

motMisc: Análisis Exploratorio de las Motivaciones y las Prácticas Cuestionables de Investigación en la Comunidad Científica Hispanohablante

Este proyecto está destinado a realizar un análisis exploratorio de las prácticas cuestionables de investigación en el ámbito de la investigación producida en países hispanohablantes. Todos los datos, brutos y procesados, de este proyecto están disponibles para que la comunidad científica pueda cotejarlos y re-analizarlos a conveniencia. 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) de este proyecto es el: <https://doi.org/10.17605/OSF.IO/S2TGW>

La información que aparece en este documento puede consultarse tanto en formato PDF (<https://osf.io/v59dw/>), en formato html (<https://osf.io/k78gf/>), así como en formato .docx legible por Microsoft Word (<https://osf.io/mxk42/>).

El mapa de variables para este archivo en formato PDF 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). De esta manera, los análisis que hemos realizado podrán reproducirse, cotejarse o extenderse (Berg, 2018).

Este proyecto se compone de tres secciones principales: Carpetas, Archivos y Referencias. En la sección *Carpetas* se listan y describen las cuatro carpetas o directorios (dat, stx, tbl y fig) que integran el proyecto. En la sección *Archivos* se describen los archivos que están vinculados al proyecto. Los archivos del proyecto se agrupan en cuatro categorías diferentes. A saber, *archivos de datos* (que contienen datos brutos o procesados), *archivos de código* (que contienen en código fuente de R que ejecuta los análisis o que generan los gráficos del proyecto), *tablas* (que contienen tablas de resúmenes estadísticos) y *figuras* (que contienen los archivos de figuras generadas en el proyecto). Cada uno de los archivos mencionados están descritos.

El archivo de datos raíz 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 intermedios 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 transformaciones de los datos y los análisis de datos correspondientes.

tbl

Esta carpeta contiene las tablas producidas en el proceso de análisis de datos. También contiene archivos de texto con resultados que han sido utilizados para generar las tablas que aparecen en el manuscrito. Es decir, esta carpeta también contiene resultados de análisis que son susceptibles de aparecer reflejados en tablas.

fig

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

Archivos

Archivos de datos

Archivos de datos iniciales

Con estos archivos de datos, y utilizando los archivos de código suministrados, se pueden realizar todos los análisis de datos que se han realizado en este proyecto.

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*.

imp.csv Contiene la base de datos lista para ser procesada.

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

coment.csv Contiene las respuestas a la pregunta de texto abierto en el que se solicita a los participantes que aporten cualquier comentario que consideren oportuno. Junto a cada comentario aparece el código de identificación del participante.

Archivos de datos intermedios

Estos son los archivos de datos intermedios que son generados a lo largo del proceso de análisis de datos. Estos archivos son invocados en fases intermedias del análisis de datos.

dat_2.csv Este archivo es análogo al archivo *dat.csv* pero añade la variable *pci* que representa el número total de prácticas cuestionables de investigación (PCI) que manifiesta haber realizado cada participante.

dat_3.csv Este archivo añade 27 nuevas variables a las contenidas en el archivo *dat_2.csv*. Estas nuevas variables (*pp1*, *pp2*, ..., *pp28*, omitiendo la *pp20*) caracterizan a cada participante teniendo en cuenta si consideran cuestionables cada una de las 27 prácticas cuestionables de investigación (*pci1*, *pci2*, ..., *pci28*, omitiendo la *pci27*) y si ellos mismos las han llevado a cabo.

ms.RData Contiene tablas que analizan la relación que se establece entre las motivaciones de investigación y las prácticas cuestionables de investigación analizadas. Para más información al respecto, véase la descripción del archivo *07_mot_pci.R*.

full.boot_gda.RData Contiene los datos necesarios para trazar la red dirigida acíclica no conexa e intermedia que se estima en este trabajo. Véase la descripción del archivo *09_grafos.R* para una descripción y contextualización más precisa de este archivo.

conect.boot_gda.RData Contiene los datos necesarios para trazar la red dirigida acíclica conexa y definitiva que se estima en este trabajo. Véase la descripción del archivo *09_grafos.R* para una descripción y contextualización más precisa de este archivo.

