



Facultad de Ciencias Económicas y Empresariales

ICADE

**El cambio estructural en el estigma
asociado a la salud mental derivado de
la pandemia de 2020**

Autor: Belén Mata Moína

Clave: 201905469

Director: Eduardo César Garrido Merchán

MADRID | Junio 2024

Resumen

En el siguiente trabajo se realiza un análisis genérico del estado de la salud mental en los profesionales dedicados a la tecnología, así como del estigma asociado a estas afecciones. Se estudia tanto la evolución en la existencia de patologías mentales, como la forma de abordarlo de las personas que las sufren con sus empresas y entornos cercanos. Posteriormente, se profundiza en la evolución del estigma asociado a esta tipología de enfermedades y la disposición de los pacientes a compartir con sus familiares y amigos información sobre esta materia. Concretamente, se demuestra la existencia de un cambio estructural causado por el impacto de la pandemia del año 2020 y se rechaza la hipótesis nula planteada que la negaba. Para ello se utiliza la prueba de Chow y se interpretan los valores derivados de esta. Los resultados obtenidos se validan utilizando métodos estadísticos que comprueban su fiabilidad. Para ello se utilizan diversas métricas de bondad del ajuste que permiten una verificación completa del nivel de acoplamiento de los modelos utilizados a los datos reales. La realización de este experimento y la claridad de sus resultados permite la extracción de unas conclusiones nítidas y robustas. Por último, se exponen líneas de investigación futuras en las que en este estudio no se ha podido profundizar.

Palabras clave: salud mental, estigma, sector tecnológico, cambio estructural, prueba de Chow.

Abstract

In the following work, a generic analysis of the state of mental health in professionals dedicated to technology, as well as the stigma associated with these conditions, is carried out. The evolution of the existence of mental pathologies is studied, as well as the way in which the people who suffer from them approach it with their companies and close environments. The evolution of the stigma associated with this type of disease and the willingness of patients to share information on this subject with their family and friends is analyzed in depth. Specifically, the existence of a structural change caused by the impact of the 2020 pandemic is demonstrated and the null hypothesis that denied it is rejected. For this purpose, the Chow test is used and the values derived from it are interpreted. The results obtained are validated using statistical methods that test their reliability. Various goodness-of-fit metrics are used to allow a complete verification of the level of coupling of the models used to the real data. The performance of this experiment and the clarity of its results allow clear and robust conclusions to be drawn. Finally, future lines of research, which this study has not been able to explore in depth, are presented.

Key words: mental health, stigma, technology sector, structural change, Chow test.

Índice de contenido

1. Introducción	7
2. Estado del Arte	9
3. Alcance del trabajo fin de grado	11
3.1 Objetivos	11
3.2 Hipótesis	11
3.3 Restricciones	12
3.4 Asunciones	12
4. Marco teórico	13
4.1 OSMI Mental Health	13
4.2 Prueba de Chow	14
4.3 Análisis previo de los datos	15
4.3.1 Análisis general	15
4.3.2 Salud mental en el ámbito laboral	20
4.3.3 Salud mental en el ámbito personal	24
5. Experimento	26
5.1 Bondad del ajuste	28
6. Resultado	35
7. Conclusiones	36
8. Futuras líneas de investigación	37
9. Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado	38
10. Referencias bibliográficas	39

Índice de figuras

Figura 1: Porcentaje de profesionales en puestos tecnológicos que padece un trastorno de salud mental.....	18
Figura 2: Distribución por sexo y edad de los individuos que padecen un trastorno de salud mental.....	19
Figura 3: Porcentaje de personas con un trastorno de salud mental dividido por sexo..	20
Figura 4: Porcentaje de cobertura de salud mental de profesionales tecnológicos.	21
Figura 5: Porcentaje de conocimiento de las opciones dentro de su seguro médico de la salud mental de profesionales tecnológicos.....	21
Figura 6: Porcentaje de empresas que han mencionado la salud mental a sus empleados.	22
Figura 7: Porcentaje de profesionales tecnológicos que han discutido su salud mental con sus empleadores.	23
Figura 8: Porcentaje de discusión de la salud mental con el empleador distribuido por sexo.....	23
Figura 9: Tabla de contingencia con las variables conv_sm_antiguo_emp y conv_sm_emp.....	24
Figura 11: Disposición al debate de la salud mental con el entorno cercano en 2017. ...	24
Figura 10: Disposición al debate de la salud mental con el entorno cercano en 2017. ...	24
Figura 12: Evolución de la variable entorno_familiar. 2017 a 2021.	25
Figura 13: Resultados de la regresión model1.....	32
Figura 14: Resultados de la regresión model2.....	32
Figura 15: Resultados de la regresión full_model.....	33

Todas las figuras expuestas en el presente trabajo son de elaboración propia.

Índice de ecuaciones

Ecuación 1: estadístico F para prueba de Chow	14
Ecuación 2: regresión lineal múltiple	27
Ecuación 3: R cuadrado	29
Ecuación 4: R cuadrado ajustado.....	29
Ecuación 5: AIC	30
Ecuación 6: BIC	31

1. Introducción

La salud mental ha ganado gran relevancia en los últimos años. Difícil de describir y comprender por su amplitud y ambigüedad, es definida por la OMS como “*un estado de bienestar mental que permite a las personas hacer frente a los momentos de estrés de la vida, desarrollar todas sus habilidades, poder aprender y trabajar adecuadamente y contribuir a la mejora de su comunidad*” (World Health Organization: WHO, 2019).

Recientemente, este término está presente en entrevistas, podcasts, periódicos, declaraciones públicas e incluso planes gubernamentales. Pero ¿Por qué ahora? Es innegable que la pandemia de coronavirus vivida en 2020 trajo consigo grandes cambios que siguen afectando a la sociedad. El aislamiento, duelo, desconocimiento y caos que sembró a su paso, trastocaron sin duda la vida de muchas personas. Las situaciones vividas, fuera de los esquemas de cualquier ciudadano, aumentaron los niveles de estrés de la población en un contexto de bajo contacto social, lo cual produjo un incremento de los problemas de salud mental, siendo los más comunes la depresión y la ansiedad (Mouzo, 2021).

Sin embargo, no todas las consecuencias de la pandemia impactaron de forma negativa a la sociedad. Si algo bueno surgió a raíz de la pandemia en materia de salud mental fue una gran reducción del estigma que favorece el debate abierto y da paso a una mayor investigación, diagnóstico y prevención de este tipo de enfermedades y sus síntomas. Dejar atrás el tabú que muchas veces acompaña a los problemas psicológicos hace que sea más fácil acceder a aquellas personas que lo sufren, ofreciéndoles asistencia de forma exhaustiva y estudiando cada caso con mayor detalle y conocimiento. Es por ello por lo que cada vez existen más datos sobre estas condiciones, que reflejan la situación de multitud de personas afectadas tanto por trastornos severos como pueden ser esquizofrenias, depresión mayor o trastorno obsesivo compulsivo; como otros a los cuales la psicóloga Marian Rojas Estapé denomina “*patologías de la vida corriente*”. Con este término se refiere a un malestar continuado, caracterizado por la presencia de síntomas leves presentes en algunas enfermedades mentales, y que, sin suponer impedimento para realizar actividades cotidianas, genera una gran incomodidad en el individuo.

Para lograr deshacerse del mencionado tabú, eterno acompañante de las condiciones de salud mental, muchas personas, profesionales de este campo o no, así como instituciones y asociaciones, ponen en marcha campañas y proyectos de

sensibilización e información con públicos muy diversos. Sin embargo, muchas veces es difícil medir si estas acciones son fructíferas y parametrizar su impacto.

Uno de los principales motivos para la realización de este trabajo es intentar aportar algo nuevo en este campo, demostrando desde una visión más académica si verdaderamente ese cambio deseado ocurre en la sociedad y se puede observar en datos reales. Añadir nuevos enfoques y puntos de vista a una materia que aún necesita mucha investigación es una razón de peso para el desarrollo del presente trabajo.

Para facilitar la lectura y seguimiento de este estudio, primero se explica el estado de la cuestión planteada, ofreciendo una revisión de los artículos y publicaciones disponibles sobre el tema a tratar. A continuación, se definen los objetivos, hipótesis, asunciones y restricciones que delimitan el trabajo. Posteriormente, se explican algunos conceptos útiles para una mejor comprensión del trabajo y se realiza un análisis general de los datos que aporta una visión general de la situación actual de la salud mental en el sector tecnológico. Después se realiza el experimento, se validan y se exponen los resultados obtenidos, así como las conclusiones que éstos permiten extraer. Por último, se sugieren líneas de investigación en las cuales no se ha podido profundizar en esta ocasión, pero cuyo desarrollo se considera muy interesante para futuros trabajos.

2. Estado del Arte

Sumándose a la mayor presencia en prensa y otros medios de comunicación mencionada anteriormente, la cantidad de fondos, tiempo y profesionales dedicados a la salud mental va en aumento. Sin embargo, la información disponible y conocimiento sobre el tema es notablemente inferior a muchas otras materias. Las patologías mentales forman parte de una ciencia inexacta, de difícil estudio y generalización. La forma de abordarlas es en sí un propio reto por dos motivos principales: El primero es que existe un gran estigma alrededor de este tipo de afecciones. La Organización Panamericana de la Salud define el estigma como *“marca que excluye a una persona de las demás y que disminuye su valor en el grupo social al que pertenece. También se refiere a la actitud y los comportamientos negativos hacia las personas con problemas por consumo de sustancias y de salud mental”*(s.f.). La actitud mencionada se puede presentar en las formas de estigma público y autoestigma como bien se describe en el artículo “Estigma y enfermedad mental” para la revista Norte de Salud Mental (Muñoz & Uriarte, 2006). La existencia de dicha distinción y actitudes cohíbe a las personas que sufren de alguna condición mental, reduciendo la probabilidad de que estas se abran con otras, y por tanto, reciban ayuda.

