



UNIVERSITAT
POLITÈCNICA
DE VALÈNCIA



Escola Tècnica
Superior d'Enginyeria
Informàtica

Escola Tècnica Superior d'Enginyeria Informàtica
Universitat Politècnica de València

Herramienta de propósito general para la aplicación de métodos de inferencia causal

TRABAJO FIN DE GRADO

Grado en Ingeniería Informática

Autor: Victoria Beltrán Domínguez

Tutor: Eva Onaindía de la Rivaherrera

Curso 2020-2021

Agradecimientos

Me gustaría agradecer a Eva Onaindía de la Rivaherrera por haberme guiado y orientado durante la realización de este trabajo de fin de grado. También dar las gracias a la empresa SCILING, especialmente a Vicent Alabau, tanto por introducirme en este tópico como por el apoyo ofrecido. Finalmente, me gustaría agradecer a todas las personas que me han apoyado durante el proceso, tanto mis padres, como mi pareja y amigos.

Resum

Aquest projecte té el propòsit de desenvolupar una eina d'inferència causal de propòsit general per a facilitar l'estudi causal a molts usuaris sense necessitat de conèixer cap llenguatge de programació. Per tal d'aconseguir-ho, s'ha integrat una sèrie de funcionalitats en un entorn que permeten crear, editar i analitzar diagrames causals. D'aquesta manera, es pot modelar un diagrama causal mitjançant l'eina, estudiar les seues implicacions, carregar dades per a la seua anàlisi i aplicar mètodes d'inferència causal amb tota la informació disponible. Aquestes funcionalitats s'han integrat en l'entorn original a través d'un framework denominat Flask. Al seu torn, aquest framework fa ús d'una llibreria de Microsoft que proporciona diferents mètodes per a realitzar inferència causal. L'eina s'ha aplicat en un context mèdic i en altres dominis d'aplicació per a exemplificar com s'ha de procedir a l'hora d'identificar i quantificar les diferents relacions causals que s'estableixen entre les variables del domini.

Paraules clau: inferència causal, diagrama causal, correlació, eina d'anàlisi

Resumen

Este proyecto tiene el propósito de desarrollar una herramienta de inferencia causal de propósito general para facilitar el estudio causal a muchos usuarios sin necesidad de conocer ningún lenguaje de programación. Para ello, se ha integrado una serie de funcionalidades en un entorno que permiten crear, editar y analizar diagramas causales. De esta manera, se puede modelar un diagrama causal mediante la herramienta, estudiar sus implicaciones, cargar datos para su análisis y aplicar métodos de inferencia causal con toda la información disponible. Estas funcionalidades se han integrado en el entorno original a través de un framework denominado Flask. A su vez, este framework hace uso de una librería de Microsoft que proporciona diferentes métodos para realizar inferencia causal. La herramienta se ha aplicado en un contexto médico y en otros dominios de aplicación para ejemplificar cómo se debe proceder a la hora de identificar y cuantificar las diferentes relaciones causales que se establecen entre las variables del dominio.

Palabras clave: inferencia causal, diagrama causal, correlación, herramienta de análisis

Abstract

The aim of this project is to develop a general-purpose causal inference tool to facilitate people carry out a causal analysis without need to know any programming language. To this end, we have integrated a series of functionalities in an environment so as to create, edit and analyze causal diagrams. In this way, the tool enables defining a causal diagram, studying its implications, uploading data to analyze them and making causal inference with all the available information. These functionalities have been integrated into the original environment through a framework called Flask. In turn, this framework makes use of a Microsoft library that provides different methods to perform causal inference. The tool has been applied in a medical context and other application domains with the purpose to exemplify how one should proceed for identifying and quantifying the existing causal relationships between the variables of the domains.

Key words: causal inference, causal diagram, correlation, analysis tool

Índice general

Índice general	VII
Índice de figuras	IX
Índice de tablas	X

1	Introducción	1
1.1	Motivación	2
1.2	Objetivos	2
1.3	Estructura de la memoria	3
1.4	Convenciones	3
2	Estado actual de herramientas de análisis causal	5
2.1	Aplicaciones de la inferencia causal	5
2.2	Herramientas para la inferencia causal	6
3	Análisis del problema	9
3.1	Conceptos básicos de la causalidad	9
3.1.1	Relación causal	9
3.1.2	Inferencia causal	11
3.1.3	Efecto promedio del tratamiento en la población: <i>ATE</i>	13
3.1.4	Intervención	13
3.2	Diagramas causales	16
3.2.1	Importancia de los diagramas causales	16
3.2.2	Definición de caminos	16
3.2.3	Manipulación de caminos	18
3.2.4	Implicaciones de los caminos en la inferencia causal	20
3.2.5	Cómo crear nuestro propio diagrama causal	22
3.3	Inferencia causal en cuatro pasos	22
3.3.1	Modelización	23
3.3.2	Identificación	23
3.3.3	Estimación	26
3.3.4	Refutación	27
4	Herramienta propuesta: CAUSAINF	29
4.1	Diseño de la herramienta CAUSAINF	29
4.2	Implementación de la herramienta CAUSAINF	33
4.2.1	Requisitos del sistema	35
5	Estudios experimentales	37
5.1	Estudio aplicado al ámbito médico	37
5.1.1	Efecto de <i>Fumar</i> en <i>Cáncer de Pulmón</i>	38
5.1.2	Efecto de <i>Nacido un día par</i> en <i>Cáncer de Pulmón</i>	42
5.1.3	Efecto de <i>Fatiga</i> en <i>Accidente de Coche</i>	43
5.1.4	Efecto de <i>Presión de Grupo</i> y <i>Dedos Amarillos</i>	45

5.2	Estudio aplicado al ámbito psicológico	46
5.2.1	Efecto de <i>Enfermedad Mental en Empleo</i>	47
5.2.2	Efecto de <i>Discapacidad</i> sobre <i>Empleo</i>	50
5.2.3	Efecto de <i>Enfermedad Mental en Hospitalización</i>	51
5.2.4	Efecto de <i>Género</i> en <i>Vivir Padres</i>	52
6	Conclusiones, relación con asignaturas y trabajos futuros	53
6.1	Conclusiones	53
6.2	Relación del trabajo desarrollado con los estudios cursados	53
6.3	Trabajos futuros	54

Índice de figuras

1.1	Conclusiones incorrectas acerca de la causalidad. Fuente	1
3.1	X (Hace frío) e Y (Hay luna llena) no están relacionados de ninguna manera .	10
3.2	X (Llueve) causa Y (La hierba se moja)	10
3.3	X (Despertarse con dolor de cabeza) e Y (Dormir con zapatos) tienen una causa común	11
3.4	X (Llueve) e Y (Se activa el aspersor) tienen una consecuencia común	11
3.5	Representación simbólica	12
3.6	Intervención sobre ingesta diaria de café	15
3.7	Relaciones entre A, B, C, D, E y F	17
3.8	Camino directo	19
3.9	Camino directo ajustado	19
3.10	Camino trasero	19
3.11	Camino trasero ajustado	19
3.12	Camino cerrado	20
3.13	Camino abierto	20
3.14	Ejemplo camino combinado	20
3.15	Ejemplo sesgo por selección	21
3.16	Proceso causal desglosado en cuatro pasos	23
3.17	Diagrama causal para el estudio de la Obesidad en la Muerte prematura .	24
3.18	Diagrama causal para el estudio de la Obesidad en la Muerte prematura .	25
4.1	Interfaz de CAUSAINF	30
4.2	Botones insertados en la página principal	31
4.3	Diálogo fase de identificación	31
4.4	Diálogo fase de estimación	31
4.5	Diálogo efecto del tratamiento en el resultado	32
4.6	Diálogo fase de refutación	32
4.7	Diseño simplificado de CAUSAINF	33
4.8	Esquema de la infraestructura de CAUSAINF	34
5.1	Red bayesiana causal de LUCAS.	38
5.2	Diagrama causal representativo del efecto de <i>Fumar en Cáncer de Pulmón</i> .	39
5.3	Conjuntos de ajuste disponibles para computar el efecto de <i>Fumar en Cáncer de Pulmón</i>	40
5.4	Efecto de <i>Fumar en Cáncer de Pulmón</i>	41
5.5	Refutación del modelo construido para el efecto de <i>Fumar en Cáncer de Pulmón</i>	41
5.6	Diagrama causal representativo del efecto de <i>Nacido un día par en Cáncer de Pulmón</i>	42

5.7	Refutación del modelo construido para el efecto de <i>Nacido un día par en Cáncer de Pulmón</i>	43
5.8	Diagrama causal representativo del efecto de <i>Fatiga y Accidente de Coche</i>	44
5.9	Refutación del modelo construido para el efecto de <i>Fatiga en Accidente de Coche ajustando en Cáncer de Pulmón</i>	45
5.10	Diagrama causal representativo del efecto de <i>Presión de Grupo en Dedos Amarillos</i>	46
5.11	Diagrama causal representativo del efecto de <i>Enfermedad Mental en Empleo</i>	49
5.12	Diagrama causal representativo del efecto de <i>Discapacidad en Empleo</i>	50
5.13	Diagrama causal representativo del efecto de <i>Enfermedad Mental en Hospitalización</i>	51
5.14	Diagrama causal representativo del efecto de <i>Género en Vivir Padres</i>	52

Índice de tablas

5.1	Efecto causal de <i>Fatiga en Accidente de Coche</i> aplicando diferentes conjuntos de ajuste y métodos de estimación	44
5.2	Descripción variables utilizadas en este estudio	48

CAPÍTULO 1

Introducción

El ser humano siempre ha tenido la necesidad de buscar las causas de los sucesos que le rodean. Ya en la antigüedad se intentaba comprender y justificar cuáles eran las causas de diferentes sucesos como la lluvia, el fuego o las tormentas. A medida que la historia avanzaba y se iba comprendiendo el mundo que nos rodeaba, se ha intentado concebir las causas de sucesos cada vez más complejos, llegando a intentar descubrir incluso las causas de diferentes enfermedades.

La inferencia causal surge así con el objetivo de esclarecer si existe una relación causa-efecto entre dos eventos. Formalmente, la inferencia causal es un procedimiento cuyo objetivo es determinar el efecto de un suceso en otro manteniendo todo lo demás constante.

Determinar el efecto de una acción no es una tarea sencilla y muchas veces es infravalorada. Generalmente, sacamos conclusiones sobre la causalidad entre dos sucesos constantemente sin tener ninguna evidencia. Por ejemplo, en la sentencia “gracias a esta pastilla ya no tengo dolor de cabeza” estamos reafirmando la existencia de una relación causa-efecto entre tomarse una pastilla y que te deje de doler la cabeza. Sin embargo, no sabemos qué habría pasado si no nos la hubiéramos tomado. Y es que fijándonos aún más en este particular caso, estudiar la relación causa-efecto entre dos sucesos implica sacar conclusiones de datos que no han sucedido, preguntas hipotéticas del estilo: “¿Qué habría pasado si no me hubiera tomado la pastilla?”. La inferencia causal pretende solventar toda esta problemática mediante evidencia matemática con el fin de poder computar el efecto de un suceso en otro, y esclarecer así las diferentes causas.

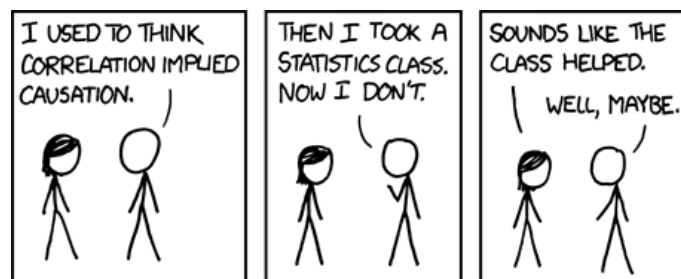


Figura 1.1: Conclusiones incorrectas acerca de la causalidad. [Fuente](#)

La inferencia causal no solo es una potente herramienta para la toma de decisiones en múltiples dominios (epidemiología, economía, política, psicología, etc), sino que también se está estudiando la posibilidad de utilizarlo para mejorar los sistemas clásicos de predicción e inteligencia artificial.

