<u>Tema 6:</u> Introducción al muestreo.

6.1.- Conveniencia y limitaciones del muestreo.

Como ya se sabe, los responsables de marketing necesitan contar con información que les permita tomar decisiones lo más acertadamente posible y así reducir el riesgo en éstas. Para ello resulta imprescindible obtener información acerca de características o parámetros propios de la ***población*** que se necesite estudiar.

Una población es el conjunto de elementos (personas, empresas, familias, etc.) que comparten una serie de características y que representan el universo cuyo estudio es objeto del trabajo de investigación.

En IC, los parámetros o variables que se desea determinar de la población son típicamente números, tales como por ejemplo, la proporción de consumidores que son leales a una marca de dentrífico, o la audiencia de una emisora de radio concreta. En general, la información sobre dicha población se podría obtener de dos maneras: bien realizando un censo, esto es, estudiando a todos los elementos que la componen, o bien trabajando con muestras.

Por tanto, un censo es la enumeración completa de todos los elementos de una población. En caso de realizar un censo, los parámetros de la población podrían calcularse directamente después de haber estudiado su valor a todos los individuos. Por contra, una muestra es un sub-grupo de la población seleccionado y estudiado para obtener los parámetros de ésta a través de la inferencia estadística. Las conclusiones que se hagan de la muestra tras el análisis, se supone que serán válidas para toda la población, si se ha tenido cuidado de escoger una muestra representativa.

La decisión de qué resulta conveniente, si realizar un censo o emplear muestras, depende de varias cuestiones:

- Tamaño de la población: Si la población es muy grande, como es el caso de la mayoría de los mercados de consumo, un censo resulta inabordable. Por ejemplo, un estudio para conocer el nivel de satisfacción de los consumidores de rollos de fotos implicaría entregar uno o varios como prueba antes de recoger el veredicto. Considerando a toda la población esto sería demasiado caro. Por contra, en el caso de poblaciones pequeñas, como ocurre en determinados mercados industriales, por ejemplo, no merece la pena establecer procedimientos de selección de muestras.

- Variabilidad de los miembros de la población: Por otro lado, si la(s) característica(s) o variable(s) que se quiere(n) estudiar tiene(n) grande(s) variacion(es), este hecho constituye una razón más, -aunque no es determinante-, para realizar un censo. Este vuelve a ser el caso de los mercados industriales, donde por ejemplo, se concentra un número pequeño de empresas dispares, tanto grandes empresas como PYMEs, que tienen cuotas de producción, amplitudes de gama o necesidades de suministro, por ejemplo, muy diferentes.

- Razones temporales: Sin embargo, dicho lo anterior, es importante saber que un censo puede provocar errores de tipo sistemático. Si la población es grande y/o está muy dispersa, solamente el tiempo que llevaría estudiar a la población haría inoperante el trabajo: se tardaría demasiado, las primeras observaciones quedarían obsoletas, etc.

- Razones de definición: Frecuentemente se desconoce la existencia de otros elementos del colectivo que se explora. Por ejemplo, ¿cómo es posible saber de antemano *quiénes* constituyen la población de clientes potenciales de un comercio? Por ejemplo, no habría forma de distinguirles del resto de los habitantes de la región o ciudad.

Por contra, cuando se sepa bien quiénes son los individuos que integran la población, será posible, en principio, trabajar con censos.

- Tipos de error que se pueden cometer: En cualquier trabajo de Investigación Comercial se pueden cometer dos tipos de error:

- Errores muestrales: No pueden ser eliminados debido a que son inherentes a la naturaleza de la variable a medir. Se deben al hecho de estar trabajando con muestras en vez de con toda la población. No obstante, siguiendo procesos de inferencia estadística, sí se pueden acotar.

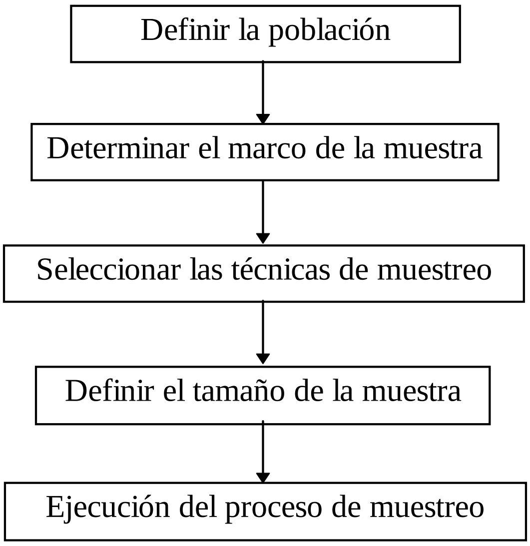
- Errores ajenos al muestreo: También se les denomina con el nombre de errores sistemáticos. No tienen que ver necesariamente con el hecho de trabajar con muestras, sino con fallos tales como una mala definición del problema de investigación, fallos en la redacción de las preguntas del cuestionario, fallos en la mala selección de los miembros de una muestra, errores de análisis, etc. Debido al hecho de que en los censos se estudiaría a toda la población, si ésta fuera muy grande, el tiempo que su estudio llevaría, junto con otros problemas mencionados, tales como la indefinición de todos los miembros de la población, darían lugar a errores sistemáticos importantes. Precisamente en estos casos, para evitar errores de este tipo se acepta asumir errores aleatorios y trabajar con muestras.

