

El Muestreo Alfabético visto desde el Muestreo Estratificado: comentarios y consecuencias

Lourdes Ramos¹
Juan Piccini²
Alejandra Villar³

Resumen. En Archivología se llama Muestreo Alfabético a una de las herramientas usadas para tomar muestras de poblaciones documentales donde cada documento está asociado a una persona. Se ha señalado como principal desventaja de esta herramienta la falta de garantías en cuanto a que la muestra sea representativa. En este artículo veremos como subsanar dicha desventaja. El proceso nos llevará a su vez a considerar al Muestreo Alfabético como un caso del Muestreo Estratificado. En la Sección 1 mostramos algunas peculiaridades del muestreo en poblaciones documentales. En la Sección 2 definimos el Muestreo Alfabético y discutimos algunos aspectos del mismo. En la Sección 3 mostraremos como mejorar dicho método. En la Sección 4 repasamos el Muestreo Estratificado, la Sección 5 conecta ambas técnicas y la Sección 6 presenta las conclusiones.

Palabras clave: archivos; muestreo; evaluación documental

**Title: Alphabetical Sampling seen from Stratified Sampling:
comments and consequences.**

Abstract. Alphabetical sampling is a tool used by archivists to take samples from populations of records associated to individuals. One major drawback of this tool is the fault of guarantees about sample

1 Archivóloga, Docente G° 3, Profesora Encargada de Curso de la Asignatura Selección Documental de la carrera Archivología de la Escuela Universitaria de Bibliotecología y Ciencias Afines, Universidad de la República. louramosv@gmail.com

2 Master en Ingeniería Matemática, Licenciado en Matemáticas. Docente G°3, Profesor del Instituto de Matemáticas y Estadística de la Facultad de Ingeniería, Universidad de la República. jepiccini@gmail.com

3 Archivóloga, Docente G° 3, Profesora de la Asignatura Introducción a la Archivología de la carrera Archivología de la Escuela Universitaria de Bibliotecología y Ciencias Afines, Universidad de la República. mava1963@gmail.com

representativeness. We show how to improve the method and how this leads us to include this tool in the Stratified Sampling frame, simplifying the theoretical framework. This paper is organized as follows: Section 1 shows some peculiarities of sampling in archives. In Section 2 we define Alphabetical Sampling and discuss some aspects of the same. Section 3 shows how to improve it. In Section 4 we define Stratified Sampling. Section 5 connects both methods. Section 6 contains the conclusions.

Keywords: Archives; Sampling; Documentary Rating.

Introducción

Cuando se desea estudiar una cierta característica de una población, generalmente no es posible o es prohibitivo estudiar a la totalidad de dicha población, por lo se elige una muestra de la misma con la intención de estudiar en ella la característica de interés, para luego extrapolar a la población lo aprendido allí.

Para ello la muestra debe ser “representativa”, término un tanto elusivo que expresa el deseo de que se “comporte igual” o que “sea igual” a la población, en lo que refiere a la característica bajo estudio. El muestreo será aleatorio o no aleatorio según el azar sea o no parte importante en la selección de la muestra.

En archivología, elegir una muestra generalmente implica la eliminación del resto de la población que suministró la muestra. Por tanto debe ser representativa, en el sentido de que contenga los rasgos esenciales de la población a la que pertenecía y que permita eventualmente reconstruir en lo esencial a esa población que ya no podrá ser estudiada.

En palabras de Félix Hull (1981, 13-14)

Según una útil definición general y no técnica del muestreo, se trata de obtener “una pequeña parte tomada de un conjunto y que ponga de manifiesto las cualidades de la masa”, y se dice que el hecho de tomar una muestra permite “obtener una experiencia representativa” o “someter a prueba las cualidades” del objeto de tal operación. De ahí se desprende que, en un sentido archivístico, el muestreo es la selección deliberada de determinados documentos (expedientes, volúmenes o cualesquiera otros formatos normales de una clase o serie) como ejemplo de esa clase o serie y con fines de conservación permanente... Por supuesto, en cierto sentido toda selección, en cualquier nivel, constituye una forma de muestreo, y esta operación empieza a efectuarse en cuanto se toma la decisión de conservar

menos de la totalidad... En otras palabras, el muestreo no es una verificación que impone la administración que creó el registro sino una técnica relacionada con las necesidades en materia de consulta e investigación... Expresándolo de un modo muy general, una muestra puede perseguir dos objetivos: o bien intentar conservar lo que, por consideraciones subjetivas, parece ser el elemento más significativo o importante de la serie o bien apuntar a hacer una selección totalmente objetiva en la cual el contenido real se ciña... En todos los casos, el hecho de tomar una muestra de una clase o serie de registros implica, y normalmente entraña, la destrucción del resto. Esta es la línea divisoria que separa al investigador del archivero, y debido a ella el muestreo resulta una técnica de archivos difícil y peligrosa... Cuando un especialista estime que necesita una muestra para sus fines estadísticos, escogerá, con arreglo a una base apropiada, una parte del conjunto, pero dejando que ese conjunto se reconstituya y vuelva a utilizarse una y otra vez. En cambio, cuando es el archivero quien toma la muestra, deja la porción escogida como conjunto futuro: el resto ha desaparecido y no podrá emplearse nunca más.

