

Clase 4

Muestreo



 **Le Sphinx**

CAMPUS

El pack Sphinx dedicado a la enseñanza de métodos y técnicas para encuestas

Objetivos

Objetivos metodológicos

- Comprender el impacto de las opciones tomadas al elegir la población que va a ser encuestada, en relación a la calidad y fiabilidad de los resultados
- Comprender la noción de representatividad
- Saber evaluar la fiabilidad y la exactitud de los resultados de la encuesta en relación con el tamaño y la composición de la población estudiada
- Dominio de las técnicas de composición y ajuste de muestras aleatorias y empíricas

Objetivos Sphinx

- No se necesita el uso de software

Tabla de contenidos

1. Muestra y estimación estadística
2. Métodos de muestreo
3. Ajuste de la muestra

1. Muestra y estimación estadística

1.1 Cuestionario y muestra

- Se lleva a cabo una encuesta o estudio estadístico sobre una **población objetivo**. Entrevistar a todos es como hacer un censo.
- A menudo es muy difícil realizar un censo, es decir, reunir la información deseada de **TODOS** los miembros de la población objetivo, dentro de un plazo y un presupuesto razonables.

Por ejemplo, para una encuesta sobre las intenciones de voto para las elecciones municipales de Córdoba, la población matriz está formada por todos los cordobeses inscritos en los censos electorales en el momento de las elecciones, que representan actualmente 784.000 personas. Obtener su opinión en pocos días sería muy complicado y sobre todo muy caro.



- Si un censo es imposible o demasiado costoso, se realiza una **encuesta**. Es decir, simplemente se recoge información de una **subpoblación** de la población objetivo, que se espera que sea un "modelo reducido" = la muestra.

1.2 Estimación estadística

- Debido a que la muestra es más pequeña, los resultados observados en la muestra son sólo una **estimación**.
- La **precisión** de una muestra se caracteriza por el "error de estimación" que introduce. El error de estimación de la encuesta varía según las siguientes fórmulas, que son diferentes según se observe un promedio o un porcentaje de la muestra.

Parámetros a estimar	Fórmulas	Comentarios	Ejemplos
Una media	$e = 1,96 \times (\sigma/\sqrt{n})$	σ es la desviación estándar y n es el tamaño de la muestra	Se entrevista a una muestra de 500 españoles y se registra una media de 200 minutos (¡por día!) de tiempo de visión de la televisión, con una desviación estándar de 45 minutos. $Error = 1,96 \times (45/\sqrt{500}) = 3,94$ minutos
Un porcentaje (o una proporción)	$e = 1,96 \times \sqrt{[(p \times 1 - p) / n]}$	p es el porcentaje observado en la muestra y n es el tamaño de la muestra	Se encuestó a una muestra de 1.500 españoles, de los cuales el 30% estaba satisfecho con la acción del gobierno. $Error = 1,96 \times \sqrt{[(0,3 \times 0,7) / 1500]} = 0,023$ o sea 2,3 puntos

1.3 Intervalo de confianza

- A partir del error de estimación calculado previamente, puede definirse el **intervalo de confianza** del resultado de la muestra observada. El intervalo de confianza (I) es un rango de valores dentro del cual el resultado real, es decir, el de la población objetivo, es seguro (95%).