Resumiendo, el siguiente cuadro recoge las circunstancias para las que resulta más apropiado bien realizar censos o bien trabajar con muestras:

	Censo	**Muestra**
Tamaño de la población	Pequeño	Grande
Variabilidad de la población	Pequeña	Grande
Tiempo disponible	Abundante	Escaso
Definición de la población	Buena	Mala o inexistente
Error aleatorio	Es inexistente	En principio tolerable y acotado
Errores sistemáticos	Mayor riesgo	Menor riesgo

6.2.- El proceso del muestreo.

En concreto, las etapas seguidas en el muestreo se pueden resumir en cinco:



1º. Definir la población:

En primer lugar, será necesario definir la población objeto del estudio de manera tal que quede bien claro quiénes deben ser incluidos y quiénes no. Esta población deberá ser definida en términos de cuatro parámetros que son: (1) los elementos; (2) las unidades de muestra; (3) la extensión o zona y (4) el tiempo.

Un **elemento** es el objeto acerca del cual o a partir del cual se desea obtener una información. En IC, el elemento de la población suele ser el entrevistado. Por otro lado, una **unidad muestral** es un elemento, o una unidad que contiene un elemento, y que está disponible para la selección del proceso de muestreo en algún momento particular. Sirve como paso previo para identificar y acceder al elemento de la población. Por ejemplo, una óptica desea recoger información sobre la demanda de productos y servicios de su sector, y tiene como público objetivo (posible clientela) a cualquier persona con problemas de visión. Puede que las entrevistas se realicen de forma personal y a pie de calle, en cuyo caso el elemento de la población, es decir, cualquier persona que tenga problemas de visión, es a la vez el elemento muestral, esto es, la persona a la que se accede.

Sin embargo, si el método de recogida de la información es mediante visitas a domicilios escogidos aleatoriamente, la unidad de muestreo será entonces el domicilio, y el elemento de la población será cada una de las personas que viviendo en dicho domicilio tengan problemas de visión, o alguna de ellas escogida a través de un procedimiento aleatorio.

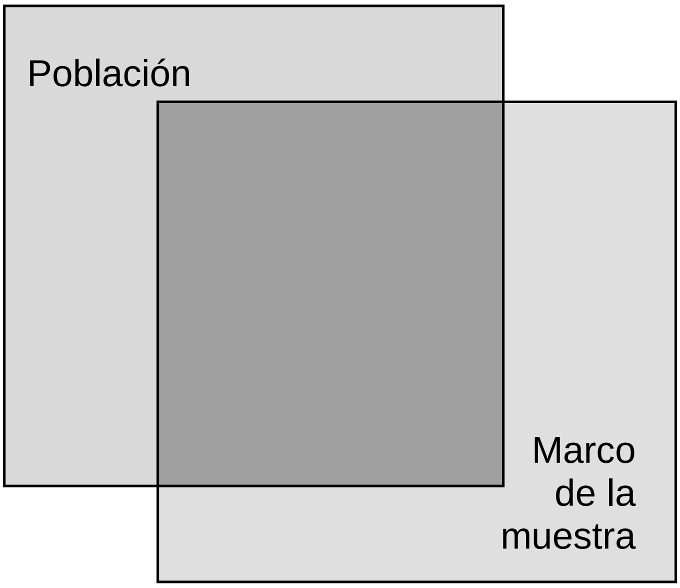
Otro ejemplo es el caso de un sondeo realizado por una empresa fabricante de maquinaria agrícola a los diferentes representantes comerciales, a saber, empresas concesionarias que venden maquinaria agrícola y prestan servicios (garantía, reparaciones, venta de repuestos, etc). Queremos saber si dichas empresas realizan esfuerzos comerciales para vender correctamente (tanto los productos de la marca en cuestión como los de otras marcas). En tal caso, el elemento será la propia empresa, en tanto que la unidad de muestreo será, el/la responsable que responda al cuestionario, por ser el medio para obtener la información de la empresa. Finalmente, la población es el conjunto de empresas concesionarias.

Como tercer ejemplo, sea el caso de un estudio sobre la demanda de productos de carnicería según las siguientes formas de venta: carnicerías tradicionales, mercados (carnicerías ubicadas en los mismos), supermercados, grandes almacenes (ej. el super de El Corte Inglés) e Hipers de la CAV durante el mes de diciembre de 1997. Queremos ver aspectos sobre las diferencias en las compras realizadas en cada tipo de establecimiento por los consumidores. Por tanto: los elementos de la muestra son las personas que hacen sus compras de productos cárnicos en cualquiera de dichas formas de venta; las unidades de muestra son los establecimientos a los que se acude para observar y/o realizar las entrevistas; la extensión es la CAV, y el tiempo de recogida de información es el mes de diciembre de 1997.

2º. Determinar el marco de muestreo o marco de la muestra:

Un marco de muestreo es la representación de los elementos de la población objetivo. Materialmente, es una lista de los elementos de la población, o bien un conjunto de instrucciones para identificar a ésta. Por ejemplo, para las entrevistas telefónicas en una provincia, un marco de muestreo es el listín telefónico de la misma. Si hay muchos números que no aparecen, el marco de la muestra podría consistir en el procedimiento automático y aleatorio (*Random-Digit Dialing*) seguido para generar una lista de teléfonos que compondrán la muestra.