Archivos de código

master_motMisc.R

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

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 (se encuentra aquí en formato PDF: <https://osf.io/hmzdu/>, aunque puede obtenerse en otros formatos como se indica en la introducción de este documento).

02_recode.R

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

- La variable *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). Así, las nuevas categorías y etiquetas asignadas son las siguientes (entre paréntesis aparecen las etiquetas que se utilizan en los gráficos generados):
 - 1: Ciencias de la vida y médicas (VM)
 - 2: Ciencias sociales y del comportamiento (SC)
 - 3: Ciencias naturales e ingeniería (NI)
 - 4: Arte y humanidades (AH)
- Las variables *acreedu*, *acreinv*, *grupoinv*, *nh1...nh5*, *m1...m10*, *p1...p28* y *mp1...mp28* son recodificadas para que la respuesta *Sí* se asocie con el valor 1 y la respuesta *No* se asocie 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 genera resultados relativos a la prevalencia. Genera los siguientes archivos:

- Tabla *tb.pv.g.csv* contiene la prevalencia global de cada práctica cuestionable de investigación (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 llevado a cabo prácticas cuestionables de investigación (pci) en función del área de conocimiento. Cada columna de esta tabla está etiquetada con alguno de estos prefijos *p.* (referido a la prevalencia de la práctica cuestionable de investigación), *ll.* (límite inferior del intervalo de confianza al 95% para la prevalencia de la práctica cuestionable de investigación) y *ul.* (límite superior del intervalo de confianza al 95% estimado para la prevalencia de la práctica cuestionable de investigación). Por su parte, el sufijo de cada columna está referido a las áreas de conocimiento consideradas siguiendo la propuesta hecha por Gopalakrishna et al (2022) y utilizando las siguientes etiquetas:
 - vm: Ciencias de la vida y médicas
 - sc: Ciencias sociales y del comportamiento
 - ni: Ciencias naturales e ingeniería
 - ah: Arte y humanidades
- 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. Los p-valores estimados para cada correlación aparecen corregidos teniendo en cuenta comparaciones múltiples utilizando el método de Holm.
- Una serie de gráficos de caja y de dispersión que se almacenan en la carpeta de *fig* en formato PDF.
- Este archivo también genera el archivo de datos *dat_2.csv* que se almacena en la carpeta *dat*. Este archivo es exactamente igual al archivo *dat.csv* pero añade la variable *pci* que representa el número total de prácticas cuestionables de investigación (PCI) que manifiesta haber realizado cada participante.

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 archivo *tb.pv.area.csv*.

06_motpub.R

Este archivo analiza las distribuciones de frecuencias de las variables *nh1*, *nh2*, ..., *nh5* de manera univariadamente, en función del perfil y en función del área de conocimiento. También analiza la variable *pubnohh* de manera univariada, en función del perfil y en función del área de conocimiento. Los análisis llevados a cabo se graban en el archivo *nh_m.txt* dentro de la carpeta *tbl*.

07_mot_pci.R

Este archivo estima las correlaciones (los correspondientes intervalos de confianza al 95% y los *p*-valores para la estimación habiendo aplicado el método de Holm para corregir la tasa de Error Tipo I) que se establecen entre cada motivación (*m1*, *m2*, ..., *m10*) y las prácticas cuestionables de investigación (*p1*, *p2*, ..., *p28*) incluidas en el estudio (la *p20* ha sido excluida de este análisis porque su frecuencia absoluta es igual a cero en la muestra). El archivo produce diez tablas (una por cada motivación, *m1.csv*, *m2.csv*, ..., *m10.csv*) que se almacenan en la carpeta *tbl*. Cada tabla contiene cinco columnas cuyas etiquetas y significado son los siguientes:

- primera columna: representa el nombre de cada una de las prácticas cuestionables de investigación (*p1.cor, p2.cor, ..., p28.cor*).
- r: coeficiente de correlación estimado para la motivación analizada y la correspondiente práctica cuestionable de investigación.
- X1: límite inferior del intervalo de confianza para la correlación estimada al 95%.
- X2: límite superior del intervalo de confianza para la correlación estimada al 95%.
- p.val: *p*-valor estimado para la correlación habiéndose corregido la tasa de Error Tipo I utilizando el método de Holm.

