

Guía de uso de los datos de la encuesta de desinformación científica en España

Diciembre de 2022

Esta guía contiene información sobre los datos de la encuesta sobre desinformación científica en España llevada a cabo 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 y proporcionar nuevos 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, y tercero se aportan algunas indicaciones de cómo proceder con el análisis de los datos teniendo en cuenta las características del diseño muestral.

Citar los datos:

Cabrera-Álvarez, Pablo; Díaz-Catalán, Celia (2022). *Desinformación científica en España*. FECYT. [Dataset]. Versión de 16 de diciembre de 2022.

1. Contenido de la carpeta

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

Archivo	Descripción
cuestionario_encuesta-desinformacion_fecyt_2022.pdf	Documentación. Cuestionario completo de la encuesta en formato pdf.
guia_encuesta-desinformacion_fecyt_2022.pdf	Documentación. Guía para el uso y análisis de los datos.
registro_encuesta-desinformacion_fecyt_2022.xlsx	Documentación. Listado de variables y valores del archivo de datos.
encuesta-desinformacion_fecyt_2022.sav	Microdatos de la encuesta en formato SPSS.
encuesta-desinformacion_fecyt_2022.dta	Microdatos de la encuesta en formato Stata.
encuesta-desinformacion_fecyt_2022.csv	Microdatos de la encuesta en formato CSV.

2. Resumen del diseño de la encuesta

La encuesta Desinformación y Ciencia se realiza en el contexto del proyecto IBERIFIER¹, el observatorio de medios digitales de España y Portugal, impulsado por la Comisión Europea y vinculado al *European*

¹ <https://iberifier.eu/observatorio/>

Digital Media Observatory (EDMO), cuyo objetivo fundamental es analizar el ecosistema de medios digitales y hacer frente a la problemática de la desinformación.

FECYT, como organización de interfaz entre ciencia y sociedad, ejerce un importante papel en la mejora de la cultura y la comunicación científicas, contribuye en el observatorio IBERIFIER² con este estudio que analiza aspectos tales como los hábitos de consumo de información científica, la confianza en los diferentes medios, la actitud hacia la desinformación y los factores relacionados con la capacidad de identificar y propagar la desinformación científica. Los siguientes párrafos incluyen un resumen del diseño de la encuesta, aunque una versión más detallada se puede encontrar en el anexo del informe de la investigación (Díaz-Catalán y Cabrera-Álvarez, 2022).

La población objetivo de la encuesta incluye a los adultos de 15 o más años residentes en España y fue administrada por teléfono a una muestra final de 2.100 entrevistados. La muestra de números de teléfonos, en total 109.207, estuvo compuesta por teléfonos fijos (70%) y móviles (30%). El trabajo de campo fue llevado a cabo por la empresa Sigma Dos entre los días 22 de junio y 15 de julio de 2022. En este período, cada número de teléfono recibió, en promedio, 3,3 llamadas y la duración media de la entrevista fue de 18 minutos.

La selección de la muestra, estratificada por comunidad autónoma y tamaño de hábitat, se realizó en dos etapas. En la primera se generaron números de teléfono fijo y móviles de forma aleatoria; en el caso de los fijos, fueron estratificados por comunidad y tamaño de hábitat. En la segunda etapa, una vez que se había establecido contacto con un número de teléfono fijo correspondiente a un hogar, se procedió a seleccionar a la persona entrevistada empleando cuotas de sexo y edad. En el caso de las llamadas a teléfono móviles, la selección de la persona se realizó con cuotas de sexo y edad.

Una vez concluido el trabajo de campo, la muestra fue calibrada con el fin de generar un coeficiente de ponderación para ser utilizado en el análisis de los datos. La calibración consiste en generar un coeficiente de ponderación que, una vez aplicado, hace coincidir el perfil de la muestra con el de la población con respecto a una serie de variables que están presentes en los datos poblacionales y en la encuesta. Las variables empleadas en la calibración fueron una combinación de las variables sexo, edad (seis grupos: 15-24, 25-34, 35-44, 45-54, 55-64, 65+) y nivel educativo (cuatro grupos: primarios o menos, secundarios de primer ciclo, secundarios de segundo ciclo, superiores) y, por otro lado, comunidad autónoma. Los totales poblacionales fueron tomados de la EPA, una encuesta que por su diseño y tamaño muestral puede actuar como referente para la encuesta.

El cuestionario se compone de siete bloques que se especifican a continuación en la Tabla 2. En la documentación de la encuesta se encuentra una copia del cuestionario.

TABLA 2. CONTENIDO DEL CUESTIONARIO DEL ESTUDIO

Bloque	Contenido
1	Hábitos de consumo de información sobre ciencia
2	Credibilidad y propagación de la desinformación
3	Factores relacionados con la credibilidad percibida y la propensión a la desinformación
4	Percepción acerca de la desinformación
5	Actitudes hacia la desinformación y la ciencia
6	Confianza interpersonal y en las instituciones
7	Demográficas

²Coordinado desde la Universidad de Navarra, está integrado por doce universidades, cinco organizaciones de verificación y agencias de noticias, y seis centros de investigación multidisciplinar.

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

El diseño de la encuesta incluye la estratificación de la muestra por comunidad autónoma y tamaño de hábitat y el uso de cuotas de sexo y edad para seleccionar a la persona entrevistada. Además, tras un análisis preliminar del perfil sociodemográfico de la muestra se detectó la sobrerrepresentación de los grupos con un mayor nivel educativo. Para corregir esta desviación se generó un coeficiente de ponderación que ajusta el perfil de la muestra con respecto a las variables sexo, grupos de edad, nivel educativo y comunidad autónoma, de forma que el tamaño de estos grupos en la muestra es idéntico al peso que tienen en la población.

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 seguridad 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.

Por un lado, usar la ponderación (`peso`) en el análisis es necesario para mitigar el sesgo que puede introducir en la estimación la falta de respuesta a la encuesta y la selección introducida por el sistema de cuotas. 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 unos errores típicos más altos, que reflejan la incertidumbre que genera la falta de respuesta a la encuesta y la selección que produce el uso de las cuotas.

Por el otro, la inclusión de la estratificación (`estrato`) 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.

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, nivel educativo 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

Díaz-Catalán, C. y Cabrera-Álvarez, P. (2022). *Desinformación Científica en España. Informe de Resultados*. Fundación Española para la Ciencia y la Tecnología (FECYT).

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>