Frecuentemente resulta posible componer u obtener una lista de elementos de la población, pero puede que dicha lista omita algunos de los elementos de la población y/o incluya algunos elementos que no pertenezcan a la misma. Por ejemplo, en el padrón municipal de un ayuntamiento podrían haber desfases en el sentido de que hay personas que se han mudado a otro municipio y sin embargo siguen empadronadas, o hay personas que habiendo venido a vivir todavía no se han empadronado. Por tanto, en el marco de muestreo, puede que *no estén todos los que son, ni sean todos los que están.*



Si las discrepancias entre la población y el marco muestral son pequeñas, podrían ignorarse. En otros casos, por contra, los investigadores deberían reconocer y tratar dicho error. Hay varias maneras de hacer esto. Una de las más sencillas consiste en redefinir la población en términos del marco muestral. Si por ejemplo, la población anteriormente era considerada como el conjunto de vecinos de una ciudad, ahora será 'el conjunto de vecinos empadronados' en la misma. Otra manera consiste en examinar aquellos elementos escogidos para ver si cumplen con las características exigibles a los elementos de la población: en el caso anterior, esto se traduce a comprobar si la persona a entrevistar sigue viviendo en la ciudad (al visitarle al domicilio para realizarle la entrevista).

3º. Seleccionar una técnica de muestreo:

En el siguiente apartado se exponen las técnicas básicas de muestreo, aunque ahora sólo vamos a establecer la clasificación básica. En función de ésta, hay dos formas genéricas de hacer muestras:

a).- Muestreo no probabilístico:

La elección de la muestra se hace sin seguir una norma prefijada, de cualquier manera y de forma cómoda. El responsable de la investigación es el que dice la forma de escoger los elementos de la muestra y su composición.

De la más sencilla a la más compleja, las técnicas de muestreo no probabilístico más importantes son las siguientes:

a.1).- Muestreo de conveniencia:

Es la técnica más sencilla, económica y que consume menor tiempo. La persona que realiza las entrevistas se encarga de seleccionar según su criterio propio a los elementos que han de conformar la muestra. No es recomendable para realizar trabajos que intenten hacer inferencia hacia toda la población, sino sólo para trabajos exploratorios que sirvan para generar ideas o elaborar hipótesis, y también para el caso de sondeos pilotos.

a.2).- Muestreo de bola de nieve:

Resulta muy adecuado para poblaciones que son pequeñas y al mismo tiempo difíciles de determinar. Consiste en entrevistar a un número de personas que componen la población, y a través de éstas indagar por otras que se encuentren en las mismas circunstancias, es decir, que formen parte de dicha población. Tiene como ventajas su coste relativamente bajo y su facultad de dar con poblaciones relativamente raras. Por ejemplo, en el caso del marketing industrial, para conocer las relaciones de compra-venta industrial, donde preguntando a una de las dos partes, se conoce a su socio de negocios habitual. También para grupos de población especiales, como por ejemplo personas que comparten un hobby determinado.

a.3).- Muestreo por cuotas:

Se trata de una técnica que tiene dos fases diferenciadas. La primera consiste en el desarrollo de proporciones o cuotas de sub-grupos de la población. Estas cuotas dependen de características típicamente demográficas que el investigador selecciona en función a su juicio y experiencia: variables tales como el sexo, la edad o el lugar de residencia se emplean como criterio. Normalmente estas cuotas son asignadas respetando la proporción que guardan a escala poblacional. Ello se hace con objeto de evitar sesgos desde el punto de vista de la composición.

Tras la selección de los criterios y la asignación de cuotas, el entrevistador se encargará de escoger a las personas respetando las cuotas que le han sido asignadas. Tiene por tanto la gran ventaja de dar con una muestra directamente y sin requerir un marco muestral, por lo que desde el punto de vista del coste económico y temporal resulta un procedimiento económico.

Sin embargo, su inconveniente está en que permiten al entrevistador aplicar su criterio propio en la selección de la muestra, lo cual provoca que ésta se haga en función a su comodidad más que a la conveniencia de veracidad. Por ejemplo, para entrevistas a pie de calle, si piden entrevistar a hombres y mujeres de entre 25 y 45 años, acabaríamos escogiendo a los/as oficiales de la OTA, y a los/as encargados/as de negocio.

b).- Muestreo probabilístico o aleatorio:

Se escogen las unidades de la muestra mediante un proceso aleatorio o de azar, con lo cual podremos calcular de antemano la probabilidad de cada una de las unidades, la precisión y los errores cometidos.

b.1) Muestreo probabilístico con reemplazamiento:

También se llama *muestreo aleatorio simple*. Las unidades son devueltas a la población una vez analizadas, y por tanto, pueden volver a formar parte de la muestra. Esto implica que la probabilidad de extracción de un elemento no depende de la extracción anterior de otros elementos.

Cuando la población es muy grande la extracción de algunos elementos no afecta a la probabilidad de salida de los que quedan. Por tanto, este caso se puede asimilar como un muestreo aleatorio simple.

b.2) Muestreo aleatorio sin reemplazamiento:

Cada individuo muestreado no es reintegrado al colectivo, y por consiguiente, ya no tendrá probabilidad de volver a ser elegido para la muestra. En consecuencia, la probabilidad de extracción de un elemento sí depende aquí de la extracción de elementos anteriores.

Las fórmulas empleadas para el muestreo sin reemplazamiento son más complejas matemáticamente que las del muestreo con reemplazamiento. Por esta razón, para poblaciones suficientemente grandes, el muestreo sin reemplazamiento se asume igual al muestreo con reemplazamiento.

4º. Determinar el tamaño de la muestra:

El tamaño de la muestra el número de elementos de la población que componen la muestra de manera que sea estadísticamente representativa. Determinarlo implica consideraciones de tipo cuantitativo y cualitativo. Todo lo relativo a esta fase se explicará en otro apartado de este mismo tema.

