



# Diseño muestral y ponderación ENAJ

## Condición de elegibilidad.

Son elegibles para participar en la ENAJ todas las personas entre 12 y 35 años que residen en viviendas particulares situadas en todo el territorio nacional.

## Diseño muestral

Debido a que no existe un marco muestral actualizado de personas jóvenes entre 12 y 35 años para seleccionar la muestra de la ENAJ, se recurrió a seleccionar una submuestra aleatoria de la Encuesta Continua de Hogares (ECH) en donde se tiene información actualizada de jóvenes que residen en hogares particulares. Por lo tanto, el diseño muestral de la ENAJ es aleatorio y en dos fases de selección, donde cada una de las fases implica a su vez, varias etapas de selección.

### 1. Primera fase

La primera fase corresponde a la muestra de la Encuesta Continua de Hogares (ECH) del año 2017 y primer semestre del año 2018. Los hogares (y todos los integrantes que lo componen) son seleccionados bajo un diseño estratificado, aleatorio y en dos etapas de selección. En la primera etapa se seleccionan unidades primarias de muestreo (UPM) que corresponden a conglomerados de zonas censales, zonas censales (UPM) bajo un diseño con probabilidades proporcional al tamaño (PPS) utilizando como medida de tamaño (MOS) la cantidad de viviendas particulares según el Censo 2011. En la segunda etapa, dentro de cada UPM, se seleccionan cinco viviendas con igual probabilidad de selección.

Como ya se dijo, el diseño muestral de la ECH (primera fase) es aleatorio, estratificado, por conglomerados y en dos etapas de selección. En tanto, el diseño de la segunda fase (muestra ENAJ) es en tres etapas de selección.

### 2. Segunda fase

El diseño de la segunda fase se realizó utilizando la información recabada, en campo, para los hogares (y sus integrantes) incluidos en la muestra de la primera fase (ECH) y teniendo en cuenta el nivel de aperturas requerido para las estimaciones de los distintos indicadores de la ENAJ.

Dentro de cada hogar de la muestra de la ENAJ, una única persona elegible es seleccionada para contestar la encuesta.



El diseño de la segunda fase es aleatorio, estratificado y en cuatro etapas de selección. En la primera etapa, los departamentos (UPM) son estratificados por regiones geográficas.

<b>Región (estrato)</b>	<b>Departamentos</b>
Metropolitana	Montevideo – Canelones – San José
Frontera con Brasil	Artigas – Cerro Largo – Rivera – Treinta y Tres
Costa Este	Canelones – Maldonado – Rocha
Litoral Sur	Colonia – Soriano – San José
Centro	Durazno – Tacuarembó
Centro Sur	Flores – Florida – Lavalleja
Litoral Norte	Paysandú – Río Negro - Salto

Dentro de cada una de las regiones (a excepción de la región Metropolitana) se sortearon departamentos con probabilidad proporcional al tamaño en términos de la cantidad de jóvenes según estimaciones provenientes de la ECH.

En la segunda etapa, para cada uno de los departamentos seleccionados (a excepción de la región Metropolitana y Montevideo), se selecciona al azar bajo un PPS localidades (USM) en donde las localidades son estratificadas por urbanicidad. En una tercera etapa, se seleccionan hogares (UTM) elegibles con igual probabilidad de selección. Finalmente, en la última etapa, se selecciona una persona elegible (UUM) en cada uno de los hogares seleccionados en etapa anterior.

El tamaño de muestra teórico se situó en 10.658 personas con el objetivo de obtener aproximadamente 8.000 personas elegibles que efectivamente respondieran. La muestra teórica se dividió de forma aleatoria en sub muestras o replicas. Las réplicas se fueron utilizando hasta alcanzar los tamaños de muestra esperados por estrato. Debido a que las réplicas son construidas al azar el uso o no de alguna no le quita aleatoriedad a la muestra.

