

# GUÍA DE USO DE LOS DATOS DE LA ENCUESTA SOBRE PERCEPCIÓN SOCIAL DE LA CIENCIA Y LA TECNOLOGÍA 2022 [EDICIÓN 1.0]

Marzo de 2023

Esta guía contiene información sobre los datos de la Encuesta de Percepción de la Ciencia y la Tecnología (EPSCT) de 2022 impulsada por la Fundación Española de la Ciencia y la Tecnología (FECYT). Los microdatos se hacen públicos con el fin de aportar transparencia a la investigación, así como proporcionar nuevas fuentes de datos a científicos y organizaciones. La guía se divide en tres secciones, primero se describe el contenido del archivo comprimido; segundo, se presenta un resumen del diseño metodológico de la encuesta; tercero, se aportan algunas recomendaciones sobre el análisis de los datos teniendo en cuenta las características del diseño muestral.

Citar los datos:

Encuesta de percepción social de la ciencia y la tecnología en España (EPSCT). Fundación Española para la Ciencia y la Tecnología (FECYT). Microdatos 2022 – edición 1.0  
<https://doi.org/10.58121/msx6-zd63>

## 1. Contenido de la carpeta

TABLA 1. CONTENIDO DE LA CARPETA CON LOS DATOS DE LA ENCUESTA DE DESINFORMACIÓN

Archivo	Descripción
cuestionario_epsct-2022_fecyt_2023.pdf	Documentación. Cuestionario completo de la encuesta en formato pdf.
guia_epsct-2022_fecyt_2023.pdf	Documentación. Guía para el uso y análisis de los datos.
registro_epsct-2022_fecyt_2023.xlsx	Documentación. Listado de variables y valores del archivo de microdatos.
microdatos-epsct-2022_fecyt_2023.sav	Microdatos de la encuesta en formato SPSS.
microdatos-epsct-2022_fecyt_2023.dta	Microdatos de la encuesta en formato Stata.
microdatos-epsct-2022_fecyt_2023.csv	Microdatos de la encuesta en formato CSV.

## 2. Resumen del diseño de la encuesta

La EPSCT es un estudio bianual impulsado por la FECYT cuya primera edición tuvo lugar en 2002. El objetivo principal de la encuesta es profundizar en el conocimiento de las relaciones entre ciencia, tecnología y sociedad; y analizar la percepción de la ciudadanía sobre los avances científicos y tecnológicos, y sobre la capacidad de éstos para la mejora de la calidad de vida de la población. Los siguientes párrafos presentan un resumen del diseño de la presente edición de la encuesta. Para una descripción más detallada se puede consultar el informe anual.

La encuesta, dirigida a la población general residente en España de 15 o más años, fue administrada mediante entrevista personal en el domicilio asistida por ordenador a una muestra final de 6.054 entrevistados. El trabajo de campo fue llevado a cabo por la empresa Ikerfel entre los días 3 de octubre y 11 de noviembre de 2022, siendo la duración media de la entrevista de 25 minutos.

La selección de la muestra se realizó en varias etapas. En la primera se procedió a seleccionar una muestra de municipios (conglomerados) con unas probabilidades de selección proporcionales al tamaño del conglomerado. Esta primera fase del muestreo fue estratificada por la variable provincia y tamaño de hábitat. En la segunda etapa, dentro de los municipios que formaban parte de la muestra, se seleccionaron una o más secciones censales al azar. Seguidamente, en cada sección censal, los/as entrevistadores/as emplearon un sistema de rutas aleatorias para seleccionar a los hogares participantes donde, por último, se emplearon cuotas de sexo y edad para elegir a la persona entrevistada.

Una vez concluido el trabajo de campo, se generó un coeficiente de ponderación para equilibrar la muestra dentro de cada estrato formado por las variables comunidad autónoma, tamaño de hábitat y grupos de edad. Para generar este coeficiente de ponderación se recurrió a los datos del Padrón ofrecidos por el Instituto Nacional de Estadística (INE) con fecha de 1 de enero de 2021. Con el fin de ampliar la información recogida en el estudio, la muestra de 6.054 entrevistados fue dividida de forma aleatoria en tres grupos. Cada submuestra recibió un cuestionario con elementos comunes – preguntados a los integrantes de las tres muestras – y algunos específicos. Para un mayor detalle se puede consultar el cuestionario. En los microdatos la variable submuestra identifica las diferentes muestras.

En la presente edición del estudio se ofrecen una serie de parámetros que pueden ayudar a comprender la recogida de los datos. Estos parámetros son la fecha y hora de inicio (z1\_1 y z2\_2) y final de las entrevistas (z2\_1 y z2\_2), la duración de la entrevista en segundos (z3), un identificador del entrevistador que la realizó (entrevistador) y el número de intentos de contacto con el hogar previo a la entrevista (z5).