5º. Ejecución del proceso de muestreo:

La ejecución del procedimiento del muestreo viene condicionada por la población, el marco muestral, la unidad muestral, la técnica escogida y el tamaño definido. Si por ejemplo en un trabajo se define el hogar como el elemento, habrá que detallar qué se entiende por *hogar* en términos operativos, esto es, la vivienda. Por otro lado, habrá que establecer procedimientos para determinar la unidad muestral, es decir, la persona que en concreto nos vaya a contestar las cuestiones referidas a su hogar.

Sobre algunos aspectos relativos a la ejecución del proceso de muestreo, veremos el último apartado de este tema.

6.3. El muestreo aleatorio simple (M.A.S.).

6.3.1. Conceptos previos.

Es el procedimiento más elemental de muestreo probabilístico, aunque en la práctica es el más difícil de aplicar, puesto que requiere que se cumplan unas condiciones muy rigurosas. De todas maneras, sirve de modelo de orientación para las situaciones reales.

El M.A.S. es un muestreo con reemplazamiento en el que todas las unidades que componen la población tienen la misma probabilidad de ser elegidas. Partiendo de una población de **N** elementos identificados, se procede a escoger aleatoriamente la muestra de **n**.

El trabajo con las fórmulas del M.A.S. requiere dominar una terminología básica, que es la que a continuación vamos a indicar:

- Variables: Tal y como se definía en el tema anterior, en IC las variables pueden ser de tres tipos: métricas o cuantitativas, ordinales y nominales. Como un ejemplo de las primeras están el número de unidades consumidas de un producto, la edad, la renta, el gasto en ocio, etc. Ejemplos de variables nominales son, en caso de ser o no consumidor de un producto: el sexo, la marca que se consume, etc.

- Parámetros: Un parámetro es todo valor que describa de manera resumida la población. Por ejemplo, la media es un parámetro: la edad media de los estudiantes de una clase de alguna forma resume las edades de todos.

Como parámetro de las variables nominales se emplea la proporción: por ejemplo, el porcentaje de fumadores de la muestra, la proporción de favorables a una marca, etc.

Los parámetros de la población son los valores desconocidos que es preciso hallar a través de la inferencia estadística, esto es, mediante los estimadores.

- Estadístico o estimador: Es una función de los valores muestrales, una descripción resumida de la muestra, como la media de la renta de las personas de la muestra o la proporción de personas de la muestra que consumen un producto, o tienen coche, etc.

A diferencia de los parámetros, los estadísticos son aleatorios: no todas las muestras proporcionan el mismo valor para un estadístico. Esto es así porque se trabaja con muestras y no con toda la población como si de un censo se tratara. Por ejemplo, en una clase de 20 personas, si se quiere obtener una muestra de 5 para estudiar la estatura media, el número posible de combinaciones que se podrían obtener es de:

$$\binom{N}{n} = \frac{N!}{n!(N-n)!} = \frac{20.19.18.17.16}{1.2.3.4.5} = 15.504$$

Son 15.504 combinaciones de muestras posibles, de las que la estatura media que se obtenga variará algo.

6.3.2. Determinación de los estimadores en el M.A.S.

Los parámetros de la población que típicamente se busca estimar a través de la muestra son:

- Medias

- Totales

- Proporciones

a) Estimación de medias:

Si la variable a medir está distribuida en la población con media μ y varianza σ^2_1 , se tiene:

$$\mu = \frac{1}{N}\sum_{i=1}^{N} X_i \qquad \sigma_1^2 = \frac{1}{N}\sum_{i=1}^{N}(X_i - \mu)^2 = \frac{1}{N}\sum_{i=1}^{N} X_i^2 - \mu^2$$

Se busca estimar μ. Así, dada una muestra **n** ($x1_1$, $x1_2$, $x1_3$,. . . , $x1_n$), obtendremos la **media muestral**, que es una variable aleatoria:

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ Esta media muestral $\bar{x}$ será:

- Insesgada: $E(\bar{x}) = E(\sum_{i=1}^{n}\frac{x_i}{n}) = \frac{1}{n}\sum_{i=1}^{n}E(x_i) = \frac{1}{n}n\mu = \mu$

- Consistente, pues si la muestra es de tamaño **N** resultará ser:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n=N} x_i = \mu$$

También en el caso de poblaciones finitas, la varianza de $\bar{x}$ viene dada por:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_1^2}{n}\frac{N-n}{N-1}$$ donde $\frac{N-n}{N-1}$ es el *factor de corrección para poblaciones finitas*. Cuando **n** tiende a **N**, $\sigma_{\bar{x}}^2$ tiende a cero.

Con el objeto de simplificar las fórmulas, se acostumbra a emplear la cuasi-varianza poblacional:

$$\sigma^2 = \frac{\sum_{i=1}^{N}(X_i - \mu)^2}{N-1}$$ $\rightarrow$ cuando N es grande coincide con $\boldsymbol{\sigma^2_1}$

Por tanto, $\sigma^2 = \sigma_1^2 \frac{N}{N-1}$

En consecuencia: $\sigma_{\bar{x}}^2 = \frac{\sigma_1^2}{n}\frac{N-n}{N-1} = \frac{\sigma^2}{n}\frac{N-n}{N} = \frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)$

Ahora, la cuasi-varianza poblacional es desconocida. Por tanto, habrá que estimarla. Un estimador insesgado de σ^2 es la cuasi-varianza muestral $\mathbf{S^2}$:

$$S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Por tanto: $\hat{\sigma}_{\bar{x}}^2 = \frac{S^2}{n}\left(1-\frac{n}{N}\right)$

Como $\bar{x}$ es una v.a., cabrá preguntarse cuál es su distribución. Para esto, se pueden considerar dos alternativas:

1ª) La población objetivo es normal. Por tanto, al ser $\bar{x}$ una combinación lineal de variables normales, también será normal.