La segunda de las razones que convierte a la salud mental en un campo complicado de abordar es su estrecha relación con el suicidio. Muchas veces se evita mencionar todo aquello que se pueda relacionar con este acto por el conocido “efecto llamada”, una reacción de imitación de aquello que se ve o escucha que provocaría que más personas decidiesen suicidarse.

Pese a las dos razones planteadas, muchas organizaciones y profesionales del ámbito psiquiátrico y psicológico optan por la visibilización de las enfermedades de salud mental, estudian el estigma asociado a ellas, la importancia de su detección y su adecuada prevención. Ejemplo de ello es el artículo “Estigma: barrera de acceso a servicios en salud mental”, el cual concluye que es necesario promover una mayor sensibilización de la población ante las enfermedades mentales para favorecer la integración de aquellos que las padecen (Campo-Arias et al., 2014).

Otros autores, como Ximena Palacios-Espinosa en su artículo El inestimable costo del estigma de la salud mental además de denunciar la situación, estudian el origen humano del estigma, a la vez que invitan a la lucha contra el mismo por parte de cada individuo (2021).

Aparte de los estudios individuales por parte de profesionales, existen también informes y guías creadas por instituciones tan relevantes como la Organización Mundial de la Salud que ofrecen claves para entender el estado actual de la salud mental global, así como recomendaciones para transformar su situación, como ocurre en el Informe mundial sobre salud mental.

Pese al repertorio de artículos, informes y asociaciones dedicadas al estudio de este campo tan interesante, resulta difícil encontrar aplicadas técnicas tanto econométricas como de machine learning a datos de esta tipología, normalmente por su difícil estandarización y recogida. Esa escasez aporta doble valor al presente trabajo, dotándolo de una originalidad especial y arrojando nuevas ideas acerca de un tema sobre el cual todavía hacen falta muchos enfoques para alcanzar el nivel de investigación y conocimiento que se posee de muchas otras áreas sanitarias.

Para la adecuada consecución del este estudio, se hacen necesarios unos objetivos claros que sirvan de guía en el transcurso de la investigación, así como un conocimiento exhaustivo de las limitaciones que lo definen.

3. Alcance del trabajo fin de grado

Es importante tener un conocimiento adecuado del marco que delimita el trabajo, pues permite la correcta definición de sus objetivos e hipótesis, estableciendo la meta que se pretende conseguir y en cuya consecución enfocar todos los recursos disponibles. Además, es necesario el entendimiento de las restricciones que acotan la investigación, así como la aclaración de las asunciones necesarias para que el estudio cobre sentido.

3.1 Objetivos

En el siguiente trabajo se pretende estudiar el estado de la salud mental en el ámbito tecnológico, deteniéndose en el impacto de la pandemia sufrida en 2020 en la visibilización y reducción del estigma que acompaña a la misma. Previo a su realización, se han definido los siguientes objetivos:

- Obtener una visión general del estado de la salud mental en los trabajadores del sector tecnológico.
- Plantear una hipótesis clara y suficientemente precisa.
- Demostrar mediante la prueba de Chow la existencia o carencia de un cambio estructural en 2020 en la predisposición al diálogo acerca de salud mental en entornos cercanos.
- Analizar el nivel de ajuste de los modelos plantados con las métricas pertinentes
- Interpretar los resultados obtenidos para poder rechazar o aceptar la hipótesis nula planteada con la suficiente significancia estadística.

3.2 Hipótesis

- Hipótesis nula: la prueba de Chow no demostrará ningún cambio estructural en los datos, por lo que un único modelo se podrá utilizar para la serie temporal completa.
- Hipótesis alternativa: el experimento revelará un definido cambio a partir del año 2020 que implicará que la pandemia supuso un cambio estructural.

3.3 Restricciones

- Restricción 1: dado el tiempo, recursos económicos y sensibilidad de los datos a tratar, no es posible la creación de una base de datos con una encuesta propia.
- Restricción 2: Las bases de datos públicos en materia de salud mental supone una limitación dada su estructura y escasez. En ocasiones, existen campos interesantes no estudiados por falta de variables específicas.
- Restricción 3: Al ser un tema en el que no se ha profundizado en el doble grado que he estudiado, mi conocimiento en lo relativo a los cambios estructurales y el test de Chow es reducido, lo cual puede limitar el alcance del trabajo.
- Restricción 4: El tiempo dedicado a la realización de este trabajo ha estado afectado por la asistencia obligatoria a clases y prácticas curriculares a tiempo completo, así como el cumplimiento de los plazos de entrega de los trabajos fin de grado. Esto disminuye el número de horas disponibles para dedicar a la búsqueda de información y desarrollo de la propuesta.

3.4 Asunciones

- Asunción 1: El número de años de los cuales se disponen datos es suficiente para llevar a cabo el experimento.
- Asunción 2: Los datos utilizados son fiables y no están altamente sesgados.
- Asunción 3: En caso de existir un cambio estructural en 2020, su causante será la pandemia de coronavirus producida ese año.

4. Marco teórico

Como se menciona anteriormente, la salud mental comprende una gran variedad de patologías y afecciones, además de ser un campo de difícil estudio y tratamiento dada su complejidad y falta de tangibilidad. Existen muchas asociaciones y organizaciones no gubernamentales cuyo objetivo es aportar información y apoyo a las personas afectadas por este tipo de trastornos, así como visibilizar su situación y luchar por una mayor investigación y destinación de fondos. Ejemplos en España son la asociación Diversamente o la fundación Soycomotu. Algunas de estas instituciones están centradas en grupos poblacionales concretos, como es el caso de OSMI (siglas de Open Source Mental Illness) Mental Health, la entidad de la cual provienen los datos utilizados en el presente trabajo.

4.1 OSMI Mental Health

OSMI Mental Health es una organización sin ánimo de lucro enfocada en la salud mental en el sector tecnológico. Su cometido es tanto sensibilizar a la población acerca de esta cuestión como ofrecer recursos a las personas pertenecientes al campo tecnológico que sufran algún tipo de enfermedad o trastorno mental, así como síntomas de estos. Fue fundada en 2013 a raíz de las declaraciones de Edward Finkler, un desarrollador que comenzó a hablar públicamente de salud mental en el entorno tecnológico a partir de su propia experiencia (*About OSMI :: Open Sourcing Mental Health - Changing How We Talk About Mental Health In The Tech Community - Stronger Than Fear*, s. f.). Sus palabras generaron un movimiento en esta comunidad del que surgió OSMI Metal Health. De la labor de esta organización, es especialmente relevante para este trabajo una encuesta pública que OSMI emite cada año desde hace al menos 6. Este cuestionario contiene preguntas acerca de la salud mental de los profesionales y la forma en que lo abordan tanto en su entorno laboral como personal. Los resultados obtenidos son publicados para promover su estudio y generar una mayor divulgación que conciencie acerca de la situación con respecto a este tema en los profesionales dedicados a la tecnología. Gracias a la disponibilidad de estos datos, se utilizan en este trabajo las respuestas reales extraídas de las encuestas mencionadas, lo cual permite estudiar la situación en este sector en un rango de cinco años. Se han seleccionado los datos de los años 2017 a 2021 porque es particularmente interesante estudiar el impacto de la reciente pandemia mundial de coronavirus. No es descabellado

pensar que en 2020 suceda un cambio estructural que altere las variables de estos datos y para comprobarlo se utiliza una prueba denominada test de Chow.

4.2 Prueba de Chow

Gregory Chow es un econométrico chino que ha hecho importantes aportaciones a lo largo de su carrera entre las cuales destaca la prueba de Chow (American Economic Association, s. f.). Esta prueba es mundialmente conocida y se utiliza para demostrar si a un conjunto de datos correspondiente a una serie temporal se le ajusta mejor una o varias líneas de regresión. Es decir, pretende comprobar si existe un punto de inflexión en algún momento que afecte paralelamente a todas las variables, dividiendo así los datos en subgrupos claros. Ese momento de cambio suele deberse a factores externos con un gran impacto en múltiples campos, como pueden ser desastres naturales, crisis económicas o guerras, aunque, dependiendo de los datos, puede ser causado por hitos menores altamente relacionados con las variables estudiadas. Un ejemplo de este último caso sería la difusión de una fotografía comprometedor (hito) impactando en el apoyo y opinión pública acerca de una celebridad. Esta prueba es comúnmente utilizada para la demostración de cambios estructurales (alteraciones generales que impactan de manera transversal una economía) aunque puede usarse en otro tipo de situaciones. Para estudiar la ocurrencia de cambios estructurales existen otro tipo de tests, como el de Supremum Wald, pero se decide optar por la prueba de Chow debido a su combinación de simplicidad, popularidad y eficiencia.

En el caso de este trabajo, se pretende comprobar si en el año 2020, debido a la pandemia, los datos extraídos de la información pública de OSMI Mental Health sufren un cambio estructural. Siendo esto cierto, se rechazaría la hipótesis nula planteada previamente, demostrando que en efecto este fenómeno sucede y son necesarios dos modelos de regresión para explicar de manera más precisa el dataset existente.

