

Propuesta de diseño estadístico: Encuesta longitudinal de personas en pobreza en España

Presentación

La presente propuesta expone un diseño estadístico inicial para la implementación de una encuesta longitudinal dirigida a personas en situación de pobreza en España. Para su elaboración se recurrió a tres fuentes principales: i) los textos de referencia de Groves et al. (2011), *Survey Methodology*, y de Peter Lynn (2009), *Methodology of Longitudinal Surveys*, que ofrecen los fundamentos teóricos y técnicos en la construcción de encuestas y estudios longitudinales; ii) el diseño de la [Encuesta Longitudinal Social de Chile \(ELSOC\)](#) desarrollada por el COES, utilizada como modelo comparativo y de orientación metodológica; y iii) los datos de la [Encuesta de Condiciones de Vida del INE](#) en España, que proporcionan información clave sobre la magnitud y características de la población en situación de pobreza según ingresos. A partir de estas referencias, el documento presenta los elementos esenciales para estructurar un instrumento longitudinal enfocado en el estudio de esta población específica.

Supuestos

Para alinear la propuesta con escenarios operativos realistas, ajustamos los parámetros de precisión y diseño:

- **Margen de error por CCAA:** ± 6 p.p. (antes: ± 5 p.p.), **95% c.i.**, **varianza máxima** ($p = 0,5$).
- **Efecto de diseño (Deff):** **1,6–1,7** (antes: 1,5–2,0), priorizando **pocos casos por sección** y mayor **dispersión** para contener el Deff.
- **Atrición total** en tres olas: **20% \rightarrow retención acumulada** ($= 0,80$).
- Se muestran **dos estrategias** de levantamiento:
 - **A) Sin refresco:** inflar **Ola 1** para que, con 80% de retención, la **Ola 3** alcance el tamaño objetivo por CCAA.
 - **B) Con refresco:** mantener el **mismo tamaño por ola**; reponer en Olas 2–3 asumiendo **retención por ola** ($\approx \sqrt{0,80} \approx 0,894$).

Nota de coherencia de totales. La suma por CCAA (**9.563.461**) difiere levemente del total nacional entregado (**9.578.080**; diferencia **0,15%**). Las tablas usan la **suma CCAA** para mantener consistencia fila-total; el impacto en tamaños es **no material**.

Población objetivo

Personas de 16+ años residentes en hogares particulares en España que se encuentran en situación de pobreza de acuerdo con el indicador acordado (p. ej., AROPE estimado desde la ECV/INE). La población de referencia se define en la Ola 1 y las mismas personas son seguidas en dos olas adicionales (panel de tres olas anuales). En longitudinal, explicitar si los análisis se refieren a la intersección (respondieron en 1–3) o a subpoblaciones por ola; esto condiciona la construcción del peso longitudinal y la interpretación de cambios (Lynn enumera los pasos clave del diseño longitudinal: definir población en el tiempo, unidades a seguir y reglas frente a cambios).

Unidad de muestreo

Muestreo por etapas: (1) Secciones censales (PPT), (2) Viviendas dentro de sección (sistemático), (3) Hogar (si hay más de uno), (4) Persona elegible (en pobreza) por cribado.

Unidad de análisis y de información

Unidad de análisis: individuo. En cada ola se levanta un módulo de hogar (composición e ingresos) para reconstituir la situación de pobreza del entrevistado. La unidad de información es el propio entrevistado/a (con información adicional del hogar para cálculos de pobreza).

Movilidad (movers): se siguen las personas muestrales allí donde residan (definir reglas operativas); Lynn (2009) destaca la necesidad de reglas frente a cambios de unidad (p. ej., mudanzas, composición del hogar).

Marco muestral

Preparación marco muestral

- Opción A: ABS (Address-Based Sampling) con cribado en la Ola 1 para identificar hogares con personas en pobreza.
- Opción B (dual frame): combinar ABS + registros administrativos (p. ej., IMV y rentas mínimas autonómicas) para mejorar cobertura de subpoblaciones con baja incidencia y reducir el esfuerzo de cribado en algunas CCAA. La literatura longitudinal resalta la utilidad de vincular datos administrativos para complementar cobertura, mantenimiento y evaluación muestral.
- Incidencia territorial: el perfil regional de la pobreza muestra mayores tasas en el sur (Andalucía, Extremadura, Castilla-La Mancha) frente al norte, con EAPN-ES documentando diferencias a partir de ECV/INE; esto es útil para planificar cribado y carga por CCAA.