2ª) La población objetivo no es normal: Por el teorema central del límite, supondremos que $\bar{x}$ es asintóticamente normal.

Por tanto, de manera exacta o aproximada, podremos suponer que:

$$\bar{x} \quad \in \quad N\left(\mu\,;\ \frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)\right)$$

Si la población fuera infinita:

$$\lim_{N\to\infty} \sigma_{\bar{x}}^2 = \lim_{N\to\infty}\left(\frac{\sigma^2}{n}\left(1-\frac{n}{N}\right)\right) = \frac{\sigma^2}{n} \quad ; \quad \hat{\sigma}_{\bar{x}}^2 = \frac{S^2}{n}$$

Y en definitiva: $\bar{x} \in N\left(\mu ; \frac{\sigma^2}{n}\right)$

b) Estimación de los totales:

El total poblacional viene dado por $T = \sum_{1}^{N} X_i$ → representa la suma de todos los valores poblacionales. Por ejemplo, el total del consumo de un producto en una población será la suma de los consumos de cada miembro de ésta.

Fácilmente se ve que:

$$T = \frac{N}{N} \sum_{i=1}^{N} X_i = N\mu$$

Dado que $\bar{x}$ es un estimador insesgado de μ, un estimador insesgado de **T** será

$$t = N\,\bar{x} = \frac{N}{n} \sum_{i=1}^{n} x_i$$

donde el término N/n se denomina *factor de elevación* y es el recíproco de la fracción de muestreo.

Dado que los valores $\mathbf{x_i}$ se suponen independientes, la varianza del total $\boldsymbol{\sigma_t^2}$ será:

$$\sigma_t^2 = N^2 \sigma_{\bar{x}}^2 = \frac{N^2\, \sigma^2}{n}\left(1 - \frac{n}{N}\right) = N(N-n)\frac{\sigma^2}{n}$$

Si la precisión de la estimación se mide en desviación típica, resulta que la precisión del total es N veces menor que la de media (es decir, la imprecisión es N veces mayor).

Un estimador insesgado de $\boldsymbol{\sigma_t^2}$, esto es $\hat{\sigma}_t^2$, será:

$$\hat{\sigma}_t^2 = \frac{N^2\, S^2}{n}\left(1 - \frac{n}{N}\right) = N(N-n)\frac{S^2}{n}$$

Al igual que con la media muestral, podemos asimilar que los totales siguen una distribución normal:

$$t \in N\left(N\mu\,;\ \frac{N^2\sigma^2}{n}\left(1-\frac{n}{N}\right)\right)$$

Lógicamente, con los totales no cabe hablar de poblaciones matemáticamente infinitas.

c) Estimación de proporciones:

El caso más sencillo en la estimación de porcentajes poblacionales, es cuando dividimos la población en dos clases exhaustivas y excluyentes (compra o no compra, hombre o mujer, etc). Así, podremos emplear variables discretas y dicotómicas.

Un parámetro importante que recoge esta distribución es la proporción **P**, respecto del total, de elementos que toman el valor 1. Esta proporción será:

$$P = \frac{1}{N}\sum_{i=1}^{N} X_i = \frac{1}{N}X \quad \text{siendo} \quad X = \sum_{i=1}^{N} X_i$$

El complemento de la proporción **P** será **Q** = 1 **-** **P**

El valor **X** es el número de elementos de la población que toman el valor 1, y recibe el nombre de *total de clase*.

Extraída una muestra de extensión **n** ($x1_1$, $x1_2$, $x1_3$,. . . , $x1_n$), un estimador de la proporción es la proporción muestral:

$$p = \frac{1}{n}\sum_{i=1}^{n} x_i$$

que resulta ser un estimador insesgado y consistente de **P**.

La varianza poblacional es:

$$\sigma_P^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - P)^2 = \frac{1}{N}\left[\sum_{i=1}^{N} x_i^2 - 2P\sum_{i=1}^{N} x_i + NP^2\right] = P - 2P^2 + P^2 = P(1-P) = PQ$$

Ahora, de aquí se deduce que la varianza de la proporción **P**, esto es $\sigma_P{}^2$, viene de sustituir $\boldsymbol{\sigma_1^2}$ por **PQ** en la fórmula de $\sigma_{\bar{x}}^2$. Así:

$$\sigma_p^2 = \frac{PQ}{n}\left(\frac{N-n}{N-1}\right)$$

La varianza muestral es:

$$S_1^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - p)^2 = p(1-p) = pq$$

que resulta ser un estimador sesgado de **PQ**. Un estimador insesgado de la varianza poblacional es:

$$S^2 = \frac{N-1}{N}\frac{n}{n-1}\,pq$$

De aquí obtenemos que un estimador insesgado de la varianza de la proporción **p** será:

$$\hat{\sigma}_p^2 = \frac{N-n}{N}\frac{pq}{n-1} = \frac{pq}{n-1}\left(1-\frac{n}{N}\right)$$