El cuestionario se compone de siete bloques (Tabla 2). Entre la documentación de la encuesta se encuentra una copia del cuestionario.

TABLA 2. CONTENIDO DEL CUESTIONARIO DEL ESTUDIO

Bloque	Contenido
A	Interés en la ciencia
B	Imagen social de la ciencia y la tecnología
C	Confianza en la ciencia
D	Imagen de la profesión científica
E	Alfabetización científica
F	Comunicación
G	Creencias y valores
H	Salud
I	Vacunas
J	Covid-19
K	Cambio climático
L	Actitudes y posicionamientos de la ciudadanía
M	Apoyo institucional a la ciencia
N	Negacionismo
S	Demográficas

### 3. Recomendaciones para el análisis de los datos de la encuesta

El diseño de la muestra incluye la estratificación de la muestra por provincia y tamaño de hábitat, la selección de conglomerados (municipios), el uso de rutas aleatorias para la selección de los hogares y el uso de cuotas de sexo y edad para seleccionar a la persona entrevistada. Además, se ofrece un coeficiente de ponderación que ajusta el perfil de la muestra con respecto a las variables grupos de edad, comunidad autónoma y tamaño de hábitat, de forma que el tamaño de estos grupos en la muestra es idéntico al peso que tienen en la población.

A pesar de que el diseño muestral de la EPSCT combina elementos probabilísticos – selección de municipios y secciones – y no probabilísticos – selección de los entrevistados empleando un sistema de cuotas –, a la hora de analizar los datos con e inferir los resultados a la población, se recomienda interpretar los resultados con precaución ya que es probable que los errores estándar de las estimaciones estén infraestimados. De la misma forma, y para mitigar este defecto parcialmente, se recomienda incluir en el análisis las características del diseño muestral de la parte probabilística, sobre todo los conglomerados, como se describe a continuación.

A la hora de analizar los datos, estos elementos del diseño deben ser tenidos en cuenta ya que afectan tanto a las estimaciones (porcentajes, medias, coeficientes...) como a la precisión con la que se puede inferir de la muestra a la población (errores típicos). Para ello, en el fichero de datos se pueden encontrar la variable peso, que corresponde al coeficiente de ponderación, y la variable estrato, que refleja la estratificación de la muestra, y la variable conglomerado, que representan el conglomerado (municipio) seleccionado en la primera fase del muestreo.

Por un lado, usar la ponderación (variable peso) en el análisis es necesario para mitigar el sesgo que puede introducir en la estimación la falta de respuesta a la encuesta. Ignorar la ponderación en el análisis de los datos implica asumir que la muestra se diseñó usando un muestreo aleatorio simple y que se obtuvo una tasa de respuesta del 100%, dos supuestos que no se cumplen en este estudio. Como resultado de utilizar la ponderación, es muy probable que la media, porcentajes o coeficientes del análisis sean diferentes a los que se obtienen al ejecutar el análisis sin ponderar. Además, el uso de la ponderación resta precisión a la estimación, lo que resulta en un incremento en el valor de los errores típicos, que reflejan la incertidumbre derivada de la falta de respuesta a la encuesta.

Por el otro, la inclusión de la estratificación (estrato) y el identificador de los conglomerados (conglomerado) en el análisis de los datos afecta únicamente a los errores típicos. La estratificación puede tener un efecto positivo, reduciendo los errores típicos, si las variables por las que se estratificó la muestra están relacionadas con la variable que se pretende estimar. Por su parte, el uso de conglomerados en el proceso de selección tiende a incrementar el error típico de las estimaciones. Una duda recurrente es la conveniencia de utilizar la ponderación cuando se analizan subpoblaciones, por ejemplo, las mujeres o aquellos que están más interesados temas científicos. En estos casos se debe utilizar la ponderación de la misma forma que se emplea en los análisis de la muestra completa. Por ejemplo, en el caso del análisis de las mujeres, la ponderación ajusta la composición de la muestra en términos de edad, tamaño de hábitat y comunidad de residencia. Los principales programas de análisis estadísticos incorporan comandos para tener en cuenta el diseño muestral a la hora de analizar los datos. Un amplio resumen de estas herramientas se puede encontrar en el artículo de West *et al.* (2018).

## Referencias

West, B., Sakshaug, J. y Aurelien, G. (2018). Accounting for Complex Sampling in Survey Estimation: A Review of Current Software Tools. *Journal of Official Statistics*, 34(3) 721-752.  
<https://doi.org/10.2478/jos-2018-0034>