Una vez introducido el tema principal de este proyecto, procederemos a indicar los motivos por los que se ha escogido este tema. Seguidamente, se presentarán los objetivos que se busca conseguir y un pequeño índice general comentando los contenidos de cada capítulo. Por último, se incluye un apartado en el que se comentarán brevemente los convencionalismos. En este trabajo no se ha incluido un glosario de términos porque todos los términos necesarios son presentados y explicados.

1.1 Motivación

En este apartado se expondrán los motivos que han llevado a la realización de este proyecto.

Durante mi experiencia de prácticas en la empresa SCILING, mi tutor Vicent Alabau me presentó el mundo de la inferencia causal como un tema que había hecho *boom* en los últimos años y que podría llegar a tener relevancia en diferentes ámbitos.

De esta manera, la motivación profesional que propulsó la realización del proyecto fue el aprendizaje en un campo potencialmente en auge y con múltiples aplicaciones. La primera aplicación consiste en ayudar a mejorar la capacidad de toma de decisiones. La segunda aplicación reside en su uso potencial en el campo del *machine learning* para reducir el sesgo entre los datos del mundo real y los datos de entrenamiento utilizados con el fin de construir modelos más robustos y generalizables.

A nivel personal, nunca había oído hablar de inferencia causal, así que el tema me llamó la atención. Me empecé a realizar preguntas del estilo: ¿cómo es posible sacar conclusiones sobre cosas que no han pasado?, ¿cuál es la diferencia principal respecto a la inferencia estadística?, ¿cuáles son los límites de la aplicación de la inferencia causal?, etc. Todas estas preguntas se convirtieron en curiosidad e interés acerca del tema en cuestión.

De esta manera, contacté con mi tutora Eva Onaindia, profesora y profesional que ha publicado algunos estudios realizando inferencia causal, y se propuso hacer un prototipo de una herramienta que facilitara el proceso de inferencia causal a personas no especializadas en el tema (aunque sí con un mínimo de conocimientos) del que la empresa pudiera hacer uso.

1.2 Objetivos

En este apartado, se pretende presentar los objetivos planteados para este proyecto:

- Ser capaz de interpretar los conceptos teóricos subyacentes de la inferencia causal.
- Desarrollar una herramienta que permita aplicar diferentes métodos de la inferencia causal.
- Encontrar al menos dos conjuntos de datos procedentes de diferentes dominios.
- Realizar un estudio con cada conjunto de datos encontrado para ejemplificar el uso de la herramienta.

1.3 Estructura de la memoria

En este apartado se procede a presentar un índice comentado de la estructura de la memoria:

- **Capítulo 1: Introducción**

En este capítulo se presenta el tema principal de esta memoria, así como la motivación que ha propulsado la realización del proyecto y los objetivos perseguidos con él. También se define la estructura de la memoria y las convenciones que se seguirán a lo largo de esta.

- **Capítulo 2: Estado actual de herramientas de análisis causal**

En este capítulo se resalta la relevancia de la inferencia causal en determinados ámbitos y se menciona alguna de sus principales aplicaciones. Asimismo, se presenta un breve análisis de las capacidades de las herramientas existentes de inferencia.

- **Capítulo 3: Análisis del problema**

En este tercer capítulo, se presentan los conceptos básicos necesarios con el fin de realizar inferencia causal. Particularmente, primero se explican las bases de la inferencia causal y, seguidamente, explicamos los diagramas causales y sus propiedades. Finalmente, se explica de qué manera podemos realizar inferencia causal con todo lo explicado anteriormente.

- **Capítulo 4: Herramienta propuesta: CAUSAINF**

En este cuarto capítulo, se presenta el diseño, la implementación y los requisitos de la herramienta propuesta como solución CAUSAINF.

- **Capítulo 5: Estudios experimentales**

En el quinto capítulo, se realizan varios estudios experimentales para ejemplificar el uso de la herramienta. Particularmente, realizaremos primero un estudio del ámbito médico y otro del ámbito psicológico.

- **Capítulo 6: Conclusiones, relación con asignaturas y trabajos futuros**

En este último capítulo, se presentan las conclusiones del proyecto, destacando los objetivos que se han conseguido. Además, se comenta la relación de este con las asignaturas cursadas por el alumno y, para finalizar, se presentan posibles líneas futuras de trabajo.

1.4 Convenciones

En este apartado, simplemente se pretende destacar la normativa de marcado que lleva asociado un significado adicional:

- Las palabras extranjeras, acrónimos en inglés y nombres de variables se remarcarán en cursiva.
- Los conceptos clave serán remarcados en negrita.
- En los números, se utilizarán puntos con el fin de separar los enteros de los decimales.

CAPÍTULO 2

Estado actual de herramientas de análisis causal

En este capítulo se resalta la relevancia de la inferencia causal en determinados ámbitos y se menciona alguna de sus principales aplicaciones. Asimismo, se presenta un breve análisis de las capacidades de las herramientas existentes de inferencia causal.

2.1 Aplicaciones de la inferencia causal

Realizar conclusiones sobre el efecto que tiene un suceso en otro es una tarea complicada. Por ejemplo, existe una gran asociación entre la cafeína y la longevidad. Una persona longeva que toma regularmente café podría llegar a preguntarse: “¿habría muerto ya si no hubiera tomado cafeína regularmente?”. ¿Podemos esclarecer las causas de su longevidad, y si la ingesta de cafeína es una de ellas?. El problema es que no sabemos qué habría pasado si no hubiera tomado cafeína regularmente, y si en dicho caso habría adoptado un nivel de vida más sedentario, haciendo más fácil el desarrollo de enfermedades. Quizá la ingesta de café no ha tenido absolutamente nada que ver, y simplemente es genético. Quizás, existe algún otro hábito que ha causado su longevidad. Establecer el efecto de la ingesta regular de cafeína en la longevidad no es un cometido simple. Para ayudar a la tarea de descubrir la relación entre dos sucesos, aparece la inferencia causal, un proceso que nos permite determinar el efecto de un evento o suceso particular en otro. La inferencia causal se emplea en muchos ámbitos con el fin de estudiar las relaciones entre diferentes sucesos, contribuyendo así a la mejora de la toma de decisiones. En este campo destacan científicos como Judea Pearl, James Robins o Miguel Hernán. Existen múltiples ejemplos de aplicaciones de inferencia causal.

Actualmente, una de las aplicaciones más relevantes de la inferencia causal es para gestionar la pandemia mundial provocada por el Coronavirus. Específicamente, se ha empleado inferencia causal para realizar estudios acerca del impacto de las políticas impuestas por los gobiernos en la propagación del Coronavirus con el fin de ayudar a la formulación de políticas en el futuro [1]. En este estudio se analizan las diferentes políticas impuestas por los gobiernos relacionadas con el distanciamiento social (estado de alarma, reuniones, cierre de escuelas, etc.), las reaperturas (sitios de entretenimiento, restaurantes, etc.) y el uso obligatorio de las mascarillas con el propósito de medir el impacto de estas actuaciones en el número de muertes y contagios por Coronavirus.

En esta misma línea de investigación, también se ha publicado un estudio en el que se aplica la inferencia causal para estimar si la reducción de la movilidad social contribuye a la reducción del número de muertos [2]. Los resultados concluyen que la movilidad reducida provoca una reducción de las muertes, pero resaltan la existencia de factores no valorados.

Aparte de su aplicación en el ámbito epidemiológico y político, también se ha aplicado para medir el impacto socio-económico del propio virus, por ejemplo en África [3]. Los resultados revelan que las regiones Africanas han sido más afectadas que otras regiones, puesto que las políticas de distanciamiento social han afectado tanto a la interacción social como a las actividades económicas.

Por último, en [4] se plantea descubrir y cuantificar las relaciones entre las características de la pandemia (por ejemplo, número de infecciones y muertes) y el sentimiento general en *Twitter*. En los resultados de este estudio se encuentran relaciones causales entre el anuncio del confinamiento y el sentimiento general en la plataforma o entre el número de muertes y la actividad en la red social.

Asimismo, la inferencia causal se ha aplicado también en diferentes estudios en el ámbito psicológico. Por ejemplo, se ha utilizado inferencia causal para descubrir la relación entre ser hijo único y salud mental [5]. En el citado estudio, realizado a partir de una encuesta nacional anual de las familias e individuos en China, se determinó que, efectivamente, ser hijo único tiene efectos negativos pequeños pero significativos en la salud mental.

Aparte de su empleo en el ámbito de la toma de decisiones, muchos expertos como Yoshua Bengio aseguran que las teorías que se desarrollan en este campo se pueden integrar con otros campos de la Inteligencia Artificial y el *machine learning* con el fin de reducir el sesgo entre los datos utilizados en entrenamiento y el mundo real, obteniendo así un modelo más generalizado con mejores resultados. Particularmente, algunos trabajos revisan conceptos fundamentales de la inferencia causal y los relacionan con problemas abiertos de *machine learning* [6], exponiendo así el modo en el que la inferencia causal puede contribuir a la investigación en este campo.

Existen muchos campos de la informática en los que se está aplicando inferencia causal de manera innovadora. Por ejemplo, en [7] se utiliza la inferencia causal para distinguir clics intencionados de clics influenciados por *clickbait*s y poder así proveer al usuario con mejores recomendaciones.

Judea Pearl, uno de los grandes propulsores de esta ciencia, argumenta en su libro que es clave reemplazar el razonamiento asociativo por razonamiento causal, para que las máquinas puedan realizarse preguntas del mundo contrafactual [8]. Este, afirma: "Si queremos que las máquinas razonen sobre intervenciones e introspecciones, debemos invocar modelos causales. Las asociaciones no son suficientes, y esto es un hecho matemático, no una opinión.". Pearl incluso argumenta que el razonamiento causal podría proveer a las máquinas con inteligencia al nivel humano.

2.2 Herramientas para la inferencia causal

En el anterior apartado, se han descrito algunas de las principales aplicaciones de la inferencia causal. En esta sección se apuntan las principales librerías o herramientas existentes en este campo junto con sus principales ventajas y limitaciones.

- *DoWhy*: [DoWhy](#) es una librería en *Python* que se apoya en un lenguaje unificado para realizar métodos de inferencia causal. Esta librería plantea el proceso de la inferencia causal como una secuencia de cuatro pasos, dando al usuario un control superior en todas las operaciones realizadas.
- *CausalInference*: [CausalInference](#) es un paquete de software en *Python* que implementa varios métodos estadísticos y econométricos utilizados en el campo de la inferencia causal, evaluación de programas o análisis del efecto del tratamiento. Esta librería aporta un gran valor estadístico, pero restringe al usuario en el proceso de inferencia causal, puesto que se realiza todo en un paso, y no permite realizar asunciones.
- *CausalImpact*: [CausalImpact](#) es un paquete en *R* desarrollado por *Google* para realizar inferencia causal utilizando modelos de series de tiempo estructurales bayesianos. Básicamente, este paquete te permite crear conjuntos de datos o cargarlos y realizar análisis sobre ellos.
- *Causalml*: [Causalml](#) es un paquete en *Python* que proporciona un conjunto de métodos de inferencia causal que utilizan algoritmos de aprendizaje automático basados en investigaciones recientes. Este permite al usuario estimar diferentes efectos a partir de datos experimentales u observacionales.
- *Econml*: [EconML](#) es un paquete de *Python* para estimar efectos de tratamiento heterogéneos a partir de datos de observación utilizando el aprendizaje automático.

A pesar de la multitud de librerías existentes, se ha detectado una escasez de herramientas capaces de simplificar los procesos necesarios para realizar un estudio causal sin tener conocimientos expertos en el campo. Las herramientas existentes en el campo de la causalidad se asientan sobre todo en librerías de *Python* o *R* y asumen cierto grado de conocimiento en causalidad y programación. También nos hemos percatado de que muchas de estas librerías que ayudan a la causalidad esconden de alguna manera las asunciones de las que el usuario está partiendo, al no indicarlas explícitamente. Por ejemplo, *CausalInference* permite realizar distintos métodos de inferencia causal pero no se muestran las asunciones realizadas. También encontramos otras librerías como *DoWhy*, que aunque sí muestran las asunciones de sus cálculos en la terminal, no lo hacen de manera clara y cómoda para el usuario. No se han encontrado herramientas visuales de inferencia causal que sean populares.