Sabemos que **p** no sigue una distribución normal. No obstante, al ser **p** una combinación lineal de v.a., por aplicación del teorema central del límite, para tamaños muestrales grandes (superiores a 30), la distribución de **p** tiende a ser normal. Es decir:

$$p \quad \in \quad N\left(P\,;\,\frac{PQ}{n}\left(\frac{N-n}{N-1}\right)\right)$$

Si la población es infinita, tendremos que:

$$p \quad \in \quad N\left(P\,;\,\frac{PQ}{n}\right)$$

6.3.3. Tamaño de la muestra en el M.A.S.

Sabemos que el tamaño de la muestra es un factor muy importante en el muestreo, porque tiene implicaciones técnicas y económicas importantes en el coste, tiempo y precisión del trabajo de investigación. A continuación, vamos a determinar los tamaños muestrales necesarios para cada uno de los tres parámetros anteriormente especificados (media, proporción y total).

a). Tamaño óptimo para el estimador de la media:

Si el intervalo de confianza para la media poblacional es:

$$P\left[-Z_{\alpha/2} \le \frac{\overline{x}-\mu}{\frac{\sigma^2}{n}\frac{N-n}{N}} \le Z_{\alpha/2} \right] = 1-\alpha$$

Sea el error aleatorio *e*:

$$e = \overline{x} - \mu$$
$$e = Z_{\alpha/2}\ \frac{\sigma^2}{n}\ \frac{N-n}{N}$$

Despejando, se tiene que:

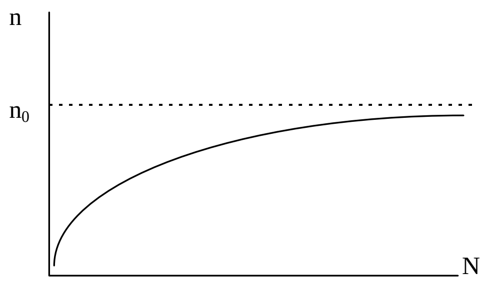
$$n = \frac{\sigma^2\ N\ Z^2{}_{\alpha/2}}{e^2 N + Z^2{}_{\alpha/2}\ \sigma^2}$$

Operativamente, dado que σ^2 es desconocida, se sustituye por su estimador insesgado S^2:

$$n = \frac{S^2\ N\ Z^2{}_{\alpha/2}}{e^2 N + Z^2{}_{\alpha/2}\ S^2}$$

Si la población se asumiera como infinita, la relación quedará así:

$$\mathrm{Lim}_{N\to\infty}(n) = \frac{S^2\ Z^2{}_{\alpha/2}}{e^2} = n_0$$



b). Tamaño óptimo para el estimador del total:

Operando de la misma forma, se obtiene el tamaño mínimo del total:

$$n = \frac{S^2\ N^2\ Z^2{}_{\alpha/2}}{e^2 + Z^2{}_{\alpha/2}\ S^2 N}$$

c). Tamaño óptimo para el estimador de la proporción:

$$n = \frac{p\,q\ N\ Z^2{}_{\alpha/2}}{e^2(N-1) + p\,q\ Z^2{}_{\alpha/2}}$$

Si la población se puede admitir como infinita, entonces el tamaño óptimo será:

$$n_0 = \frac{p\,q\ Z^2{}_{\alpha/2}}{e^2}$$

[**Ejemplos y ejercicios:** Realizar fotocopia]

6.3.4. Elección definitiva de los elementos que compondrán la muestra.

Una vez conocido el tamaño óptimo (mínimo) de la muestra **n**, se realiza la elección material de los elementos. Para ello, y partiendo de trabajar con un marco de la muestra o listado, se le asigna a cada elemento un número correlativo, y a continuación se eligen los **n** elementos a partir de una tabla de números aleatorios.

La elección material es muy sencilla. Por ejemplo, suponer que de un marco muestral de 800 elementos se deben escoger 10. Para ello, comenzando en la primera fila y primera columna de la tabla de números aleatorios, se eligen los tres úlimos dígitos de cada número. Los elementos correspondientes a los números aleatorios serán los que conformen la muestra. Así, tomados de la tabla que se adjunta, los elementos son: 480, 368, 130, 167, 570, 562, 301, 579, 475 y 553. Los números que pasen de 800 se descartan.

6.4. Otros tipos de muestreo probabilístico.

A continuación veremos varios tipos de muestreo probabilístico:

6.4.1. Muestreo estratificado aleatorio.

Se trata de un tipo de muestreo en el que los elementos son clasificados en sub-grupos de unidades con características homogéneas llamadas ***estratos***. Luego, de cada estrato se obtiene una muestra aleatoria representativa del mismo. Con ello, se gana exactitud al particionar el colectivo de acuerdo a algún criterio socioeconómico.

Por ejemplo, si en una clase quisiéramos determinar la estatura media, podríamos hacerlo tomando una muestra de 10 personas indistintamente de cualquier otra variable, y hallar la estatura media. O también, y de manera más precisa, podríamos separar a las personas con arreglo a variables que sabemos que inciden en la estatura, como por ejemplo, la edad (si se trata de niños, no en el caso de adultos) o el sexo (con personas adultas). A continuación, obtendríamos una muestra de cada estrato y hallaríamos la estatura media de cada uno. De esta manera se gana en precisión.

Para su correcta realización, el muestreo estratificado debe cumplir varias condiciones importantes:

- En primer lugar, los estratos formados deberán ser mutuamente excluyentes y colectivamente exhaustivos, esto es, no podrán haber elementos que pertenezcan simultáneamente a más de un estrato, y la partición deberá recoger a toda la población.

