservadora, ya que tales expresiones son cotas superiores para sus homólogas en poblaciones finitas. No obstante, este comportamiento atenta claramente contra el criterio de tomar decisiones óptimas. En términos del tamaño de la muestra, se buscaría una cantidad tal que consiga la máxima precisión con el mínimo coste (por ejemplo, Sukhatme, 1953; Raj, 1968; Cochran, 1976; Stuart, 1984; Kalton, 1987 o Salgado, 1990). Considérese que "El costo de los procesos puede ser mucho más determinante que cualquier consideración teórica" (Silva, 1993:142).

En concreto, cabe señalar tres situaciones problemáticas en relación a la finitud de las poblaciones:

1. Existe una clara confusión entre la finitud de la población y los modelos de muestreo *con* y *sin reposición*.

2. El hábito y la comodidad allanan el camino para la utilización de las expresiones algebraicas que corresponden al contexto de las poblaciones infinitas. Estas

expresiones son siempre más fáciles y cómodas de aplicar, gracias a que consideran menos elementos para el cálculo y resultan ser más breves. Considerando que la decisión sobre qué técnicas estadísticas utilizar depende muchas veces de la comodidad de su uso (Lebart y otros, 1985) más que a su adecuación, resulta comprensible el que algunos investigadores caigan en la trampa de escoger las expresiones más sencillas precisamente por su sencillez, inclusive cuando puede llegar a insistirse en ello desde la literatura específica (Kish, 1965; Veres, 1981).

3. Ya que, en la práctica social, no existen poblaciones con infinita cantidad de elementos, corren algunas informaciones acerca de *a partir de qué tamaño puede considerarse una población como prácticamente infinita.*

2. ¿A PARTIR DE QUÉ TAMAÑO UNA POBLACIÓN ES "PRÁCTICAMENTE" INFINITA?

Si bien se ha señalado ya la inexistencia de poblaciones estrictamente infinitas en el entorno social, puede ocurrir, no obstante, que la población cuente con un número de unidades tal que la diferencia entre los resultados que generan las expresiones para situaciones de finitud o infinitud sea despreciable. El problema que surge entonces es ¿Cuándo una población estrictamente finita puede ser considerada como prácticamente infinita?. Para responder a la pregunta con algún viso de utilidad, es necesario circunscribir las deducciones a un contexto concreto. Para ello, los contenidos que siguen se refieren a la estimación de una media aritmética en el muestreo aleatorio simple. Esta decisión nos permitirá expresar la *infinitud práctica* de las poblaciones en función del nivel de significación y del error de precisión o radio del intervalo de estimación.

Para ello, denominamos:

n_N tamaño de muestra para una población finita

n_∞ tamaño de muestra para una población infinita

N tamaño finito de la población, y

$$k = \frac{n_N}{n_\infty} \qquad [1]$$

En principio, el supuesto "población infinita" es asumible únicamente cuando la discrepancia

$$dn = |n_N - n_\infty| = 0$$

No obstante, es razonable admitir una distancia $dn \neq 0$ siempre que su valor no exceda algún umbral δ predeterminado como el máximo asumible. Es decir, el calificativo "infinita" para una población sería aceptable únicamente cuando $dn \leq \delta$.

Sin embargo, la distancia absoluta *dn* no es una buena medida para evaluar la discrepancia, considerando su sensibilidad a los tamaños muestrales. Mejor estrategia es operar con un índice relativo: la razón entre ambos tamaños (razón de similitud, k). De esta forma y dado que

$$0 \leq n_N \leq n_\infty$$

dividiendo todos los términos entre n_∞ y según [1]:

$$0 \leq k \leq 1$$

De la expresión para el cálculo del tamaño de la muestra para la estimación de medias en el muestreo aleatorio simple (1):

$$n_N = \frac{N}{\frac{N-1}{n_\infty}+1} = k n_\infty \rightarrow$$

$$\rightarrow N = k n_\infty \left[\frac{N-1+n_\infty}{n_\infty}\right] = k(N+n_\infty-1) \rightarrow$$

$$\rightarrow N-kN = (n_\infty-1)k \rightarrow N = (n_\infty-1)\frac{k}{1-k} \qquad [2]$$