Aunque podemos discutir sobre el grado de irreversibilidad que tiene el tomar una muestra (esto se cumple cuando se toma una muestra de una población de series cerradas, no tanto para series abiertas, por cuanto el sistema sigue generando documentos similares a aquellos que son eliminados, por lo que la población documental se regenera de forma similar a una población biológica), nadie objetará la necesidad de que la muestra sea representativa. En el caso archivístico se tiene además la dificultad adicional de no saber a ciencia cierta (cuando la muestra se toma para servir de base a futuras investigaciones), cuáles son los rasgos o características distintivas de la población que será eliminada, exceptuando esa muestra.

En la próxima sección veremos lo que en Archivística se denomina “Muestreo Alfabético” y extraeremos algunas conclusiones que nos servirán como insumo para mejorar la calidad de las muestras obtenidas por este método. Como resultado llegaremos al “Muestreo Alfabético Mejorado”.

El Muestreo Alfabético

Supongamos un conjunto de documentos asociados a personas. Se elige una o varias letras del alfabeto y solamente se conservarán aquellos documentos donde el apellido de la persona comienza con una de las letras seleccionadas.

Citando a Hull nuevamente:

el método adoptado en Colonia para el muestreo de registros de la administración social consiste en conservar todos los expedientes de

las personas cuyo apellido empieza por H. Se nos dice que este método proporciona una muestra de un 8.5 %, y que ha sido recomendado como estadísticamente válido. Se alude, no obstante, a las objeciones que ha suscitado, sobre todo porque elimina de la muestra a casi todos los apellidos de origen romance. Después de recibir ese informe, se hizo un experimento similar en los archivos del Departamento de Servicios Sociales del Kent County Council (Inglaterra). Escogiendo la letra H se obtendría una muestra de un 6 %, pero con ello se omitirían todos los individuos procedentes de las Antillas, que constituyen un elemento particularmente importante de esa serie. Escogiendo la letra A o la letra O –en ambos casos hubiera quedado incluido un importante elemento antillano- se obtendría una muestra de un 14 % o de un 12 % respectivamente. Se observó algo similar en el Canadá: para poder recoger a los inmigrantes vietnamitas en una muestra basada en la distribución alfabética hubiera habido que emplear la letra N. Pese a la validez aparente de la elección inicial para la localidad en cuestión, se constata que la muestra resultante no es plenamente válida, que no es un método fácilmente transplantable y que, en otro contexto, puede omitir fácilmente factores informativos esenciales. (HULL. F. 1981, 42-43)

Observaciones

El tamaño de la muestra queda determinado por la frecuencia que tienen los apellidos que comienzan con la letra elegida (frecuencia de la letra en la población), lo cual depende de la letra por la que se haya optado. Los comentarios de Hull evidencian la existencia de una cierta información previa respecto de la estructura étnica de la población, gracias a la cual notamos que el Muestreo Alfabético no funciona como uno quisiera. Esto a su vez nos muestra que el apellido de una persona contiene información socioeconómica, étnica, religiosa, etc. Es un rasgo o variable de alto poder explicativo, dada su alta correlación con otras variables.

Constituye por tanto un buen descriptor de la población, por lo que una muestra representativa debería imitar a la población respecto de los apellidos que contiene. El Muestreo Alfabético implica que la frecuencia de la letra elegida es la medida para evaluar cuánto se parece la muestra a la población. Por tanto la mejor muestra será aquella en la que los apellidos ocurran con la misma abundancia relativa (que llamaremos frecuencia) que en la población. Esto implica tener una idea de cómo es el perfil o distribución de las letras en la población, para así tomar una muestra que imite dicho perfil, lo que necesariamente requiere un cierto conocimiento de la población a muestrear.

Los comentarios de F. Hull ponen de manifiesto que dicho conocimiento está presente, al menos en cierta medida, por algo se dan cuenta que el método no funciona bien. Con una sencilla investigación

previa (como por ejemplo contar en la guía telefónica de la localidad o entorno cuantos apellidos comienzan con cada letra, o cuantas páginas hay para cada inicial), puede conocerse la distribución de las letras en la población (esto es, qué porcentaje respecto del total representa cada inicial), para luego elegir la muestra de acuerdo a la información recogida. El paso siguiente es tomar una muestra que imite a la población respecto del rasgo elegido, que en este caso es la abundancia relativa de cada inicial en los apellidos en la población.

En la siguiente sección seguiremos hasta su conclusión lógica las observaciones que desprendemos de los comentarios de Hull sobre las experiencias con el Muestreo Alfabético. Esto nos conducirá a una forma mejorada de Muestreo Alfabético que ilustraremos a través de un ejemplo tomado de la realidad.

Muestreo Alfabético Mejorado: ejemplo de trabajo

Trabajaremos con una población formada por 701 documentos (exámenes de un curso de matemática universitario). Estos exámenes se guardan por dos años y luego se eliminan. Transcurrido ese plazo deseamos conservar una muestra de los mismos, de modo que la misma sea representativa de la población de documentos (por ejemplo, que los estudiantes que hicieron dichos exámenes sean representativos de la población estudiantil). El criterio será alfabético: se conservarán solamente aquellos exámenes (de ahora en más, documentos) de estudiantes (de ahora en más, personas) cuyo apellido comienza con una determinada letra. Esta puede ser única (p. ej. conservaremos los documentos de personas cuyo apellido comienza con “B”) o puede ser un conjunto de letras (conservaremos documentos de personas cuyos apellidos comienzan con “B”, “E” o “L”, por ejemplo). Dicha letra o grupo de letras puede ser elegida al azar o mediante algún criterio no aleatorio. Mostraremos las consecuencias que cada variante puede acarrear y cómo mejorarlas.

