



# Capítulo 10

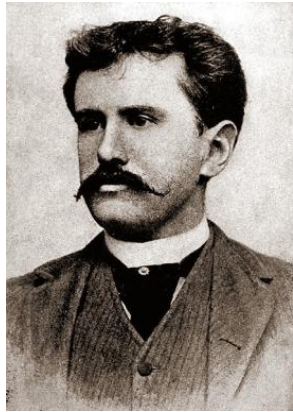
## Teoría y práctica del muestreo

Investigación de mercados  
Prof. Verónica Rosendo Ríos

## CAPÍTULO 10. TEORÍA Y PRÁCTICA DEL MUESTREO

---

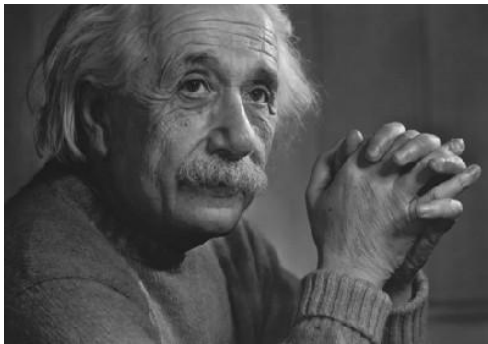
*“Un voto solo muestra en qué dirección sopla el aire”*



O. HENRY

William Sydney Porter (1862-1910), known by his pen name O. Henry, was an American writer.”

*“Haz todo lo más simple posible, pero no más simple”*



ALBERT EINSTEIN

Albert Einstein (1879-1955) was a German-born theoretical physicist who developed the general theory of relativity, effecting a revolution in physics.

## CONTENIDOS

1. Población y muestra
2. Tipos de muestreo: probabilístico y no-probabilístico
3. Errores
4. El tamaño muestral

## ★ POBLACIÓN:

“Colección de elementos que comparten características comunes y componen el **UNIVERSO** del problema de investigación”.

### Censo:

“Enumeración de los elementos de la población”.

## ★ MUESTRA:

“Subgrupo de la población que ha sido seleccionado para participar en el estudio”.

### Unidad muestral

“Elementos de la población que se van a tomar como muestra”.

- **Muestra vs. población:**
  - ✓ Presupuesto y límites de tiempo
  - ✓ Tamaño
  - ✓ Varianza
  - ✓ Coste de los errores muestrales
  - ✓ Coste de errores no muestrales
  - ✓ Confidencialidad

## EL PROCESO DE MUESTREO

Definir la  
población  
objetivo



Determinar el  
marco muestral



Seleccionar la  
técnica de  
muestreo



Estimar el  
tamaño  
muestral



Ejecutar el proceso  
muestral

## 1. Definir la población

✓ Precisión

✓ La población objetivo debe definirse en términos de:

- Elementos. -Unidades muestrales.

- Extensión. -Límite temporal.

## 1. Definir la población

Imagine que una empresa está considerando realizar una investigación de mercado que evalúe la respuesta del cliente a una nueva marca de desodorante para hombres.

¿Quién debería ser parte de la población objetivo?





## 2. Determinar el marco muestral

“Un marco muestral es una representación de los elementos de una población objetivo”

Ej.: Directorio telefónico...

- ✓ ¡¡¡Es importante comprobar la precisión!!!
- ✓ Ajustar el **error del marco muestral**:
  - ✓ Redefinir la población en términos del marco muestral.
  - ✓ Filtrar ejemplos por factores demográficos...
  - ✓ Realizar un esquema de ponderación.

## 3. Seleccionar la técnica de muestreo

### ★ ¿Bayesiano o tradicional?

- Bayesiano.
- Tradicional.

### ¿Con o sin remplazo?

- Con remplazo.
- Sin remplazo.

### ¿Probabilística vs. no-probabilística?

## 4. Determinar el tamaño muestral

### - Tamaños muestrales utilizados en estudios similares.

Tipo de estudio	Tamaño mínimo	Rango típico
Identificación del problema de investigación (ej. Mercado potencial de un producto)	500	-
Investigación para solventar problema (ej. Precio)	200	-
Estudios de test de producto	200	-
Estudios de test de mercado	200	-
TV, radio, publicidad (por anuncio testado)	150	-
Auditorías de mercado	10 comercios	10-20 comercios
Grupos focales	2 grupos	8-12 grupos

- **Tasas de incidencia.** Ej.: a veces los recursos y el tiempo son limitados.

- **Otras consideraciones.**

## 5. Ejecutar el proceso de muestreo

Especificaciones detalladas sobre aspectos del diseño muestral con respecto a:

- Población.
- Marco muestral.
- Unidad muestral.
- Técnica de muestreo.
- Tamaño muestral.