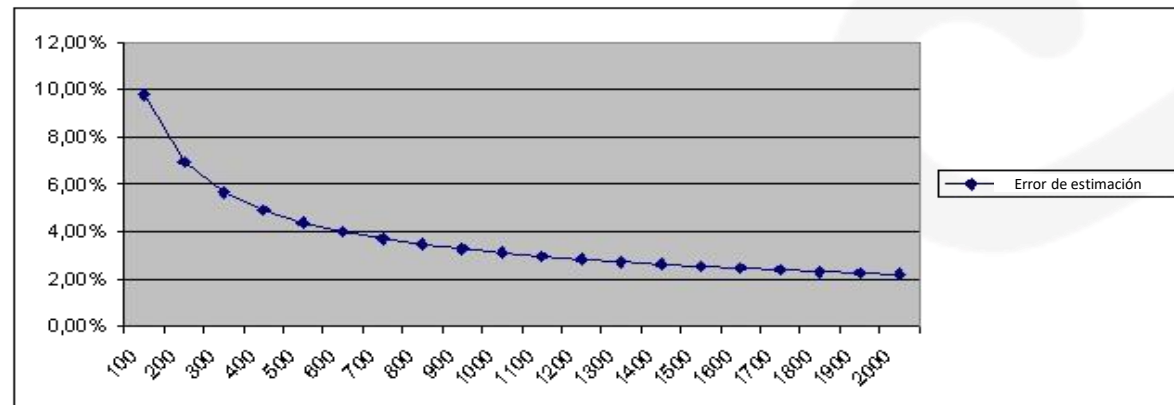
Parámetros a estimar	Fórmulas	Comentarios	Ejemplos
Una media	$I = [m - e ; m + e]$	m = muestra promedio	Para 500 españoles, el promedio de tiempo que pasan viendo la televisión es de 200 minutos por día, con una desviación estándar de 45 minutos. Error = 3,94 minutos o 4 minutos $I = [196 \text{ minutos}; 204 \text{ minutos}]$ → Para el conjunto de los españoles, la media de tiempo que pasan viendo la televisión es de entre 196 y 204 minutos al día.
Un porcentaje (o proporción)	$I = [p - e ; p + e]$	p es el porcentaje observado en la muestra	De los 1.500 españoles, el 30% dice estar satisfecho con la acción del gobierno. Error = 2,3 puntos $I = [27,7 \% ; 32,3 \%]$ → Para el conjunto de la población española, el porcentaje de personas satisfechas con la acción del gobierno se sitúa entre el 27,7 y el 32,3%.

1.4 Error máximo

- A menudo necesitamos saber el error general asociado con una encuesta. La práctica habitual es utilizar el **máximo error** asociado a una muestra determinada. Este siempre está en $p = 0,5$; es decir, cuando se intenta estimar el error para un porcentaje observado del 50%.

Por ejemplo, los resultados observados en una encuesta política basada en una muestra de 1.500 españoles deben considerarse con un margen de error máximo de 2,53 puntos.

- El error no es directamente proporcional al **tamaño de la muestra**. El simple hecho de duplicar el tamaño de la muestra no duplica la precisión. Básicamente, para duplicar la precisión, el tamaño de la muestra debe multiplicarse por cuatro. Las encuestas nacionales suelen incluir muestras de mil personas. Esta es la mejor relación precisión/tamaño, que en realidad equivale a la relación calidad/precio de una muestra para un instituto de encuestas.



2. Métodos de muestreo

- Más allá de su precisión, el objetivo de la constitución de una muestra es lograr la **representatividad**.
- Para obtener una muestra representativa, distinguimos:
 - métodos **aleatorios** (o probabilísticos) que se basan en teorías probabilísticas de las matemáticas
 - métodos **empíricos** en los que las muestras se construyen extrayendo varias categorías de individuos de la población estudiada, según objetivos predefinidos

2.1 Métodos aleatorios (1/2)

- Los métodos de muestreo aleatorio son teóricamente los más eficientes.
- De hecho, se basan en la teoría matemática de la probabilidad. En ella se establece que, con un **método de selección** que garantiza a cada individuo de la población objetivo la **misma probabilidad de ser entrevistado**, la muestra producida será necesariamente representativa.
- No obstante, seguirá existiendo un error de estimación, que puede calcularse y que depende principalmente del tamaño de la muestra (véase la sección 1.2).
- Pregunta clave: ¿Cómo se puede implementar un método de selección completamente aleatorio?

2. 1 Métodos aleatorios (2/2)

- Es esencial disponer de un **marco de muestreo**: la lista completa de todos los miembros de la población objetivo y establecer un sorteo aleatorio, por ejemplo se puede hacer en una hoja de cálculo o a través de Sphinx. También es posible definir un "patrón" y elegir por ejemplo *un individuo cada quince o veinte líneas*.
- Algunos investigadores considerarán que la selección aleatoria de la calle o de una guía telefónica puede sustituir al sorteo aleatorio, pero no se cumple la condición básica, ya que no todos los individuos de la población objetivo tienen exactamente la misma probabilidad de ser seleccionados.
- Sin un marco de muestreo, será prácticamente imposible extraer una muestra mediante un método aleatorio.