Previamente se contaron cuantos apellidos comenzaban con cada letra. Aunque los documentos son en soporte papel, en este caso concreto (exámenes) se contaba además con el acta de inscriptos de la Bedelía (en soporte electrónico), por lo que no hubo necesidad de un conteo físico. En esta población de 701 documentos, hay 50 correspondientes a personas cuyos apellidos comienzan con “A”, 56 documentos de personas cuyos apellidos comienzan con “B”, 70 que comienzan con “C”, etc.

Si dividimos cada cantidad entre 701 tendremos la frecuencia con que aparece cada letra, que mostramos en la Tabla N° 1:

Inicial	A	B	C	D	E	F	G	H	I	J	K	L	M
Cantidad	50	56	70	39	12	34	57	9	9	6	5	36	66
Frecuencia	7.1 %	7.9 %	9.9 %	5.5 %	1.7 %	4.8 %	8.1 %	1.2 %	1.2 %	0.8 %	0.7 %	5.1 %	9.4 %
Inicial	N	O	P	Q	R	S	T	U	V	W	X	Y	Z
Cantidad	12	16	61	4	57	50	18	1	26	3	0	2	2
Frecuencia	1.7 %	2.2 %	8.7 %	0.5 %	8.1 %	7.1 %	2.5 %	0.1 %	3.7 %	0.4 %	0%	0.2 %	0.0 %

Tabla N° 1. Composición de la población respecto de las iniciales de los apellidos

Así, por ejemplo aproximadamente el 7% de los apellidos (50 en 701) comienzan con “A”, el 8% (aproximadamente) de los apellidos (56 en 701) comienzan con “B”, aproximadamente el 10% de los apellidos (70 en 701) comienzan con “C”, etcétera. No encontramos apellidos que comiencen con “X” en esta población de 701 documentos. Esto no es raro, dado que la población del Uruguay es mayoritariamente de ascendencia europea, sobre todo española e italiana. Si graficamos las frecuencias, obtenemos el siguiente patrón o perfil (Gráfico N° 1):

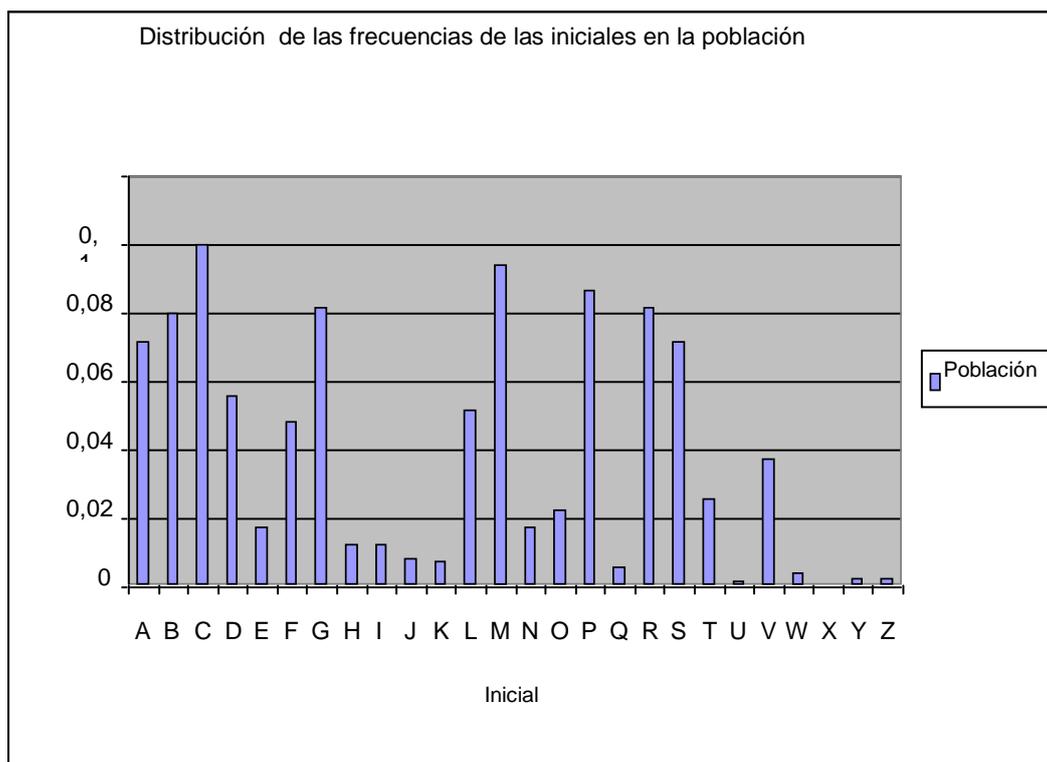


Gráfico N° 1. Distribución de las frecuencias de las iniciales de los apellidos

Si reordenamos de mayor a menor, tenemos el Gráfico 2:

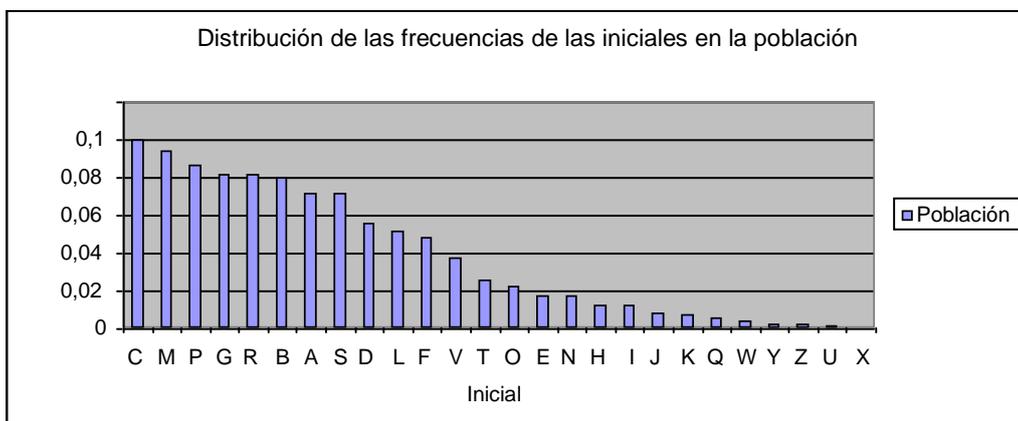


Gráfico N° 2. Distribución de las frecuencias de las iniciales de los apellidos en orden decreciente

Vemos que en nuestra población de 701 documentos, los de las personas cuyos apellidos comienzan con “C” son los más abundantes (70 en 701, un 10% del total), seguidos por los que comienzan con “M” (66 en 701, un 9.4% del total), luego los que comienzan con “P”, etc.

Supongamos que aplicamos el Muestreo Alfabético “Estándar”, esto es, se decide conservar solamente los documentos de personas cuyo apellido comienza con una determinada letra. Dicha letra se elige al azar y resulta ser la “L”. Hay 36 documentos de los 701 que comienzan con “L”, por lo que nuestra muestra supone un 5% de los 701 documentos. Cuando comparamos nuestra muestra con la población, vemos que su distribución de frecuencias (perfil, de ahora en más) no se parece demasiado al de la de la población, como puede apreciarse en el Gráfico N° 3:

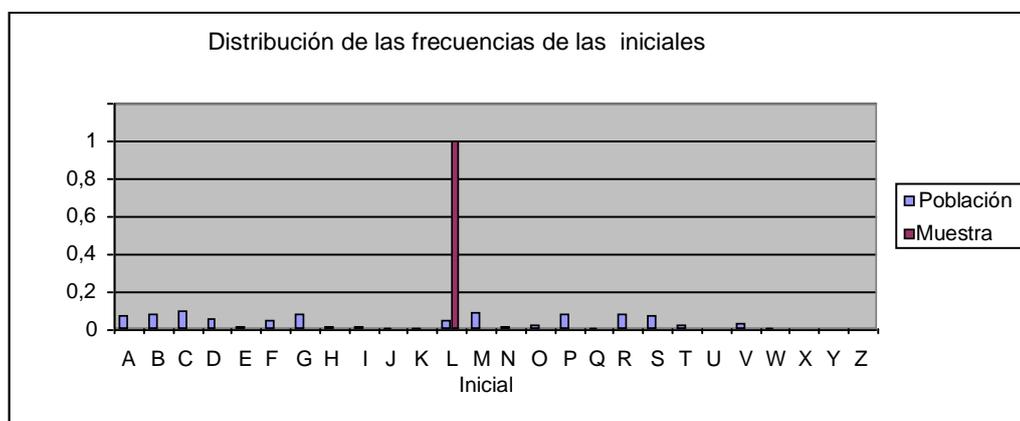


Gráfico N° 3. Comparación población – muestra eligiendo la inicial al azar

Si deseamos que la muestra sea “representativa” y pretendemos lograr esto mediante el criterio de guardar solamente documentos de personas cuyos apellidos comiencen con una determinada letra, lo más razonable sería elegir documentos de personas cuyos apellidos comiencen con la letra “C”, por cuanto es el caso más común en la población. Por tanto la muestra se parecería a la población porque en ambas la letra “C” es la que aparece con mayor frecuencia (10% en la población, 100% en la muestra). El tamaño de esta muestra ahora representa el 10% de la población.

Al elegir la letra “C” frente a cualquier otra mejora el parecido de la muestra con la población, sin embargo aún no luce como la población a la cual se supone debe imitar, como muestra el Gráfico N° 4:

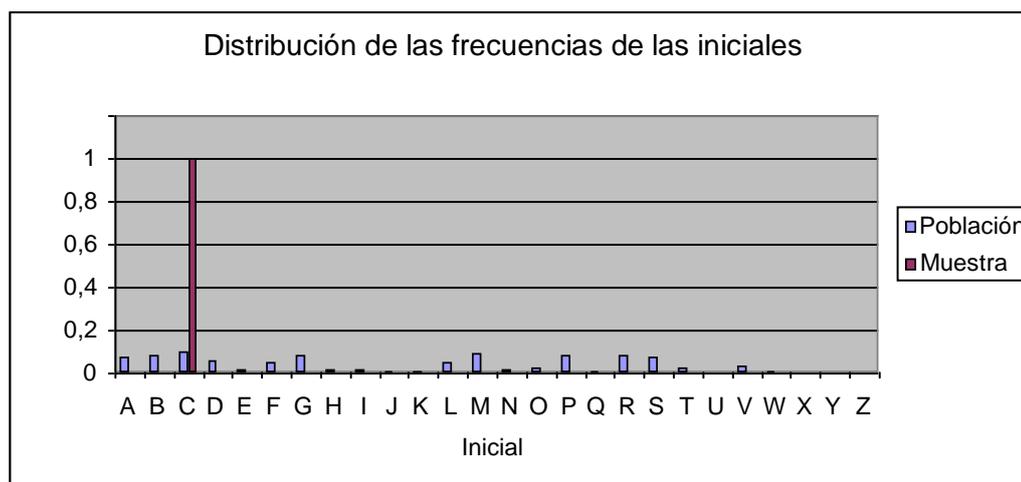


Gráfico N° 4. Comparación población- muestra eligiendo la inicial de mayor frecuencia

Podemos mejorar la representatividad de la muestra (que se parezca más a la población) si logramos que tenga el mismo perfil de distribución de las letras que el que tiene la población total.