Selección de unidades

1. Secciones censales: selección PPT en cada estrato.
2. Viviendas/direcciones: muestreo sistemático con arranque aleatorio.
3. Cribado: cuestionario filtro para identificar elegibles (personas en pobreza).
4. Persona: selección aleatoria (por ejemplo método Kish).

En las olas siguientes, se reentrevistará a las personas muestrales (panel fijo). ELSOC mantiene el diseño de seguimiento entre olas sobre la base muestral original.

Diseño muestral

- Tipo de muestreo: Probabilístico estratificado y por conglomerados.
 - Estratificación explícita por CCAA; dentro de cada CCAA, tamaño de hábitat (urbano/intermedio/rural) para control de varianza.
 - Conglomeración en secciones (trade-off costo/Deff).

Lynn (2009) detalla los aspectos fundamentales: tamaño, clustering, tratamiento de movers, estratificación, varianzas y efectos de diseño.

Dominio de estudio

- Total nacional de personas en pobreza.
- Sexo (hombres/mujeres) a nivel nacional.
- Cada CCAA (17 dominios).

(Si posteriormente se exigiera precisión por sexo dentro de cada CCAA, esto duplicaría fuertemente los tamaños.)

Cálculo y distribución del tamaño muestral

Parámetros comunes:

$$z = 1,96, \quad p = 0,5, \quad e = 0,06$$
$$n_{\text{SRS}} = \frac{z^2 p(1-p)}{e^2} \approx 266,78 \text{ por dominio (antes de Deff/FPC)}$$

Se aplica **Deff** ({1{,}6, 1{,}7}) y **FPC** con el **N de población pobre** por CCAA.

Esquemas de operación:

- **A) Sin refresco:** (Ola 1 = $\lceil \frac{n_{\text{por ola}}}{0,80} \rceil$).
- **B) Con refresco:** (Ola 1 = n por ola); en Olas 2-3 reponer **top-up** (= $n_{\text{ola}} - \lceil n_{\text{ola}} \cdot \sqrt{0,80} \rceil$).

Tablas por CCAA — $Deff = 1,6 (\pm 6\%)$

CCAA	Pobres (N)	n por ola	A) Ola1 sin refresco	B) Ola1 con refresco	B) Refresco O2	B) Refresco O3
01 Andalucía	2.520.504	427	534	427	45	45
02 Aragón	204.090	426	533	426	44	44
03 Asturias, Principado de	157.497	426	533	426	44	44
04 Balears, Illes	139.190	426	533	426	44	44
05 Canarias	550.733	427	534	427	45	45
06 Cantabria	102.217	426	533	426	44	44
07 Castilla y León	442.461	427	534	427	45	45
08 Castilla - La Mancha	576.615	427	534	427	45	45
09 Cataluña	1.033.578	427	534	427	45	45
10 Comunitat Valenciana	1.319.183	427	534	427	45	45
11 Extremadura	290.037	427	534	427	45	45
12 Galicia	384.228	427	534	427	45	45
13 Madrid, Comunidad de	1.002.325	427	534	427	45	45
14 Murcia, Región de	407.808	427	534	427	45	45
15 Navarra, Comunidad Foral de	96.323	425	532	425	44	44

CCAA	Pobres (N)	n por ola	A) Ola1 sin refresco	B) Ola1 con refresco	B) Refresco O2	B) Refresco O3
16 País Vasco	209.402	426	533	426	44	44
17 Rioja, La	62.892	424	530	424	44	44
18 Ceuta	28.780	421	527	421	44	44
19 Melilla	35.598	422	528	422	44	44
Total	9.563.461	8.092	10.122	8.092	846	846