Por otro lado, en el muestreo aleatorio simple para la estimación de una media aritmética, el tamaño de la muestra se calcula según:

$$n_\infty = \left(\frac{Z_{\alpha/2}\,\sigma}{e_p}\right)^2$$

donde $Z_{\alpha/2}$ es la distancia estandarizada que corresponde a una probabilidad α, σ es la desviación tipo poblacional y e_p es el error de precisión o radio del intervalo de

(1) Las expresiones de cálculo están basadas en Azorín y Sánchez Crespo (1986), si bien pueden encontrarse en cualquier texto sobre muestreo.

estimación. Incluyendo esta expresión en [2] e indicando el error de precisión estandarizado $\left(Z_e = e_p / \sigma\right)$, se obtiene:

$$N = \left(\frac{Z_{\alpha/2}^2}{Z_e^2} - 1\right)\frac{k}{k-1} \qquad [3]$$

Para estudiar el comportamiento de [3] se requieren expresiones concretas para $Z_{\alpha/2}$ y para Z_e. En el primer caso, los niveles de significación más usuales son .05 y .01 (por ejemplo, Rosnow y Rosenthal, 1989; Cohen, 1992; Manzano, 1997). Por otro lado, a la hora de traducir estos valores de probabilidad a distancias estandarizadas $Z_{\alpha/2}$ y si bien es habitual considerar la distribución muestral normal, contemplaremos también el caso de total desconocimiento (acotación de Chebyshev). Con ello, las distancias estandarizadas correspondientes se muestran en el cuadro 1.

Cuadro 1
VALORES DE $Z_{\alpha/2}$

	Normal	*Chebyshev*
α = .01	2.5758	10.0000
α = .05	1.9600	4.4721

Con respecto a Z_e, una estrategia satisfactoria es la consideración de un error de precisión que se corresponda aproximadamente con 1/10 o 1/20 de la desviación tipo (Manzano, 1996), en cuyo caso y respectivamente, Z_e podría acotarse entre .1 y .05

Tabla 1
TAMAÑOS POBLACIONALES "PRACTICAMENTE" INFINITOS

Ze=.1				*Ze=.05*				
Normal		*Chebyshev*		*Normal*		*Chebyshev*		
$\alpha \rightarrow$.05	*.01*	*.05*	*.01*	*.05*	*.01*	*.05*	*.01*	*k*
3448	5962	17991	89991	13821	23876	71990	359991	.90
3874	6698	20212	101101	15527	26824	80877	404434	.91
4406	7618	22988	114988	17660	30508	91987	459988	.92
5091	8801	26558	132844	20402	35246	106271	531415	.93
6003	10379	31317	156651	24058	41562	125315	626650	.94
7280	12587	37980	189981	29177	50405	151978	759980	.95
9196	15899	47975	239975	36855	63669	191972	959974	.96
12389	21420	64633	323300	49652	85777	258629	1293297	.97
18775	32461	97949	489948	75246	129991	391943	1959941	.98
37932	65584	197896	989890	152027	262634	791879	3959857	.99

Con los valores especificados para las dos distancias estandarizadas implicadas en el cálculo de un tamaño para la muestra, la expresión (3) suministra los resultados que figuran en la tabla 1. En ésta, se han considerado valores para k entre .99 y .9. Con ello, se tiene por ejemplo que una población de 7280 elementos puede ser considerada ya como prácticamente infinita en el caso habitual de una estimación de medias o proporciones con α=.05, un nivel de exigencia moderado para la precisión (Z_e=.1) y una razón k=.95.