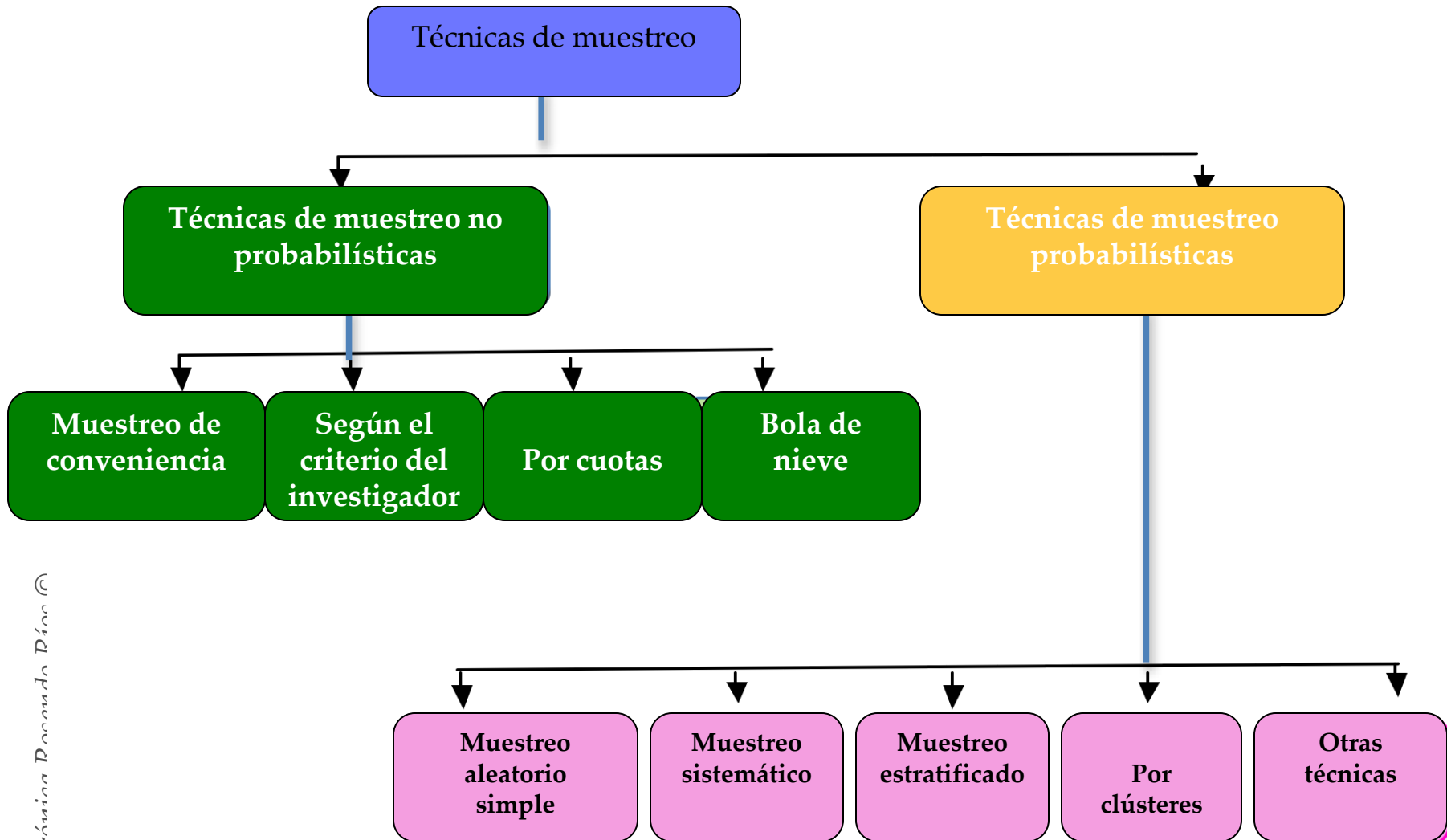
...

### **-Consideraciones:**

Ej.: Si los hogares constituyen la unidad muestral, se requiere una definición específica de hogar y de los procedimientos en caso de casas vacías...

# TIPO DE MUESTRA: PROBABILÍSTICA Y NO-PROBABILÍSTICA

Muestra: diseño y procedimientos



## ★ MUESTREO NO-PROBABILÍSTICO:

### 1. Muestreo de conveniencia:

- Intenta obtener una muestra de elementos de conveniencia.
- A veces los elementos se seleccionan por estar en el lugar correcto en el momento adecuado. Ej.: En el centro comercial, estudiantes, entrevistas en la calle...

**Ventajas:** Es el menos costoso y lento.

**Desventajas:** No es representativo. No debe utilizarse en investigaciones conclusivas.

## ★ MUESTREO NO-PROBABILÍSTICO:

### 2. Muestreo según el criterio (*Judgmental sampling*):

Los elementos poblacionales son seleccionados por el investigador de acuerdo con su criterio.

Por ejemplo, los **tests de mercado** seleccionados para determinar el potencial de un nuevo producto, seleccionar una muestra de personas en una empresa, testigos expertos que se utilizan en los juicios, tiendas comerciales seleccionadas para probar un nuevo sistema de merchandising.

## ★ MUESTREO NO-PROBABILÍSTICO:

### 3. Muestreo por cuotas

4. Es un tipo de muestreo según el criterio restringido de **dos fases**.
  - ✓ La **primera fase** consiste en desarrollar **categorías de control o cuotas**, de los elementos de la población.
  - ✓ En la **segunda fase**, los elementos muestrales se seleccionan por conveniencia, según el criterio del investigador.



# TIPO DE MUESTRA: PROBABILÍSTICA Y NO-PROBABILÍSTICA

## ★ NO-PROBABILÍSTICO:

### 3. Muestreo por cuotas

Ej.: Una empresa puede estar interesada en llevar a cabo un estudio para determinar la lealtad de los consumidores adultos con respecto a un periódico. 300.000 habitantes constituyen la población adulta. Se selecciona una muestra por cuotas de 1.000 adultos. Las características de control son sexo, edad y raza. Las cuotas se asigna de la siguiente manera:

Ejemplo de muestreo por cuotas

Control	Características	Composición de la población	Composición de la muestra	Columna1
		Porcentaje	Porcentaje	Número
Sexo	Hombre	46	46	460
	Mujer	54	54	540
		100	100	1000
Edad	18-30	24	24	240
	31-45	41	41	410
	46-60	15	15	150
	Mayor 60	19	19	190
		100	100	1000
Raza	Blanca	65	65	650
	Negra	29	29	290
	Otra	6	6	60
		100	100	1000

## ★ MUESTREO NO-PROBABILÍSTICO:

### 3. Muestreo por cuotas

#### ✓ Desventajas

- No hay garantía de que la muestra sea representativa.
- Si una característica que es importante para el problema se ha obviado, la muestra por cuotas no será representativa.
- A menudo se omiten características de control relevantes porque puede haber dificultades asociadas con muchas de las características de control.
- Debido a que los elementos en cada cuota se seleccionan según la conveniencia o el juicio del investigador, muchas fuentes de sesgo de selección están potencialmente presentes.
- Los entrevistadores pueden ir a áreas seleccionadas donde es más probable que se encuentren encuestados elegibles, o pueden evitar erróneamente áreas o personas que no consideran adecuadas. El muestreo por cuotas no permite determinar el error de muestreo.