Esto es, si en nuestra muestra el 10% de los documentos son de personas cuyos apellidos comienzan con “C”, el 9.4% de los documentos son de personas cuyos apellidos comienzan con “M”, el 8.7% con “P”, etc., entonces dicha muestra será representativa del total de 701 documentos, en el sentido que imitará la conducta de la población en cuanto a la abundancia relativa de apellidos que comienzan con cada letra.

¿Qué tamaño debería tener la muestra? En el Muestreo Alfabético Estándar, el tamaño viene dado por la abundancia de la letra elegida. Si nos basamos en la letra “C”, podríamos tomar una muestra del 10% y llenarla siguiendo el criterio del párrafo anterior. Tendríamos entonces una muestra de 70 documentos, de los cuales el 10% (7 documentos) se elegirían entre

los documentos correspondientes a personas cuyos apellidos comienzan con “C”, por cuanto el 10% de la población tiene apellidos que comienzan con “C”; el 9.4% (6.6 documentos que redondeamos a 7) deberían corresponder a personas cuyos apellidos comienzan con “M”, por cuanto las personas cuyo apellido comienza con “M” representan el 9.4% del total de 701 documentos, etc.

La Tabla N° 2 muestra cuantos documentos de cada inicial hacen falta para que tengan el mismo peso o porcentaje (sobre un total de 70) que el que tiene la inicial en la población entera de 701 documentos.

C	M	P	G	R	B	A	S	D	L	F	V	T
7	7	6	6	6	6	5	5	4	4	3	3	2
O	E	N	H	I	J	K	Q	W	Y	Z	U	X
2	1	1	1	1	1	0	0	0	0	0	0	0

Tabla N° 2

Por tanto tomamos nuestra muestra comenzando por las iniciales de mayor frecuencia y nos detenemos cuando alcanzamos los 70 documentos. Llamémosle a esta forma de tomar la muestra “llenado de mayor a menor”. Entonces, nuestra muestra de 70 documentos estará formada como muestra la Tabla N° 3:

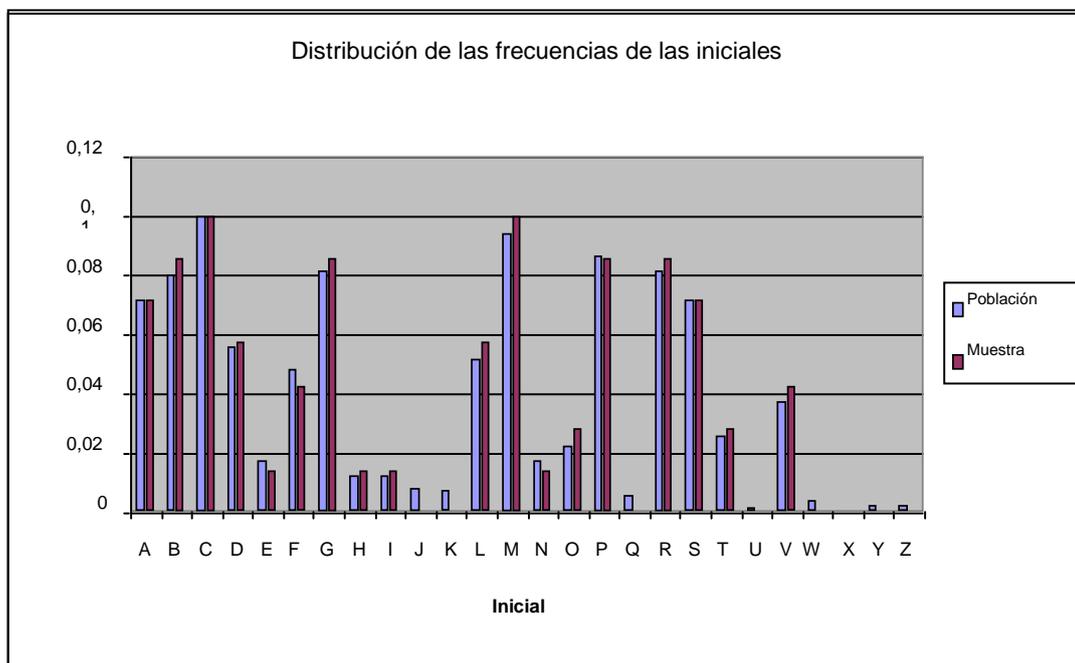
C	M	P	G	R	B	A	S	D	L	F	V	T	O	E	N	H	I
7	7	6	6	6	6	5	5	4	4	3	3	2	2	1	1	1	1

Tabla N° 3. Composición de la muestra de 70 documentos, llenado de mayor a menor

Tenemos una muestra de 70 documentos donde 7 de ellos (el 10%) corresponden a personas cuyos apellidos comienzan con “C”, 7 (aproximadamente el 9.6%) a personas cuyos apellidos comienzan con “M”, etc. Para cada letra los documentos se pueden elegir al azar o en forma sistemática. En total estamos eligiendo documentos de personas cuyos apellidos comienzan con una de 18 letras posibles: C, M, P, G, R, B, A, S, D, L, F, V, T, O, E, N, H, I.