De esta manera, este proyecto se propone desarrollar una herramienta que permita estudiar las relaciones en los datos de manera dinámica y visual, sin necesidad de aprender a programar, constituyendo así un medio para que empresas o usuarios con cierto nivel de conocimiento en inferencia causal y causalidad puedan centrarse en lo que es realmente importante.

CAPÍTULO 3

Análisis del problema

En la mayoría de trabajos se realiza primeramente una exposición de la teoría o estado del arte en el que se fundamenta el trabajo y posteriormente se presenta el problema a resolver. Sin embargo, en este proyecto los conceptos que definen la teoría de la causalidad constituyen los componentes principales del problema que se quiere abordar. Por este motivo, en este capítulo se presentarán los principales ingredientes de la causalidad y grafos causales y destacaremos al final de cada sección las nociones y componentes que necesitaremos para desarrollar la herramienta que se verá en el capítulo 4.

Este capítulo se estructura del siguiente modo: comenzaremos en la sección 3.1 realizando una introducción a los conceptos fundamentales de la causalidad. Seguidamente se presentarán los diagramas causales y su utilidad en el campo de la causalidad en 3.2 y, finalmente, se expondrán los pasos básicos necesarios para realizar inferencia causal.

Antes de empezar, destacamos que por la complejidad y amplitud de la teoría causal, se han utilizado principalmente los libros *Causal Inference: What If* (Hernan M.A. y Robins J.M) [9] y *Causal Inference Course Lectures de Brady Neal*[10] como fuentes de información.

3.1 Conceptos básicos de la causalidad

En esta sección se introducen los conceptos que constituyen los principios básicos de la teoría de la causalidad. Los conceptos relevantes que se presentan en esta sección y que se utilizarán posteriormente en el Capítulo 4 se resaltan en negrita junto con su definición y concepción.

3.1.1. Relación causal

Dos variables (o eventos) tienen una relación causal cuándo la ocurrencia de una de ellas causa la ocurrencia de la otra. Por ejemplo, existe una relación causal entre calentar un helado y que este se derrita porque es algo que pasa inequívocamente, es decir, si un helado se calienta, este siempre se derretirá bajo determinadas condiciones.

Es importante distinguir entre **causalidad** y **correlación**. Dos variables están correlacionadas o asociadas cuándo los valores de una variable varían sistemáticamente con respecto a los valores homónimos de la otra. Por ejemplo, existe correlación entre el número de personas ahogadas y la consumición de helado porque cuando hace buen tiempo, ambas ocurren más frecuentemente y, por lo tanto, varían de manera sistemática.

La correlación mencionada anteriormente entre *personas ahogadas* y *consumición de helados*, sin embargo, no implica causalidad puesto que observar que dos variables se mueven conjuntamente no significa necesariamente que una variable sea la causa de la otra. Por eso se suele decir que **correlación no implica causalidad**. Extraer conclusiones incorrectas de causalidad cuando se observa correlación es un error lógico muy común y algo que debemos tener en cuenta a lo largo de este proyecto.

Una vez diferenciados ambos conceptos, vamos a ejemplificar el tipo de relación que podemos encontrar entre dos variables X e Y . Para facilitar la visualización de las diferentes situaciones, utilizaremos un diagrama causal o DAG. Estos se explicarán detalladamente en la siguiente sección, así que por ahora sólo debemos saber que:

- Los nodos representan variables o eventos.
- Las aristas representan relaciones causales.
- Las aristas son dirigidas y su dirección indica el flujo de causalidad. De esta manera, una arista saliendo de X en dirección Y significa que X causa un efecto en Y (y no al revés).

Una vez remarcado el funcionamiento básico de los diagramas causales, pasamos a presentar las diferentes relaciones entre X e Y :

1. **X e Y no están relacionados de ninguna manera:** X e Y son dos eventos completamente independientes. Ninguno causa al otro y no existe correlación entre ellos.

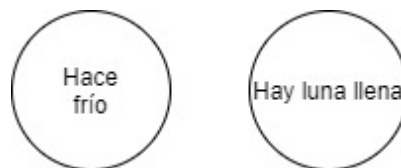


Figura 3.1: X (Hace frío) e Y (Hay luna llena) no están relacionados de ninguna manera

Un ejemplo podría ser la Figura 3.1, puesto que la luna llena no causa más frío ni viceversa. Tampoco existe ninguna correlación entre ambas puesto que hay luna llena cada mes independientemente de la temperatura que haya.

2. **X causa Y :** la ocurrencia de X causa la ocurrencia de Y de modo que existe tanto una relación causal como correlación entre ambas variables.

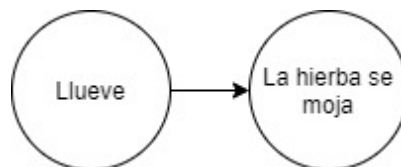


Figura 3.2: X (Llueve) causa Y (La hierba se moja)

Un ejemplo podría ser la Figura 3.2, puesto que si llueve, la hierba se va a mojar, y por lo tanto entre ellas existe una relación causal.

Otra posible relación entre X e Y es que Y cause X . La situación es idéntica a la anterior, simplemente hemos intercambiado la dirección de la flecha.

3. **X e Y tienen una causa común:** existe un evento que causa tanto X como Y. En este caso, no existe una relación causal entre X e Y pero sí existe correlación ya que al existir una causa común, ambas variables se mueven conjuntamente.

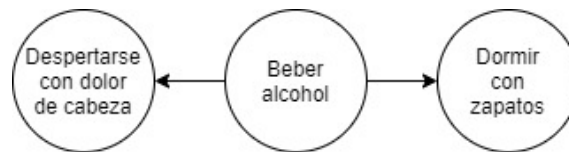


Figura 3.3: X (Despertarse con dolor de cabeza) e Y (Dormir con zapatos) tienen una causa común

El ejemplo de la Figura 3.3 muestra que *beber alcohol* causa tanto *despertarse con dolor de cabeza* como *dormir con zapatos*. Existe correlación entre *despertarse con dolor de cabeza* y *dormir con zapatos*, puesto que ambos tienen un padre o causa común y, por lo tanto, ocurren conjuntamente. Sin embargo, no existe ninguna relación causal entre ellos.

Dentro de este caso también entraría el ejemplo anteriormente mencionado sobre *número de personas ahogadas* y *consumición de helado* porque estas dos variables tienen una causa común que es *el buen tiempo*, y por lo tanto cuando esta causa ocurre, afecta a sus descendientes haciendo que varíen de forma sistemática.

4. **X e Y tienen una consecuencia común:** tanto la ocurrencia de X como la de Y tiene la misma consecuencia. En este caso no existe relación causal ni correlación entre X e Y porque son dos eventos independientes con la misma consecuencia.



Figura 3.4: X (Llueve) e Y (Se activa el aspersor) tienen una consecuencia común

Un ejemplo podría ser el que se muestra en la Figura 3.4, donde se observa que tanto el evento *Llueve* como el evento *Se activa el aspersor* tienen en común la consecuencia *La hierba se moja*. Sin embargo, que tengan la misma consecuencia no conlleva ninguna implicación y ambas variables son independientes.

Las cuatro situaciones descritas son las que identifican la relación existente entre dos variables X e Y. Cabe destacar que entre estas cuatro posibilidades, sólo la segunda situación (X causa Y, Figura 3.2) representa una relación causal entre X e Y.

3.1.2. Inferencia causal

La inferencia causal es el proceso que nos permite identificar y cuantificar relaciones causales. En este apartado, sólo se pretende presentar los elementos principales de la inferencia causal y, en secciones posteriores, una vez se disponga de los conocimientos necesarios, se describirá el proceso que seguiremos para aplicar la inferencia causal.

De este modo, introducimos previamente un ejemplo para facilitar el entendimiento de los conceptos que se van a explicar. Imaginemos que un individuo tiene dolor de cabeza, se toma una pastilla y al cabo de cuatro horas se le pasa:

- ¿Siente alivio a causa de haberse tomado una pastilla?
- ¿Qué habría pasado si no se lo hubiera tomado?

Estas son preguntas que sólo podemos responder a través de la inferencia causal y para ello debemos comprobar si existe una relación causal entre *tomar pastilla* y *dolor de cabeza*. En este ejemplo podemos observar diferentes conceptos:

- **Unidad:** Es el objeto básico en el estudio del efecto del tratamiento. Una unidad puede ser un objeto físico, una entidad, un paciente, un individuo, una colección de personas de una clase, una asociación, etcétera. En este ejemplo, la unidad se corresponde con el individuo o la persona.
- **Tratamiento:** Se refiere a la acción particular que se aplica a la unidad. En nuestro ejemplo, el tratamiento se corresponde con tomarse una pastilla. Suele representarse con la variable T .
- **Resultado:** Evento o variable cuya respuesta al tratamiento queremos estudiar. En nuestro ejemplo, el resultado se corresponde con dejar de doler la cabeza al cabo de cuatro horas. Suele representarse con la variable Y .
- **Resultado potencial.** Para cada par $\langle \text{unidad}, \text{tratamiento} \rangle$ es el valor obtenido en la variable resultado cuando se aplica el tratamiento a la unidad. En nuestro ejemplo, si nos deja de doler la cabeza, el resultado potencial será uno, mientras que si el dolor no cesa en cuatro horas, el resultado potencial será cero. Existen diferentes notaciones que debemos tener en cuenta para esta variable:
 - $Y_i(\text{do}(T = 1))$ representan el resultado potencial en una unidad i cuando se aplica el tratamiento. En nuestro ejemplo, el resultado potencial será igual a uno porque al tomar la pastilla, al individuo le ha dejado de doler la cabeza. Destacar que el operador *do* implica intervenir activamente en el sistema y será explicado posteriormente en 3.1.4. De momento, simplemente debemos interpretarlo como el acto de realizar una acción.
 - $Y_i(\text{do}(T = 0))$ representan el resultado potencial en una unidad i cuando el tratamiento no se aplica. En nuestro ejemplo, no sabemos si el resultado potencial sería uno o cero ya que una vez que la persona se toma la pastilla, no sabemos qué habría pasado si no se la hubiera tomado.

De esta manera, para cada individuo sólo podremos observar un resultado potencial.

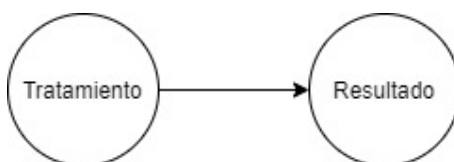


Figura 3.5: Representación simbólica

La inferencia causal tiene como objetivos:

- comprobar si las variables tratamiento y resultado tienen una relación causal (tal como se muestra en la Figura 3.5)
- calcular el efecto causal o magnitud por la que el resultado cambia al aplicar el tratamiento.

En nuestro ejemplo no sólo queremos verificar si la variable *tomarse una pastilla* tiene una relación causal en la variable *dejar de doler cabeza al cabo de cuatro horas*, sino que además, queremos saber cuál es la diferencia que existe entre aplicar o no aplicar el tratamiento; es decir, entre la presencia o ausencia de *tomar una pastilla* en *dejar de doler cabeza al cabo de cuatro horas*.

3.1.3. Efecto promedio del tratamiento en la población: ATE

El efecto que causa el tratamiento en el individuo está definido como la diferencia de los resultados potenciales entre aplicar el tratamiento y no aplicarlo. Como se ha comentado anteriormente, no es posible observar ambos resultados potenciales para un mismo individuo o unidad, puesto que una vez se ha aplicado el tratamiento, ya no es posible saber cuál sería el resultado potencial de no aplicar el tratamiento, y viceversa. De esta manera, en vez de calcular el efecto del tratamiento para un individuo particular, nos fijaremos en el efecto promedio del tratamiento en la población. Esta medida se corresponde con el concepto **population Average Treatment Effect** (ATE) o efecto medio del tratamiento en la población y se identifica con la siguiente fórmula, donde \mathbb{E} se corresponde al valor esperado o esperanza:

$$ATE = \mathbb{E}[Y|do(T = 1) - Y|do(T = 0)] = \mathbb{E}[Y|do(T = 1)] - \mathbb{E}[Y|do(T = 0)] \quad (3.1)$$

Así, el efecto medio del tratamiento se calcula como la diferencia entre el valor esperado del resultado potencial bajo tratamiento y el valor esperado del resultado potencial bajo ningún tratamiento. Cabe destacar que la fórmula del ATE se corresponde con un **estimando causal** o, en otras palabras, una expresión que contiene términos causales como el operador *do*.