Para calcular el estadístico de Chow que determina si un acontecimiento supone un cambio estructural significativo en los datos, se utiliza la siguiente fórmula:

$$F_{\text{chow}} = \frac{(RSS_{\text{full}} - (RSS_1 + RSS_2))/k}{(RSS_1 + RSS_2)/(n - 2k)}$$

Ecuación 1: estadístico F para prueba de Chow

Donde:

- RSS_{full} : Suma de los residuos al cuadrado para el modelo completo.

- RSS1: Suma de los residuos al cuadrado para el primer modelo.
- RSS2: Suma de los residuos al cuadrado para el segundo modelo.
- k: Número de características (predictores) en la matriz de características.
- n: Tamaño total de la muestra (número de observaciones).

Si el valor obtenido tras sustituir todos los componentes de la fórmula es considerablemente mayor a 1, la prueba habrá demostrado que las diferencias entre los modelos son significativas. Es decir, que, dentro de un modelo, la división en dos subgrupos con dos modelos diferentes permitiría un mayor ajuste y reflejaría mejor los datos reales.

Anterior al experimento en el que se utiliza este tipo de prueba, es importante conocer en profundidad los datos con los que se trabaja, así como la transformación de las variables necesarias para llevar a cabo con éxito el test planteado.

4.3 Análisis previo de los datos

4.3.1 Análisis general

En este trabajo se utilizan 5 archivos csv que contienen los datos obtenidos a partir de la encuesta anual sobre salud mental que emite la organización no gubernamental OSMI. Los datos pertenecen a los años 2017, 2018, 2019, 2020 y 2021 por lo que su análisis permite una visión clara de la situación pre-pandemia y post-pandemia.

Para comenzar, se analiza cada csv de forma independiente extrayendo la información más útil de cada uno de ellos y descartando aquellas variables que no se necesitan para un manejo más fácil y rápido del dataset. Python es la herramienta seleccionada para llevar a cabo este estudio. Dada su versatilidad e intuitiva interfaz, se considera idónea para realizar la tarea planteada, permitiendo plasmar los resultados en gráficas para una mejor interpretabilidad. Después, se estudia la evolución histórica de los parámetros más significativos mediante la cual se puede observar si, a priori, se rechaza la hipótesis nula. Durante todo este proceso se obtiene una visión general del estado de la salud mental en el ámbito tecnológico y la apertura a compartir este tipo de problemas tanto en el trabajo como en el ámbito familiar.

Como primer paso, se comprueba el tamaño de cada dataset (tanto filas como columnas) y se eliminan las variables con valores nulos o N/A en más de una 90% de las respuestas.

A continuación, se muestran todas las variables que quedan en el dataset y se puede identificar, mediante sus claves (que son las preguntas de la encuesta), cuáles son aquellas que pueden ofrecer más información y cuya evolución en los cinco años estudiados puede ser de mayor interés. Se reduce el tamaño del dataframe prescindiendo de las variables no deseadas y se renombran las variables indicando claves más cortas e intuitivas que agilicen el procesado de los datos, evitando tener que referirse a la pregunta completa a la cual corresponde cada parámetro cada vez que se quiere utilizar esa variable. Tras este proceso, las variables de las que dispone cada uno de los dataframes son:

- autonomo: variable cualitativa que contiene las respuestas a la pregunta ‘Are you self-employed?’
- num_emp: variable que contiene las respuestas a la pregunta ‘How many employees does your company or organization have?’
- empresa_tec: variable cualitativa dicotómica que contiene las respuestas a la pregunta ‘Is your primary role within your company related to tech/IT?’
- puesto_tec: variable que contiene las respuestas a la pregunta ‘Does your employer provide mental health benefits as part of healthcare coverage?’
- beneficio: variable cualitativa que contiene las respuestas a la pregunta ‘Do you know the options for mental health care available under your employer-provided health coverage?’
- mencion_sm_empleador: variable que contiene las respuestas a la pregunta ‘Has your employer ever formally discussed mental health (for example, as part of a wellness campaign or other official communication)?’
- info_opciones_sm: variable que contiene las respuestas a la pregunta ‘Does your employer offer resources to learn more about mental health disorders and options for seeking help?’
- oferta_recursos_sm: variable que contiene las respuestas a la pregunta ‘Does your employer offer resources to learn more about mental health disorders and options for seeking help?’

- conv_sm_empleador: variable que contiene las respuestas a la pregunta ‘Have you ever discussed your mental health with your employer?’
- conv_sm_comp: variable que contiene las respuestas a la pregunta ‘Have you ever discussed your mental health with coworkers?’
- imp_saludfisica: variable que contiene las respuestas a la pregunta ‘Overall, how much importance does your employer place on physical health?’
- imp_saludmental: variable que contiene las respuestas a la pregunta ‘Overall, how much importance does your employer place on mental health?’
- conv_sm_antiguo_emp: variable que contiene las respuestas a la pregunta ‘Did you ever discuss your mental health with your previous employer?’
- conv_sm_antiguo_comp: variable que contiene las respuestas a la pregunta ‘Did you ever discuss your mental health with a previous coworker(s)?’
- problema_sm_act: variable que contiene las respuestas a la pregunta ‘Do you *currently* have a mental health disorder?’
- problema_sm_pasado: variable que contiene las respuestas a la pregunta ‘Have you had a mental health disorder in the past?’
- tratamiento_sm: variable que contiene las respuestas a la pregunta ‘Have you ever sought treatment for a mental health disorder from a mental health professional?’
- historial_familiar: variable que contiene las respuestas a la pregunta ‘Do you have a family history of mental illness?’
- entorno_familiar: variable que contiene las respuestas a la pregunta ‘How willing would you be to share with friends and family that you have a mental illness?’
- edad: variable que contiene las respuestas a la pregunta ‘What is your age?’
- sexo: variable que contiene las respuestas a la pregunta ‘What is your gender?’
- pais: variable que contiene las respuestas a la pregunta ‘What country do you *live* in?’

Posteriormente se tratan algunas de las variables que se utilizarán en el análisis para facilitar su manejo.

Variable sexo – al visualizar los valores que toma esta variable, se obtienen multitud de respuestas diferentes. Para facilitar el posterior análisis de los datos se agrupan las respuestas en tres grupos: Hombre, Mujer y Otro.

Variable edad – utilizando la función *.describe*, se visualiza el número de observaciones, su media, su desviación estándar, el valor mínimo de los datos, su percentil 25, su mediana, su percentil 75, el valor máximo obtenido, el nombre de la variable y su tipo (en este caso, float). Con esta función se puede tener una rápida visión general de los datos que permite identificar posibles irregularidades. Este es el caso en los datos de los años 2019 y 2021. En el primero, se obtiene un valor mínimo de cero. Esto supondría que alguien con 0 años de edad ha respondido la encuesta, lo cual es imposible. Para solventar este problema, se sustituyen aquellos valores menores a 18 y mayores a 65 con el valor de la media. En los datos del año 2021, se obtiene un valor máximo de los datos de 223, por lo que se aplica la misma regla mencionada para regularizar los datos y evitar desviaciones causadas por outliers.

Variable puesto_tec – tras visualizar los sumatorios de las observaciones que toman valor 1 (afirmativo) y 0 (negativo), se decide prescindir de aquellas observaciones que hayan tomado valor 0 en esta variable por ser un número demasiado reducido de individuos. Se reduce así el dataframe únicamente a las respuestas de los individuos que ostentan un puesto relacionado con la tecnología, lo cual permite que los datos sean más homogéneos y comparables entre sí.

Después de haber comprobado que las variables que se necesitan para un primer análisis general del dataset contienen datos coherentes y útiles, se hace una aproximación general que ofrece una imagen a grandes rasgos de la situación.

En primer lugar, se genera un gráfico de tarta en el que se visualiza el porcentaje de respuestas de cada uno de los cuatro tipos existentes (Yes, No, I don't know y Possibly) a la pregunta “Do you currently have a mental health disorder?”. En el caso de 2021 la situación era la siguiente:

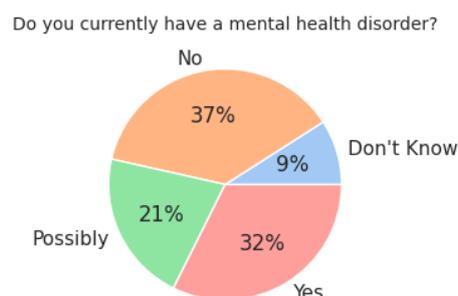


Figura 1: Porcentaje de profesionales en puestos tecnológicos que padece un trastorno de salud mental.

Aunque se observa una mayoría de respuestas negativas ante esta pregunta, la diferencia con las respuestas positivas no es grande y son muchos los individuos que dudan acerca del estado de su salud mental.