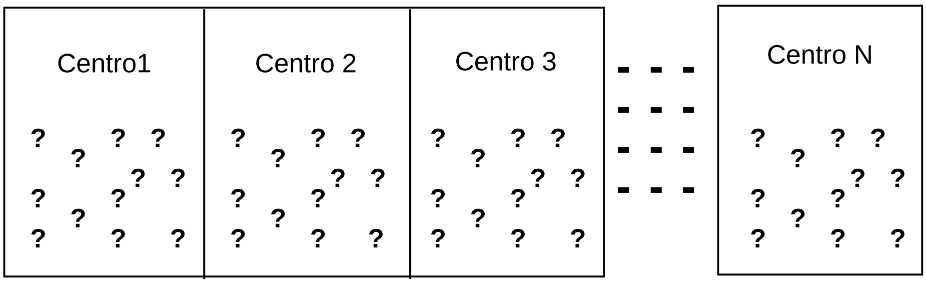
- El criterio para la selección de las variables que nos sirvan para realizar la partición deberá dar con estratos que sean lo más homogéneos internamente y heterogéneos externamente que se pueda. Es decir, los miembros de cada estrato deberán parecerse entre ellos y diferenciarse de los elementos de los otros estratos.

-En tercer lugar, los criterios empleados tienen que estar bastante vinculados a las características que se estudien. Por ejemplo, si quisiéramos estratificar a la población de clientes actuales de canales de TV por satélite en España, tal vez el criterio de la zona geográfica no ayude a diferenciar bien los gustos hacia la programación en general. Si los dividiésemos en ‘Clientela-Zona-Norte’ y ‘Clientela-Zona-Sur’ (para la mitad norte y sur del Estado, respectivamente) veríamos que no hay una separación en el tipo de programas que se piden, el horario, el ritmo de consumo (para la programación ‘a la carta’), etc.

- Finalmente, las variables escogidas deberán ser fáciles de medir y aplicar. Por ejemplo, variables tales como las demográficas (que son objetivas y generales) son fáciles de reconocer, en contra de otras, tales como los gustos hacia el producto que se estudia, o las aficiones de la persona entrevistada, para cuyo reconocimiento haría falta realizar preguntas.

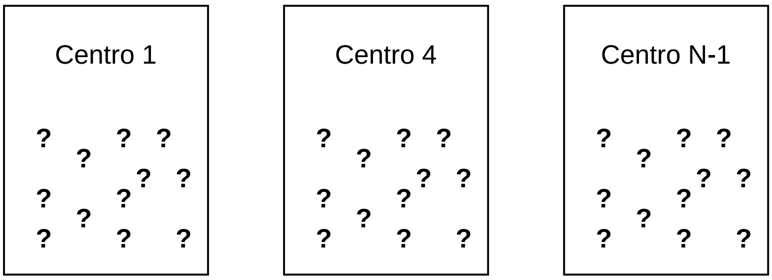
6.4.2. Muestreo por conglomerados o áreas.

Es una técnica útil cuando no se cuenta con listados de los elementos de la población directamente, sino como pertenecientes a alguna unidad intermedia para la que sí es posible confeccionar un listado o relación. Por ejemplo, si un banco está interesado en conocer la aceptación de sus servicios dirigidos a jóvenes universitarios, no podrá tener acceso a un listado de clientes potenciales de otra forma más que consultando a los propios centros a través de sus listas de matriculados. Por tanto, extendiendo el ámbito a todo el Estado, el instituto de investigación hará primero una muestra de centros universitarios, y a partir de ésta, censará a los jóvenes de aquéllos centros escogidos.

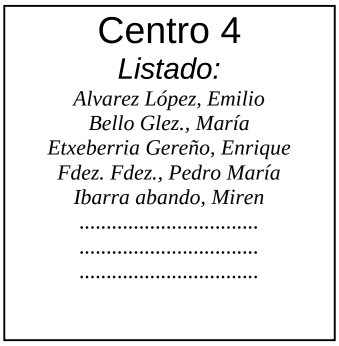


No sabemos *quiénes* son los estudiantes, aunque sí sabemos cuáles son todos los centros universitarios del Estado (unidad de muestreo)**.**

1º Escogemos un número representativo de centros.



2º Censamos a todos los elementos (estudiantes) de cada centro escogido.



6.4.3. Muestreo por etapas.

Es otra modalidad de muestreo igual a la anterior en casi todo. Se particiona a la población en unidades muestrales a las que se elige aleatoriamente, con la diferencia de que para pasar de la unidad muestral a los elementos definitivos (en el ejemplo de arriba, de los centros a los estudiantes) no se realiza un censo en cada uno de éstos, sino que se vuelve a realizar una elección aleatoria de elementos en cada unidad muestral, es decir, en el ejemplo anterior, una vez escogidos los centros de estudios, se elegiría aleatoriamente a un número de estudiantes de cada uno. Finalmente, con estudiantes escogidos aleatoriamente de todos los centros, se forma la muestra.

Cada partición de la población de una manera conforma una etapa. Así, el ejemplo ilustrado arriba es el de un muestreo bietápico, ya que en una primera etapa se accede a los centros, y en la segunda a los estudiantes.

Tanto el muestro por conglomerados como el de etapas tienen como aplicación poder acceder a poblaciones cuyos elementos no se conocen. Sin embargo, les diferencia el hecho de que el muestreo por conglomerados intenta ganar mayor precisión al censar a las personas que componen la unidad muestral, en tanto que el muestreo por etapas busca la economicidad al volver a extraer una muestra de cada unidad intermedia.