## ★ MUESTREO NO-PROBABILÍSTICO:

### 3. Muestreo por cuotas

#### ✓ Ventajas

- Intenta obtener muestras representativas a un coste relativamente bajo.
- Sus principales ventajas son el ahorro en costes y la mayor conveniencia para los entrevistadores dado que seleccionan los elementos para cada cuota.
- Recientemente, se han impuesto controles más estrictos a los entrevistadores, así como procedimientos de entrevistas que tienden a reducir el sesgo de selección y se han sugerido pautas para mejorar la calidad de las muestras por cuotas.
- En determinadas condiciones, el muestreo por cuotas obtiene resultados parecidos a los del muestreo probabilístico convencional.

## ★ MUESTREO NO-PROBABILÍSTICO:

### 4. Muestreo de diseño de bola de nieve

- Se selecciona un grupo inicial de encuestados, generalmente al azar.
- Después de ser entrevistados, se les pide a estos encuestados que identifiquen a otras personas que pertenecen a la población objetivo de interés.
- Los encuestados posteriores se seleccionan según las **referencias**. Al obtener referencias de referencias, este proceso puede llevarse a cabo en **oleadas**, lo que da lugar al efecto bola de nieve.
- Un **objetivo principal** del muestreo de bola de nieve es estimar las características que son raras en la población.
- Por ejemplo, usuarios de servicios sociales particulares, como cupones de alimentos cuyos nombres no pueden ser revelados; grupos censales especiales; miembros de poblaciones minoritarias dispersas.
- La técnica de la bola de nieve se usa en la investigación industrial entre compradores y vendedores para identificar parejas de compradores y vendedores.

## ★ MUESTREO PROBABILÍSTICO:

- Las unidades muestrales se seleccionan **al azar**.
- Dado que los elementos se seleccionan al azar, es posible determinar el nivel de precisión de la muestra (el error de las estimaciones).
- Se pueden calcular los intervalos de confianza, que contienen el verdadero valor de la población con un determinado nivel de incertidumbre. Esto le permite al investigador hacer inferencias o proyecciones sobre la población objetivo de la que se tomó la muestra.

## ★ MUESTREO PROBABILÍSTICO:

Las técnicas de muestreo probabilístico se clasifican en base a (Malhotra, 2012):

- ✓ Elementos *vs.* clústeres.
- ✓ Unidades de igual probabilidad *vs.* desigual probabilidad.
- ✓ Selección no-estratificada *vs.* estratificada.
- ✓ Selección aleatoria *vs.* sistemática.
- ✓ Técnicas de una fase *vs.* multifases.

Todas las combinaciones posibles de estos cinco aspectos dan como resultado 32 técnicas diferentes de muestreo probabilístico. Las más típicas son: **muestreo aleatorio simple, muestreo sistemático, muestreo estratificado y muestreo por clústeres.**

## ★ MUESTREO PROBABILÍSTICO:

### 1. Muestreo aleatorio simple (SRS)

La muestra se selecciona por un procedimiento aleatorio de un marco muestral. (Este método es similar al de una lotería en la que se sacan nombres).

- Por ejemplo, imagine que se selecciona un tamaño muestral de 30 a partir de un marco muestral que contiene 800 elementos. Se podría utilizar una hoja de Excel para realizar una selección aleatoria.

## ★ MUESTREO PROBABILÍSTICO:

### 1. Muestreo aleatorio simple (SRS)

- Ventajas del SRS:

Se pueden generalizar los resultados y proyectarlos a la población objetivo; la mayoría de las inferencias estadísticas se basan en SRS.

- Limitaciones:

- Es difícil construir un marco muestral que permita extraer una muestra aleatoria simple.
- En segundo lugar, SRS puede dar como resultado grandes muestras o muestras distribuidas en áreas geográficas extensas.
- Tercero, SRS a menudo da lugar a una baja precisión.
- Cuarto, SRS puede generar o no una muestra representativa.



## ★ MUESTREO PROBABILÍSTICO:

### 2. Muestreo sistemático

- La muestra se elige seleccionando un punto de partida aleatorio y luego seleccionando cada elemento  $-ith$  en sucesión del marco muestral.
- Por ejemplo, hay 100.000 elementos en la población, y se requiere una muestra de 1.000. En este caso, el intervalo muestral,  $i$ , sería 100. Se seleccionaría un número aleatorio entre 1 y 100. Si, por ejemplo, este número fuese 23, la muestra constaría de los elementos 23, 123, 223, 323, y así sucesivamente.

## ★ MUESTRA PROBABILÍSTICA:

### 2. Muestreo sistemático

- El muestreo sistemático es similar al SRS en que cada elemento poblacional tiene el mismo número de posibilidades de ser seleccionado.
- Se diferencia del SRS en que solo las muestras de tamaño  $n$  que se pueden extraer de la población tienen una probabilidad de selección conocida e igual. Las muestras restantes de tamaño  $n$  tienen una probabilidad cero de ser seleccionadas.
- En el muestreo sistemático, el investigador asume que los elementos de la población están ordenados de algún modo.

## ★ MUESTREO PROBABILÍSTICO:

### 3. Muestreo estratificado

- Es un **proceso de dos etapas**.
- Los estratos deben ser **mutuamente exclusivos** y **colectivamente exhaustivos**.
- Posteriormente, **se seleccionan los elementos de cada estrato aleatoriamente**, normalmente usando el SRS.
- Técnicamente, solo se debe emplear SRS al seleccionar los elementos de cada estrato. En la práctica, a veces se utilizan muestreos sistemáticos y otros procedimientos de muestreo probabilístico.
- El muestreo estratificado difiere del muestreo por cuotas en que los elementos de la muestra se seleccionan probabilísticamente en lugar de basarse en la conveniencia o el juicio del investigador.

## ★ MUESTREO PROBABILÍSTICO:

- 3. Muestreo estratificado
- Variables de estratificación.
- Los elementos dentro de un estrato deben ser lo más homogéneos posible, pero los elementos en diferentes estratos deben ser lo más heterogéneos posible.
- Ej.: Usuarios o no de tarjetas de crédito, tamaño de la empresa o tipo de industria o sector.

## ★ MUESTREO PROBABILÍSTICO:

- 3. Muestreo estratificado: características
  - ✓ ¿Cuántas variables se deben incluir?
  - ✓ ¿Muestreo proporcional o desproporcionado?