Notemos que en la muestra no aparecen documentos de personas cuyos apellidos comienzan con letras de frecuencia menor que la I, como J, K, Q, W, X, Y, Z, U. Esto no supone una gran pérdida, porque en la población de 701 documentos, el 96.85% de los mismos corresponden a personas cuyos apellidos comienzan con una de las 18 letras C, M, P, G, R, B, A, S, D, L, F, V, T, O, E, N, H, I, por lo que nuestra muestra deja fuera casos de frecuencia baja, que son una minoría dentro de la población.

Si representamos el perfil de esta muestra (en rojo) y lo comparamos con el de la población de la cual fue extraída (en azul), vemos que la misma imita bien el perfil poblacional. (Gráfico N° 5)



Supongamos que nos decidimos por una muestra del 5% (35 documentos). Seguimos el mismo procedimiento: primero convertimos los porcentajes que cada inicial tiene en la población a su equivalente cuando el total son 35 documentos. Como el 10% de 35 es 3,5, hemos redondeado a 4, de modo similar el 9.4% de 35 es 3,3, que redondeamos a 3, por lo que esta muestra tendrá 4 documentos de personas con inicial C, 3 documentos de personas con inicial M, etc.

La tabla N° 4 muestra cuantos documentos de cada inicial hacen falta para que tengan el mismo peso o porcentaje (sobre un total de 35) que el que tiene la inicial en la población entera de 701 documentos.

C	M	P	G	R	B	A	S	D	L	F	V	T
4	3	3	3	3	3	2	2	2	2	2	1	1
O	E	N	H	I	J	K	Q	W	Y	Z	U	X
1	1	1	0	0	0	0	0	0	0	0	0	0

Tabla N° 4

En ambas muestras al comenzar el llenado desde las frecuencias mayores hacia las menores, el corte se da en las frecuencias más bajas, por lo que tendremos muestras que se ajustan al perfil poblacional sobre todo en las frecuencias altas. Esto es, la composición de la muestra imita bien los apellidos más frecuentes, no tanto los poco frecuentes. Puede argumentarse que precisamente los casos “raros” (de baja frecuencia) son los que pueden tener más importancia, y que es preferible incluirlos a expensas de los casos más comunes, quienes de cualquier modo estarán representados.

Si deseamos una muestra que además capture aquellos casos poco frecuentes o “raros”, lo que haremos es simplemente comenzar a llenar la muestra comenzando desde las frecuencias bajas (llenado de menor a mayor). Esto garantiza que dichos casos poco o muy poco frecuentes sean recogidos en la muestra, a expensas de casos que de todas formas estarán contemplados por tener frecuencias mayores. Naturalmente, si tomamos una muestra de mayor tamaño podremos incluir todos los casos. De hecho, puede fijarse de antemano el tamaño necesario de la muestra para que todos los casos cuya frecuencia es mayor que un determinado umbral estén contemplados.

Veamos nuevamente las muestras al 10% (70 exámenes) y al 5% (35 exámenes) si ahora nos preocupa no perder los casos raros. Esta nueva muestra de 70 documentos (que ahora llenamos de menor a mayor) aparece en la tabla N° 6:

C	M	P	G	R	B	A	S	D	L	F	V	T	O	E	N	H	I	J
6	7	6	6	6	6	5	5	4	4	3	3	2	2	1	1	1	1	1

Tabla N° 6. Composición de la muestra de 70 documentos,
llenado de menor a mayor

Casi no hay cambios respecto de la muestra de 70 documentos obtenida llenando de mayor a menor: ingresa un documento de una persona cuyo apellido comienza con “J” y sale el de una persona cuyo apellido comienza con “C”. El perfil de la muestra sigue siendo muy parecido al de la población, mejorando en las letras de baja frecuencia y empeorando un poco en las de mayor frecuencia, como puede verse en el Gráfico N° 7 (comparemos con el Gráfico N° 5):

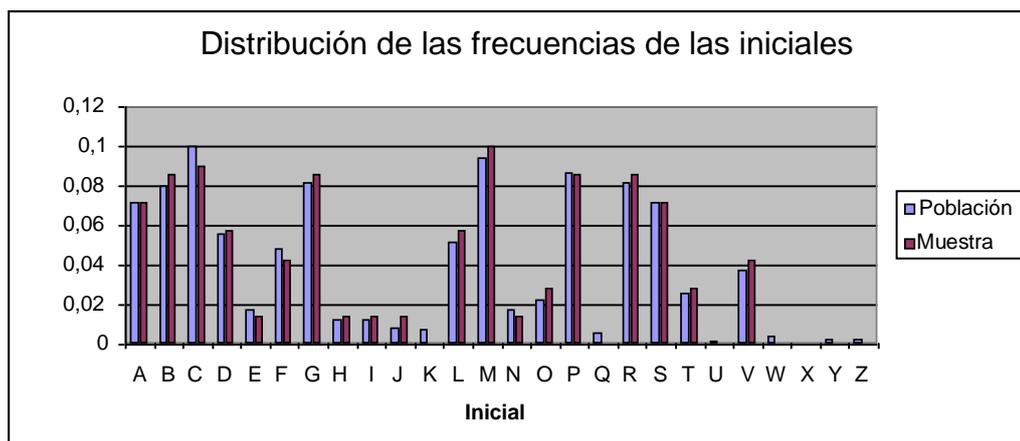


Gráfico N° 7. : Perfiles Población – Muestra
(70 documentos, llenado de menor a mayor)

La muestra de 35 exámenes sufre cambios similares: ingresa una “N” y sale una “C”, quedando como indica la Tabla N° 7 (comparemos con la tabla N° 5):