El ATE nos indica cómo se comporta el resultado al aplicar el tratamiento en la población. Por ejemplo, imaginemos que estamos estudiando el efecto de fumar en el cáncer de pulmón. Un ATE de 0.55 significaría que si tenemos una población de 1.000.000 personas en el grupo fumador y otra población de 1.000.000 personas en el grupo no fumador, esperamos que existan 550.000 personas más (un 55 % de ellas) que desarrollen cáncer de pulmón en el grupo que fuma que en el grupo que no fuma.

3.1.4. Intervención

Como se ha explicado anteriormente, la inferencia causal pretende comprobar si las variables tratamiento y resultado tienen una relación causal así como calcular el efecto causal o magnitud por la que el resultado puede cambiar al aplicar el tratamiento. Tal como se explica en [11], deberemos ser capaces de intervenir activamente en el sistema para poder contestar preguntas del estilo:

- ¿Ayuda la medicación X a tratar la enfermedad Y?
- ¿Puedo incrementar mis ventas cambiando el precio de mi producto por X?
- ¿Puedo incrementar mi felicidad adoptando un perro?
- ¿Tomar café de forma diaria incrementa mi longevidad?

En la inferencia causal, para contestar este tipo de preguntas normalmente se realizan intervenciones. Realizar una intervención en una variable significa fijar su valor. Al hacerlo, estamos activamente cambiando todo el sistema y las variables restantes pueden cambiar como consecuencia de ello.

Conviene resaltar primero la diferencia entre **intervenir** una variable y **condicionar** en una variable: mientras que al intervenir fijamos el valor de esta, al condicionar simplemente estamos mirando el subconjunto de datos que cumple con ese valor (pero sin forzar a que la variable adopte dicho valor).

La manera más efectiva de realizar una intervención es a través de **pruebas controladas aleatorizadas**. Por ejemplo, supongamos que queremos saber si tomar café puede incrementar la longevidad de la población. Para realizar una prueba controlada aleatorizada, podríamos coger a cada individuo de la población y, tirando una moneda, asignar si debe tomar café de forma diaria o bien tiene prohibido el café. Al cabo de cien años, podemos estudiar si las personas que tomaron café de forma diaria han sido en general más longevas que las que lo tenían prohibido, y de ahí podríamos calcular el efecto de la ingesta diaria de café en la longevidad. De este modo, se sabría con seguridad que el efecto es causado por el café y no por algún otro motivo que no hayamos tenido en cuenta porque los grupos han sido formados de manera completamente aleatoria, y por lo tanto, cualquier tendencia visible debería ser atribuible al café (la única diferencia entre los grupos).

En otras palabras, estas pruebas consisten en dividir a la población en dos grupos de manera aleatoria, aplicar el tratamiento en un grupo (grupo tratado), no aplicarlo en el otro (grupo de control) y analizar el impacto del tratamiento a través de los resultados obtenidos. Así, el impacto identificado es realmente atribuible al tratamiento y no a otras causas, porque al haber hecho idénticos los grupos de manera previa, lo único que difiere entre ellos es el hecho de tener o no tener el tratamiento.

Sin embargo, las pruebas controladas aleatorizadas no siempre son realizables porque pueden existir problemas éticos [10]. Por ejemplo, si se está estudiando el efecto de fumar en el cáncer de pulmón no se puede escoger aleatoriamente a unas personas y obligarlas a fumar para ver si finalmente desarrollan cáncer. También pueden dar lugar a problemas de realización; por ejemplo si se quiere estudiar el efecto del comunismo contra el capitalismo en la economía, no se puede asignar de forma aleatoria los países que van a tener un sistema comunista y cuáles uno capitalista; así como problemas de imposibilidad (por ejemplo, no se puede cambiar el ADN de una persona para estudiar el efecto de este en el cáncer de mama).

Dicho lo anterior, es evidente que en algunas situaciones es imposible obtener datos a partir de una prueba controlada aleatorizada, por lo que hay que recurrir a trabajar con datos observacionales. Así, es necesario estudiar en detalle y entender lo que se está haciendo al realizar una intervención en una variable para poder simularla.

Retomamos el ejemplo previamente comentado acerca del efecto de la *ingesta diaria de café* en la *longevidad* y vamos a asumir que *estilo de vida activo* es un antecesor común

a *ingesta diaria de café* y *longevidad*. Supongamos que se realiza una intervención sobre *ingesta diaria de café*. Al fijar el valor de esta variable de manera aleatoria, la estamos desconectando de sus antecesores, tal y como se muestra gráficamente en la Figura 3.6, puesto que ya no depende de nadie más:



Figura 3.6: Intervención sobre ingesta diaria de café

Al intervenir la variable *ingesta diaria de café*, se borran todas las aristas entrantes de dicho nodo, haciendo así que todo el sistema siga actuando con normalidad, excepto la propia variable intervenida, que ya no está respondiendo a sus antecesores. Debemos recalcar la asunción de **modularidad**, que es indispensable a la hora de hacer intervenciones puesto que implica una intervención en una variable T sólo cambia las relaciones con sus padres (siendo estas eliminada). Si esta asunción no se respetase, la intervención en una variable provocaría cambios en las otras variables, haciendo así un grafo completamente diferente. Modularidad implica que los cambios siempre son locales. Además, por la asunción local de Markov, que dice que dados los padres de un nodo, este es independiente de todos sus no-descendientes, sabemos que lo único que ha cambiado en este sistema es la distribución de los valores de *ingesta diaria de café*, puesto que antes se correspondía a $P(\text{Ingesta diaria café} | \text{Estilo vida activo})$ y ahora se corresponde a nuestra intervención $P_{\text{manipulada}}(\text{Ingesta diaria café})$.

Matemáticamente, dado que realizamos una intervención sobre la variable X (*ingesta diaria de café*), la distribución de Y (*longevidad*) se representa como $P(Y | do(X))$ (distinguiible del simple condicionamiento $P(Y | X)$). Aplicando una serie de propiedades estadísticas en la expresión $P(Y | do(X))$, esta se puede transformar en términos de probabilidades condicionales corrientes tal que:

$$P(Y | do(X)) = \sum_z P(Y | X, Z = z) P(Z = z) \quad (3.2)$$

Esta expresión se conoce como la **formula de ajuste**, donde Z representa el conjunto de padres del tratamiento. Esto es así porque en una intervención, al fijar el valor de X , rompemos la conexión entre X y sus padres, por lo que los padres no pueden influir directamente en X . Sin embargo, aún pueden influenciar otras variables. Al promediar todos los valores posibles de los padres, se tienen en cuenta todos estos efectos de forma natural.

Aplicando la fórmula de ajuste a nuestro ejemplo, obtenemos:

$$P(\text{Longevidad} = 1 | do(\text{Café})) = [P(\text{Longevidad} = 1 | \text{Café}, \text{Activo} = 1) P(\text{Activo} = 1) - P(\text{Longevidad} = 1 | \text{Café}, \text{Activo} = 0) P(\text{Activo} = 0)]$$

La fórmula de ajuste nos permite pasar de un **estimando causal** como puede ser el ATE a un **estimando estadístico**. En otras palabras, pasamos de una expresión que con-

tiene términos causales a una que sólo contiene términos estadísticos, permitiéndonos así realizar intervenciones sin necesidad de hacer una prueba controlada aleatorizada, ya que podemos calcular el ATE a partir de datos observacionales. La aproximación del estimando estadístico utilizando los datos corresponderá al **valor estimado**.

Una vez hemos presentado las bases y los conceptos fundamentales de la inferencia causal, pasamos a explicar una pequeña noción de grafos causales.

3.2 Diagramas causales

Los diagramas causales son grafos dirigidos acíclicos (DAGs) donde los nodos representan variables o eventos y las aristas o arcos dirigidos representan relaciones causales. Un DAG contiene la causa común de cada par de nodos presentes, si esta existe. El diagrama debe ser dirigido porque se asume que un nodo padre es una causa directa de todos sus nodos hijos y acíclico porque una variable no puede causarse a sí misma.

En esta sección, se explica en primer lugar la importancia de los diagramas causales. Seguidamente, se introduce el concepto de camino y los tipos de caminos básicos que se pueden identificar en los diagramas. Posteriormente se estudian las implicaciones de la detección de caminos en la inferencia causal y, finalmente, se presentan unas directrices para crear nuestro propio diagrama causal.

3.2.1. Importancia de los diagramas causales

Según [9], los diagramas causales son una herramienta que nos permite codificar fácilmente nuestros conocimientos y nuestras asunciones sobre el dominio. En otras palabras, un diagrama casual nos permite representar nuestras creencias sobre la forma en la que se han generado los datos a analizar. En un diagrama causal reflejamos nuestras asunciones sobre los componentes que forman parte del diagrama y, aunque no parezca intuitivo, también se reflejan información sobre los componentes que no forman parte del diagrama.

Si una variable está representada en el diagrama causal, se asume que de alguna manera es importante. Por el contrario, si una variable no está representada, estamos asumiendo que esa variable no es relevante para nuestros propósitos (o no sabemos de su existencia o no sabemos cómo medirla). Asimismo, un arco dirigido de A a B significa que asumimos que A tiene un efecto en B mientras que la inexistencia de un arco entre A y B indica que asumimos que ninguna variable causa a la otra.

Estas asunciones codificadas en el diagrama causal no sólo permiten estudiar de manera gráfica los tipos de relaciones entre las variables (causales o no causales), sino que a su vez ayudan a inferir independencias y dependencias condicionales en las variables representadas. Toda esta información se extrae del diagrama causal a través de los caminos.

3.2.2. Definición de caminos

En este apartado se presenta una definición de camino y se explica los tipos de caminos que existen. Para facilitar el entendimiento se utilizará el diagrama causal de la Figura 3.7, que representa las diferentes relaciones entre las variables A , B , C , D , E y F . La

presentación del ejemplo de la Figura 3.7 así como algunos conceptos fundamentales se apoyan en diversa documentación [12].

Un camino entre dos nodos X e Y es una secuencia de arcos que conecta el nodo X con el nodo Y . Esta secuencia debe cumplir una serie de propiedades:

1. El primer arco de la secuencia debe tener como nodo inicial X .
2. El último arco de la secuencia debe tener como nodo final Y .
3. Para cada arco, el nodo final debe coincidir con el inicial del siguiente arco en la secuencia.
4. Todos los arcos de la secuencia deben existir en el diagrama, sin importar la dirección de estos (la dirección del arco no se utiliza a la hora de definir caminos en el mismo). Por ejemplo, si tenemos un arco dirigido $W \rightarrow Z$, podemos utilizar tanto el arco $W \rightarrow Z$, como el arco $Z \rightarrow W$ en la secuencia.
5. No se puede visitar ningún nodo más de una vez.

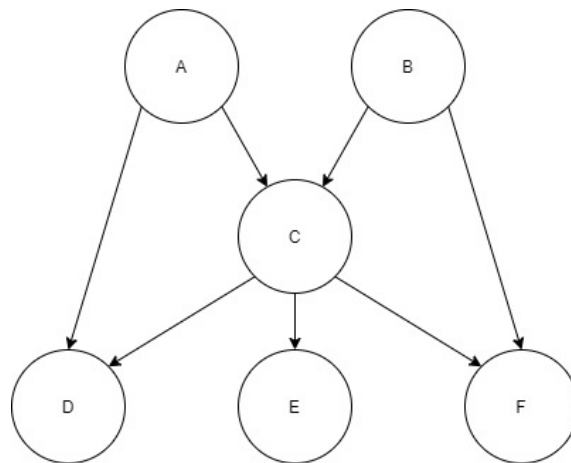


Figura 3.7: Relaciones entre A, B, C, D, E y F

Por ejemplo, en la Figura 3.7 encontramos cuatro caminos entre A y F :

- $A \rightarrow C \rightarrow F$
- $A \rightarrow C \leftarrow B \rightarrow F$
- $A \rightarrow D \leftarrow C \rightarrow F$
- $A \rightarrow D \leftarrow C \leftarrow B \rightarrow F$

Algunas secuencias de arcos que no se corresponderían con caminos entre A y F podrían ser $A \rightarrow C \rightarrow E \leftarrow C \rightarrow F$ (porque estamos visitando el nodo C más de una vez) o $A \rightarrow C \rightarrow E \rightarrow F$ (porque no existe ningún arco entre E y F).