Utilizando la función Facetgrid de la librería seaborn se obtiene el siguiente gráfico para mostrar la distribución de las respuestas positivas por sexo y edad. La situación de 2021 era la siguiente:

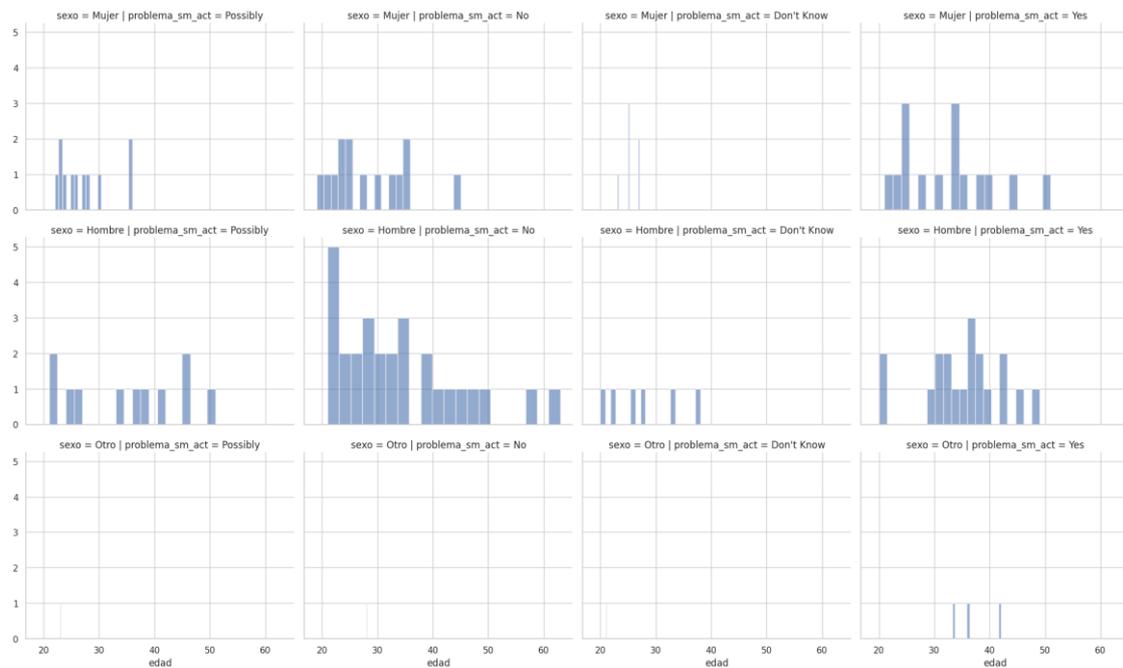


Figura 2: Distribución por sexo y edad de los individuos que padecen un trastorno de salud mental.

Se observa que la mayoría de las respuestas positivas se acumulan en hombres de mediana edad. Sin embargo, este dato no es significativo puesto que en el sector tecnológico existe una amplia mayoría de trabajadores hombres de entre 30 y 50 años de edad.

Para comprender el porcentaje real de casos positivos de problemas de salud mental distribuido por sexo, lo cual aporta una mayor información, se obtiene el siguiente gráfico:

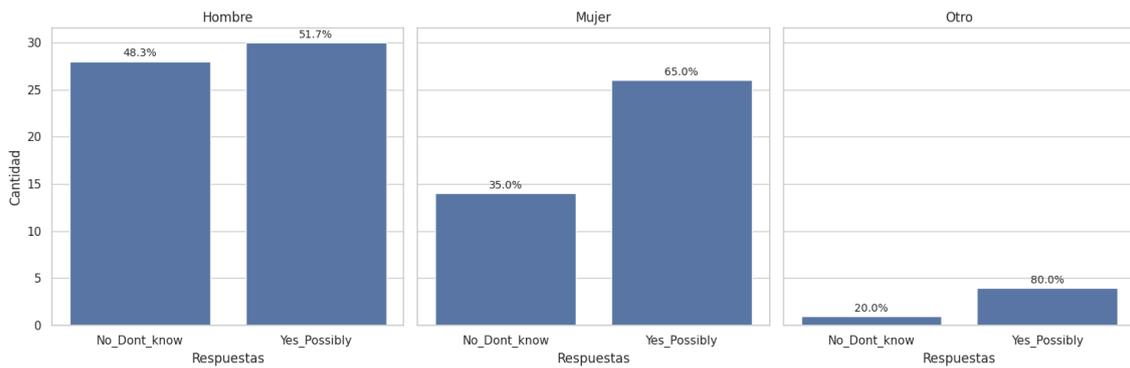


Figura 3: Porcentaje de personas con un trastorno de salud mental dividido por sexo.

Se puede observar que, aunque un menor número de mujeres han respondido a la encuesta, habitualmente un mayor porcentaje de ellas sufren trastornos de salud mental.

4.3.2 Salud mental en el ámbito laboral

Para analizar el nivel de implicación de las empresas en materia de salud mental, la accesibilidad a recursos y la facilidad de discusión de este tema, se filtran primero algunas variables para tener dataframes lo más útiles posibles.

Variable autónomo – se puede observar que el número de respuestas con valor 1 es muy escaso, lo cual significa que la gran mayoría de los encuestados no son autónomos. Dado que en este estudio se pretende medir el impacto de la pandemia de coronavirus en el ámbito empresarial, se eliminan aquellas observaciones de individuos que hayan manifestado que son autónomos.

Variable empresa_tec – se comprueba que el número de observaciones con valor 0 en esta variable es muy reducido. Para obtener una muestra de individuos cuyas empresas son similares, se eliminan aquellas las respuestas de los encuestados que no trabajan en una empresa tecnológica.

Después se visualizan los datos de cada dataset para obtener respuesta a tres preguntas principales:

- 1) ¿Cuál es el porcentaje de cobertura de los trabajadores tecnológicos en sus empresas en materia de salud mental?

Se crea para cada año un gráfico de tarta que indica el porcentaje de respuestas entre “Yes,” , “No”, “I don’t know” y “Not eligible for coverage/NA” (Si, No, No lo sé y No aplica). La situación en 2021 era la siguiente:

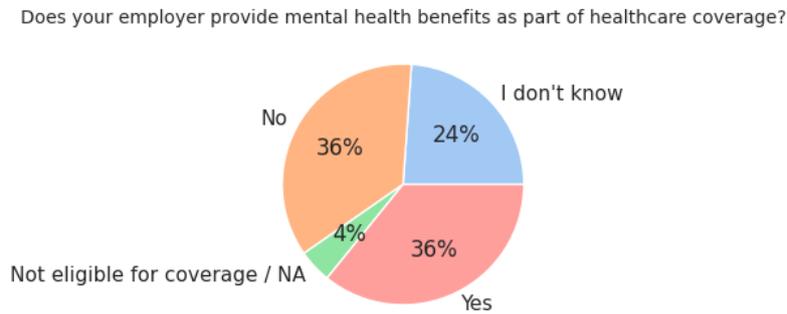


Figura 4: Porcentaje de cobertura de salud mental de profesionales tecnológicos.

Sorprendentemente, el porcentaje de personas que creen estar cubiertas por su seguro en este tipo de casos disminuye a lo largo de los años, siendo un 54% las respuestas afirmativas (“Yes”) en 2017 y alcanzando un 36% en 2021.

2) ¿Conocen los trabajadores las opciones en materia de salud mental recogidas en su cobertura sanitaria?

Existe una gran variedad de recursos que las empresas pueden disponer al servicio de los trabajadores. Desde psicólogos y grupos de apoyo, hasta cobertura de medicamentos. A la pregunta acerca del conocimiento de las opciones disponibles en sus empresas, las respuestas eran las siguientes en 2021:

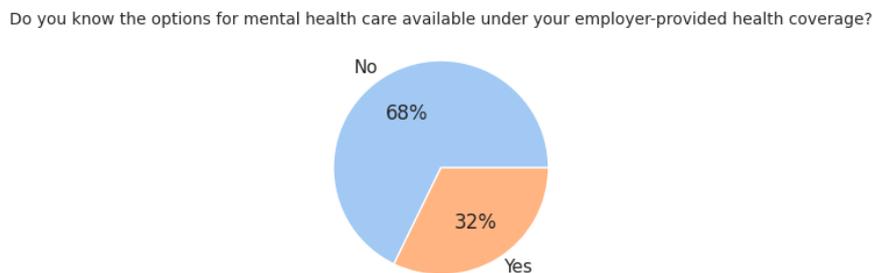


Figura 5: Porcentaje de conocimiento de las opciones dentro de su seguro médico de la salud mental de profesionales tecnológicos.

Como ocurre con la pregunta anterior, los resultados sorprenden. Es lógico pensar que, con la creciente reducción del estigma y aumento de visibilidad de la salud mental, las empresas la incluyan en sus seguros e informen a los trabajadores de sus

opciones. Sin embargo, observamos que el porcentaje de individuos que desconocen las diferentes posibilidades en esta materia aumenta a lo largo de los años, siendo un 55% en 2017 y alcanzando un 68% en 2021.

3) ¿Es frecuente que se mencione la salud mental por parte de las empresas?

Las empresas pueden hacer llegar mensajes relacionados con la salud mental en campañas de visibilidad y concienciación, para la promoción del bienestar de los trabajadores, o simplemente como iniciativa del propio departamento o equipo. Ante esta situación las respuestas de los individuos encuestados en 2021 eran:

Has your employer ever formally discussed mental health (for example, as part of a wellness campaign or other official communication)?

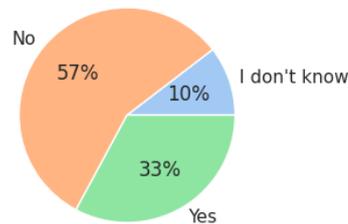


Figura 6: Porcentaje de empresas que han mencionado la salud mental a sus empleados.

En este caso, la evolución de las respuestas sí va acorde con una tendencia positiva hacia un menor estigma. El número de respuestas “No” en 2017 alcanzaba el 66%, habiéndose reducido en 10 puntos porcentuales en 2021.

4) ¿Hablan los trabajadores de su salud mental con sus managers?

Por último, se considera interesante valorar los datos relativos a las conversaciones de los trabajadores con sus managers. Es importante saber si los mensajes de las empresas y asociaciones se hacen efectivos abriendo el debate en el día a día y dejando atrás el histórico tabú que acompaña a la salud mental. Los resultados obtenidos en 2021 fueron:

Have you ever discussed your mental health with your employer?