C	M	P	G	R	B	A	S	D	L	F	V	T	O	E	N
3	3	3	3	3	3	3	3	2	2	2	1	1	1	1	1

Tabla N° 7. Composición de la muestra de 35 documentos,
llenado de menor a mayor

El Gráfico N° 8 muestra que de nuevo el perfil mejora en las letras de menor frecuencia a expensas de empeorar un poco en las letras de mayor frecuencia (comparemos con el Gráfico N° 6):

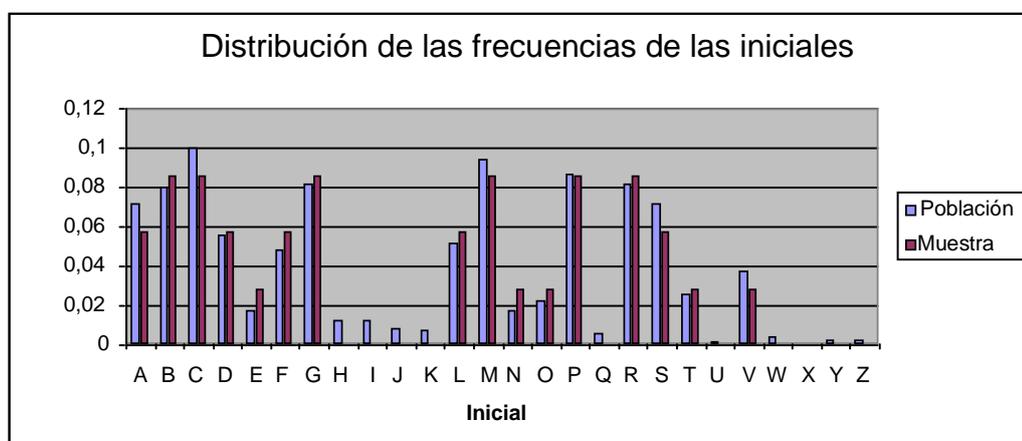


Gráfico N° 8. Perfiles Población – Muestra
(35 documentos, llenado de menor a mayor)

Terminemos esta sección con un breve resumen:

Los apellidos concentran información sobre poblaciones. Parece razonable que para capturar dicha información en una muestra, ésta deba imitar a la población en cuanto a la estructura de apellidos, reflejada en la frecuencia relativa de éstos. Esto es, que los apellidos ocurran en la muestra con la misma frecuencia con que lo hacen en la población.

El Muestreo Alfabético Estándar es un primer intento para lograrlo. El Muestreo Alfabético Mejorado que acabamos de mostrar logra muestras muy superiores en cuanto a su parecido con la población, por lo que podemos suponer que tendrán un mayor potencial para permitir reconstruir dicha población.

En la siguiente sección veremos el Muestreo Estratificado, mostrando sus dos principales variantes mediante ejemplos. Esto nos permitirá apreciar la conexión con el Muestreo Alfabético Mejorado que acabamos de mostrar, el cual quedará entonces como un caso particular del Muestreo Estratificado, al igual que el Muestreo Alfabético Estándar.

El Muestreo Estratificado

Siguiendo los conceptos de J. Piccini, definiremos el Muestreo Estratificado y daremos algunos ejemplos sencillos como ilustración (2011 p. 39-40).

El Muestreo Estratificado asume la división previa de la población bajo estudio en grupos o clases (llamadas estratos) que se suponen homogéneos respecto de alguna característica de interés. Se decide primero cuántas unidades de cada estrato compondrán la muestra (se asigna una cuota a cada estrato). Habitualmente se hace de dos formas:

Por Asignación Proporcional: el tamaño del estrato dentro de la muestra es proporcional al tamaño del estrato dentro de la población.

Por Asignación Óptima: la muestra recogerá más individuos de aquellos estratos que tengan más variabilidad interna.

Una vez decidida la representación que cada estrato tendrá en la muestra, resta ver cómo elegir en cada estrato los individuos que conformarán la muestra. Suelen usarse dos métodos:

Dentro de cada estrato se puede usar la técnica de Muestreo Sistemático. También puede usarse la técnica de Muestreo Aleatorio Simple dentro de cada estrato.

El Muestreo Estratificado se recomienda cuando la población tiene marcadas diferencias, así al agrupar en estratos más homogéneos se gana en precisión. Este tipo de muestreo pretende asegurar la representación de cada grupo (estrato) en la muestra, tiende a garantizar que la muestra represente adecuadamente a la población en función de la o las variables

seleccionadas, pero previamente se ha de conocer la distribución en la población de las variables utilizadas para la estratificación.

Para ilustrar veamos un ejemplo de naturaleza archivística. Supongamos un fondo documental de una Universidad con un total de 1500 metros lineales de documentos, de los cuales el 25% son de naturaleza contable, el 40% de naturaleza administrativa no contable y el 35% restante son de naturaleza académica. Aquí las unidades o individuos serán los documentos que conforman dicho fondo. Se decide conservar una muestra de 100 mts. lineales de documentos.

Mediante Asignación Proporcional:

Si se desea una muestra de 100 metros lineales de documentos, 25 de ellos (el 25% de la muestra de 100) se elegirán de entre los metros lineales de los documentos contables, 40 metros (el 40% de 100) de entre los metros lineales de los documentos administrativos no contables y 35 metros (el 35% de 100) de entre los metros lineales de documentos académicos, por cuanto esos son los pesos relativos de cada estrato en la población de la cual se extrae la muestra.