Las diez tablas anteriormente descritas también son almacenadas en un archivo de datos específico para R en el archivo *ms.RData* que se guarda en la carpeta *tbl*.

08_perfil_pci.R

Este archivo analiza las prácticas cuestionables de investigación (pci) teniendo en cuenta si los participantes las realizan, o no, y si las consideran, o no, como malas prácticas en la investigación científica. Lo que se hace, en cierto modo, es computar una serie de variables que posicionan a cada participante en un grupo de “perfil de integridad académica”. Para ello, se crean 27 nuevas variables (*pp1, pp2, ..., pp28*) que identifican cuatro perfiles diferentes de participantes (el archivo *dat_3.csv* intermedio que contiene estas nuevas variables es almacenado en la carpeta *dat*). Los cuatro perfiles, así como las etiquetas que los identifican, que se identifican para cada una de las prácticas cuestionables de investigación son los siguientes:

- BP: referido a “buenas prácticas de investigación”, ya que el participante considera que la práctica correspondiente de investigación es cuestionable y no la lleva a cabo.
- PC: referido a “práctica cuestionable de investigación”, ya que el participante considera que la práctica de investigación es cuestionable y la lleva a cabo.
- HD: referido a “honestidad despreocupada”, ya que el participante considera que la práctica cuestionable de investigación no es, de hecho, cuestionable y tampoco la lleva a cabo.
- CD: referido a “práctica cuestionable despreocupada”, ya que la persona lleva a cabo la práctica cuestionable de investigación y, de hecho, no la considera cuestionable.

Todos los resultados de los análisis realizados utilizando estos perfiles se almacenan en la carpeta *tbl*. Estos son los archivos de resultados generados y las descripciones que permiten interpretarlos:

- Las frecuencias absolutas de cada uno de estos perfiles para cada práctica cuestionable de investigación, así como las correspondientes frecuencias relativas, se aparecen en el archivo *ppci.csv*. El prefijo del nombre de la columna (justo antes del punto) identifica al perfil (bp: buenas prácticas, cd: cuestionable despreocupada, hd: honesta despreocupada, y pc: práctica cuestionable) y el sufijo (después del punto) identifica si se trata de una frecuencia absoluta (f) o relativa (p, de porcentaje).
- Este archivo también genera documentos que analizan la relación que se establece entre los perfiles previamente mencionados para cada práctica cuestionable de investigación en función del área de conocimiento del participante y su perfil académico. Se generan los siguientes archivos de resultados intermedios:
 - *fa_perfil_a.txt*: contiene, para cada práctica cuestionable de investigación, la tabla de contingencia que confronta el área de investigación y los perfiles de prácticas descritas previamente.
 - *fa_perfil_p.txt*: contiene, para cada práctica cuestionable de investigación, la tabla de contingencia que confronta el perfil académico del participante y los perfiles de prácticas descritas previamente.
 - *aso_perfil_a.txt*: contiene, para cada práctica cuestionable de investigación, los estadísticos de asociación para la relación que se establece entre el área de investigación y los perfiles de prácticas descritas previamente.

- *aso_perfil_p.txt*: contiene, para cada práctica cuestionable de investigación, los estadísticos de asociación para la relación que se establece entre el perfil académico del participante y los perfiles de prácticas descritas previamente.

Los datos relevantes derivados de todos estos archivos aparecen reflejados en el archivo *aso_ap.csv*. El prefijo del nombre de cada columna (lo que aparece antes del punto) se refiere a la χ^2 de Pearson (ji), a los grados de libertad de χ^2 (gl), al p -valor relativo a χ^2 (p.val), al coeficiente de contingencia (cc) y a la V de Cramer (v). El sufijo del nombre de cada columna (lo que aparece tras el punto) sirve para hacer referencia a la variable “área de conocimiento” del participante (a) y “perfil académico” del participante (p).

Todos estos datos, intermedios y colapsados, están también contenidos, como objetos diferenciados de R, en el archivo *perfil_pci.RData*.

09_grafos.R

Este archivo produce una serie de grafos y los representa utilizando diferentes algoritmos. Para los dos grafos finales seleccionados (uno dirigido acíclico y otro no dirigido) estima sus correspondientes estadísticos de centralidad, intermediación y conglomeración.