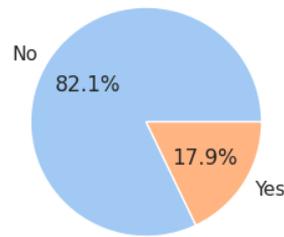


Figura 7: Porcentaje de profesionales tecnológicos que han discutido su salud mental con sus empleadores.

De nuevo, una evolución poco favorable. En 2017 los encuestados que respondía “No” eran un 65%, llegando a ser un 82% en 2021. Para un estudio más exhaustivo que muestre si existen grandes diferencias en la propensión a hablar con su manager dependiendo del sexo de la persona se genera el siguiente gráfico:

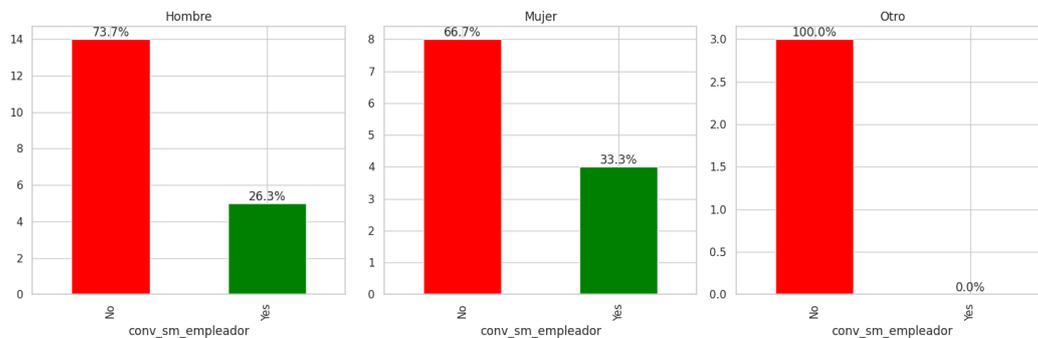


Figura 8: Porcentaje de discusión de la salud mental con el empleador distribuido por sexo.

Como se puede observar, las diferencias en los porcentajes de hombres y mujeres no son representativas. Aunque las mujeres presentan en cierto modo mayor grado de facilidad para tratar temas de salud mental con sus managers, es común a todos los sexos una aversión a este tipo de conversaciones.

En los datasets utilizados también se dispone de una variable que registra si los individuos encuestados habían discutido alguna vez la salud mental con sus empleadores anteriores. Para comprobar si aquellas personas que habían mantenido antes una conversación eran más propensas a dialogar con sus empleadores actuales sobre su salud mental, se crea una tabla de contingencia que contiene los campos “Si” y “No” para las variables `conv_sm_empleador` y `conv_sm_antiguo_emp`. Sin embargo, los resultados son poco concluyentes dado que las respuestas positivas en ambas variables son muy escasas y la mayoría de los datos se concentran en la respuesta “No”. Como ejemplo, esta es la matriz obtenida para el año 2021:

conv_sm_antiguo_emp	No	Yes
conv_sm_empleador		
No	31	3
Yes	5	3

Figura 9: Tabla de contingencia con las variables conv_sm_antiguo_emp y conv_sm_emp.

En este genérico análisis, los datos muestran una alarmante realidad: la, visibilización y adecuada oferta de recursos ante los problemas de salud mental, los proyectos se quedan en las palabras de las empresas tecnológicas. Las cifras expuestas reflejan que, aunque las empresas tecnológicas sí hacen un esfuerzo por tener más presente la salud mental, no se aprecian grandes mejoras en cuanto a medidas efectivas que afecten de manera directa a los individuos de este sector como pueden ser conversaciones reales con sus managers o una mayor cobertura en sus seguros sanitarios.

4.3.3 Salud mental en el ámbito personal

En cuanto a la predisposición a discutir abiertamente problemas relacionados con la salud mental en el entorno familiar (recogido en la variable bajo la pregunta “How willing would you be to share with friends and family that you have a mental illness?”), se han encontrado los siguientes resultados:



Figura 11: Disposición al debate de la salud mental con el entorno cercano en 2017.

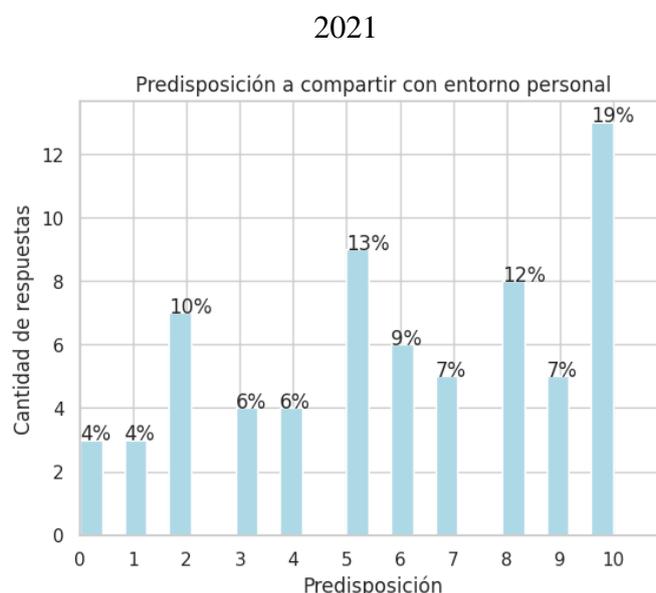


Figura 10: Disposición al debate de la salud mental con el entorno cercano en 2021.

Analizando los dos gráficos encima de estas líneas, la tendencia no es clara. Si bien el porcentaje de respuestas de algunos valores como el 10 aumenta considerablemente, también lo hacen algunos otros como el 2 o el 7 por lo que extraer conclusiones es difícil. Para una mejor visualización se decide representar los valores en un gráfico de área que recoja el porcentaje de respuestas de cada valor en la serie temporal. Se obtiene el siguiente gráfico:

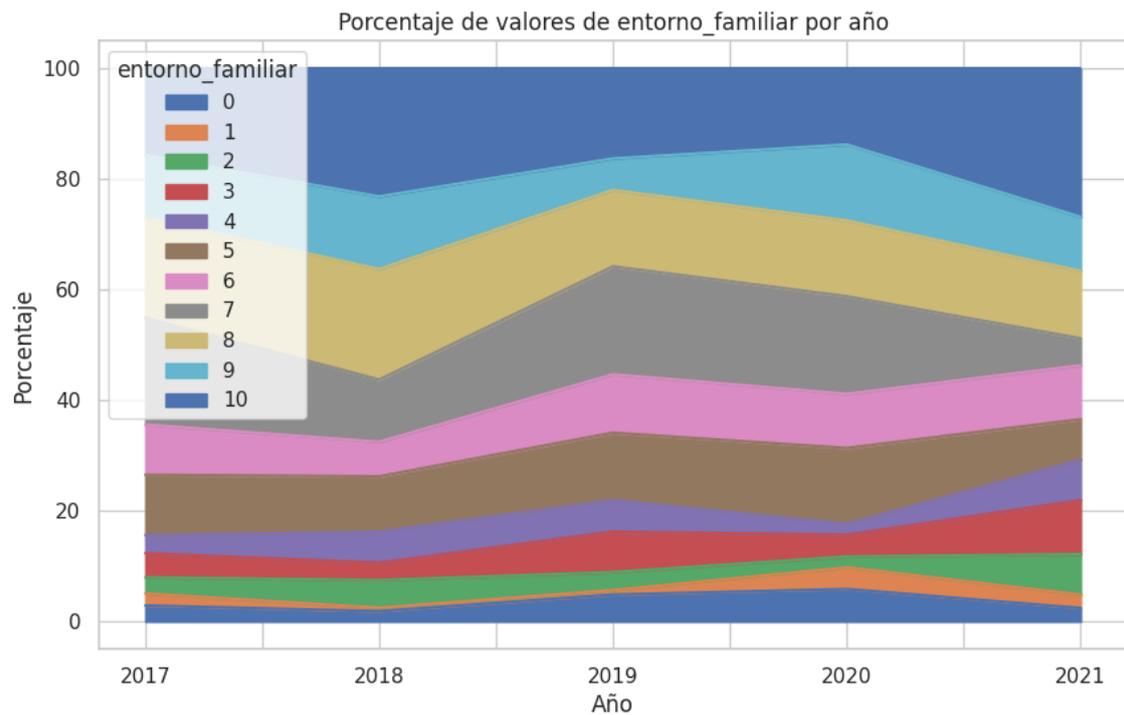


Figura 12: Evolución de la variable entorno_familiar. 2017 a 2021.

En el gráfico que recoge las respuestas de todos los años, se aprecia un claro aumento de respuestas con el número 10, lo cual implica que están completamente dispuestas a debatir su salud mental con su entorno cercano. Aun así, el resto de las respuestas fluctúa bastante, por lo que de nuevo las conclusiones no son claras.

Este análisis general de los datos permite obtener una visión global de la situación de la salud mental en el sector tecnológico, pero no ofrece un argumento claro que permita rechazar o aceptar la hipótesis nula planteada al inicio del trabajo. Para poder rechazar con un alto nivel de confianza (se ha elegido un 95%) la hipótesis nula, se realiza un experimento con una prueba de Chow, el cual ofrece unas conclusiones más definidas y estadísticamente robustas.

5. Experimento

En este trabajo se pretende comprobar si efectivamente existe un menor estigma alrededor de la salud mental después de la pandemia de coronavirus en 2020. Para ello se utiliza la prueba de Chow. Este test nos permite rechazar o no la hipótesis nula planteada anteriormente.

