

Diseño estadístico: Encuesta sobre las Implicaciones Psicosociales de la Pobreza, Riqueza y Desigualdad Socioeconómica (PORIDE)

Mario Sainz Martínez¹

¹Departamento de Psicología Social, Facultad de Psicología, Universidad de Granada, Spain

Presentación

Esta propuesta expone el diseño estadístico 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 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 criterios operativos realistas y control explícito del Error Total de Encuesta (TSE), adoptamos los siguientes supuestos y metas de calidad (Groves et al., 2011; Lynn, 2009):

- **Población objetivo:** Personas de 16+ años en **pobreza por ingresos** en España (ámbito nacional), definidas al momento del muestreo. La población de referencia se fija en Ola 1; las mismas personas se siguen en Olas 2 y 3 (panel fijo).
- **Estrategia muestral: Muestreo aleatorio simple (SRS)** de individuos, **sin estratificación ni conglomeración**
 - **Implicación:** $Deff = 1$ como supuesto operativo (no hay inflación de varianza por diseño).
 - **Nota:** si los ajustes por no respuesta/calibración generan variabilidad de pesos, se monitorizará el **deff por pesos** para contener pérdidas de eficiencia (Groves et al., 2011).
- **Dominios de estimación:** Publicación en **Total nacional y Sexo** (Mujeres/Hombres). No se exige precisión por CCAA en este diseño.
- **Precisión objetivo:** Margen de error $\leq \pm 5$ p.p., 95% CI, peor caso $p = 0.5$:

$$n_{ME} = \frac{z_{0.975}^2 p(1-p)}{e^2} = \frac{1.96^2 \cdot 0.25}{0.05^2} \approx 385$$

Para garantizar ± 5 p.p. **por sexo en ola 3**, el tamaño total n_3 debe satisfacer

$$n_3 \cdot \pi_f \geq 385 \wedge n_3 \cdot \pi_m \geq 385,$$

con $\pi_f \approx 0.528$ y $\pi_m \approx 0.470$ (proporciones observadas en la población pobre).

- **Poder estadístico (ola 3):** ≥ 0.80 para detectar $\delta \approx 8$ p.p. de diferencia por sexo, con proporciones cercanas a 0.5 y prueba bilateral $\alpha = 0.05$:

$$n_{\text{por sexo}} \approx \frac{2\bar{p}(1-\bar{p}) \left(z_{1-\alpha/2} + z_{1-\beta} \right)^2}{\delta^2} = \frac{2 \cdot 0.25 \cdot (1.96 + 0.84)^2}{0.08^2} \approx 613$$

Meta conservadora: fijamos ola 3 total en $n_3 \approx 1,226$ (≈ 613 por sexo), lo que cumple poder ≥ 0.80 y supera el mínimo de precisión (385/sexo).

- **Atrición y horizonte longitudinal:** Tres olas, sin refresco, con atrición total del 30% entre Ola 1 y Ola 3.
 - **Retención acumulada (O1→O3):** 0.70.
 - **Retención por intervalo (O1→O2 y O2→O3):** $r = \sqrt{0.70} \approx 0.8367$.
- **Dimensionamiento por ola (consistente con poder y precisión):**

$$n_1 = \left\lceil \frac{n_3}{0.70} \right\rceil, \quad n_2 \approx \lceil n_1 \cdot r \rceil, \quad n_3 \approx \lceil n_2 \cdot r \rceil$$

Con $n_3 \approx 1,226 \Rightarrow n_1 \approx 1,752$, $n_2 \approx 1,466$, $n_3 \approx 1,226$.

- **Ponderación y control de sesgo:** Peso base SRS; ajuste por no respuesta (clases o propensión) y calibración a **Total** y **Sexo** por ola. El sesgo de no respuesta depende de $\text{cov}(y, p)$ (no de la tasa bruta); se monitorizará junto con el deff por pesos.