El archivo comienza estimando un gráfico no dirigido utilizando el algoritmo EBICglasso de regularización implementado en el paquete *qgraph* de R (el grafo estimado se almacena en la carpeta *fig* como archivo PDF con el nombre *nc.gnd.pdf*). Seguidamente se eliminan los nodos inconexos al grafo principal y se obtiene un nuevo grafo no dirigido (que se almacena en la carpeta *fig* en formato PDF y con el nombre *c.gnd.pdf*).

Las variables incluidas en este grafo no dirigido son posteriormente utilizadas para generar un grafo dirigido acíclico utilizando el paquete *bnlearn*. Para estimar el grafo no dirigido acíclico se utilizan el algoritmo grow-shrink (Scutari, 2010), y 1000 remuestreos aleatorios a partir de los datos observados Nagarajan et al, 2013. El código que genera este modelo dirigido está comentado para que el cómputo no consuma tiempo excesivo. Por tanto, los datos necesarios generados, y que permiten especificar la red, son guardados en el archivo *full_boot_gda.RData*. Esos datos son cargados por el archivo para estimar los gráficos y estadísticos correspondientes. La red generada por este proceso se guardan en formato PDF en los archivos *nc.gda.pdf* (usando el paquete *qgraph*) y *gp.nc.gda.pdf* (usando la función *graphviz.plot()* incluida en el paquete *Rgraphviz* y que ha sido incorporada en *bnlearn*). Las fuerzas de asociación para la red generada se encuentran en el archivo *pesos.full.gda.txt* (almacenado en la tabla *tbl*) para los enlaces cuya fuerza de asociación son mayores a 0.85 y cuya dirección es estable en más del 50% de las redes estimadas.

Como el grafo generado no es conexo, se toman las variables que forman parte del grafo conexo principal y se estima una nueva red utilizando el procedimiento de remuestreo anteriormente descrito. Este procedimiento genera las redes que se dibujan en los archivos *c.gda.pdf* y *gp.c.gda.pdf*. El código que genera los datos de remuestreo están comentados pero fueron guardados en el archivo *connect_boot_gda.RData*, que son cargados por el archivo para trazar los grafos y estimar los estadísticos correspondientes. Todos los grafos mencionados hasta el momento se guardan en la carpeta *fig*. Las fuerzas de asociación para la red generada se encuentran en el archivo *pesos.connect.gda.txt* (almacenado en la tabla *tbl*) para los enlaces cuya fuerza de asociación son mayores a 0.85 y cuya dirección es estable en más del 50% de las redes estimadas.

Este archivo también genera dos tablas que contienen estadísticos de resúmenes para las redes (no dirigida y dirigida acíclica) finales. La tabla contenida en el archivo *cent.t.gnd.csv* corresponde al grafo no dirigido y contiene las siguientes columnas:

- *inter.gnd*: representa el coeficiente de intermediación de cada nodo de la red.
- *grado.ponderado.gnd*: representa el grado de cada nodo teniendo en cuenta que los enlaces de la red son ponderados.
- *ws.c*: representa el coeficiente de conglomeración usando el método propuesto por Watts y Strogatz, (1998).

Por su parte, la tabla contenida en el archivo *cent.t.gda.csv* corresponde al grafo no dirigido acíclico y contiene las siguientes columnas:

- *inter.gda*: coeficiente de intermediación de cada nodo.
- *grado.in*: grado de entrada para cada nodo. Es decir, el número de arcos que apuntan hacia ese nodo.
- *grado.out*: grado de salida para cada nodo. Es decir, el número de arcos que salen de ese nodo.
- *grado.t*: grado total para cada nodo. Es decir, el número de conexiones que tiene cada nodo independientemente de la dirección de la conexión.
- *ws.c*: coeficiente de conglomeración para cada nodo calculado usando el método propuesto por Watts y Strogatz, (1998).

Finalmente, este archivo también traza los dos grafos estimados finales (el no dirigido y el dirigido) en dos dispositivos gráficos de R.

10_descriptivo.R

Este archivo ejecuta un conjunto de análisis que son de apoyo para el desarrollo narrativo del manuscrito producido para reportar los resultados de la investigación. Aunque los análisis nucleares de la investigación han sido producidos con los archivos de código precedentes, este archivo está concebido para producir los análisis de datos que se irán presentando a lo largo del manuscrito. Por ejemplo, los análisis estadísticos que son necesarios para describir la muestra de participantes utilizada se ejecutan con el código contenido en este archivo.