La implementación de este experimento se lleva a cabo en Python y en el siguiente link [TFG-Belen-Mata en github](#) se puede encontrar el código necesario para su desarrollo.

Para realizar el experimento, se seleccionan cuatro de las variables existentes en los dataframes de los años 2017, 2018, 2019, 2020 y 2021. Éstas son:

- ‘beneficio’: indica si los individuos disponen de beneficios como parte de su cobertura médica en materia de salud mental. Las respuestas obtenidas son “Si”, “No”, “No lo sé” y “No aplica”. Para poder utilizarla se transforman en 0 y 1 utilizando la función `.map`
- ‘mencion_sm_empleador’: indica si la empresa de la persona que responde la encuesta ha mencionado alguna vez la salud mental, bien sea como parte de una campaña de visibilización u otro tipo de evento. Toma los valores “Si”, “No” y “No lo sé”, que también han sido transformados en 0 y 1 para poder usar esta variable en el modelo.
- ‘problema_sm_act’: indica si el encuestado sufre actualmente una enfermedad de salud mental. A las respuestas obtenidas, se les han asignado los siguientes valores para facilitar su uso: “Si” = 1, “No” = 0, “No lo sé” = 0.5, “Posiblemente” = 0.75
- ‘entorno_familiar’: indica la predisposición por parte del individuo encuestado a compartir con su entorno familiar que padece un trastorno de salud mental. No necesita transformación en los valores que toman las observaciones que recoge porque son números del 1 al 10.

Después de transformar las variables que se decide utilizar para el experimento, se eliminan los valores nulos mediante el uso de `.dropna`.

A continuación, se dividen los datos en dos grupos para poder verificar si existe un cambio estructural causado por la pandemia o no. Se separan los datos en previos a 2020 y posteriores a 2020.

Ya segregados los datos, se define como variable dependiente para ambos grupos la variable “entorno_familiar” mencionada anteriormente. La ecuación planteada para la cual se buscan los coeficientes de regresión óptimos tendrá la siguiente forma:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$$

Ecuación 2: regresión lineal múltiple

Donde:

- Y es la variable dependiente.
- X_1, X_2, \dots, X_k son las variables explicativas.
- $\beta_0, \beta_1, \dots, \beta_k$ son los coeficientes de regresión.
- ε es el término de error

Una vez estos pasos están completos, se utiliza el Método de Cuadrados Ordinarios (OLS) para encontrar los parámetros desconocidos en los dos modelos resultantes de la división de los datos. Este método comúnmente utiliza la regresión lineal y busca los parámetros que resulten en una línea recta que mejor se ajuste a los datos (Zdaniuk, B, 2014). Para hallar los mejores coeficientes de regresión, este método minimiza la suma de los cuadrados de las diferencias entre los valores observados y los predichos por la línea de regresión.

Después de ajustar los dos modelos de regresión a los cuales hemos denominado modelo1 (con los datos previos a 2020) y modelo2 (con los datos posteriores a 2020), se utiliza el mismo método para ajustar un modelo a la serie temporal completa, conteniendo los datos desde 2017 a 2021. A este modelo en el código de Python utilizado se le da el nombre de `model_full`.

Seguidamente, se someten los modelos al test o prueba de Chow, pero para ello, es necesario calcular su suma de cuadrados de los residuos (SCE) de cada uno de los modelos obtenidos. Esta representa la variabilidad de la variable dependiente que no se explica mediante el modelo de regresión (es decir, mediante las variables independientes) y se suele representar como: $SCE = \sum e^2$, donde (e) representa las diferencias entre el valor observado y el estimado por el modelo para cada valor de todo el conjunto de datos (Sanjuán, 2022). Los valores obtenidos, junto con el número de parámetros de la regresión, ofrecen toda la información necesaria para llevar a cabo la prueba mencionada anteriormente.

Tras aplicar la fórmula perteneciente al test de Chow, se obtiene el estadístico F. Esta métrica resulta de la comparación de la suma de cuadrados de los residuos entre un modelo (el `full_model`) con coeficientes iguales, en este caso, antes y después de 2020; y otro modelo (compuesto por `modelo1` y `modelo2`) que tiene coeficientes diferentes antes y después de 2020. La cercanía o lejanía del valor obtenido a 1 permite determinar si existe o no una diferencia significativa entre los coeficientes de los modelos de regresión planteados. En el contexto de los datos estudiados en el presente trabajo, el estadístico F derivado de la prueba de Chow toma un valor de 2.90, lo cual demuestra que existe una diferencia significativa en el ajuste del modelo que engloba todo el conjunto de datos si éste es dividido en dos modelos, uno que contenga los datos previos a 2020 y otro con aquellos posteriores a ese año.

Para determinar la significancia estadística de la prueba realizada se utiliza el p valor. Esta métrica indica la probabilidad de obtener un valor igual o mayor que el resultado observado si la hipótesis nula fuese cierta (Beers, 2024). En este trabajo se establece un nivel de confianza del 95%. Esto quiere decir que $\alpha=0.05$ (nivel de significación), es decir, que si el p valor obtenido es inferior a 0.05, se puede rechazar con evidencia estadística la hipótesis nula. El p valor asociado al estadístico F obtenido en este trabajo es de 0.03, lo cual implica que la hipótesis nula planteada al inicio de la investigación es rechazada. Dadas las variables utilizadas, la interpretación de estos valores sugiere que existe la evidencia suficiente como para afirmar que, en efecto, el año 2020 supuso un cambio significativo en la predisposición de las personas con enfermedades mentales a compartirlo con su entorno familiar, y por tanto, en la normalización y reducción del estigma asociado a la salud mental.

5.1 Bondad del ajuste

Existen multitud de métricas que permiten determinar en qué medida un modelo se ajusta o no a los datos reales estudiados. A ese nivel de acoplamiento se le denomina bondad del ajuste.

En este trabajo se utilizan cinco métricas diferentes para asegurar que los modelos son precisos, y por tanto, fiables. Lo que se pretende comprobar, es que el `modelo1` y `modelo2`, son más representativos que el `full_model`, sugiriendo así, que la división de los datos teniendo como punto de corte el año 2020, está justificada; y que la

utilización de dos modelos en vez de uno es beneficiosa en términos de representación de los datos reales.

R cuadrado

El también llamado coeficiente de determinación es un estadístico que mide la proporción de varianza en la variable dependiente (en este caso 'entorno_familiar) explicada por las variables dependientes (en este caso 'beneficio', 'mencion_sm_empleador' y 'problema_sm_act') en un modelo de regresión (Taylor, 2023). Es calculado con la siguiente fórmula:

$$R^2 = 1 - \frac{SSR}{SST}$$

Ecuación 3: R cuadrado

Donde:

- SSR es la suma de los residuos cuadrados (explicada anteriormente porque es necesario su cálculo para realizar la prueba de Chow)
- SST es la suma total de los cuadrados (representa la variabilidad de los datos con respecto a la media de todos los valores observados).

El resultado obtenido puede tomar valores entre 0 y 1, siendo más representativo el modelo cuanto más se acerque a 1 el estadístico.

R cuadrado ajustado

Esta métrica surge para evitar uno de los problemas del estadístico anterior conocido como "sobrestimación del modelo": R cuadrado no penaliza la incorporación de variables no explicativas en el modelo, es decir, cuantas más variables se añadan al modelo, mayor será el R cuadrado obtenido, independientemente de que esas variables sean representativas o no (Calzada, 2019). Para su cálculo se utiliza la siguiente fórmula:

$$R_{\text{ajustado}}^2 = 1 - \left(\frac{(1 - R^2) \cdot (n - 1)}{n - k - 1} \right)$$

Ecuación 4: R cuadrado ajustado

Donde:

- (R^2) es el coeficiente de determinación sin ajustar.
- (n) es el número total de observaciones en la muestra.
- (k) es el número de variables independientes

Se puede apreciar que la fórmula penaliza la inserción de variables independientes en el modelo al incluir k en el denominador. Esta característica, la convierte en una métrica más fiable que la anterior para comprobar el ajuste de un modelo a los datos reales utilizados. Asimismo, esta métrica permite la comparación de modelos con un número de variables independientes diferentes. Esta cualidad, aunque en este caso no aplica, facilita la comparación de modelos en multitud de estudios.

R cuadrado ajustado toma siempre valores inferiores a los de R cuadrado dada su condición de penalización y, al contrario que R cuadrado, puede ser negativo en determinadas ocasiones (Sanjuán, 2022a).

Log-verosimilitud

Esta métrica se utiliza para la comparación de modelos de regresión al ser maximizada para obtener los parámetros óptimos que multiplican a las variables independientes. Puede tomar valores desde menos infinito hasta más infinito y cuanto mayor sea su valor, mejor ajuste tienen el modelo (Bobbitt, 2024). A pesar de que este método posee unas propiedades matemáticas que lo hacen altamente fiable, su uso no está tan extendido debido, entre otras razones, a que requiere el mismo número de variables en los modelos que compara, lo cual impide su uso en muchas investigaciones.

AIC

El criterio de información de Akaike (AIC) es otra de las métricas utilizadas para comparar y seleccionar modelos en función de su representatividad. Una característica positiva de este método es que considera un balance entre el ajuste del modelo y su complejidad, evaluando los modelos de forma relativa y penalizando el sobreajuste (Benites, 2021).

Es calculado con la siguiente fórmula:

$$AIC = -2 \log(\mathcal{L}) + 2k$$

Ecuación 5: AIC

Donde:

- (k) es el número de parámetros en el modelo estadístico.
- (L) es el máximo valor de la función de verosimilitud para el modelo estimado.