## ★ MUESTREO PROBABILÍSTICO:

### 4. Muestreo por clústeres (o conglomerados)

- La población objetivo se divide primero en subpoblaciones o agrupaciones mutuamente excluyentes y colectivamente exhaustivas o clústeres.
- Luego se selecciona una muestra aleatoria de conglomerados, basada en una técnica de muestreo probabilístico como SRS.
- Muestreo por clústeres de una etapa.
- Muestreo por clústeres de dos etapas.
- Muestreo clústeres de probabilidad proporcional al tamaño (PPS).
- Muestreo polietápico.

## ★ MUESTREO PROBABILÍSTICO:

### Tipos de muestreo por clústeres



## ★ MUESTREO PROBABILÍSTICO:

### 4. Muestreo por clústeres

- Ventajas: El muestreo por clústeres tiene dos ventajas principales:
- Viabilidad y bajo coste.
- Limitaciones:

Por ejemplo, los hogares en un bloque tienden a ser similares en lugar de distintos

- Puede resultar difícil calcular e interpretar estadísticas basadas en clústeres.



# TIPO DE MUESTRA: PROBABILÍSTICA Y NO-PROBABILÍSTICA

## Ventajas y limitaciones de las principales técnicas de muestreo

Técnica	Ventajas	Limitaciones
<b>Muestreo no probabilístico</b>		
Muestreo por conveniencia	Menos caro, requiere menos tiempo, más conveniente	Sesgo de selección, muestra no representativa, no recomendado para investigación descriptiva o causal
Muestreo por criterio	Bajo coste, conveniente, no requiere mucho tiempo	No permite generalizar, subjetivo
Muestreo por cuotas	La muestra se puede controlar para ciertas características	Sesgo de selección, no asegura representatividad
Bola de nieve	Puede estimar características especiales	Requiere mucho tiempo
<b>Muestreo probabilístico</b>		
Muestreo aleatorio simple	Fácilmente entendible, resultados proyectables Puede aumentar la representatividad, más fácil de implementar que el SRS, marco muestral no necesario	Difícil construir un marco muestral, caro, baja precisión, no asegura representatividad
Muestreo sistemático		Puede disminuir la representatividad Difícil seleccionar variables de estratificación importantes, no es posible estratificar muchas variables, caro
Muestreo estratificado	Incluye todas las subpoblaciones importantes, precisión Fácil de implementar, efectivo con respecto al coste	
Muestreo por cluster		impreciso, difícil de interpretar los resultados

## ★ OTRAS TÉCNICAS DE MUESTREO PROBABILÍSTICO:

### 1. Muestreo secuencial

Los elementos de la población se muestrean secuencialmente:

- El tamaño muestral no se conoce de antemano.
- En cada etapa, esta regla o criterio indica si el muestreo debe continuar o si se ha obtenido suficiente información.
- El muestreo secuencial se utiliza, por ejemplo, para determinar las preferencias de dos alternativas competitivas.

Ej.: Para establecer las preferencias del consumidor entre un modelo estándar y un modelo de luxe o sobre dos marcas competitivas.

## ★ OTRAS TÉCNICAS DE MUESTREO PROBABILÍSTICO:

### 2. Muestreo doble (o muestreo de dos fases)

Ciertos elementos poblacionales se muestrean dos veces.

El proceso puede extenderse a tres o más fases, y las diferentes fases pueden tener lugar simultáneamente o en diferentes momentos.

El muestreo doble **puede resultar útil cuando no existe un marco muestral disponible** para seleccionar las unidades muestrales pero se sabe que los elementos del marco muestral están contenidos dentro de un marco de muestreo más amplio.

## ★ OTRAS TÉCNICAS DE MUESTREO PROBABILÍSTICO:

### 2. Muestreo doble (o muestreo de dos fases)

Ejemplo:

Un investigador quiere seleccionar hogares, en una ciudad determinada, que consuman zumo de manzana. Los hogares de interés están incluidos en el conjunto de todos los hogares, pero el investigador no sabe cuáles son.

Al aplicar el muestreo doble, el investigador obtendría un marco de muestreo de todos los hogares en la primera fase. Esto se construiría desde el directorio de la ciudad o se compraría. Luego se tomaría una muestra de hogares, utilizando un muestreo aleatorio sistemático para determinar la cantidad de zumo de manzana consumido.

En la segunda fase, los hogares que consumen zumo de manzana serían seleccionados y estratificados de acuerdo con la cantidad de zumo de manzana consumido. Luego se tomaría una muestra aleatoria estratificada y se realizarían preguntas detalladas sobre el consumo de zumo de manzana.

## ★ Muestreo probabilístico *vs.* no-probabilístico

- ✓ Tipo de investigación
- ✓ Generalizaciones
- ✓ Precisión

## ★ Muestreo probabilístico *vs.* no-probabilístico

- ✓ Variabilidad en la población
- ✓ Consideraciones estadísticas
- ✓ Consideraciones operacionales

# TIPO DE MUESTRA: PROBABILÍSTICA Y NO-PROBABILÍSTICA

## Condiciones que favorecen el uso del muestreo probabilístico *vs.* no-probabilístico

Factores	Condiciones que favorecen el uso de	
	Muestreo no probabilístico	Muestreo probabilístico
Tipo de investigación	Exploratoria	Conclusiva
Relativa magnitud de los errores muestrales y no muestrales	Los errores no muestrales aumentan	Los errores muestrales aumentan
Generalización	No permite proyecciones estadísticas	Permite proyecciones estadísticas
Precisión	En caso de errores no muestrales (altos), el juicio del investigador permite mayor control	Eliminación del sesgo de selección y posibilidad de calcular el error muestral
Variabilidad de la población	Población homogénea	Población heterogénea
Consideraciones estadísticas	Desfavorables	Favorables
Consideraciones operacionales	Favorables	Desfavorables