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) en la columna derecha de la Tabla 2.

tb.pv.area.csv

Contiene el porcentaje de ocurrencia de cada práctica cuestionable de investigación (*pci*) en función del área de conocimiento. Esta tabla es análoga a la columna derecha de la Tabla 2 presentada en el artículo artículo de Gopalakrishna et al (2022).

matcor.txt

Matriz de correlaciones para la relación que se establece entre las prácticas cuestionables de investigación (*pci*) y demás variables del estudio. Los *p*-valores reportados están corregidos con el método de Holm.

mh_m.txt

Este archivo contiene los comentarios, el código fuente y los resultados correspondientes que se obtienen para el análisis de las frecuencias, absolutas y relativas, de las variables *nh1*, *nh2*, ..., *nh5* y *pubnohh*, tanto univariadamente como en función del perfil y del área de conocimiento.

m1.csv, m2.csv, ..., m10.csv

Estas diez tablas contienen la estimación de la correlación (con sus correspondientes intervalos de confianza al 95% y sus p -valores corregidos con el método de Holm) que se establece entre cada motivación de investigación ($m1, m2, \dots, m10$) y las prácticas cuestionables de investigación ($p1, p2, \dots, p28$).

ms.RData

Contiene las tablas *m1.csv, m2.csv, ..., m10.csv* en forma de objetos (*data frames*) legibles por R.

aso_ap.csv

Contiene estadísticos relevantes para interpretar la relación que se establece entre el área de conocimiento, el perfil académico y los perfiles de malas prácticas para cada práctica cuestionable de investigación. Véase la descripción del archivo *08_perfil_pci.R* para más información contextualizada al respecto.

aso_perfil_a.txt

Contiene estadísticos que valoran la asociación que se establece, para cada práctica cuestionable de investigación, entre el área de conocimiento del participante y los perfiles de prácticas cuestionables de investigación. Véase la descripción del archivo *08_perfil_pci.R* para más información contextualizada al respecto.

aso_perfil_p.txt

Contiene estadísticos que valoran la asociación que se establece, para cada práctica cuestionable de investigación, entre el perfil académico del participante y los perfiles de prácticas cuestionables de investigación. Véase la descripción del archivo *08_perfil_pci.R* para más información contextualizada al respecto.

fa_perfil_a.txt

Contiene, para cada práctica cuestionable de investigación, la tabla de contingencia que compara el área de conocimiento con los perfiles prácticas cuestionables de investigación. Véase la descripción del archivo *08_perfil_pci.R* para más información contextualizada al respecto.

fa_perfil_p.txt

Contiene, para cada práctica cuestionable de investigación, la tabla de contingencia que compara el perfil académico con los perfiles prácticas cuestionables de investigación. Véase la descripción del archivo *08_perfil_pci.R* para más información contextualizada al respecto.

fabs.txt

Contiene, para cada práctica cuestionable de investigación, la frecuencia absoluta de cada perfil de prácticas cuestionables de investigación. Véase la descripción del archivo *08_perfil_pci.R* para más información contextualizada al respecto.

frel.txt

Contiene, para cada práctica cuestionable de investigación, la frecuencia relativa (en forma de porcentaje) de cada perfil de prácticas cuestionables de investigación. Véase la descripción del archivo *08_perfil_pci.R* para más información contextualizada al respecto.

perfil_pci.RData

Este archivo contiene todos los objetos de R generados con el archivo *08_perfil_pci.R*. Para una descripción más detallada del mismo, se sugiere consultar la descripción de ese archivo de código que aparece más arriba.

ppci.csv

Contiene, para cada práctica cuestionable de investigación, las frecuencias absolutas y relativas de cada perfil de prácticas cuestionables de investigación. Véase la descripción del archivo *08_perfil_pci.R* para más información contextualizada al respecto.

cent.t.gda.csv

Contiene los estadísticos de centralidad y de conglomeración para la red dirigida acíclica seleccionada en el estudio. En el archivo *09_grafos.R* se proporciona información más detallada y contextualizada sobre este archivo.

