motMisc: Análisis Exploratorio de la Relación que se establece entre las Motivaciones y las Prácticas Cuestionables de Investigación

La web de este proyecto se ubica en la siguiente dirección de la Open Science Framework: <https://osf.io/s2tgw/> El Identificador Digital del Objeto (DOI) electrónic es el: 10.17605/OSF.IO/S2TGW

El archivo de datos de este proyecto es *exp.csv*. En este archivo aparecen los datos brutos del proyecto. Se encuentra ubicado en la carpeta *dat*.

# Carpetas

## *dat*

Esta carpeta contiene el archivo de datos bruto (*exp.csv*) así como los archivos de datos intermendios que han sido utilizados para realizar los diferentes análisis de datos.

## *stx*

Esta carpeta contiene los archivos de código fuente de R para realizar las transformacioens de los datos y los análisis de datos correspondientes.

## *tbl*

Esta carpeta contiene las tablas producidas en el proceso de análisis de datos.

## *fig*

Esta carpeta contiene las figuras generadas a lo largo del proceso de análisis de datos.

# Archivos

## **Archivos de datos**

### *exp.csv*

Es el archivo principal de datos de este proyecto. Contiene la información recogida por la encuesta electrónica. Este archivo genera la base de datos *imp.csv*. El mapa de variables para este archivo se encuentra aquí: <https://osf.io/hmzdu/> Este mapa de variables ha sido diseñado para que los datos que ponemos a disposición de la comunidad científica gocen de la calidad suficiente desde el punto de vista aplicado [(Musen, 2022)](https://doi.org/10.1038/d41586-022-02820-7).

### *imp.csv*

Contiene la base de datos lista para ser procesada. El mapa de variables se encuentra el la web <https://osf.io/hmzdu/> en formato PDF y en la web <https://osf.io/7jks9/> en formato html.

### *dat.csv*

Contiene los datos listos para ser procesados y analizados tras haber recodificado convenientemente las variables que contiene.

## **Archivos de código**

### *master\_motMisc.R*

Es el archivo de código fuente principal. Desde el se ejecutan los demás archivos presentes en la carpeta *stx*.

### *01\_clean.R*

Este archivo genera el archivo *imp.csv* a partir del archivo *exp.csv* y lo almacena en la carpeta *dat*. Este código lo que hace es seleccionar y renombrar las 90 variables que se utilizan en el estudio. Véase el mapa de variables para más información (en formato PDF se encuentra aquí: <https://osf.io/hmzdu/>).

### *02\_recode.R*

Este archivo recodifica las variables del archivo de datos *imp.csv*. Estas son las recodificaciones ejecutadas:

* La variabe *pais* se recodifica a *pais.r*.
* La variable *area* original es recodificada a la variable *area.r* teniendo en cuenta las categorías planteadas en el artículo de [Gopalakrishna et al (2022)](https://doi.org/10.1371/journal.pone.0263023).
* Las variables *acreedu*, *acreinv*, *grupoinv*, *nh1…nh5*, *m1,…m10*, *p1…p28* y *mp1…mp28* fueron recodificadas para que la respuesta *Sí* se asociase con el valor 1 y la respuesta *No* se asociase con el valor 0.
* La base de datos generada (que sustituye las variables *pais* y *area* por *pais.r* y *area.r* respectivamente) es guardada en el archivo *dat.csv* dentro de la carpeta *dat*.

### *03\_prev.R*

Contiene el código que generar resultados relativos a la prevalencia. Genera los siguientes archivos:

* Tabla *tb.pv.g.csv* contiene la prevalencia global de cada pci así como su intervalo de confianza al 95%.
* Tabla *tb.pv.area.csv* en la que se muestra el porcentaje de veces que se reconoce haber haber llevado a cabo prácticas cuestionables de investigación (pci) en función del área de conocimiento.
* Tabla *matcor.txt* en la que aparece la matriz de correlaciones en la que se estudia la relación que se establece entre el número total de prácticas cuestionables de investigación (pci) y otras variables cuantitativas del estudio.
* Una serie de gráficos de caja y de dispersión que se almacenan en la carpeta de *fig* en formato PDF.

### *04\_fc\_pv\_g.R*

Este archivo contiene el código de la función pv.g() que computa la prevalencia y los intervalos de confianza para un conjunto de pci. Produce la tabla contenida en el archivo *tb.pv.g.csv*.

### *05\_fc\_pv\_area.R*

Este archivo contiene el código de la función pv.area() que computa la prevalencia y los intervalos de confianza para un conjunto de pci en función del área de conocimiento. Produce la tabla contenida en el arcnivo *tb.pv.area.csv*.

## **Tablas**

### *tb.pv.g.csv*

Contiene la prevalencia global para cada práctica cuestionables de investigación (pci) con su correspondiente intervalo de confianza al 95%. Estos datos son análogos a los presentados por [Gopalakrishna et al (2022)](https://doi.org/10.1371/journal.pone.0263023) en la columna derecha de la Tabla 2

### *tb.pv.area.csv*

Contiene el porcentaje de ocurrencia de cada pci en función del área de conocimiento. Esta tabla es análoga a parte de la presentada en el artículo artículo de [Gopalakrishna et al (2022)](https://doi.org/10.1371/journal.pone.0263023) en la Tabla 2.

### *matcor.txt*

Matriz de correlaciones para la relación que se establece entre la variable PCI y demás variables del estudio.

## **Figuras**

### *hist\_pci.pdf*

Es el histograma de la variable PCI, número total de prácticas cuestionables de investigación llevadas a cabo por los participantes del estudio.

### *box\_pci\_area.pdf*

Es el gráfico de caja de la variable PCI en función del área de conocimiento de los participantes.

### *box\_pci\_insti.pdf*

Gráfico de caja para la variable PCI en función del tipo de institución.

### *pci\_edex.pdf*

Gráfico de dispersión de la relación que se establece entre el número de PCI y los años de experiencia docente.

### *pci\_exin.pdf*

Gráfico de dispersión de la relación que se establece entre el número de PCI y los años de experiencia investigadora.

### *pci\_autprin.pdf*

Gráfico de dispersión de la relación que se establece entre el número de PCI y el número de artículos publicados como autor/a principal.

### *pci\_njcr.pdf*

Gráfico de dispersión de la relación que se establece entre el número de PCI y el número de artículos publicados en revistas indexadas en la JCR (*Journal Citations Reports*).

# Referencias

* Gopalakrishna, G., Gerben ter Riet, G., Gerko Vink, G., Stoop, I., Wicherts, J. M., Bouter, L. M. (2022). Prevalence of questionable research practices, research misconduct and their potential explanatory factors: A survey among academic researchers in The Netherlands, *PlosONE*, <https://doi.org/10.1371/journal.pone.0263023>
* Musen, M. A. (2022, Septiembre 5). Without appropriate metadata, data-sharing mandates are pointless. *Nature, 609*, 222. <https://doi.org/10.1038/d41586-022-02820-7>