# TIPO DE MUESTRA: PROBABILÍSTICA Y NO-PROBABILÍSTICA

**Tabla comparativa de las diferencias entre los métodos muestrales probabilísticos y no probabilísticos**

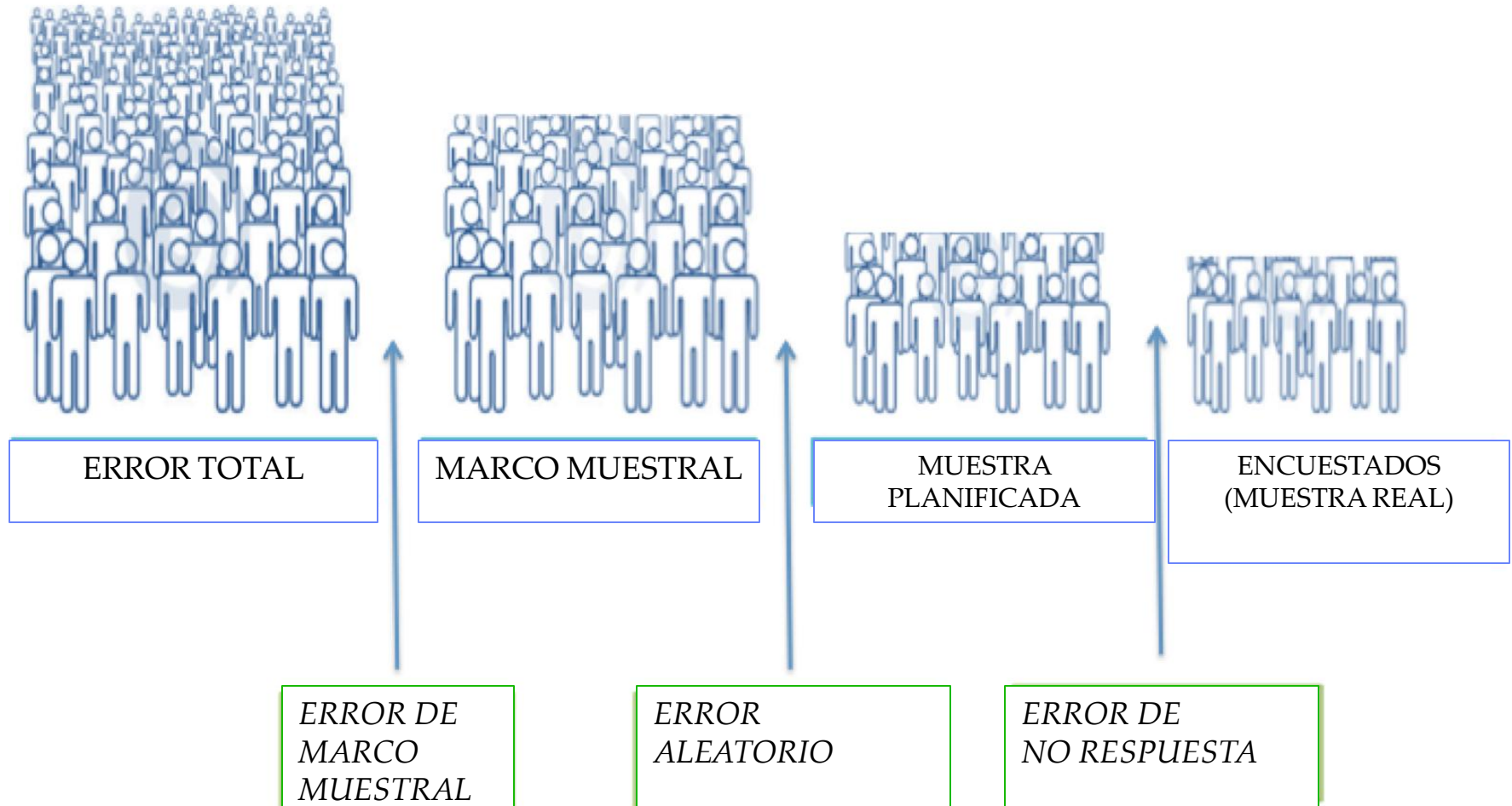
<b>Factores comparativos</b>	<b>Muestreo probabilístico</b>	<b>Muestreo no probabilístico</b>
Lista de los elementos poblacionales	Se necesita la lista completa	no necesaria Se necesitan detalles de hábitos, actividades, etc
Información de las unidades muestrales	Se identifica cada unidad	Se necesitan detalles de hábitos, actividades, etc
Habilidades muestrales requeridas	Se requieren habilidades	pocas habilidades requeridas
Tiempo	Requiere tiempo	Poco tiempo
Coste por unidad de muestra	Moderado a alto	Bajo
Estimaciones de parámetros poblacionales	Sin sesgo	Sesgadas
Representatividad de la muestra	Buena	Indeterminada, dudosa
Precisión y fiabilidad	Calculada con intervalos de confianza	Desconocidos No se dispone de métodos estadísticos
Medida del error muestral	Medidas estadísticas	

Fuente: *Hair et al.* (2006).



# EL ERROR DE MUESTREO

Figura de los errores asociados con el muestreo



## ESTADÍSTICOS

El enfoque estadístico para determinar el tamaño muestral considerado aquí se basa en la **inferencia estadística tradicional**. En este enfoque, el nivel de precisión se especifica de antemano. Se basa en la construcción de intervalos de confianza alrededor de las medias o proporciones de la muestra.

✓ **Los estadísticos muestrales o estadísticos** son las medidas numéricas (tales como la media, la varianza, etc.) cuando se aplican a la muestra. Cuando estas medidas numéricas se aplican a la población, se denominan parámetros poblacionales o simplemente **parámetros**. La media estadística muestral ( $\bar{X}$ ) estima la media poblacional  $\mu$ ; la varianza estadística muestral ( $S^2$ ) estima la varianza poblacional  $\sigma^2$ .

✓ **Distribución muestral:** El uso principal de la distribución muestral y los conceptos relacionados con ella es ayudar a predecir cuánto está de cerca el estadístico muestral del parámetro poblacional, y con qué probabilidad.