Una vez se ha presentado la definición de camino, se exponen los distintos tipos que podemos encontrar:

- **Camino causal:** Un camino es causal si sigue la dirección marcada en las aristas. En otras palabras, es un camino en donde todos los nodos padre son visitados antes que sus hijos. En los caminos definidos anteriormente entre A y F , solamente $A \rightarrow C \rightarrow F$ es causal.
- **Camino no causal:** Un camino no es causal si en algún momento vamos en contra de la dirección marcada en las aristas. En otras palabras, es un camino en donde en algún momento se visita un hijo antes que a su padre. Un ejemplo entre A y F podría ser $A \rightarrow C \leftarrow B \rightarrow F$.

Un nodo es un **colisionador** en un camino cuando se visita después de uno de sus padres y antes de otro. Por ejemplo, en el camino $A \rightarrow C \leftarrow B \rightarrow F$, C es un colisionador, porque viene precedido por A y sucedido por B , ambos padres del nodo C . Nótese que aunque C sea un colisionador en el camino descrito anteriormente, en el camino $A \rightarrow C \rightarrow F$ no es un colisionador. Esto significa que un colisionador se define respecto a caminos particulares. A continuación se define otra tipología de caminos que utiliza el concepto de colisionador:

- **Camino abierto:** Un camino está abierto si no existe ningún colisionador en él. Por ejemplo, el camino $A \rightarrow C \rightarrow F$ está abierto. Según [13], si existe un camino abierto entre dos nodos, las variables que estos representan pueden ser estadísticamente dependientes.
- **Camino cerrado:** Un camino está cerrado si existe algún colisionador en él. Por ejemplo, el camino $A \rightarrow D \leftarrow C \leftarrow B \rightarrow F$ está cerrado. Si todos los caminos entre dos nodos cualquiera están cerrados, estos están **d-separados**. Según [12], podemos deducir que si dos nodos están d-separados en el grafo, las variables que estos representan son independientes estadísticamente.

De esta manera, un camino puede ser causal o no causal y puede estar abierto o cerrado.

3.2.3. Manipulación de caminos

La naturaleza causal o no causal de un camino no se puede alterar. Sin embargo, sí que se puede manipular su estado (abierto o cerrado). Poder cambiar el estado de un camino de abierto a cerrado y viceversa nos ayudaría a controlar las relaciones estadísticas entre las respectivas variables. En este apartado se explica cómo podemos modificar el estado de un camino, para lo cual se presentan tres diagramas causales fundamentales en los que se identificará el camino existente y se procederá a su modificación. Antes de empezar con la explicación, cabe destacar que indicaremos los caminos con líneas discontinuas. Los caminos causales serán dibujados en verde, mientras que los no causales se representarán en rojo. Además, los caminos abiertos serán dibujados de nodo origen a nodo destino, mientras que los caminos cerrados serán abruptamente cerrados con una equis a mitad de camino.

El primer diagrama causal que vamos a estudiar tiene la forma $X \rightarrow Y \rightarrow Z$. En este diagrama, se reconoce un camino causal abierto entre X y Z , tal y como se ve en la Figura 3.8. Este camino se llama **camino directo**. Para cambiar el estado del camino a cerrado, se puede condicionar en un nodo intermedio del camino. En nuestro ejemplo, tal como se

muestra en la Figura 3.9, condicionar en Y cierra el camino. Al condicionar en Y , estamos fijando su valor, haciendo estadísticamente independientes X y Z .

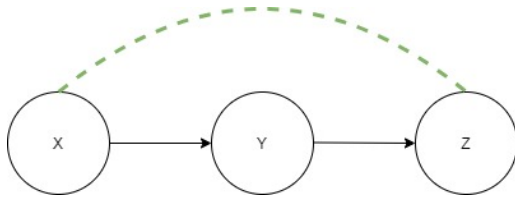


Figura 3.8: Camino directo

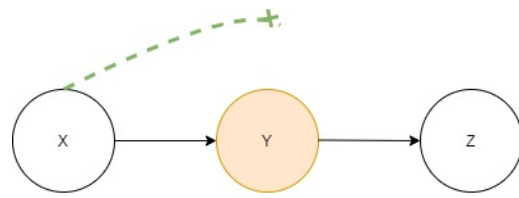


Figura 3.9: Camino directo ajustado

El segundo diagrama causal que vamos a estudiar tiene la forma $X \leftarrow Z \rightarrow Y$. En este se reconoce un camino no causal abierto (Figura 3.10). Este camino se denomina **camino trasero** o *backdoor path* y se forma a través de la presencia de antecesores comunes. Para cambiar el estado del camino a cerrado, deberíamos condicionar en una de las causas comunes presentes en el camino. En nuestro caso, condicionar en Z bastará para cerrar el camino, tal como se muestra en la Figura 3.11. Al condicionar en Z , estamos fijando su valor, haciendo estadísticamente independientes X e Y .

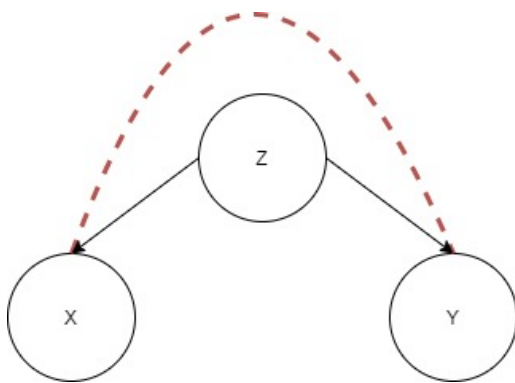


Figura 3.10: Camino trasero

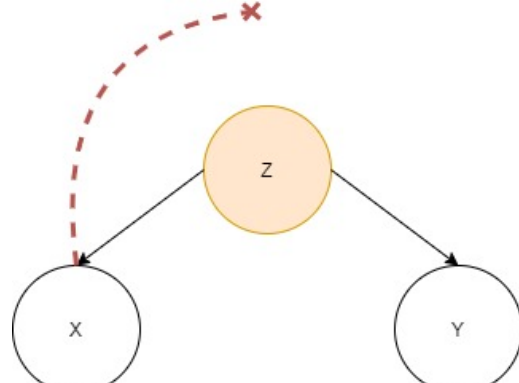


Figura 3.11: Camino trasero ajustado

Finalmente, el tercer diagrama causal por analizar tiene la forma $X \rightarrow Z \leftarrow Y$. En este se reconoce un camino no causal cerrado (ver Figura 3.12). Para cambiar el estado del camino a abierto, deberíamos condicionar en el colisionador (en este caso Z), tal como se muestra en la Figura 3.13. Al condicionar en Z , estamos fijando su valor, haciendo estadísticamente dependientes X e Y .

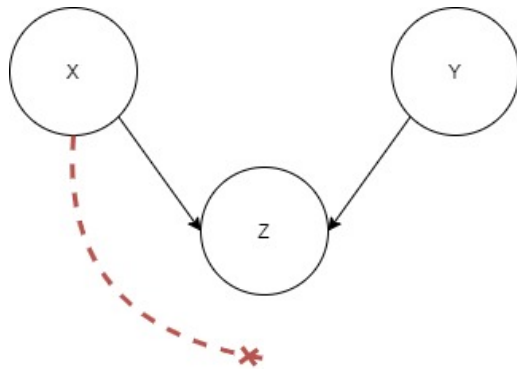


Figura 3.12: Camino cerrado

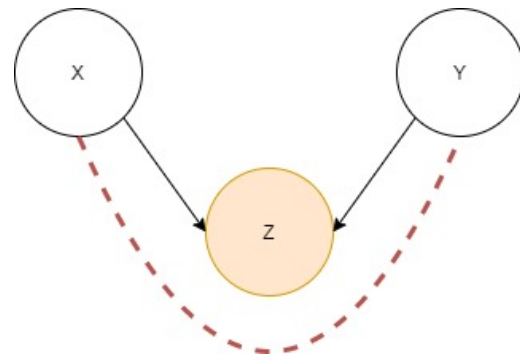


Figura 3.13: Camino abierto

Con estos tres ejemplos, escenificamos la manipulación básica de los caminos.

3.2.4. Implicaciones de los caminos en la inferencia causal

Una vez explicada la manipulación de caminos, sólo nos queda presentar su utilidad en la inferencia causal. Como ya se ha explicado anteriormente, el objetivo de la inferencia causal es encontrar el efecto del tratamiento en el resultado. Este cálculo no es trivial ya que pueden existir múltiples factores que influyen en ambas variables, dando así una falsa sensación de causalidad. Gracias a la identificación y manipulación de caminos en diagramas causales, no sólo se puede obtener una visualización de estas influencias, sino que también se consigue anularlas en la mayoría de los casos. Así, a continuación se presenta un ejemplo a través del cual que se ejemplificará la utilidad de los caminos en la inferencia causal.

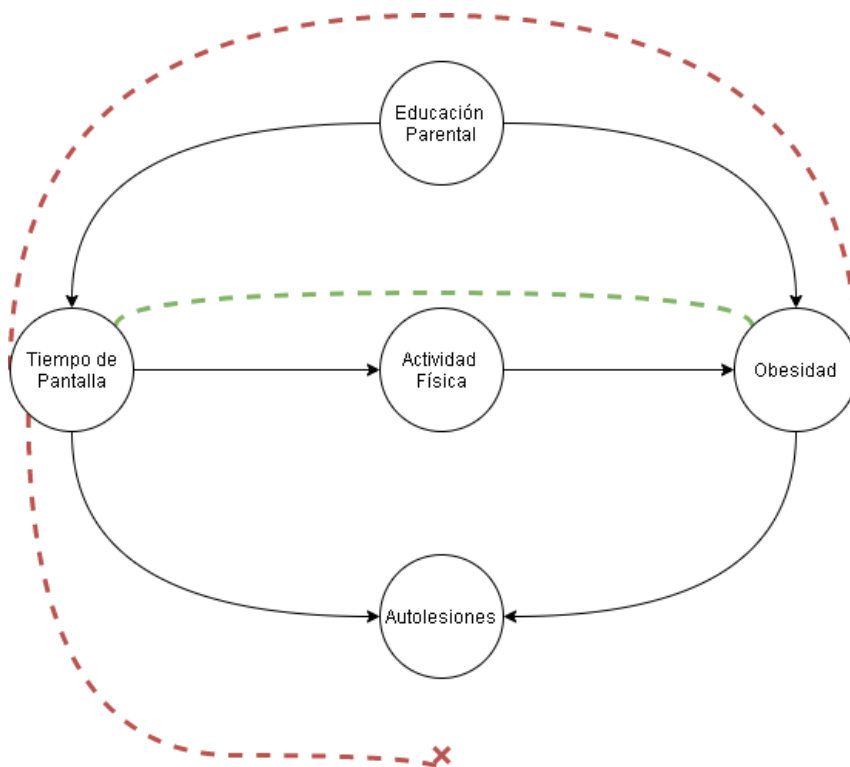


Figura 3.14: Ejemplo camino combinado

Imaginemos que queremos estudiar el efecto del *Tiempo de pantalla* en la *Obesidad*. En el diagrama causal de la Figura 3.14 encontramos tres caminos distintos entre *Tiempo de pantalla* y *Obesidad*: un camino directo (abierto y causal) a través de *Actividad Física*, un camino trasero (abierto y no causal) a través de *Educación Parental*, y finalmente un camino cerrado no causal a través de *Auto lesiones*. La presencia de caminos abiertos entre *Tiempo de pantalla* y *Obesidad* significa que estos pueden ser estadísticamente dependientes. Sin embargo, el camino abierto a través de *Educación Parental* no es causal y puede introducir sesgos en nuestros resultados. Al condicionar en *Educación Parental* se cierra el camino trasero, consiguiendo así que el único camino abierto entre el tratamiento y el resultado sea causal. De esta manera, los cambios en *Obesidad* sólo pueden ser atribuibles a *Tiempo de pantalla*.