2. 2 Métodos de cuotas (1/2)

- La muestra se constituye mediante el establecimiento de "**cuotas**" proporcionales al peso relativo de cada categoría de individuos en la población objetivo: hombres, mujeres, jóvenes, ancianos, etc.
- Por lo tanto, la composición de la población objetivo se debe conocer de antemano y, si es posible, los detalles del desglose de los individuos dentro de las diferentes categorías. Esto hace posible llevar a cabo "cuotas cruzadas".

Por ejemplo, hay 4.848 hombres de 18 a 29 años en la población objetivo o el 11,4% de los habitantes de Córdoba son de 18 años o más. Por lo tanto, será necesario reclutar a 114 jóvenes de 18 y 29 años en Córdoba para esta encuesta.

Métodos de cuotas- Ejemplo

Porcentaje	Hombres	Mujeres	%
18-29 años	11,4%	11,2%	22,6 %
30-39 años	8,5%	7,9%	16,4 %
40-49 años	6,8%	7,6%	14,4 %
50-59 años	6,4%	7,5%	13,9 %
60-69 años	5,1%	6,9%	12 %
70 años o más	7,6%	13,1%	20,7 %
%	45,8 %	54,2 %	100 %



Muestra	Hombres	Mujeres	n
18-29 años	114	112	226
30-39 años	85	79	164
40-49 años	68	76	144
50-59 años	64	75	139
60-69 años	51	69	120
70 años o más	76	131	207
	458	542	1 000

2. 2 Métodos de cuotas (2/2)

- ¿Qué criterios deben elegirse y cuántos para una muestra de cuota? En realidad, establecer cuotas no es tan fácil. En efecto, una población objetivo puede describirse sobre la base de miles de características.
- En la mayoría de los casos, se utilizan 3 o 4 criterios para completar las cuotas. Esto es realista y suficiente. Se dará prioridad a aquellos:
 - que se conocen de estudios anteriores o se supone que tienen la mayor influencia en el tema que se estudia
 - de los que se conoce con mayor precisión la distribución sobre la población objetivo
 - que pueden ser fácilmente identificadas al entrevistar a los encuestados.

En un estudio para las elecciones municipales, por ejemplo, además del grupo de edad y el género, se habría razonado también teniendo en cuenta la distribución de la población según los principales distritos de la ciudad que corresponden a realidades sociodemográficas relativamente diferentes (distritos de clase trabajadora, zonas residenciales, etc.) o de la reciente instalación en la ciudad (más o menos de 5 años).

3. Ajuste de la muestra

- A pesar de todos los esfuerzos, a veces es posible obtener una muestra "sesgada" cuya composición no es satisfactoria porque no corresponde con los criterios de representatividad previamente definidos. En este caso, la muestra debe ser "**ajustada**".
- Hay varios métodos de ajuste, presentaremos dos de los más comunes: la **supresión** y la **ponderación**.

3. 1 Ajuste por supresión

- Para recuperar las proporciones esperadas, las de la población objetivo, los encuestados de las categorías sobrerrepresentadas pueden ser eliminados al azar.
- El tamaño general de la muestra se reducirá en consecuencia, lo que resulta frustrante dados los esfuerzos realizados para motivar a las personas contactadas a responder y los costos que ello conlleva.
- Por otro lado, la precisión se pierde a medida que aumenta el error asociado.