6.4.4. Muestreo sistemático.

Consiste en ir eligiendo las unidades muestrales de **k** en **k** unidades tomando origen una de ellas elegida al azar entre el elemento que ocupa el primer lugar y el que ocupa el lugar **k**.

Este procedimiento necesita un listado de los elementos de la población o ‘marco de la muestra’. Si éste dispone a los elementos con arreglo a algún criterio que no tiene nada que ver con el problema que se estudia, entonces la elección de la muestra no dará lugar a errores de tipo sistemático. Por ejemplo, un listado de empresas ordenadas según el nombre, no tiene porqué dar lugar a problemas. Sin embargo, si dicha lista se ordenara según la antiegüedad, podría pensarse que entre las empresas más antiguas la proporción de PYMEs es menor. Por tanto, para el momento en que se ha escogido el número necesario, y no habiendo llegado hasta el final del marco de muestra, la posibilidad de haber cometido un error sistemático (un número desproporcionado de empresas grandes) será real.

En consecuencia, siempre que se garantice no incurrir en el error señalado arriba, el muestreo sistemático aporta los mismos resultados que el M.A.S. con la ventaja adicional de que resulta más cómodo. Otra ventaja del muestreo sistemático es que se puede emplear sin tener que conocer la composición de los elementos que componen el marco de la muestra. Por ejemplo, cada i-ésima persona que sale de una tienda podría ser entrevistada, evitando que el entrevistador inconscientemente escoja a aquéllos/as que más de su agrado le resulten para entrevistar.

6.5.- La determinación del tamaño de muestra inicial.

Lo visto hasta ahora nos sirve para poder entender cómo se sigue en teoría un proceso de muestreo. Sin embargo, la elección material de los elementos de la población, y su acceso para recoger de ellos la información que sea necesaria, trae consigo muchos detalles y problemas prácticos que será preciso resolver. Comentamos a continuación los más importantes:

El valor **n** de la muestra, que se ha comentado hasta ahora, representa el tamaño neto o final que debe alcanzarse para asegurarnos de que los parámetros son estimados con el nivel de confianza y precisión deseados. En la práctica deberá considerarse un tamaño mucho mayor de entrevistados potenciales, considerando por un lado que ciertas personas del marco no son elegibles, y por otro lado, que habrá un número de personas que no acepten realizar la entrevista o que habiéndola comenzado, no la terminen. En consecuencia, tenemos que medir dos efectos:

- El ratio de incidencia (***Ri***), que es el porcentaje de personas elegibles para participar en el estudio. Determina cuántos contactos son necesarios para formar una muestra con un tamaño dado. Los criterios que se vayan a considerar para admitir un elemento como válido o no para el trabajo, dependen del caso concreto y de la experiencia e información que se dispongan.

Por ejemplo, para conocer la opinión que los clientes tienen hacia la calidad de un nuevo detergente para lavavajillas, habrá que considerar que no todas las personas elegibles para la entrevista tienen uno. Consultando datos del sector, se sabe que la penetración del lavavajillas en los hogares es del 30%. A partir de aquí, el número de hogares a los que inicialmente habrá que preguntar es:

$$\frac{100}{30} = 3{,}34 \quad \text{veces n}$$

Al mismo tiempo, si por ejemplo el nuevo producto había salido al mercado hace menos de un mes, habrá que descartar a aquéllas personas que compraron por última vez detergente para lavavajillas hace más de un mes. Si a través de la consulta de información secundaria (por ejemplo, mirando un panel de consumidores) se sabe que un 65% de usuarios compraron alguna marca en el último mes, lo que supone ampliar la muestra un 154%:

$$\frac{100}{65} = 1{,}54 \quad \text{veces n}$$

Finalmente, si se estima que han sido considerados todos los criterios que determinan la incidencia, se procede a calcular R_i:

$$\boldsymbol{R_i} = 0'3 * 0'65 = 0'195$$

- El ratio de finalización (***Rf***), que indica las negativas recibidas por aquéllos/as que siendo aptos para figurar en la muestra no desean colaborar para realizar o acabar la entrevista.

Por tanto, teniendo en cuenta estos dos efectos, el tamaño inicial de la muestra será:

$$\begin{array}{l}\text{Tamaño inicial} \\ \text{de la muestra}\end{array} = \frac{n}{Ri * Rf}$$

La magnitud del ratio de finalización depende de la experiencia en estudios pasados, o de lo que estimen los investigadores, así como del medio de recogida de información empleado. Por ejemplo, para el caso anterior, si se tiene experiencia previa de estudios parecidos realizados a través de un ómnibus, con un 80%, por lo que se toma dicho valor como ratio ***Rf***.

En consecuencia, con estos datos el tamaño inicial de la muestra será:

$$\begin{array}{l}\text{Tamaño inicial} \\ \text{de la muestra}\end{array} = \frac{n}{Ri * Rf} = \frac{n}{0{,}195 * 0{,}8} = 6{,}41 * n$$

En otras palabras, para obtener 100 entrevistas válidas tendremos que acudir inicialmente a 641 personas.

Bibliografía:

MALHOTRA, NareshK.: *Marketing Research. An Applied Orientation,* 2nd. Edition, Prentice Hall (, 1995, caps. 11 y 12

ORTEGA MARTÍNEZ, Enrique: *Manual de Investigación Comercial,* Pirámide (Madrid), 1992, caps.15 y 16.

SERRANO GÓMEZ, Francisco: *Marketing para economistas de empresa,* Colección Universidad, ESIC (Madrid), 1990, cap.10.