En otras palabras, estamos interesados en cerrar todos los caminos no causales abiertos que existen entre el tratamiento y el resultado, pues estos representan influencias que queremos anular. En caso de no cerrar estos caminos se podrían introducir dos tipos de sesgos en nuestros resultados: sesgo por confusión y sesgo por selección.

- El **sesgo por confusión** está causado por caminos traseros que no se han cerrado.
- El **sesgo por selección**, a su vez, está causado por colisionadores que han sido condicionados. El sesgo por selección es particularmente peligroso porque las propias personas que conducen el estudio pueden no darse cuenta de que lo están induciendo. Por ejemplo, imaginemos que estudiamos el efecto de tomar pastillas anti-conceptivas en la trombosis. Para ello, escogemos como población del estudio cien mujeres que han estado hospitalizadas por trombosis. Los médicos del hospital confirman que las mujeres que presentan síntomas de trombosis son más propensas a estar hospitalizadas si además toman las pastillas anticonceptivas. Por lo tanto, el resultado que se obtendrá del estudio con esas cien mujeres seleccionadas presentará sesgo por selección, al haber escogido mujeres que estando hospitalizadas, es más probable que tomen o hayan tomado pastillas anticonceptivas, tal como se muestra en la Figura 3.15.

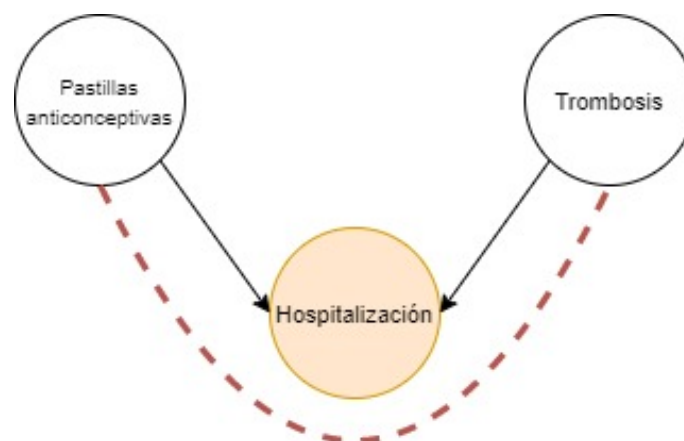


Figura 3.15: Ejemplo sesgo por selección

Para finalizar, simplemente destacar que los diagramas causales no sólo son una forma de estructurar el conocimiento del dominio y codificar asunciones, sino que también se pueden utilizar como instrumento para identificar el efecto del tratamiento. Para más información, consultar el apartado 3.3.2.

3.2.5. Cómo crear nuestro propio diagrama causal

Esta sección ofrece unas pautas sobre cómo construir diagramas causales [14]. Cabe destacar que los diagramas causales son una herramienta que nos ayuda a encontrar fuentes de sesgo al hacer inferencia causal y en ningún momento pretenden ser una representación exacta del mundo. A continuación, se presentan unos pasos que describen el proceso de creación de un diagrama causal:

1. Añadir las variables tratamiento y resultado.
2. Añadir causas comunes de todos los pares de variables presentes en el diagrama. Cada vez que se inserta una variable nueva en el grafo, deberemos repetir este paso. Para decidir qué causa qué, deberemos utilizar nuestro conocimiento experto del dominio. Si este es insuficiente para excluir un efecto, deberemos incluirlo. Sin embargo, si nuestro conocimiento experto nos permite excluir algunas flechas, las omitiremos.
3. Añadir variables representativas de los criterios de selección del estudio (si existen). En otras palabras, identificar posibles factores que hayan podido hacer que la población comparta algún tipo de característica. Este paso está enfocado en identificar posibles sesgos por selección.
4. En caso de que exista algún camino que no se pueda cerrar, añadir variables que puedan facilitar ese cómputo. Añadiendo alguna variable mediadora, se puede conseguir cerrar el camino o cuantificar el camino directo.

3.3 Inferencia causal en cuatro pasos

Una vez presentados los conceptos fundamentales de la inferencia causal así como las bases de los diagramas causales, en esta sección explicamos un procedimiento genérico para realizar inferencia causal, utilizando como base todo lo explicado anteriormente. El procedimiento consta de cuatro pasos tal y como se muestra en la Figura 3.16: modelización, identificación, estimación y refutación.

A grandes rasgos, la primera fase o **fase de modelización** se centra en diseñar el diagrama causal correspondiente. En la **fase de identificación**, partiendo de un estimando causal (que generalmente será el ATE), y teniendo en cuenta el diagrama causal previamente diseñado, se obtiene un estimando estadístico. En la tercera fase o **fase de estimación** utilizaremos ese estimando estadístico junto con los datos disponibles para calcular el valor estimado. Finalmente, en la **fase de refutación** utilizaremos ciertos métodos para poner a prueba la validez del modelo construido.

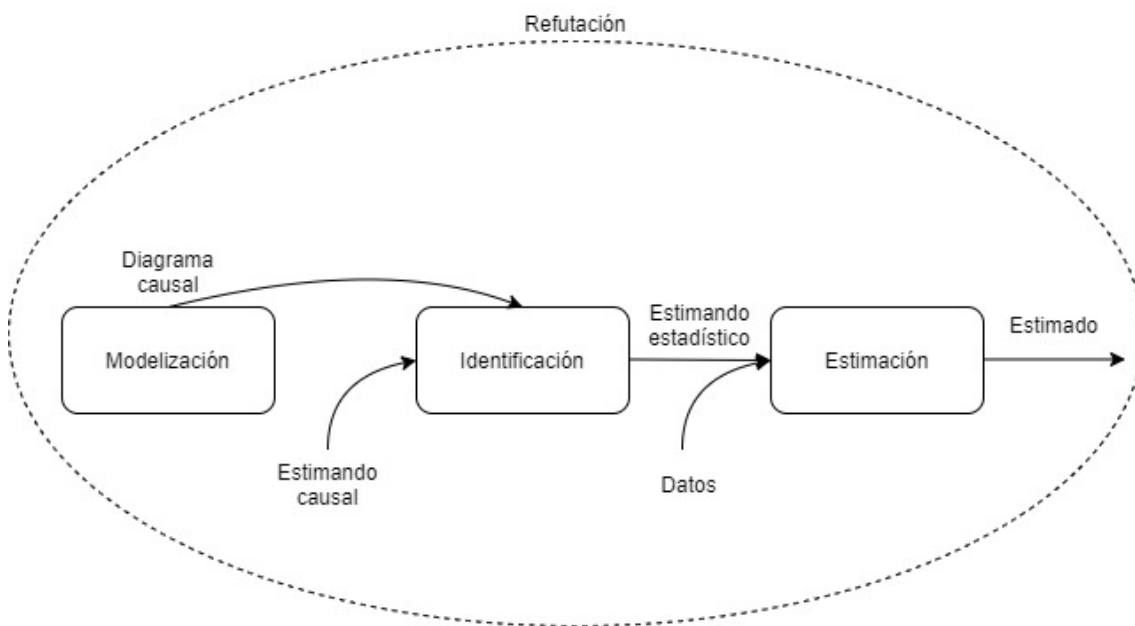


Figura 3.16: Proceso causal desglosado en cuatro pasos

3.3.1. Modelización

El proceso de modelización consiste en utilizar el conocimiento experto del dominio para construir el diagrama causal. El apartado 3.2 explica el concepto de diagrama causal, así como su importancia y cómo construirlo entre otras cosas. En el diseño del diagrama causal debe incluirse las variables tratamiento y resultado. Además, se considerará que dos variables mantienen una relación causal si no podemos descartar que no tengan relación alguna. En otras palabras, si no es posible asegurar que no existe una relación entre dos variables, se asumirá que estas variables están relacionadas y se incluirá una arista entre ellas. Se considera que los contenidos de la sección previamente mencionada son directrices suficientes para realizar esta primera fase.

3.3.2. Identificación

Antes de definir en qué consiste la fase de identificación, recordemos nuestro objetivo: se pretende aplicar inferencia causal para cuantificar el efecto medio del tratamiento en el resultado (ATE) visto en la Fórmula 3.1 con el fin de descartar o corroborar una relación causal entre el tratamiento y el resultado. Sin embargo, tal y como indican los operadores *do* de la expresión a estimar, para computarlo deberíamos realizar intervenciones en la población. En esta fase de la inferencia causal, se pretende encontrar una expresión equivalente al ATE que no contenga intervenciones. En otras palabras el proceso de identificación consiste en convertir la expresión causal que se quiere cuantificar (particularmente el ATE) en una expresión estimable. Esto se conseguirá a través del diagrama causal definido en la fase de modelización. A continuación, se introducen dos métodos gráficos que, en caso de ser aplicables, ayudan a convertir el ATE en una expresión sin intervenciones:

- **Ajuste de caminos traseros o *backdoor criterion*:** El ajuste de caminos traseros o *backdoor criterion* es un método gráfico de identificación cuyo objetivo es cerrar

todos los caminos traseros para cuantificar sólo aquellos que son causales. La intuición de este método se expone en la sección 3.2.4. Esta forma de identificación puede aplicarse si se cumplen dos condiciones:

1. Debe existir un conjunto de variables W cuyo condicionamiento cierre todos los caminos traseros entre nuestra variable **tratamiento** y nuestra variable **resultado**.
2. El conjunto de variables W no debe contener ningún descendiente de la variable **tratamiento**.

Si el conjunto de variables W satisface ambas condiciones, dada la asunción de modularidad explicada en 3.1.4 podemos identificar $P(Y|do(T))$ como:

$$P(Y|do(T)) = \sum_W P(Y|T, W)P(W) \quad (3.3)$$

Esta fórmula es una variante de la fórmula 3.2 (fórmula de ajuste). La única diferencia es que en la fórmula 3.2, el conjunto de variables a condicionar se corresponde con los padres del tratamiento, mientras que en la fórmula 3.3 W es el conjunto de nodos cuyo condicionamiento cierra todos los caminos traseros. Es preferible utilizar el ajuste de caminos traseros sobre el ajuste a todos los padres del tratamiento ya que este es más flexible ante las diferentes situaciones que pueden producirse, según se detalla en [15].

A continuación se muestra un ejemplo a través del diagrama causal de la Figura 3.17 dónde se analizará el efecto de la variable *Obesidad* en *Muerte prematura*.

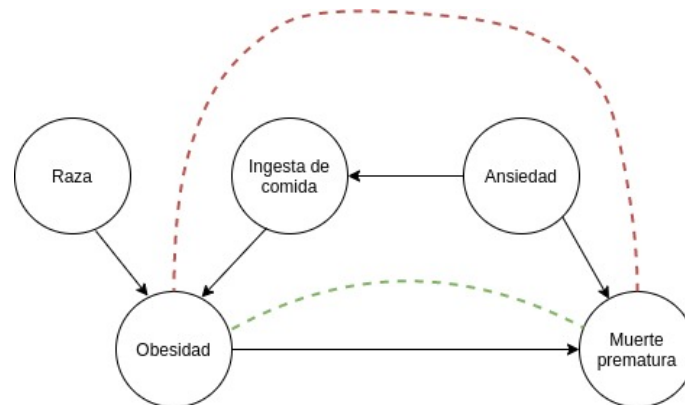


Figura 3.17: Diagrama causal para el estudio de la Obesidad en la Muerte prematura

Como se puede observar en el diagrama de la figura 3.17, existe un camino trasero de *Obesidad* a *Muerte prematura* a través de $Obesidad \leftarrow Ingesta\ de\ comida \leftarrow Ansiedad \rightarrow Muerte\ prematura$. En este caso, el conjunto de nodos $\{Raza, Ingesta\ de\ comida\}$ es un conjunto válido de variables tanto para aplicar el ajuste de caminos traseros como para aplicar la fórmula de ajuste (puesto que estos son los padres del tratamiento). Sin embargo, el ajuste de caminos traseros ofrece diferentes conjuntos de variables que satisfacen los requisitos previamente establecidos, ampliando así las alternativas de ajuste:

- $\{Ansiedad\}$

- {Ingesta de comida}
- {Ansiedad, Raza}
- {Ingesta de comida, Raza}
- {Ingesta de comida, Ansiedad}
- {Ansiedad, Ingesta de comida, Raza}

Cabe destacar que el **conjunto mínimo de ajuste** se corresponde con el mínimo conjunto de variables cuyo condicionamiento cerraría todos los caminos traseros. En este ejemplo, tanto el conjunto {Ansiedad} como el conjunto {Ingesta de comida} son conjuntos mínimos de ajuste.