	Muestra objetivo	Muestra obtenida	Muestra corregida
Diputados mayoritarios	145 (63 %)	120 (52 %)	120 (63 %)
Opositores	85 (37 %)	110 (48 %)	70 (37 %)
Total	230	230	190
Error	5 puntos		5,82 puntos

3. 2 Ajuste por ponderación

- Si bien se conservan todas las respuestas registradas, a cada encuestado se le asigna un "peso" específico según la categoría a la que pertenezca.
- Este peso es **mayor que 1 si la categoría está infrarrepresentada y menor que 1 si está sobrerrepresentada**. A efectos de los resultados, la opinión de un individuo ya no pesará 1, sino este nuevo peso calculado.
- Es difícil implementar este método de ajuste sin la ayuda de un software. Para cada categoría, se debe calcular el peso que se utilizará para el ajuste de la ponderación.

En el caso de las elecciones, en cuanto a los "opositores", el 37% fueron buscados mientras que el 47,8% fueron obtenidos. El peso se calcula simplemente dividiendo la frecuencia deseada por la frecuencia obtenida, es decir, $0,37/0,47826 = 0,7736$.

Ajuste por ponderación- Ejemplo

Categoría	Efectivo obtenido	% obtenido	% deseado	Peso utilizado para el ajuste
Opositores	110	47,826 %	37 %	0,7736
Mayoría	120	52,174 %	63 %	1,2075

Cálculo del peso para la ponderación

Opositores	Mayoría	Total	% ajustado
67 x 0,7736 = 51,83	94 x 1,2075 = 113,51	161 => 165,34	71,9 %
43 x 0,7736 = 33,26	26 x 1,2075 = 31,40	69 => 64,66	28,1 %
		230	100 %

Cálculo de la ponderación y de los resultados ponderados

¿Piensa presentarse de nuevo a las próximas elecciones generales?	
Número observado	% observado
161	70 %
69	30 %
230	100 %

Resultados iniciales

Conceptos clave

- **Muestra:** un subgrupo de la población a estudiar, seleccionado para representarla. Puede constituirse al azar o según cuotas predefinidas.
- **Estimación estadística:** Principio probabilístico que permite generalizar los resultados observados en una muestra a una población de referencia mayor, con un error de estimación calculable.
- **Representatividad:** La capacidad de una muestra - en virtud de su tamaño y composición - para proporcionar resultados que pueden ser generalizados a toda la población.
- **Muestreo aleatorio:** la extracción aleatoria (al azar) de una muestra para garantizar que reproducirá las características de la población total. Para que una extracción sea verdaderamente aleatoria, es necesario asegurarse de que todos los miembros de la población objetivo (o matriz) tengan exactamente la misma probabilidad de ser seleccionados.
- **Método de cuotas:** método de constitución empírica de una muestra mediante la extracción de varias categorías de individuos de la población estudiada, según porcentajes conocidos: por ejemplo, 45% de hombres, 20% de menores de 30 años, etc. El objetivo es reproducir las características de la población total de esta manera.
- **Intervalo de confianza:** teniendo en cuenta el error de estimación (vinculado al muestreo), el resultado de un estudio puede considerarse dentro de un intervalo, entre el valor probable más bajo y el más alto.
- **Ajuste:** Corrección de una muestra recogida para que recupere las características esperadas. Esto se hace generalmente por supresión o ponderación.

Ejercicios (1/2)

Ejercicio 4a: audiencia de radios en Andalucía

Se realizó una encuesta para medir la audiencia de radio en Andalucía, por teléfono, en una muestra representativa de 750 personas. Los resultados indican que la proporción de encuestados que han escuchado Onda Cero al menos una vez en el día o el día de antes es del 11%. Las cifras son 9,6% para COPE y 7,9% para Cadena Ser.

1. ¿Podemos concluir con certeza que Onda Cero está por delante de COPE y Cadena Ser en el ranking de audiencias de Andalucía para este período?
2. Sobre la base de los mismos resultados, ¿qué tamaño de muestra permitiría concluir que el liderazgo de Onda Cero es seguro?

Ejercicios (2/2)

Ejercicio 4b: la MAP

Una importante compañía de seguros francesa, MAP (Mutuelle d'Assurance Professionnelle), está a punto de llevar a cabo una gran encuesta nacional de satisfacción del cliente. Obviamente, desea consultar una muestra representativa de su base asegurada. Está disponible en su totalidad en formato electrónico. En este archivo, es posible distinguir la región y el tipo de seguro (automóvil, vivienda, previsión, etc.) para cada observación. En el cuadro que figura a continuación se presenta la distribución de todos los contratos por tipo (A, B, etc.) y por regiones principales.