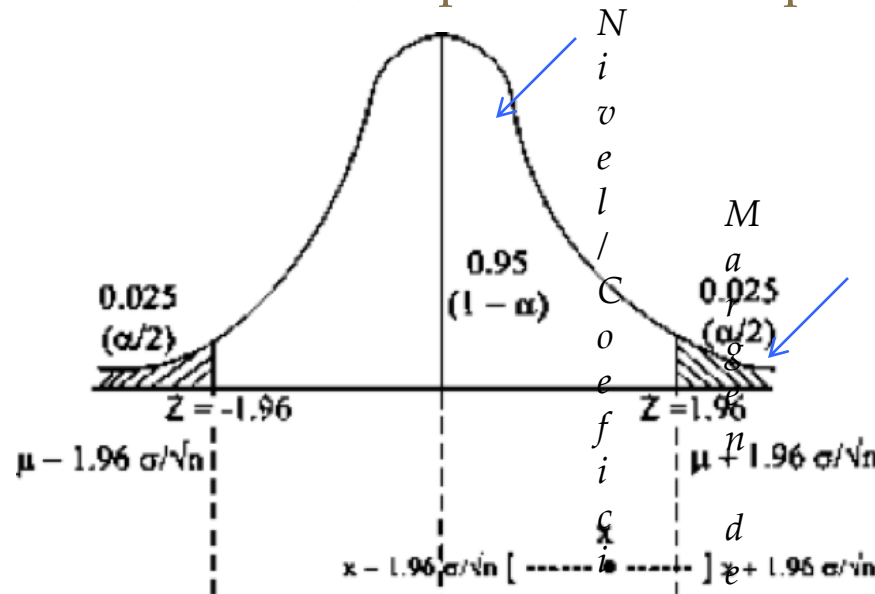
## ESTADÍSTICOS

✓ **Grados de libertad (*degrees of freedom*):** Si nos piden que elijamos tres números aleatorios  $a$ ,  $b$  y  $c$ , somos libres de elegir tres números sin ninguna restricción, en otras palabras, tenemos 3 grados de libertad. Pero si los tres números se juntan en un modelo  $a + b + c = 10$ , entonces tenemos solo 2 grados de libertad. La elección de  $a$  y  $b$  puede ser arbitraria, pero  $c$  está limitada a tomar un valor específico que satisfaga el modelo. El uso de los grados de libertad ( $df$ ) es un mecanismo compensatorio en los cálculos, específico del contexto/situación donde se aplica.

✓ **Intervalo de confianza (CI):** Una estimación de intervalo, con su nivel de confianza asociado, se denomina intervalo de confianza. El intervalo de confianza es un rango de números que probablemente contiene el parámetro poblacional desconocido, con un nivel de confianza contiguo. Es un rango estadístico de valores dentro del cual se espera que se encuentre el valor verdadero del parámetro de la población objetivo.

## ESTADÍSTICOS

- ✓ El área combinada bajo las curvas en las colas (i.e.  $1 - 0.95 = 0.05$  en el ejemplo siguiente) se denomina **nivel de significación  $\alpha$** , y/o **margen de error**.
- ✓ -El area bajo las curvas excluyendo las colas ( $1 - \alpha$ ) se denomina **coeficiente de confianza**.
- ✓ -El coeficiente de confianza x 100, expresado como porcentaje, es el **nivel de confianza**.



## El enfoque del intervalo de confianza

Por ejemplo, imagínese que un investigador toma una muestra aleatoria simple de 300 hogares para estimar los gastos mensuales de las compras en grandes almacenes y estima que el gasto mensual promedio por hogar para la muestra es de 182 €. Estudios anteriores han demostrado que la desviación estándar de la población  $\sigma$  (o  $s$ ) puede suponerse en 55 €.

Se pretende encontrar un intervalo dentro del cual se encontraría una proporción fija de la media muestral. Suponga que se quiere determinar un intervalo al rededor de la media poblacional que incluya un 95% de la media muestral, basado en la muestra de 300 hogares anterior. El 95% podría dividirse en dos partes iguales, uno por debajo y otro por encima de la media, como muestra la siguiente figura.

El cálculo del intervalo de confianza implica determinar una distancia por debajo ( L ) y otra por encima ( U ) de la media poblacional (  $\bar{x}$  ), que contenga un área específica de la curva normal. Se puede entonces fijar un intervalo de confianza del 95% al rededor de la media muestral de 182€. Como primer paso, se calcularía el **error estandar de la media (SEM)**:

•  $SEM = 55 / \sqrt{300} = 3.18$

$$\frac{s}{\sqrt{n}}$$

- De las tablas estadísticas generales se sabe que el 95% central de la distribución normal se encuentra entre  $\pm 1,96$  z valores. El intervalo de confianza del 95% se calcula mediante:  $\bar{x} \pm z \frac{s}{\sqrt{n}}$
- Calculamos:  $182 \pm 1,96 (3,18) = 182 \pm 6.23$
- Por tanto, el intervalo de confianza del 95% tiene como rango de 175,77€ a 188.23€. La probabilidad de que la media poblacional se encuentre entre 175.77€ y 188.23€ es del 95%.

## Cálculo del tamaño muestral: MEDIAS

Imaginemos ahora que el investigador desea estimar el gasto mensual del hogar en compras de tiendas departamentales de manera más precisa, de modo que **la estimación esté dentro de  $\pm 5.00$  €** del verdadero valor de la población. **¿Cuál debería ser el tamaño de la muestra?**

- ✓ Especifique el **nivel de precisión**. Esta es la diferencia máxima permisible (D) entre la media de la muestra y la media de la población. En nuestro ejemplo,  $D = \pm 5.00$  €.
- ✓ Especifique el **nivel de confianza**. Supongamos que se precisa un nivel de confianza del 95%.
- ✓ Determine el **z valor asociado al nivel de confianza** seleccionado utilizando las tablas de la distribución normal. Para un nivel de confianza del 95%, la probabilidad de que la media de la población caiga fuera del intervalo es de 0,025 ( $0,5 / 2$ ). El valor de z asociado es 1.96.

## Cálculo del tamaño muestral: MEDIAS

✓ Determine la **desviación estándar de la población**. La desviación estándar de la población puede conocerse a partir de fuentes secundarias; o podría estimarse realizando un estudio piloto. Alternativamente, podría estimarse sobre la base del juicio del investigador (ej.: 55 en nuestro caso).

$$\bullet n = \frac{\sigma^2 Z^2}{D^2}$$

En el ejemplo,  $n = \frac{55^2 (1.96)^2}{5^2} = 464.83 = 465$  (redondeado al siguiente nº entero).