- **Ajuste de caminos directos o *frontdoor criterion***: El ajuste de caminos directos o *frontdoor criterion* es un método gráfico de identificación cuyo objetivo es cuantificar el efecto de los caminos causales entre la variable tratamiento y la variable resultado utilizando un mediador entre estos. Un mediador es una variable intermedia entre el tratamiento y el resultado. Esta forma de identificación puede aplicarse si se cumplen tres condiciones:

1. Debe existir un conjunto de variables Z que intercepta todos los caminos directos entre el tratamiento y el resultado. Es decir, no debe existir ningún camino directo entre el tratamiento y el resultado que no pase por Z .
2. No debe existir ningún camino abierto entre la variable de tratamiento y el conjunto Z .
3. Todos los caminos traseros de Z al resultado serían cerrados condicionando en la variable tratamiento.

Si el conjunto de variables Z satisface todas las condiciones anteriores, y dada la asunción de modularidad explicada en 3.1.4, podemos reconocer $P(Y|do(T))$ de la siguiente manera, dónde T' se corresponderá a diferentes valores del tratamiento:

$$P(Y|do(T)) = \sum_Z [P(Z|T) \sum_{T'} P(Y|T', Z) P(T')] \quad (3.4)$$

Supongamos la situación de la Figura 3.18. Podemos observar que el efecto de *Obesidad* en *Muerte prematura* está mediado por la variable *Enfermedades*. Vamos a estudiar si en esta situación es aplicable el ajuste de caminos directos.

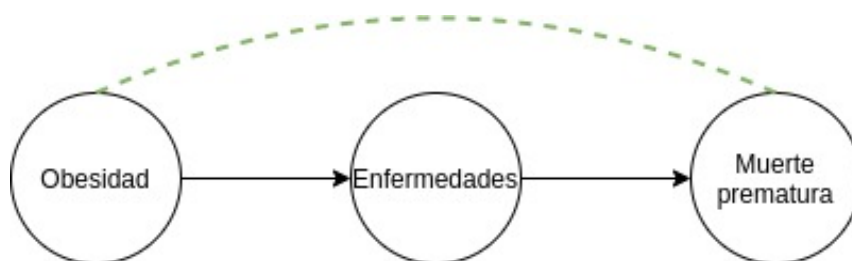


Figura 3.18: Diagrama causal para el estudio de la Obesidad en la Muerte prematura

Tomemos la variable *Enfermedades*, mediadora de *Obesidad* en *Muerte prematura*. Cumple con las tres condiciones impuestas por el ajuste de caminos directos o *frontdoor* puesto que:

1. No existe ningún camino directo entre *Obesidad* y *Muerte prematura* que no pase por *Enfermedades*.
2. No existe ningún camino trasero entre *Obesidad* y *Enfermedades*
3. Todos los caminos traseros de *Enfermedades* a *Muerte prematura* serían cerrados condicionando en la variable tratamiento (*Obesidad*).

Aplicar el criterio de *frontdoor* o caminos directos sería equivalente a calcular el efecto de *Obesidad* a *Enfermedades*, de *Enfermedades* a *Muerte prematura*, y combinar ambos.

De esta manera, se han explicado dos métodos gráficos que en caso de ser aplicables, consiguen encontrar una expresión equivalente en el *ATE* sin intervenciones: uno de ellos se basa en cerrar todos los caminos traseros entre el tratamiento y el resultado y el otro se basa en centrarse en los caminos directos entre estos. Existen otros métodos de identificación, como variables instrumentales o las reglas expresadas por Judea Pearl para realizar *do-calculus*, pero en este proyecto nos centramos exclusivamente en métodos gráficos.

3.3.3. Estimación

Una vez finalizada la fase de identificación, obtenemos una expresión estimable que es equivalente al efecto medio del tratamiento en el resultado (*ATE*) pero que no contiene intervenciones. Gracias a la ausencia de intervenciones, podemos estimar su valor a partir de los datos observados. Este valor estimado representará el efecto causal del tratamiento en el resultado. De esta forma, el proceso de estimación consiste en estimar o cuantificar la expresión obtenida en la fase de identificación utilizando los datos observados con el fin de descartar o corroborar la existencia de una relación causal entre la variable tratamiento y la variable resultado.

Existen diferentes métodos de estimación. A continuación, vamos a explicar sólo aquellos de los que se hará uso en este proyecto:

- **Regresión Lineal:** Se trata de construir un modelo de regresión lineal a partir de la expresión obtenida en el proceso de identificación para descubrir el efecto causal del tratamiento en el resultado. Si la relación entre el tratamiento y el resultado no es lineal, este método no garantiza proporcionar resultados correctos.
- **Propensity Score Matching:** Según [16], este método pretende estimar la expresión objetivo basándose en la idea de que las diferencias dadas en el resultado en individuos similares que han recibido tratamientos diferentes serán solamente debidas al tratamiento aplicado. En otras palabras, en este método pretendemos emparejar individuos con características similares pero con distinta relación con el tratamiento (para cada par, uno de ellos pertenecerá al grupo de control y el otro al grupo de tratamiento). De esta manera, encontrando individuos similares con tratamientos diferentes, conseguimos que las diferencias en el resultado no estén influenciadas por otros factores que no sean el tratamiento.

Para determinar la similitud entre dos individuos, se utiliza una medida llamada puntuación de propensión. Esta medida corresponde a la probabilidad de que un individuo reciba el tratamiento o no, dadas sus características. Este método es sólo aplicable a tratamientos binarios.

- **Propensity Score Stratification:** Este método también utiliza puntuaciones de propensión para estimar la expresión objetivo. Según [17], separamos a los individuos en diferentes grupos o estratos, teniendo en cuenta sus puntuaciones de propensión. En cada estrato o subconjunto, los individuos tendrán puntuaciones de propensión similares y por lo tanto, serán individuos con características similares. De esta manera, dentro de cada grupo o estrato podremos comparar los resultados entre los individuos que han recibido tratamiento y los individuos que no lo han recibido para computar directamente el efecto del tratamiento. Finalmente, podemos combinar los efectos del tratamiento calculados para cada estrato, para computar el efecto general del tratamiento.
- **Double Machine Learning:** Este método de estimación consiste en dos fases [10]. En la primera fase, definimos W como el conjunto de variables condicionadas en el estimando estadístico. Para ello se deben entrenar dos modelos: uno para predecir el tratamiento a partir de W y otro para predecir el resultado a partir de W . Estos dos modelos, a partir de unos valores específicos de W , se utilizarán para predecir unos valores del tratamiento T' y del resultado Y' . Una vez tengamos los dos modelos, en la segunda fase entrenamos un modelo para predecir la diferencia entre el resultado predicho y el real a partir de la diferencia entre el tratamiento predicho y el real. En otras palabras, entrenar un modelo para predecir $Y - Y'$ a partir de $T - T'$, para poder calcular así el efecto promedio del tratamiento.

Cualquiera de estos métodos nos ayudará a estimar la expresión obtenida en el apartado de identificación, obteniendo así el efecto causal del tratamiento en el resultado. Un efecto causal cercano a cero indica la inexistencia de una relación causal entre el tratamiento y el resultado.

3.3.4. Refutación

Una vez calculado el efecto causal del tratamiento en el resultado, queremos asegurarnos de que no hemos cometido ningún error básico en el proceso. Así, en la fase de refutación simplemente se tratará de ejecutar una serie de pruebas para asegurar que estamos utilizando una configuración válida. Ser capaces de pasar estas pruebas no quiere decir que el modelo generado sea correcto, simplemente es una indicación de que no tenemos ninguna prueba para desconfiar de él. Es decir, estas pruebas son necesarias pero no suficientes.

1. **Validación Bootstrap:** En esta prueba, simplemente comprobamos que el efecto calculado no cambia significativamente a pesar de utilizar muestras aleatorias e incluso duplicadas de los datos originales.
2. **Cambiar el tratamiento por un placebo:** Esta prueba consiste en sustituir los valores del tratamiento por los valores de una variable independiente aleatoria. Como el tratamiento es aleatorio, el efecto estimado debería ser cero.
3. **Añadir causa común:** Esta prueba consiste en añadir una causa común aleatoria entre el tratamiento y el resultado y volver a calcular el efecto causal entre estos. Como esta variable es estadísticamente independiente tanto del tratamiento como del resultado (puesto que es aleatoria), aunque la marquemos como causa común no debería provocar ningún cambio. Por lo tanto el efecto estimado debería mantenerse igual que el original.

4. **Validación por subconjunto:** Esta prueba consiste en recalcular el efecto al reemplazar el conjunto de valores por un subconjunto de este. Al ser un subconjunto aleatorio, el efecto debería ser el mismo.

No pasar estas pruebas implica que el modelo no es correcto y deberíamos revisarlo.

A parte de estas pruebas, también utilizaremos el *p-value*. En estadística, este valor es una probabilidad que nos indica si el resultado obtenido ha sido producto del azar del muestreo, o si los resultados son estadísticamente significativos. Lo utilizaremos de la siguiente manera: si al calcular el efecto medio del tratamiento en el resultado obtenemos un *p-value* menor o igual a 0.05 (y teniendo en cuenta también los resultados de las anteriores pruebas de refutación), consideraremos que nuestros resultados son correctos. Por el contrario, si obtenemos un *p-value* mayor que 0.05, deberemos descartar cualquier resultado obtenido y asumir que el efecto medio del tratamiento en el resultado es cero.

Herramienta propuesta: CausalInf

En este capítulo se presenta CAUSAINF, una herramienta que permite a un usuario poner en práctica los conceptos vistos en el capítulo anterior y analizar relaciones causales a partir de un conjunto de datos.

Primero se introduce brevemente la herramienta CAUSAINF y sus funcionalidades, así como el diseño general de la herramienta y las tecnologías utilizadas. Seguidamente, se detallará de forma breve algunos aspectos de la implementación, incluyendo los requisitos del sistema.

4.1 Diseño de la herramienta CausalInf

CAUSAINF constituye la solución propuesta en este trabajo con el fin de facilitar el proceso de inferencia causal a muchos usuarios. Esta herramienta, se apoya principalmente en dos tecnologías de inferencia causal: *DAGitty* y *DoWhy*. Procedemos en primer lugar a presentar brevemente estas dos tecnologías y posteriormente, se detallará el papel que desempeñan en CAUSAINF.

- *DAGitty*: *DAGitty* [18] es una herramienta software enfocada a investigadores y estudiantes de disciplinas empíricas (epidemiología, sociología, psicología, etc.) cuyo objetivo es facilitar la creación y el análisis de diagramas causales. Para ello ofrece de diversas funcionalidades:
 - La derivación de implicaciones demostrables a partir de un diagrama causal.
 - La identificación del conjunto mínimo de ajuste.
 - El diagnóstico de ajuste insuficiente o incorrecto a través de la identificación de caminos de sesgo.

Esta herramienta nos permitirá realizar satisfactoriamente la **fase de modelización** de la inferencia causal. Se ha escogido *DAGitty* sobre otras herramientas por su relación directa con el estudio causal (existen otras herramientas de ámbito más general que no proporcionan funcionalidades como la identificación del conjunto mínimo de ajuste), por su facilidad de uso y claridad, y por su adaptabilidad. Permite estudiar el diagrama causal y todas sus implicaciones de forma dinámica y visual, cambiando fácilmente las conexiones y reajustando toda la información dinámicamente. De esta manera, *DAGitty* constituye un lienzo sobre el que podemos

crear y analizar el diagrama causal objetivo. Más información sobre *DAGitty* puede consultarse en el manual de usuario [19].

- **DoWhy:** *DoWhy* [20] es una librería en Python mantenida por *Microsoft* cuyo objetivo es clarificar y facilitar el pensamiento y análisis causal. Esta librería basa el proceso de inferencia causal en las cuatro fases mencionadas en la sección 3.3: modelización, identificación, estimación y refutación.