cent.t.gnd.csv

Contiene los estadísticos de centralidad y de conglomeración para la red no dirigida seleccionada en el estudio. En el archivo *09_grafos.R* se proporciona información más detallada y contextualizada sobre este archivo.

pesos.full.gda.txt

Contiene las magnitudes de asociación que se establecen entre cada variable para el grafo dirigido acíclico no conexo estimado en una fase intermedia del análisis de datos. En el archivo *09_grafos.R* se proporciona información más detallada y contextualizada sobre este archivo.

pesos.conect.gda.txt

Contiene las magnitudes de asociación que se establecen entre cada variable para el grafo dirigido acíclico conexo estimado en este estudio. En el archivo *09_grafos.R* se proporciona información más detallada y contextualizada sobre este archivo.

pais.csv

Contiene la distribución de frecuencias absoluta (fila “pais”) y relativa (en forma de porcentajes, fila “pais.per”) de los participantes atendiendo al país en el que trabajaban cuando se recogieron los datos del estudio.

descriptivo.txt

Contiene los análisis de datos que acompañan la descripción del manuscrito.

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 las prácticas cuestionables de investigación (PCI) en función del área de conocimiento de los participantes. Las etiquetas utilizadas para identificar a las áreas son las siguientes:

- VM: Ciencias de la vida y médicas
- SC: Ciencias sociales y del comportamiento
- NI: Ciencias naturales e ingeniería
- AH: Arte y humanidades

box_pci_insti.pdf

Gráfico de caja para la variable PCI (número de prácticas cuestionables de investigación) en función del tipo de institución.

box_pci_perfil.pdf

Gráfico de caja para la variable PCI (número de prácticas cuestionables de investigación) en función del perfil del participante. El significado de las etiquetas es el siguiente:

- ED: estudiante de doctorado
- PI: personal investigador
- PD: personal docente
- PDI : personal docente e investigador

pci_edex.pdf

Gráfico de dispersión de la relación que se establece entre el número de PCI (prácticas cuestionables de investigación) y los años de experiencia docente.

pci_exin.pdf

Gráfico de dispersión de la relación que se establece entre el número de PCI (prácticas cuestionables de investigación) y los años de experiencia investigadora.

pci_autprin.pdf

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

c.gda.pdf

Grafo dirigido acíclico conexo estimado en esta investigación. Los nodos etiquetados como $p1, p2, \dots, p28$ se refieren a las prácticas cuestionables de investigación $pci1, pci2, \dots, pci28$.

c.gnd.pdf

Grafo no dirigido conexo estimado en esta investigación. Los nodos etiquetados como $p1, p2, \dots, p28$ se refieren a las prácticas cuestionables de investigación $pci1, pci2, \dots, pci28$.

gp.c.gda.pdf

Contiene el mismo grafo que el archivo *c.gda.pdf* pero trazado con función `graphviz.plot()` incluida en el paquete `Rgraphviz`.

gp.nc.gda.pdf

Contiene el grafo no dirigido conexo trazado con la función `graphviz.plot()` incluida en el paquete `Rgraphviz`.

nc.gda.pdf

Contiene el grafo dirigido acíclico no conexo.

nc.gda.pdf

Contiene el grafo no dirigido no conexo.

pais.pdf

Este gráfico es la distribución de frecuencias de los participantes atendiendo al país en el que trabajaban en el momento de la recogida de datos.

Referencias

- Berg, J. (2018, Abril 13). Obfuscating with transparency. *Science*, 360(6385), 133. <https://doi.org/10.1126/science.aat8121>
- 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, *Plos One*, 17(2), Artículo e0263023 <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>

- Nagarajan, R., Scutari, M., y Lèbre, S. (2013). *Bayesian Networks in R*. Springer. <https://doi.org/10.1007/978-1-4614-6446-4>
- Scutari, M. (2010). Learning Bayesian Networks with the bnlearn R Package. *Journal of Statistical Software*, 35(3), 1–22. <https://doi.org/10.18637/jss.v035.i03>
- Watts, D., y Strogatz, S. (1998, Junio 4). Collective dynamics of ‘small-world’ networks. *Nature*, 393, 440–442. <https://doi.org/10.1038/30918>