## Tasa de elegibilidad y respuesta

No todos los casos seleccionados originalmente en la muestra son elegibles para participar en la ENAJ. La no elegibilidad se puede deber a diversos motivos (vivienda desocupada al momento de la realización de la encuesta, el hogar que habita la vivienda es distinto al seleccionado y/o la persona elegible no se encuentra más en el hogar). A su vez, existen casos en la muestra donde no es posible determinar si es elegible o no. Dichos casos corresponden en su mayoría a replicas que no fueron utilizadas en el relevamiento de campo. En la siguiente tabla se presenta la distribución de los causales obtenidos en el relevamiento:



**Tabla 1. Distribución de los causales**

<b>Causales</b>	<b>Conteos</b>
Respuesta (R)	6.543
No respuesta (NR)	2.500
No elegible (NE)	189
Elegibilidad Desconocida (ED)	1.426
Total ( $n_{teo}$ )	10.658

La tasa de elegibilidad se situó en 98% aproximadamente y se define como:

$$\text{Tasa de elegibilidad} = \frac{n_{teo} - (ED + NE)}{n_{teo} - ED} \times 100 = \frac{10658 - (1426 + 189)}{10658 - 1426} \times 100 = 98\%$$

La tasa de respuesta se situó en el 72% aproximadamente y se define como:

$$\text{Tasa de respuesta} = \frac{ER}{ER + NR} \times 100 = \frac{6543}{6543 + 2500} \times 100 = 72\%$$

Por otra parte, la tasa de efectividad, se situó en 81.8% y se define como:

$$\text{Tasa de efectividad} = \frac{R}{n_{esp}} \times 100 = \frac{6543}{8000} \times 100 = 81.8\%$$

## Ponderación

Los ponderadores de la muestra son un componente fundamental para producir las estimaciones de los distintos indicadores de la ENAJ. Los datos de la muestra son ponderados de forma de obtener estimaciones tanto a nivel nacional como para otros dominios de estudio, como ser, sexo, tramos de edad, nivel socioeconómico, regiones y etc.

Para la determinación de los ponderadores finales se realizaron los siguientes pasos.

1. Determinación de los ponderadores originales para cada uno de los casos teniendo en cuenta el diseño de muestra.
2. Ajuste por elegibilidad desconocida
3. Ajuste por no respuesta.
4. Calibración de los ponderadores a conteos provenientes de la Encuesta Continua de Hogares.



Los ponderadores finales para cada una de las personas elegibles que respondieron vienen dados como:

$$w = d \times f_1 \times f_2 \times f_3$$

$d$  = **ponderador original (proveniente del diseño muestral)**, el cual se computa como el inverso de la probabilidad de selección teniendo en cuenta el diseño muestral. Es decir, las distintas etapas de selección y la estratificación, tanto en la ECH, como en la ENAJ.

$f_1$  = **ajuste por elegibilidad desconocida**. Consistió en distribuir la suma de los ponderadores originales de las viviendas sobre las que se desconoce su elegibilidad entre aquellas cuya elegibilidad es conocida (elegible o no elegible). Dicho ajuste se realizó a nivel de localidad para el caso del interior del país y a nivel de los estratos socioeconómicos para el caso de Montevideo y zona metropolitana según la ECH.

$f_2$  = **ajuste por no respuesta**. Se basó en la probabilidad o propensión de respuesta para cada unidad por medio de un árbol de clasificación. El método permite clasificar a las unidades de la muestra como R o NR, en base a información disponible para todas las personas de la muestra, las cuales provienen de la primera fase de muestreo (ECH). Las variables utilizadas para la construcción del árbol de clasificación fueron: sexo, edad, nivel educativo, ingreso del hogar, tamaño del hogar, región y estrato. Una vez construido el árbol, los nodos del mismo se utilizaron para construir post-estratos o clases de no respuesta, en donde, a todas las unidades dentro de un mismo nodo se le asigna la misma propensión estimada de responder. Por lo tanto, el ajuste  $f_2$ , queda determinado como el inverso de la propensión ponderada estimada de responder en el nodo.

$f_3$  = **el ajuste provenientes de la calibración**. Se utilizó la post-estratificación incompleta (raking) de forma que las estimaciones provenientes de la encuesta coinciden de forma exacta con conteos provenientes de la ECH. Para este caso se utilizaron cuatro variables: sexo, edades simples, regiones de la ENAJ y estratos de la ECH.