## Cálculo del tamaño muestral: MEDIAS

✓ Si el tamaño muestral resultante **representa un 10% o más de la población**, debería aplicarse el **factor de corrección de poblaciones finitas (fpc)**. El tamaño muestral requerido debería calcularse con la fórmula:

$$n_c = \frac{nN}{N+n-1}$$

Donde  $n$  = tamaño muestral sin fpc y  $n_c$  = tamaño muestral con fpc



## Cálculo del tamaño muestral: MEDIAS

✓ Si la **desviación estándar de la población,  $\sigma$ , se desconoce y se utiliza una estimación**, se debería re-estimar una vez que la muestra se haya seleccionado. La desviación estándar muestral,  $s$ , se utiliza como estimador de  $\sigma$ . Se debería calcular un intervalo de confianza revisado para determinar el verdadero nivel de precisión obtenido.

✓ Supongamos que el valor de 55 usado para  $\sigma^2$  era una estimación porque se desconocía el valor verdadero. Se toma una muestra de  $n = 465$ , y se calcula una media de 180 y una desviación estándar de la muestra de 50. El intervalo de confianza revisado se calcula posteriormente como:

$$\bullet \bar{X} \pm Z s_x = 180.00 \pm 1.96 \frac{50}{\sqrt{465}} = 180.00 \pm 4.55 \text{ or } 175.45 \leq \mu \leq 184.55$$

- 
- Nótese que el intervalo de confianza obtenido es más pequeño ( $\bar{X} \pm 4.55$ ) que el inicialmente planificado ( $\bar{X} \pm 5.00$ ), porque se sobre-estimó la desviación estándar poblacional, en base a la desviación estándar muestral.

## Cálculo del tamaño muestral: PROPORCIONES

Si el estadístico de interés es una proporción en lugar de una media, el cálculo del tamaño muestral es similar. Supongamos que el investigador está interesado en estimar la proporción de hogares que tienen una tarjeta de crédito asociada a una tienda comercial. Se deberían seguir los siguientes pasos:

1. Especifique **el nivel de precisión**. Supongamos que el nivel de precisión deseado es tal que el intervalo admisible se fija como  $D = p - \pi = \pm 0.05$ .
2. Especifique el **nivel de confianza**. Supongamos que se desea un nivel de confianza del 95%.
3. Determine el **z valor** asociado con el nivel de confianza. Tal como se vió, sería 1.96
4. Estime **la proporción poblacional  $\pi$** . Como se explicó, la proporción poblacional puede estimarse de fuentes secundarias, un estudio piloto, o estar basadas en el juicio del investigador. Supongamos que en base a datos secundarios el investigador estima que el 64% de los hogares de la población objetivo pose una tarjeta de crédito de un centro comercial. Por tanto,  $\pi=0.64$ .

## Cálculo del tamaño muestral: PROPORCIONES

5. Determine el tamaño muestral utilizando la fórmula de error estándar de la proporción.

$$\sigma_p = \frac{p - \pi}{z} = \frac{D}{z} = \frac{\sqrt{\pi(1-\pi)}}{n} \quad \text{o} \quad n = \frac{\pi(1-\pi)z^2}{D^2}$$

✓ En nuestro ejemplo,  $n = \frac{0.64 (1-0.64) (1.96)^2}{(0.05)^2} = 354.04 = 355$  (redondeado al siguiente  $n^\circ$ )

6. Si el **tamaño muestral resultante representa un 10% o más de la población**, debería aplicarse el **factor de corrección para poblaciones finitas (fpc)**. El tamaño muestral requerido se debería calcular en base a la fórmula:

$$n_c = \frac{nN}{N+n-1}$$

Donde  $n$  = tamaño muestral sin fpc;  $n_c$  = tamaño muestral con fpc.

## Otras técnicas de muestras probabilísticas

Hasta ahora, se ha considerado el cálculo del tamaño muestral basado en el método de inferencia estadística que utiliza la técnica de muestreo aleatorio simple. El cálculo del tamaño muestral con otras técnicas de muestras probabilísticas se basa en los mismos principios subyacentes.

- ✓ El investigador debe especificar el nivel de precisión y el grado de confianza y estimar la distribución muestral del test estadístico.
- ✓ Adicionalmente, el investigador debe tener en cuenta la **variabilidad dentro de los estratos o dentro -y entre- los clústeres o conglomerados**.
- ✓ Una vez determinado el tamaño muestral, la muestra se distribuye entre los diferentes estratos o grupos. Esto aumenta la complejidad.
- ✓ Como regla general, **para proporcionar la misma fiabilidad que proporciona el muestreo aleatorio simple, los tamaños de las muestra serían los mismos para el muestreo sistemático, más pequeños para el muestreo estratificado y mayores para el muestreo por conglomerados**.

## Ajuste del tamaño muestral determinado estadísticamente

- ✓ El tamaño muestral determinado estadísticamente es el tamaño muestral final o tamaño muestral neto que debe conseguirse para asegurar que los parámetros se estiman con el grado de precisión y nivel de confianza deseados.
- ✓ En las encuestas, este número representa el número de entrevistas que deben completarse.
- ✓ Para conseguir este tamaño muestral final, deben generalmente contactarse un número potencialmente bastante más grande de encuestados.
- ✓ En otras palabras, el tamaño muestral inicial debe ser mucho mayor porque normalmente los ratios de respuesta (y de incidencia) son menores del 100%.