En los resultados, el modelo preferido será aquel con un valor AIC más bajo, ya que indicará que se ajusta bien a los datos sin ser demasiado complejo.

BIC

El criterio de información de Bayes (BIC) o criterio de Schwarz es el último de los métodos considerado para la comparación de los modelos obtenidos en este trabajo. Fue publicado en 1987 y está íntimamente relacionado con el criterio de información de Akaike mencionado anteriormente, diferenciándose en que aplica una mayor punición al sobreajuste (Benites, 2022). Para calcular este criterio se utiliza la siguiente fórmula:

$$BIC = -2\log(L) + k\log(n)$$

Ecuación 6: BIC

Donde:

- (L) es la verosimilitud máxima del modelo.
- (k) es el número de parámetros estimados.
- (n) es el tamaño de la muestra.

De nuevo, un valor inferior en esta métrica indica un mejor modelo, más ajustado, pero sin llegar este a ser excesivo.

Una vez comprendidas las métricas de bondad del ajuste que se desea estudiar, utilizando la función `.summary()` se obtienen los datos específicos para los modelos creados en el presente estudio. Los valores resultantes son los siguientes:

OLS Regression Results

```

=====
Dep. Variable:    entorno_familiar    R-squared:        0.080
Model:           OLS                 Adj. R-squared:   0.075
Method:          Least Squares       F-statistic:      16.07
Date:            Wed, 12 Jun 2024     Prob (F-statistic): 5.00e-10
Time:            18:03:58            Log-Likelihood:   -1301.6
No. Observations: 558                AIC:              2611.
Df Residuals:    554                BIC:              2629.
Df Model:        3
Covariance Type: nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	5.6656	0.198	28.635	0.000	5.277	6.054
beneficio	0.8409	0.237	3.555	0.000	0.376	1.305
mencion_sm_empleador	0.2798	0.241	1.160	0.246	-0.194	0.753
problema_sm_act	0.9409	0.222	4.248	0.000	0.506	1.376

```

=====
Omnibus:         42.954    Durbin-Watson:      2.073
Prob(Omnibus):   0.000    Jarque-Bera (JB):   51.346
Skew:            -0.741    Prob(JB):           7.09e-12
Kurtosis:        3.117    Cond. No.           3.75
=====

```

Figura 13: Resultados de la regresión model1

OLS Regression Results

```

=====
Dep. Variable:    entorno_familiar    R-squared:        0.170
Model:           OLS                 Adj. R-squared:   0.142
Method:          Least Squares       F-statistic:      6.022
Date:            Wed, 12 Jun 2024     Prob (F-statistic): 0.000886
Time:            18:04:01            Log-Likelihood:   -220.45
No. Observations: 92                AIC:              448.9
Df Residuals:    88                BIC:              459.0
Df Model:        3
Covariance Type: nonrobust
=====

```

	coef	std err	t	P> t	[0.025	0.975]
const	5.5816	0.475	11.745	0.000	4.637	6.526
beneficio	-0.0350	0.640	-0.055	0.957	-1.306	1.236
mencion_sm_empleador	-0.1961	0.630	-0.311	0.756	-1.447	1.055
problema_sm_act	2.4530	0.586	4.185	0.000	1.288	3.618