Resumen numérico derivado:

- **Ola 1 = 1,752;**
- **Ola 2 = 1,466;**
- **Ola 3 = 1,226** (≈ 613 por sexo).
- **Total $\approx 4,444$ entrevistas**

En ola 3 se garantiza **ME** $\leq \pm 5$ p.p. por sexo y **poder** ≥ 0.80 para $\delta \approx 8$ p.p.

Universo

El universo del estudio serán los hombres y mujeres que ese encuentren en situación de pobreza en España en 2026.

Población objetivo

La población objetivo serán los hombres y mujeres mayores de 18 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) en 2026. 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).

Unidad de muestreo

La unidad de muestreo son los hombres y mujeres (individuos) mayores de 18 años que se encuentran en situación de pobreza por ingresos en seleccionados directamente en la población.

Unidad de análisis y de información

La unidad de análisis serán los individuos en situación de pobreza cumplen los requisitos para formar parte de la población objetivo. En cada ola se levanta un módulo de hogar (composición e ingresos) para reconstituir la situación de pobreza del entrevistado.

La unidad de información serán los hombres y mujeres en situación de pobreza mayores de 18 años que cumplen los requisitos para formar parte de la población objetivo y que resulten seleccionadas en el diseño muestral.

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

El marco muestral corresponde a un listado o procedimientos que permiten identificar todos los elementos con probabilidad de ser seleccionados de la población objetivo ([Groves et al., 2011](#)). Se asume muestreo probabilístico aleatorio simple (SRS) individual y cobertura nacional (sin dominios territoriales).

Conformación marco muestral

Completar. Para esto se usa info de estadísticas oficiales de cantidad y proporción de hombres y mujeres en pobreza a nivel nacional. Requisito: un marco nominal de personas en pobreza (p. ej., registros administrativos integrados y depurados o listados derivados de un operativo de cribado).

Selección de unidades

1. Construcción/obtención de listado de elegibles (personas en pobreza por ingresos).
2. Persona: selección aleatoria simple de n individuos a nivel nacional para Ola 1.
3. Seguimiento longitudinal: reentrevistar a los mismos individuos en Ola 2 y Ola 3. Corresponde a un panel fijo de tres olas (t_0, t_1, t_2) sin refresco, atracción esperada 30% por ola ($r \approx 0.70$ por transición).

Diseño muestral longitudinal

- Tipo de muestreo: Probabilístico aleatorio simple (SRS) de personas en el marco muestral. No hay estratos ni conglomerados: esto simplifica el peso base y elimina deff por conglomeración; a costa de no garantizar auto-

ponderación por subgrupos geográficos¹.

- Panel de tres olas (t_0, t_1, t_2) reentrevistando a los mismos individuos, sin refresco, atrición esperada 30% por ola ($r \approx 0.70$ por transición).
- Sin refresco: cualquier pérdida entre olas se refleja en menor tamaño en la ola siguiente.
- Plan de retención: protocolos de contacto, tracking de “movers”, modos de seguimiento flexibles y recordatorios para minimizar atrición.

Dominio de estudio

- Total nacional.
- Sexo: Mujeres y Hombres (dominios principales de publicación y precisión).

Cálculo y distribución del tamaño muestral

Parámetros poblacionales conocidos

- Total personas en pobreza: $N = 9,578,080$.
- Mujeres: 5,057,736 $\Rightarrow \pi_f \approx 0.528$.
- Hombres: 4,503,279 $\Rightarrow \pi_m \approx 0.470$.