Tablas por CCAA — $Deff = 1,7 (\pm 6\%)$

CCAA	Pobres (N)	n por ola	A) Ola1 sin refresco	B) Ola1 con refresco	B) Refresco O2	B) Refresco O3
01 Andalucía	2.520.504	454	568	454	47	47
02 Aragón	204.090	453	567	453	47	47
03 Asturias, Principado de	157.497	453	567	453	47	47
04 Balears, Illes	139.190	453	567	453	47	47
05 Canarias	550.733	454	568	454	47	47
06 Cantabria	102.217	452	565	452	47	47
07 Castilla y León	442.461	454	568	454	47	47
08 Castilla - La Mancha	576.615	454	568	454	47	47
09 Cataluña	1.033.578	454	568	454	47	47
10 Comunitat Valenciana	1.319.183	454	568	454	47	47
11 Extremadura	290.037	453	567	453	47	47

CCAA	Pobres (N)	n por ola	A) Ola1 sin refresco	B) Ola1 con refresco	B) Refresco O2	B) Refresco O3
12 Galicia	384.228	453	567	453	47	47
13 Madrid, Comunidad de	1.002.325	454	568	454	47	47
14 Murcia, Región de	407.808	454	568	454	47	47
15 Navarra, Comunidad Foral de	96.323	452	565	452	47	47
16 País Vasco	209.402	453	567	453	47	47
17 Rioja, La	62.892	451	564	451	47	47
18 Ceuta	28.780	447	559	447	47	47
19 Melilla	35.598	448	560	448	47	47
Total	9.563.461	8.600	10.759	8.600	893	893

Control por sexo — dominios nacionales (informes transversales por ola)

Dominio	Pobres (N)	n por ola (Deff=1,6)	Ola1 sin refresco (Deff=1,6)	n por ola (Deff=1,7)	Ola1 sin refresco (Deff=1,7)
Nacional – Mujeres	5.057.736	427	534	454	568
Nacional – Hombres	4.503.279	427	534	454	568

Resultado del tamaño muestral

- Deff=1,6:
 - A) Sin refresco → Ola 1 10.122 (suma CCAA) para terminar con 8.092 en Ola 3.
 - B) Con refresco → Ola 1 = 8.092; refresco por ola 846 (O2 y O3).
- Deff=1,7:
 - A) Sin refresco → Ola 1 10.759 para terminar con 8.600 en Ola 3.

- B) Con refresco \rightarrow Ola 1 = 8.600; refresco por ola 893 (O2 y O3).

Atrición (recordatorio operativo)

- Seguimiento de movers, modos mixtos en olas de seguimiento, incentivos moderados, contactos alternativos y actualización entre olas.
- En el esquema con refresco, se repone hasta el **n por ola** por CCAA en cada ola, asegurando representatividad transversal estable.

Factores de expansión (ponderadores)

- Transversal por ola:

$$w_{i,t} = w_i^{(0)} \times A_{i,t}^{(NR)} \times g_{i,t},$$

con $(w_i^{\{(0)\}=1/i})$ (todas las etapas, incluyendo **cribado** si hay dos fases), $(A_{i,t}^{\{(NR)\}})$ ajuste por no respuesta (clases/modelo), y $(g_{i,t})$ **calibración (raking/IPF)** a **totales de población pobre** por **CCAA** y **sexo** (y, si se dispone, **sexo** \times **CCAA** desde ECV/INE).

- Longitudinal (1–3):

$$w_{i,1:3} = w_i^{(0)} \times A_{i,1}^{(NR)} \times A_{i,2}^{(NR)} \times A_{i,3}^{(NR)} \times h_i,$$

donde (h_i) calibra a la población de referencia (p. ej., totales de **Ola 1** por CCAA, sexo, edad).

Recomendación para cerrar plan y presupuesto

- Elegir un **Deff operativo** (p. ej., **1,65**) según cuántos casos por sección se desplieguen (1–2 casos por sección Deff más bajo).
- Definir si **habrá refresco** (recomendado para mantener precisión transversal con menor carga inicial).
- Con esa decisión, cerrar una **tabla única final** (Ola 1 y O2–O3) y el **dimensionamiento de cribado** por CCAA en función de tasas de respuesta (RR cribado y RR cuestionario).

Referencias

Groves, R. M., Fowler Jr, F. J., Couper, M. P., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). Survey methodology. John Wiley & Sons.

Lynn, P. (2009). Methodology of Longitudinal Surveys.