El gerente de marketing le dice el tamaño de la muestra: 2.000 asegurados.

1. ¿Qué método de selección recomiendas para esta muestra?
2. ¿Cuál sería entonces el error introducido en los resultados de la encuesta?
3. Ahora supongamos que sólo una parte de la base asegurada está disponible en formato electrónico. ¿Qué método de muestreo debería utilizarse entonces? Propón la distribución de los individuos a entrevistar, basándote en la información disponible sobre la clientela de MAP.

Regiones/tipos de contrato	A	B	C	D	E	F	G	Contratos totales
Centro	23 884	31 963	52 952	25 398	60 097	31 463	17 020	242 777
Noreste	20 824	39 711	47 502	55 995	124 316	55 016	19 738	363 102
Bretaña	31 891	44 243	88 288	25 821	75 858	44 280	20 951	331 332
Alpes	56 415	94 621	154 935	39 697	121 621	85 036	41 844	594 169
Picardía	27 397	47 614	68 562	44 216	92 088	48 611	17 706	346 194
Norte	34 644	61 078	92 292	29 923	64 878	42 292	15 634	340 741
Auvernia	68 805	49 198	89 242	47 036	66 116	43 705	15 537	379 639
Provenza	45 810	53 366	77 431	39 241	57 114	42 749	32 514	348 225
Suroeste	28 353	42 258	70 323	32 240	82 883	42 010	30 249	328 316
Oeste	38 621	42 509	96 023	37 792	94 913	47 068	26 211	383 137
Isla de Francia	66 124	158 651	124 034	89 220	156 280	114 589	57 778	766 676
Total Francia	442 768	665 212	961 584	466 579	996 164	596 819	295 182	4 424 308

Más recursos

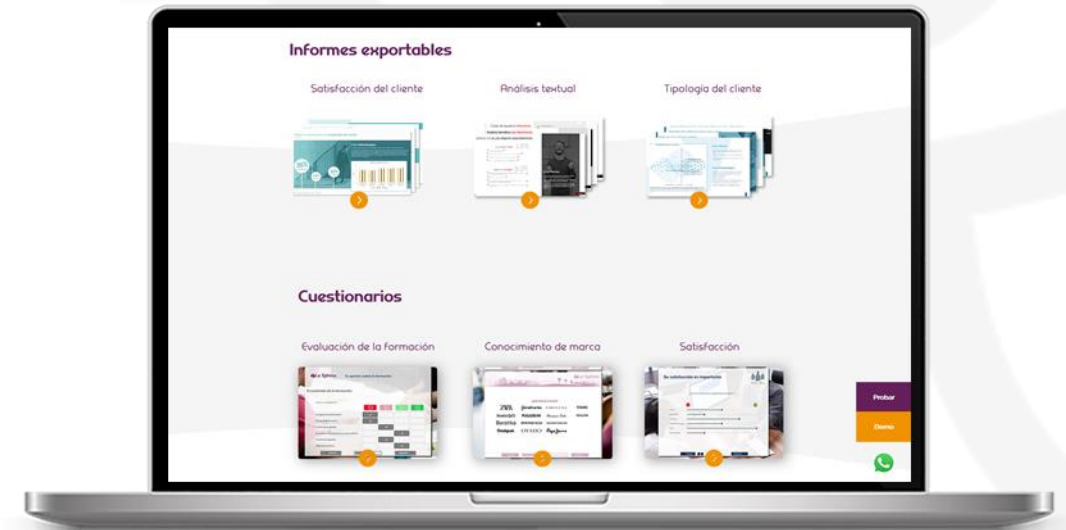
Tutoriales

www.lesphinx.es/videos



Ejemplos

www.lesphinx.es/ejemplos



Descubre iQ3

Descarga una licencia

<https://www.lesphinx.es/prueba-general>