Precisión objetivo en ola 3

Bajo SRS, $p = 0.5$ y 95% CI:

$$n_{\text{ME}} = \frac{z_{0.975}^2 p(1-p)}{e^2} = \frac{1.96^2 \cdot 0.25}{0.05^2} \approx 384.16 \Rightarrow 385.$$

Para garantizar **por sexo** en ola 3:

$$n_3 \times \pi_f \geq 385, \quad n_3 \times \pi_m \geq 385 \Rightarrow n_3 \geq \max\left(\frac{385}{\pi_f}, \frac{385}{\pi_m}\right) = \frac{385}{0.470} \approx 822.$$

Poder estadístico

Poder ≥ 0.80 para $\delta \approx 8$ p.p. (sexo, ola 3). Escenario conservador (proporciones cercanas a 0.5, tamaños similares por sexo), prueba bilateral ($alpha = 0.05$):

$$n_{\text{por sexo}} \approx \frac{2 \bar{p}(1 - \bar{p})(z_{1-\alpha/2} + z_{1-\beta})^2}{\delta^2} = \frac{2 \cdot 0.25 \cdot (1.96 + 0.84)^2}{0.08^2} \approx 613.$$

Por lo tanto, fijamos **ola 3 total** en:

¹Cuando se usan estratos o asignaciones desproporcionadas, pueden lograrse ganancias de precisión si la estratificación reduce la varianza interna; aquí se renuncia a esa ganancia para simplificar y priorizar poder por sexo, manteniendo SRS ($\text{deff} \approx 1$) (Groves et al., 2011).

$$n_3 \approx 1,226 \quad (\approx 613 \text{ por sexo}),$$

que **cumple poder ≥ 0.80** para $\delta \approx 8$ p.p. y **superá** el mínimo de precisión (385/sexo).

Dimensionamiento por ola (sin refresco; atrición total 30%)

Con **retención acumulada** ($= 0.70$) y **retención por intervalo** ($r = \sqrt{0.70} \approx 0.8367$):

$$n_1 = \left\lceil \frac{n_3}{0.70} \right\rceil = \left\lceil \frac{1,226}{0.70} \right\rceil = 1,752,$$

$$n_2 \approx \lceil n_1 \cdot r \rceil = \lceil 1,752 \cdot 0.8367 \rceil = 1,466,$$

$$n_3 \approx \lceil n_2 \cdot r \rceil = \lceil 1,466 \cdot 0.8367 \rceil = 1,226.$$

Tabla de tamaños (SRS, sin refresco, atrición total 30%) [2·3]

Ola	Total esperado	Mujeres ($\approx 52.8\%$)	Hombres ($\approx 47.0\%$)
Ola 1	1,752	≈ 925	≈ 823
Ola 2	1,466	≈ 774	≈ 689
Ola 3	1,226	≈ 647	≈ 576

- En ola 3, ambas celdas superan 613/sexo si la composición observada se aproxima a π_f, π_m .
- Si se desea exactamente 613 por sexo (sin depender de proporciones), puede estratificarse por sexo con asignación igual, manteniendo SRS dentro de cada estrato.

Verificación rápida

Precisión en ola 3 (SRS):

- Mujeres ($n \approx 647 \Rightarrow (e \approx 1.96\sqrt{0.25/647} \approx 3.85)$ p.p.)
- Hombres ($n \approx 576 \Rightarrow (e \approx 1.96\sqrt{0.25/576} \approx 4.09)$ p.p.)
- (Ambos **mejores** que ± 5 p.p.)

Poder ($\delta \approx 8$ p.p.): ($n_{\text{por sexo}} \geq 613$) garantiza $(1 - \beta \geq 0.80)$ en el peor caso; con $(n_f \approx 647, n_m \approx 576)$, el poder es ≥ 0.80 .

Atrición (prevención y medición)

- **Prevención:** protocolos de recontacto, *tracking* de movers, incentivos, modos mixtos en seguimiento, calendario flexible.
- **Medición de cambio:** reducir *seam bias* y cambios de medición con **instrumentos consistentes y dependent interviewing** cuando proceda ([Lynn, 2009](#)).
- **Sesgo de no respuesta:** depende de $\text{cov}(y, p)$, no de la tasa bruta; monitorizar perfil de respondedores vs. no respondedores ([Groves et al., 2011](#)).

Factores de expansión (ponderadores)

Peso base

Con **SRS**:

$$w_i^{(0)} = \frac{N_M}{n_1},$$

donde (N_M) es el tamaño del **marco** de personas en pobreza.