3. CONCLUSIONES

La observación de la tabla 1 debería llevar a la conclusión de que la consideración de una población finita como *prácticamente* infinita no se basa en una única cantidad sino que ésta depende de multitud de factores de los que hemos realizado una exposición únicamente en el caso de la estimación unitaria de medias en el muestreo aleatorio simple. Cabe esperar variaciones sensibles en otros contextos. La solución conservadora no es recomendable aquí, ya que genera estimaciones de varianzas y decisiones sobre tamaños muestrales innecesariamente elevados, cuando existe información suficiente (expresiones de cálculo pertinentes) como para aspirar a una solución más cercana a la óptima.

Las consecuencias de la mencionada actitud conservadora son difícilmente evaluables. Un aspecto positivo es el aumento de la precisión en las estimaciones o disminución de los errores asociados. No obstante, este supuesto efecto positivo sólo es posible si el aumento en el número de unidades seleccionadas, con respecto al tamaño óptimo, no viene acompañado por una disminución en la capacidad de control durante la investigación. El aspecto negativo viene representado por un incremento injustificado del coste total del estudio, en función de la distancia absoluta entre tamaños *dn*.

Por último, es recomendable acudir siempre a las expresiones para tamaños finitos en los estudios aplicados. Entre otras consecuencias, este hábito aumenta la probabilidad de conseguir un tamaño óptimo de muestra. En contraposición, las expresiones de cálculo son más complejas. No obstante, el tiempo implicado aún bajo el supuesto de realizar las operaciones con papel y lápiz, es despreciable con respecto al tiempo total que exige el diseño de la investigación.

REFERENCIAS

AZORÍN, F.; SÁNCHEZ CRESPO, J.L. (1986). «Métodos y aplicaciones del muestreo». *Colección Alianza Universidad Textos*. Madrid: Alianza Editorial.

COHEN, J. (1992). «Cosas que he aprendido (hasta ahora)». *Anales de Psicología*, 8, 3-17.

COCHRAN, WILLIAM G. (1976). «Técnicas de muestreo». México: Compañía Editorial Continental S.A.

KALTON, GRAHAM (1987). «Introduction to survey sampling». *SAGE University Paper. Beverly Hills*: SAGE Publications, Inc.

KISH, L. (1965). «Survey Sampling». New York: John Wiley & Sons.

LEBART, L.; MORINEAU, A.; Y FÉNELON, J. P. (1985). «Tratamiento estadístico de datos. Métodos y programas». Barcelona: Marcombo.

MANZANO, V. (1996). «Tamaño óptimo de muestra en investigación mediante encuestas. Fundamentos e implementación de un sistema de ayuda a la decisión». *Tesis doctoral no publicada*. Facultad de Psicología. Universidad de Sevilla.

MANZANO, V. (1997). «Usos y abusos del error tipo I». *Psicológica*. 18, 2, 153-169.

RAJ, DES (1968). «Sampling Theory». New York: McGraw-Hill.

ROSNOW, R.L. y ROSENTHAL, R. (1989). «Statistical procedures and the justification of knowledge in psychological science». *American Psychologist*, 44, 1276-1284.

SALGADO CARRIÓN, JOSÉ ANTONIO (1990). «La práctica del muestreo». En Ortega, E. (Ed.) *Manual de investigación comercial*. Madrid: Pirámide. Páginas 344 a 377.

SILVA AYÇAGUER, LUÍS CARLOS (1993). «Muestreo para la investigación en ciencias de la salud». Madrid: Díaz de Santos.

STUART, ALLAN (1984). «The Ideas of Sampling». High Wycombe (England): Charles Griffin & CO. Ltd.

SUKHATME, PANDURANG V. (1953). «Sampling theory with applications». Iowa: Iowa State College Press.

VERES, E. (1981). «Tamaño muestral óptimo de maximización del valor esperado de un experimento». *Estadística española*, 92,47-65.

ABOUT THE FINITE OR INFINITE POPULATIONS

SUMMARY

The expressions «finite population» and «infinite population» they are frequently used. But its use does not come accompanied of a based comprehension of what is what mean each one of these terms and in what circumstances can be considered to an infinite or finite population. In this work we intend to facilitate practical resources to be defined about the finite or infinite population, on the simple random sampling context.

Key words: finite populations, infinite populations, sampling.

AMS Classification: 62D05.