El mismo ejemplo pero con Asignación Óptima:

Supongamos que los documentos administrativos no contables presentan una mayor variabilidad o diversidad que los documentos administrativos contables y los documentos académicos, entonces se les asignará una mayor cuota dentro de los 100 metros lineales de documentos que formarán la muestra, por ejemplo eligiendo 50 metros lineales de documentos administrativos no contables en vez de 40 metros (en detrimento de los otros estratos).

En el ejemplo precedente no hay estratos vacíos y todos los estratos participan en la muestra. En la siguiente sección debilitaremos este requisito, lo que nos permitirá inscribir al Muestreo Alfabético en el marco del Muestreo Estratificado.

Si volvemos a la población de documentos de nuestro ejemplo de trabajo (donde cada documento está asociado a un apellido), los grupos homogéneos o estratos serán los formados por los documentos correspondientes a personas cuyos apellidos comienzan con la misma letra. En la siguiente sección profundizaremos en esto.

El Muestreo Alfabético como caso particular del Muestreo Estratificado

En el ejemplo de los 701 documentos, la característica de interés son los apellidos, de donde los estratos serán grupos de personas cuyos apellidos comiencen con la misma letra. Llamemos estrato "A" al grupo formado por todos los documentos de personas cuyos apellidos comienzan con la letra "A", estrato "B" al grupo formado por todos los documentos de personas cuyos apellidos comienzan con la letra "B", etc.

Tenemos que los 701 documentos se agrupan en 26 estratos. El 10% de los mismos pertenece al estrato “C” (70 documentos de 701), el 9% pertenece al estrato “M”, etc. El estrato “X” está vacío. Si nos quedamos con los estratos de mayor peso (18 estratos o 15, según la muestra represente el 10% o el 5%) y asignamos a cada estrato en la muestra el mismo peso que tiene en la población, entonces nuestra muestra tendrá el 10% de sus individuos elegidos del estrato “C”, el 9% serán elegidos del estrato “M”, etc.

Si admitimos que en el Muestreo Estratificado pueden dejarse estratos que no participen en la muestra (por tener poco peso en la población), es claro que lo que hicimos fue un Muestreo Estratificado por Asignación Proporcional, estratificando por las iniciales de los apellidos. Por tanto el Muestreo Alfabético Mejorado es en realidad un caso particular del Muestreo Estratificado, por lo que no hay necesidad de definirlo como si fuese otro tipo distinto de muestreo.

Del mismo modo, el Muestreo Alfabético Estándar puede verse como un caso –rudimentario del Muestreo Estratificado: se elige toda la muestra dentro de un mismo estrato (por ejemplo la letra H en Colonia), por lo que no es de extrañar que las muestras así obtenidas sean tan pobres en cuanto a representatividad y tan difíciles de transplantar, como señala Hull.

Conclusiones

- Los apellidos concentran información sobre poblaciones, por tanto para poder capturar esto en una muestra, la misma debería imitar a la población en cuanto a la frecuencia relativa de cada inicial. Esto nos lleva a preferir muestras cuyo perfil alfabético sea lo más parecido posible al de la población.
- Lo que en Archivología se llama “Muestreo Alfabético”, es un primer intento para lograrlo, pero no garantiza que la muestra sea representativa de la población. Los comentarios y críticas citados nos obligan a pensar con mayor profundidad en el significado de “representativo”.
- Si tomamos como atributo la primera letra de los apellidos, entonces **una definición razonable de “muestra representativa” es que la misma se asemeje a la población respecto del rasgo elegido.** El Muestreo Alfabético Mejorado logra esto y tan solo requiere de un poco de información previa, la que puede conseguirse fácilmente.

Por tanto, dichas muestras tendrán un mayor potencial para pretender reconstruir la población original, en relación a las muestras obtenidas mediante el Muestreo Alfabético Estándar.

Siguiendo el criterio conocido como “Navaja de Occam”, no es necesario seguir definiendo como si fuesen cosas distintas a lo que son casos particulares de algo más general, por lo que debería abandonarse en la literatura archivística la mención al Muestreo Alfabético, o al menos no creer que sea una técnica distinta del Muestreo Estratificado.

Referencias bibliográficas

CANAVOS, George C. Probabilidad y estadística: aplicaciones y métodos. México: Mc Graw Hill, 1988.

COCHRAN, Wiliam G. Técnicas de muestreo. México, CECSA, 1996.

DUVERGER, Maurice. Métodos de las ciencias sociales. Barcelona, Ariel, 1962.

HULL, Félix. (1981) Utilización de técnicas de muestreo en la conservación de registro: estudio del RAMP y directrices al respecto. París : UNESCO. PGI; UNISIST.

MORALES DEL CASTILLO, José Manuel. Selección y valoración. Tema 8 de Archivística. [On line] Presentación en Segundo Curso de la Diplomatura de Biblioteconomía y Documentación de la Universidad de Granada (España) <<http://www.ugr.es/~josemdc/Recursos/Tema8.pdf>> [Consulta: 01.07.2012] .

Muestreo en estadística. En: Wikipedia. 2012. <http://es.wikipedia.org/wiki/Muestreo_en_estad%C3%ADstica> [Consulta: 25.07.2012]

PERERA, Gonzalo. Probabilidad y estadística. Montevideo: Fin de Siglo, 2011.

PICCINI, Juan. (2011) El muestreo: técnica al servicio de la valoración documental. *Informatio*, N° 14/16, 2009/2011: 34-45.

SLONIM, Morris James. Muestreo: guía ágil y precisa de estadística práctica. B.A.: Americana, 1974.