```

=====
Omnibus:         3.737    Durbin-Watson:      2.054
Prob(Omnibus):   0.154    Jarque-Bera (JB):   3.723
Skew:            -0.477    Prob(JB):           0.155
Kurtosis:        2.755    Cond. No.           3.68
=====

```

Figura 14: Resultados de la regresión model2

OLS Regression Results						
=====						
Dep. Variable:	entorno_familiar	R-squared:	0.085			
Model:	OLS	Adj. R-squared:	0.080			
Method:	Least Squares	F-statistic:	19.92			
Date:	Wed, 12 Jun 2024	Prob (F-statistic):	2.34e-12			
Time:	18:04:03	Log-Likelihood:	-1526.8			
No. Observations:	650	AIC:	3062.			
Df Residuals:	646	BIC:	3079.			
Df Model:	3					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

const	5.6360	0.182	30.923	0.000	5.278	5.994
beneficio	0.7105	0.221	3.210	0.001	0.276	1.145
mencion_sm_empleador	0.1838	0.222	0.826	0.409	-0.253	0.621
problema_sm_act	1.1673	0.207	5.632	0.000	0.760	1.574
=====						
Omnibus:	47.968	Durbin-Watson:	2.057			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	57.343			
Skew:	-0.727	Prob(JB):	3.53e-13			
Kurtosis:	3.074	Cond. No.	3.65			
=====						

Figura 15: Resultados de la regresión full_model

Dados los valores observados en las figuras expuestas sobre estas líneas, se estudian los métodos de bondad del ajuste previamente explicados:

R cuadrado

Este estadístico toma el valor 0.085 en el modelo full_model, mientras que en model1 es de 0.080 y en model2 es de 0.170, es decir, de 0.125 de media. Por tanto, aunque en el caso del primer modelo parcial no se supere el valor de la métrica del modelo completo, la media de los modelos subdivididos es considerablemente superior a la de full_model.

R cuadrado corregido

Este método de bondad del ajuste resulta en un valor de 0.080 para full_model, siendo 0.075 en model1 y 0.142 en model2. De nuevo, aunque en el caso de model no se supere el valor obtenido en full_model, la media de los dos modelos (1 y 2) es de 0.109, el R cuadrado corregido con dos modelos tomando el año 2020 como punto de separación de nuevo supera al de un único modelo para toda la serie temporal.

Log-verosimilitud o log-likelihood

El valor de esta métrica es de -1526.8 en `full_model`, -1301.6 en `modell` y -220.5 en `model2`. En este caso, ambos `modell` y `2` ofrecen un valor superior, demostrando un mejor ajuste del modelo a los datos.

AIC

`Full_model` tiene un valor en este criterio de 3062 mientras que `modell` tiene un valor de 2611 y `model2` de 448. Como en el método anterior, el mejor acoplamiento tras la división del modelo en dos diferentes es evidente, corroborando que esa decisión es acertada y apropiada.

BIC

En la última métrica comprobada, `full_model` tiene un valor de 3079, `modell` de 2629 y `model2` de 459. En esta comparación, los modelos `modell` y `model2` también demuestran un mejor ajuste a los datos reales.

Comparando todas las métricas de bondad del ajuste expuestas, se confirma en todos los casos que la división de los datos en previos y posteriores al año 2020 está justificada. La representación de los datos y ajuste de los modelos es mucho mayor si se utilizan dos modelos y no una única regresión para los datos de los datasets usados en este trabajo.

Como se puede comprobar en el código expuesto previamente que permite el desarrollo de este experimento, para hacer una última comprobación de la originalidad y excepcionalidad de los resultados obtenidos, se intenta llevar a cabo la prueba de Chow con una variable más (`conv_sm_empleador`). Sin embargo, el p valor en este caso no es inferior a 0.05 por lo que la hipótesis nula no se podría rechazar con un 95% de confianza. Esto demuestra de nuevo que el experimento realizado contiene las variables adecuadas y sus resultados no son fácilmente replicables.

6. Resultado

La realización de este experimento resulta verdaderamente fructífera y exitosa. Tras la selección de las variables adecuadas, se someten los datos a una prueba de Chow, que ofrece dos valores que dan completo sentido al experimento. El primero, un estadístico F de 2.90 que indica que el ajuste a los datos reales utilizando dos modelos (uno para los datos anteriores a 2020 y otro para los datos posteriores a 2020) es mayor que el ajuste proveniente de la utilización de un único modelo con el que representar todo el conjunto de datos. Para analizar la probabilidad de ocurrencia de la obtención de esta cifra, se utiliza el p valor. Con una confianza del 95%, el p valor debe ser inferior a 0.05 para demostrar que el F estadístico no es arbitrario sino un valor con una probabilidad de obtención reducida. En este trabajo, el p valor resultante tras la realización de la prueba de Chow es de 0.03, por lo que esto permite rechazar cómodamente la hipótesis nula definida al inicio del estudio.

Adicionalmente, se mide la bondad del ajuste de los modelos obtenidos, es decir, la capacidad de representación de los datos reales de las regresiones utilizadas. Se compara el nivel de acoplamiento a los datos del modelo `full_model` (que representa toda la serie temporal en una única regresión) y los modelos `model1` y `model2` (dos regresiones resultantes de la división de los datos en el año 2020). Para llevar a cabo esta comprobación, se utilizan los métodos de R cuadrado, R cuadrado ajustado, log-verosimilitud, criterio de información de Akaike y criterio de información de Bayes. En todos los análisis se confirma que la separación de los datos disponibles tomando como punto de corte el año 2020 está justificada, ya que la utilización de dos modelos diferentes permite un mejor acoplamiento de las regresiones planteadas.

7. Conclusiones

En este trabajo se ha obtenido una visión clara del estado de la salud mental en los trabajadores del sector tecnológico. El primer análisis de los datos arroja importante información que ofrece una idea de en qué punto está actualmente la salud mental, cómo se aborda esta materia por parte de las empresas y la disposición de los trabajadores de este sector a compartir y debatir abiertamente esta cuestión con su entorno cercano.

Las principales ideas extraídas de este estudio general son:

- En 2021, un mayor porcentaje de mujeres que de hombres con puestos de trabajo relacionados con la tecnología padecía una enfermedad de salud mental.
- El porcentaje de personas cubiertas por su seguro en materia de salud mental en empresas tecnológicas ha disminuido desde 2017.
- El porcentaje de individuos que desconocen las opciones relacionadas con la salud mental dentro de su seguro médico ha aumentado desde 2017.
- Desde 2017, las empresas dan una mayor visibilidad a la salud mental. Sin embargo, los managers dentro de estas no suelen conversar con sus empleados acerca de su salud mental y el porcentaje de aquellos que sí lo hacen ha decrecido desde 2017.

En el análisis de la situación en relación con la apertura frente a personas cercanas para dialogar acerca de salud mental, no se obtiene una conclusión de evolución clara, por lo que se lleva a cabo una prueba de Chow esperando que ésta si la ofrezca. Cuando se realiza, los valores obtenidos muestran significativa evidencia de un cambio estructural en el año 2020 y se rechaza con una alta confianza estadística la hipótesis nula planteada que negaba la existencia de dicho cambio. Por lo tanto, se puede exponer que la pandemia de coronavirus producida en 2020 supuso un punto de inflexión en la forma en la que los trabajadores del sector tecnológico tratan la salud mental en ambientes familiares.

8. Futuras líneas de investigación

En este trabajo se hace evidente que, aunque el avance en la lucha contra el estigma que acompaña a la salud mental es claro, aún queda mucho camino por recorrer en lo referente a este tipo de trastornos. El presente estudio, a la vez que disipa dudas y estudia algunos campos en profundidad, abre la puerta a líneas de investigación que podrían resultar muy interesantes.

En investigaciones futuras, se podría indagar más en el impacto que tiene compartir y expresar el sufrimiento de una enfermedad mental, así como las diferencias en la propensión para hacerlo según el sexo y país de residencia de la persona que las padece. Además, sería útil la identificación de clusters o tipos de personas para estudiar cómo impactan las enfermedades de salud mental a diferentes individuos según su personalidad, cómo reaccionan ante ello, y las mejores técnicas para ayudarles, tanto en el ámbito empresarial como en sus círculos más cercanos.

9. Declaración de Uso de Herramientas de Inteligencia Artificial Generativa en Trabajos Fin de Grado

ADVERTENCIA: Desde la Universidad consideramos que ChatGPT u otras herramientas similares son herramientas muy útiles en la vida académica, aunque su uso queda siempre bajo la responsabilidad del alumno, puesto que las respuestas que proporciona pueden no ser veraces. En este sentido, NO está permitido su uso en la elaboración del Trabajo fin de Grado para generar código porque estas herramientas no son fiables en esa tarea. Aunque el código funcione, no hay garantías de que metodológicamente sea correcto, y es altamente probable que no lo sea.

Por la presente, yo, Belén Mata Moína, estudiante de Doble grado en Administración y dirección de Empresas y Business Analytics de la Universidad Pontificia Comillas al presentar mi Trabajo Fin de Grado titulado "El cambio estructural en el estigma asociado a la salud mental derivado de la pandemia de 2020", declaro que he utilizado la herramienta de Inteligencia Artificial Generativa ChatGPT u otras similares de IAG de código sólo en el contexto de las actividades descritas a continuación:

1. **Referencias:** Usado conjuntamente con otras herramientas, como Science, para identificar referencias preliminares que luego he contrastado y validado.
2. **Metodólogo:** Para descubrir métodos aplicables a problemas específicos de investigación.
3. **Traductor:** Para traducir textos de un lenguaje a otro.

Afirmo que toda la información y contenido presentados en este trabajo son producto de mi investigación y esfuerzo individual, excepto donde se ha indicado lo contrario y se han dado los créditos correspondientes (he incluido las referencias adecuadas en el TFG y he explicitado para que se ha usado ChatGPT u otras herramientas similares). Soy consciente de las implicaciones académicas y éticas de presentar un trabajo no original y acepto las consecuencias de cualquier violación a esta declaración.

Fecha: 20 junio 2024

Firma:



10. Referencias bibliográficas

- Muñoz, A. A., & Uriarte, J. J. U. (2006). Estigma y enfermedad mental. Dialnet. <https://dialnet.unirioja.es/servlet/articulo?codigo=4830167>
- Campo-Arias, A., Oviedo, H. C., & Herazo, E. (2014). Estigma: barrera de acceso a servicios en salud mental. *Revista colombiana de psiquiatría*, 43(3), 162-167.
- Huarcaya-Victoria, J. (2020). Consideraciones sobre la salud mental en la pandemia de COVID-19. *Revista Peruana de Medicina Experimental y Salud Pública/Revista Peruana de Medicina Experimental y Salud Pública*, 37(2), 327-334. <https://doi.org/10.17843/rpmesp.2020.372.5419>
- Palacios-Espinosa, X. (2021). El inestimable costo del estigma de la salud mental. *Revista Ciencias de la Salud*, 19(1), 1-4.
- American Economic Association. (s. f.). <https://www.aeaweb.org/about-aea/>
- About OSMI :: Open Sourcing Mental Health - Changing how we talk about mental health in the tech community - Stronger Than Fear. (s. f.). <https://osmihelp.org/about/about-osmi.htmlhonors-awards/distinguished-fellows/gregory-chow>
- Zdaniuk, B. (2014). Ordinary Least-Squares (OLS) Model. In: Michalos, A.C. (eds) *Encyclopedia of Quality of Life and Well-Being Research*. Springer, Dordrecht. https://doi.org/10.1007/978-94-007-0753-5_2008
- Sanjuán, F. J. M. (2022, 24 noviembre). Suma de cuadrados de los residuos (SCE). *Economipedia*. <https://economipedia.com/definiciones/suma-de-cuadrados-de-los-residuos-sce.html>
- Beers, B. (2024, 9 mayo). P-Value: What It Is, How to Calculate It, and Why It Matters. *Investopedia*. <https://www.investopedia.com/terms/p/p-value.asp>
- D'Agostino, R. (1986). *Goodness-of-Fit-Techniques* (1st ed.). Routledge. <https://doi.org/10.1201/9780203753064>
- Taylor, S. (2023, 22 noviembre). R-Squared. *Corporate Finance Institute*. <https://corporatefinanceinstitute.com/resources/data-science/r-squared/>
- Sanjuán, F. J. M. (2022a, noviembre 24). R cuadrado ajustado (Coeficiente de determinación ajustado). *Economipedia*. <https://economipedia.com/definiciones/r-cuadrado-ajustado-coeficiente-de-determinacion-ajustado.html>

- IBM Cognos Analytics 12.0.X. (s. f.). <https://www.ibm.com/docs/en/cognos-analytics/12.0.0?topic=terms-adjusted-r-squared>
- Coeficiente de determinación (R cuadrado) - MATLAB & Simulink - MathWorks España. (s. f.). https://es.mathworks.com/help/stats/coefficient-of-determination-r-squared.html?searchHighlight=r%20cuadrado%20ajustado&s_tid=srchtitle_support_results_1_r%20cuadrado%20ajustado
- Calzada, H. (2019, 27 diciembre). ¿Qué es el R cuadrado ajustado? Rankia. <https://www.rankia.mx/blog/como-comenzar-invertir-bolsa/4439147-que-r-cuadrado-ajustado>
- Rodó, P. (2022, 24 noviembre). Criterio de información bayesiano. Economipedia. <https://economipedia.com/definiciones/criterio-de-informacion-bayesiano.html>
- Benites, L. (2022, 5 enero). Criterio de información bayesiano (BIC) / Criterio de Schwarz. Statologos. <https://statologos.com/criterio-de-informacion-bayesiano/>
- Pacheco, J. (2021, 6 octubre). Criterio de información bayesiano: Definición, Qué es y Ejemplos | 2024. Economía360. <https://www.economia360.org/criterio-de-informacion-bayesiano/>
- Bobbitt, Z. (2024, 12 febrero). How to Interpret Log-Likelihood Values (With Examples). Statology. <https://www.statology.org/interpret-log-likelihood/>
- Kwong, A. S. F., Pearson, R. M., Adams, M. J., Northstone, K., Tilling, K., Smith, D., ... Timpson, N. J. (2021). Mental health before and during the COVID-19 pandemic in two longitudinal UK population cohorts. *The British Journal of Psychiatry*, 218(6), 334–343. doi:10.1192/bjp.2020.242
- Usher, K., Durkin, J., & Bhullar, N. (2020). The COVID-19 pandemic and mental health impacts. *International journal of mental health nursing*, 29(3), 315.
- Khan, K. S., Mamun, M. A., Griffiths, M. D., & Ullah, I. (2022). The mental health impact of the COVID-19 pandemic across different cohorts. *International journal of mental health and addiction*, 20(1), 380-386.
- Boden, M., Zimmerman, L., Azevedo, K. J., Ruzek, J. I., Gala, S., Magid, H. S. A., ... & McLean, C. P. (2021). Addressing the mental health impact of COVID-19 through population health. *Clinical psychology review*, 85, 102006.

- Mouzo, J., Mouzo, J., & Mouzo, J. (2021, 8 octubre). La epidemia que subyace tras la covid: los casos de depresión y ansiedad crecen más de un 25% en el mundo. El País. <https://elpais.com/ciencia/2021-10-08/la-epidemia-que-subyace-tras-la-covid-los-casos-de-depresion-y-ansiedad-crecen-mas-de-un-25-en-el-mundo.html>
- Lázaro, P. M. R., Hornero, J. P., & Roncero, C. I. (2022b). Reducing stigma-discrimination. Help seeking-empowerment. Emotional education-mental health literacy. *Nutrición Hospitalaria*. <https://doi.org/10.20960/nh.04178>