Ajuste por no respuesta (por ola)

Clases de ajuste o **modelo de propensión** con variables observables del marco (sexo, edad, historial de contacto, etc.):

$$w_i^{(\text{NR},t)} = \frac{w_i^{(t-1)}}{\hat{p}_{g(i),t}},$$

con $g(i)$ clases homogéneas o \hat{p} de un modelo (logit). Estándar recomendado en TSE ([Groves et al., 2011](#)).

Calibración (post-estratificación / raking)

Calibrar por ola a **totales externos fiables** (p. ej., **Total** y **Sexo** de la población en pobreza del año de referencia):

$$w_i^{(t)} = w_i^{(0)} \times w_i^{(\text{NR},t)} \times g_{i,t}, \quad \sum_{i \in c} w_i^{(t)} x_i = T_c.$$

Peso longitudinal (intersección 1–3)

Aplicar ajustes de **permanencia** análogos a la no respuesta y **calibrar** la submuestra intersección 1–3 a totales de referencia (p. ej., sexo/edad de Ola 1).

Operación y calidad de medición

- Mantener **modo e instrumentos** constantes entre olas; documentar cambios inevitables y su impacto.
- Protocolos de **edición/imputación** reproducibles (principios Fellegi–Holt), alineados con el diseño de captura ([Groves et al., 2011](#)).

Viabilidad y pertinencia del marco de captación

Para la recogida de datos se utilizará un panel online y ya se ha contactado con Netquest, que apoyó el reclutamiento en el estudio previo y ofrece un servicio de calidad que garantiza respuestas fiables por parte de los participantes, contando además con un panel suficientemente amplio para cubrir los tamaños por ola previstos.

Dentro de este panel, trabajaremos con la [clasificación socioeconómica EGM](#) utilizada por Netquest, que es un índice de ingreso del hogar construido mediante un modelo de regresión que integra tres bloques informativos diferenciados: (i) “Grupo” del sustentador principal, definido por la matriz Ocupación × Nivel educativo y ordenado por ingresos medios (continúa la tradición de “clase social”, pero anclada en renta observada); (ii) la actividad del sustentador (situación laboral/ocupacional vigente, distinta del “Grupo”, por ejemplo ocupado, desempleado, inactivo), y (iii) la estructura del hogar mediante la matriz Tamaño del hogar × número de perceptores de ingreso (economías de escala y capacidad de generación de renta). Previo al modelamiento, los ingresos declarados se corrigen/imputan con la estimación del entrevistador y los cortes monetarios de los siete tramos se ajustan a la distribución de sueldos del Instituto Nacional de Estadísticas. Con ello, IE2 (< €745) e IE1 (€745–€1.312) representan, por construcción, los menores niveles de ingreso, que son los de interés para esta encuesta.

Empíricamente, IE1 concentra en torno al 15–16% e IE2 al 7–8% de la población (\approx 23–24% conjunta), con mayor presencia relativa de mujeres en los segmentos de menor ingreso. Este marco justifica focalizar la captación en IE2 y, de ser necesario, completar con IE1, manteniendo selección probabilística y posterior calibración por sexo y total. Operativamente, Netquest dispone de 4.300 casos en IE1 y 1.400 en IE2 (total 5.700), stock que supera los requerimientos del diseño conservador (Ola 1 \approx 1.752, Ola 3 \approx 1.226) y respalda la viabilidad online del estudio con margen de error $\leq \pm 5$ p.p. por sexo y poder $\geq 0,80$ para detectar diferencias de ≈ 8 p.p. entre sexos, sujeto a la financiación adecuada para incentivos y logística de campo.

Referencias

- Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2011). *Survey Methodology* (2nd ed (Online-Ausg.)). Somerset: Wiley.
- Lynn, P. (Ed.). (2009). *Methodology of Longitudinal Surveys*. New York, NY: John Wiley & Sons.