## Ajuste del tamaño muestral determinado estadísticamente

- ✓ **La tasa de incidencia** se refiere a la tasa de ocurrencia o el porcentaje de personas elegibles para participar en el estudio.
- ✓ La tasa de incidencia determina la cantidad de contactos que se deben evaluar para un tamaño muestral determinado.
- ✓ Supongamos que en un estudio de ordenadores portátiles se necesita una muestra compuesta por mujeres trabajadoras con edades comprendidas entre los 25 y los 55. De las mujeres de edad entre los 20 y los 60 que pueden ser potencialmente elegibles, aproximadamente el 75% son trabajadoras freelance de entre 25 y 55 años. Esto significa que, de media se debe contactar a  $(1/0.75)$  o 1.33 mujeres para obtener un encuestado adecuado.
- ✓ Criterios adicionales para ser un encuestado elegible (por ejemplo, comportamiento de compra) harían aumentar el número de personas a contactar. Supongamos que un requisito de elegibilidad añadido es que las mujeres deben haber utilizado un ordenador portátil al menos durante los dos meses anteriores a la encuesta. Se estima que el 60% de las mujeres contactadas cumplen este criterio. Por tanto, la tasa de incidencia es de  $0.75 \times 0.60 = 0.45$ . **Así pues, el tamaño muestral final debería aumentarse por un factor de  $(1/0.45)$  o 2.22.**

## Ajuste del tamaño muestral determinado estadísticamente

- Igualmente, en el cálculo del tamaño muestral deben tenerse en cuenta **rechazos anticipados** de gente elegible. La tasa de respuesta denota el porcentaje de encuestados elegibles que completan la entrevista. Si, por ejemplo, el investigador espera una tasa de respuesta del 80% de encuestados elegibles, el número de contactos debería aumentarse por un factor de 1.2.
- La tasa de respuesta e incidencia juntas implican que el número de encuestados potenciales que han de contactarse –es decir, el tamaño muestral inicial– debería ser 2.22x1.2 ó 2.77 veces el tamaño muestral requerido. En general, si hay  $x$  factores de cualificación con una incidencia de  $Q_1, Q_2, Q_3, \dots, Q_c$ , cada una expresada como proporción, se daría:

$$\text{Tasa de incidencia} = Q_1 \times Q_2 \times Q_3 \times \dots \times Q_c$$

- **Tamaño muestral inicial** = tamaño muestral final  
tasa de incidencia x tasa de respuesta
- El número de unidades que deben ser muestreadas vendrá determinado por el tamaño muestral inicial.

## Ajuste por falta de respuesta (*non-response*)

La alta tasa de respuesta aumenta la probabilidad de que el sesgo de falta de respuesta sea sustancial. Las tasas por falta de respuesta siempre deben ser reportadas, y cuando sea posible, **los efectos de la tasa por falta de respuesta deben ser estimados.**

- Esto se puede hacer **vinculando la tasa de no respuesta a las diferencias estimadas entre los encuestados y los que no respondieron.**

Por ejemplo, se podrían extrapolar las diferencias que se han encontrado cuando los encuestados han sido contactados en un segundo intento, o se puede intentar realizar un seguimiento sobre una submuestra de personas que no han respondido. Alternativamente, puede que resulte útil estimar estas diferencias en base a otras posibles fuentes.



## Ajuste por falta de respuesta (*non-response*)

- ✓ Submuestreo (submuestra de personas que no han respondido)
- ✓ Remplazo
- ✓ Sustitución
- ✓ Estimaciones subjetivas

## Ajuste por falta de respuesta (*non-response*)

### ✓ Análisis de tendencias

	Porcentaje de respuestas	Media de gasto (dólar)	Porcentaje de olas de respuesta previas
Primer mail	12	412	-
Segundo mail	18	325	79
Tercer mail	13	277	85
No-respuesta	-57	-230	91

Ajuste por falta de respuesta (*non-response*)

✓ Ponderación

✓ Imputación



AMERICAN  
KENNEL CLUB<sup>SM</sup>

## EJEMPLO



### “American Kennel Club intenta mantener a los dueños de mascotas fuera de la caseta del perro”

La American Kennel Club (AKC) es una organización dedicada a promocionar los perros de raza así como la preocupación por su salud y bienestar entre las familias de propietarios. La compañía llevó a cabo un estudio para investigar la propiedad y aceptación de los perros en el vecindario. La AKC utilizó el muestreo por cuotas en su estudio, para comparar la actitud hacia los perros de los que son dueños frente a los que no, en base a una muestra de mil personas. En una muestra tan pequeña de la población de U.S., puede que algunos grupos no estuviesen representados, así que se introdujeron algunas cuotas en el estudio en base a la edad de los entrevistados, el sexo y las categorías geográficas. Las unidades de muestreo primaria se seleccionaron por marcación telefónica aleatoria. En la siguiente fase de selección, los investigadores se aseguraron de que los encuestados se enmarcaban dentro de alguna de las cuotas de cada grupo. Posteriormente filtraron a los encuestados en base a los que eran o no dueños de perros.

Uno de los objetivos de la encuesta era ayudar a los dueños de los perros a comprender las preocupaciones de los vecinos del barrio. La AKC pretendía proporcionar así factores claves para mejorar la educación responsable de los propietarios de perros para contribuir con la armonía del vecindario. El estudio señaló que las principales preocupaciones de los no propietarios eran los ladridos de los perros y el hecho de que los propietarios no recogiesen los “restos orgánicos”. Lisa Peterson, directora de comunicación de AKC, comentó, “Cualquiera que se esté considerando criar a un perro debería darse cuenta de que conlleva un compromiso a largo plazo de entre 10 y 15 años, además de un gasto adicional”. El estudio puso de manifiesto tanto las ventajas de poseer un animal de compañía, como los inconvenientes y obligaciones. Uno de las principales ventajas resultó ser que los propietarios se auto-describían como personas más flexibles y felices.

Fuente: Zikmund et al. (2013). “AKC Mission Statement” and “History of the American Kennel Club”, American Kennel Club, 2011.

## REFERENCIAS BIBLIOGRÁFICAS

- Brown, T. J., y Suter, T. (2012): *MR*. South Western, Cenage Learning. USA.
- Hair, J.; Bush, R., y Ortinau, D. (2006): *Marketing research. Within a changing environment*. Revised International Edition (3<sup>rd</sup> ed.). McGraw-Hill, New York, USA.
- Malhotra, N. K. (1996): *Marketing Research. An Applied Orientation*. 2<sup>nd</sup> ed. Prentice-Hall International. USA.
- Rosendo-Ríos, V.; de Esteban, J., y Antonovica, A. (2012): *MR: Development of Theoretical Concepts for Market Research I and II*. South Western, Cenage Learning. USA.
- Zikmund, W. G.; Babin, B. J.; Carr, J. C., y Griffin, M. (2013): *Business Research Methods*. 9<sup>th</sup> Edition. South Western, Cenage Learning. USA.