Se ha escogido *DoWhy* sobre otras por los siguientes motivos:

- Proporciona un marco formal para hacer asunciones explícitas. Esto lo hace a través de los diagramas causales como parámetro del modelo.
- Separa la fase de identificación y la de estimación.
- Integra la fase de estimación con librerías de *machine learning* (en este proyecto se utiliza sólo *EconML*).
- Extiende diferentes pruebas de refutación para comprobar la consistencia y robustez del modelo.

La Figura 4.1 muestra la interfaz de CAUSAINF que se ha construido sobre el código principal *HTML* de *DAGitty*. En esta interfaz se han incluido unas modificaciones, las cuales aparecen resaltadas en un recuadro rojo en la Figura 4.1 y que consisten en la inserción de dos botones, *Choose File* y *Compute ATE*, que sirven, respectivamente, para cargar archivos e iniciar un proceso de inferencia causal a partir del grafo creado y los datos subidos. En la Figura 4.2 se han ampliado los dos botones insertados. De esta manera, CAUSAINF proporciona tres funcionalidades principales:

- crear y analizar diagramas causales
- subir datos a la herramienta y realizar un proceso de inferencia causal
- seleccionar el método de identificación y estimación para obtener el efecto medio del tratamiento

The screenshot displays the CAUSAINF web interface. On the left, there is a sidebar with various settings and a legend. The main area shows a causal diagram with nodes A, B, Z, E, and D. Node A is a yellow circle, B is a blue circle, Z is a grey circle, E is a yellow circle with a play button, and D is a blue circle with an information icon. Arrows indicate causal relationships: A to Z, B to Z, E to D, and A to D. On the right, there is a panel with analysis results, including 'Causal effect identification', 'Testable implications', 'Model code', and 'Summary'. A red box highlights the 'Upload csv' and 'Compute causal effect' sections, which contain a 'Choose File' button and a 'Compute ATE' button.

Figura 4.1: Interfaz de CAUSAINF

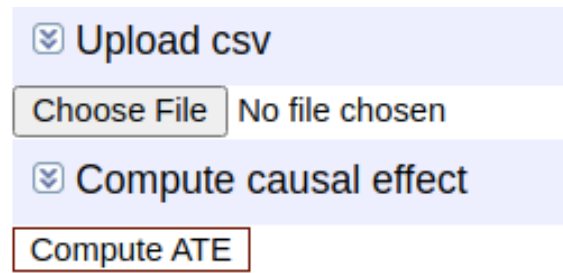


Figura 4.2: Botones insertados en la página principal

Para crear y analizar diagramas causales en CAUSAINF, simplemente es necesario crear el diagrama causal objetivo en el lienzo, marcando las variables que se corresponden con el tratamiento y con el resultado. Una vez marcadas dichas variables se mostrará información deducible a partir del diagrama, tal como el conjunto mínimo de ajuste o implicaciones estadísticas. Esta funcionalidad la proporciona directamente *DAGitty*.

El proceso para cargar datos en la herramienta se realiza a través del botón *Choose File*. Este botón, mostrará a su vez el sistema de archivos local y se seleccionará el *csv* deseado. Cabe destacar que las columnas del *csv* deben tener exactamente el mismo nombre que los nodos del diagrama causal para una correcta identificación.

Finalmente, para realizar inferencia causal en CAUSAINF, deberemos cerciorarnos previamente de dos precondiciones básicas: que el diagrama causal está construido (con las variables tratamiento y resultado marcadas), y que se han cargado los datos. Una vez se cumplen ambas condiciones, deberemos hacer clic en *Compute ATE* para iniciar el proceso de inferencia causal y aparecerán una secuencia de diálogos. El primer diálogo (Figura 4.3) nos muestra todas las opciones de la fase de identificación. El usuario debe seleccionar la opción deseada y confirmar y aparecerá el segundo diálogo (Figura 4.4), dónde se selecciona el método de estimación que se desea aplicar. Tras confirmar el método, el tercer diálogo (Figura 4.5) nos muestra el efecto medio del tratamiento en el resultado y nos da la opción de realizar pruebas de refutación. Opcionalmente, si accedemos a refutar el modelo, nos aparecerá un cuarto y último diálogo (Figura 4.6) dónde se visualizarán los resultados de las pruebas de refutación realizadas y el *p-value* obtenido.



Identification

Please, choose a method in order to identify the ATE

- Adjust for variables: Z,B
- Adjust for variables: B
- Do not adjust for any variable
- iv: A

OK

Figura 4.3: Diálogo fase de identificación



Estimation

Please, choose a method in order to estimate the ATE

- Double Machine Learning
- Linear Regression
- Propensity Score Matching
- Propensity Score Stratification

OK

Figura 4.4: Diálogo fase de estimación

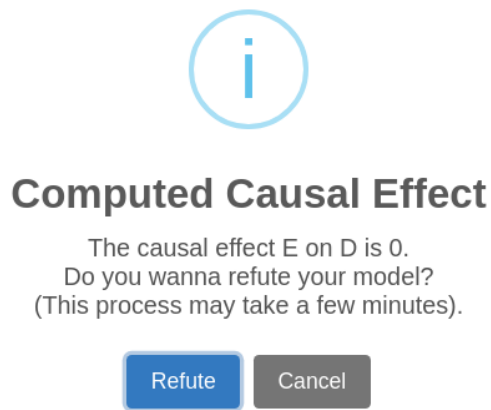


Figura 4.5: Diálogo efecto del tratamiento en el resultado

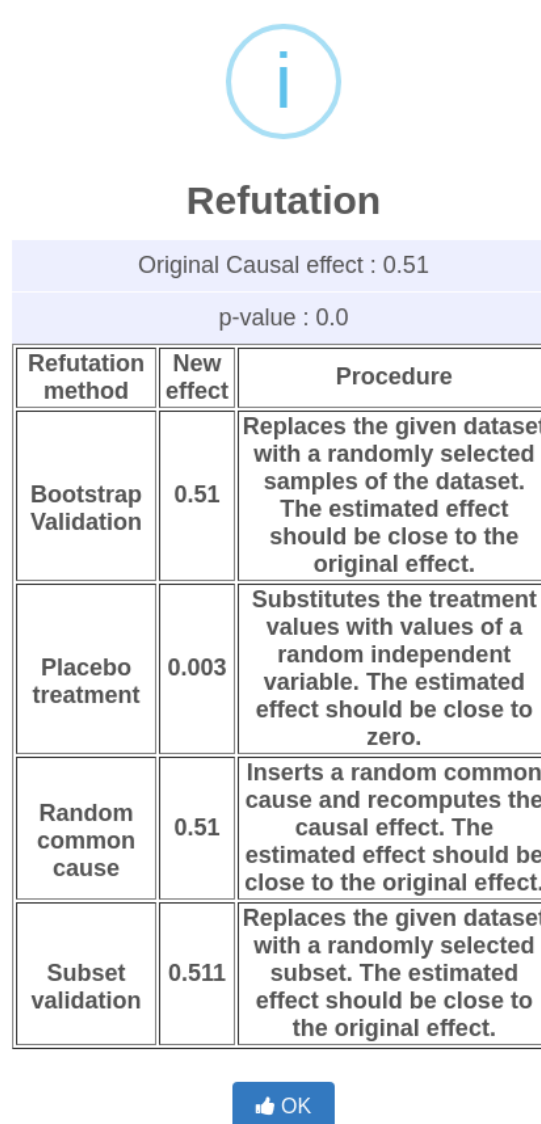


Figura 4.6: Diálogo fase de refutación

Por último, destacamos que la herramienta *DoWhy* es la que realiza los cálculos mostrados en los diferentes diálogos, conectando la interfaz de CAUSAINF con los resultados obtenidos en *DoWhy*. Es decir, cuando se hace clic en el botón *Compute ATE* se pasa el grafo, los datos y los nodos tratamiento y resultado a *DoWhy* para que esté encuentre las posibles opciones para realizar la fase de identificación. Una vez encontradas *DoWhy* las envía a *DAGitty* modificado para que las muestre como opciones al usuario, que selecciona una opción y la vuelve a enviar a *DoWhy*. Básicamente, tal y como podemos ver en la Figura 4.7, esta funcionalidad establece una conexión entre la interfaz de CAUSAINF y *DoWhy*, para realizar satisfactoriamente las cuatro fases de la inferencia causal previamente explicadas.

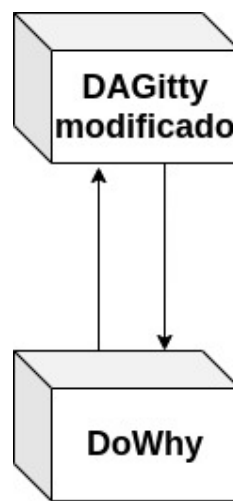


Figura 4.7: Diseño simplificado de CAUSAINF

4.2 Implementación de la herramienta CausalInf

Como se ha explicado en la sección 4.1, CAUSAINF se implementa mediante una conexión entre una versión modificada de *DAGitty* y *DoWhy*. En esta sección se aborda el proceso para crear la conexión entre estas dos herramientas.

Tal y como se puede ver en la Figura 4.8, *DAGitty* y *DoWhy* se conectan mediante *Flask*. *Flask* es un *framework* que permite crear aplicaciones web en *Python*, y se ha utilizado tanto para cargar el *frontend* como para crear todo el *backend* de la herramienta.

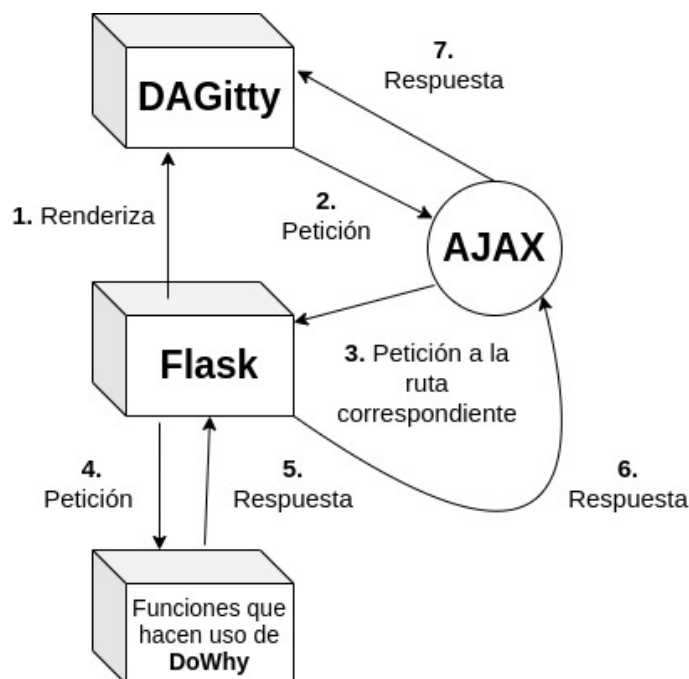


Figura 4.8: Esquema de la infraestructura de CAUSAINF

El *backend* consiste en diferentes rutas que hacen uso de *DoWhy*. Todas estas rutas son accesibles directamente desde el código *HTML* de *DAGitty* modificado gracias al uso de *Ajax* (*Asynchronous JavaScript And XML*), una librería *JavaScript* de desarrollo web que nos permite enviar y recibir datos en segundo plano y sin recargar, dinamizando así la página web. De esta manera, *Flask* renderiza la página principal y esta hace uso de las rutas anteriormente especificadas. A continuación, se explicará brevemente qué rutas existen y para qué se utilizan:

- */*: Esta ruta es la ruta raíz y es la que se ejecuta por defecto al acceder a la *url* de CAUSAINF. Se encarga de renderizar el *HTML* principal. Esta ruta es la única a la que el usuario tiene acceso directo y se corresponde al paso 1 de la Figura 4.8.
- */upload-csv*: Dado el contenido de un fichero *csv*, esta ruta guarda su respectivo *dataframe* de *pandas* en un objeto sesión, haciéndolo accesible a otras rutas. Esta ruta se activa cuando apretamos el botón *Choose File*, la cual se corresponde con los pasos 2-7 de la Figura 4.8.
- */retrieve-identification-options*: Dado el grafo, el tratamiento y el resultado, esta ruta utiliza la librería *DoWhy* para devolver todas las opciones de identificación posibles. Esta ruta se activa en la secuencia de diálogos para realizar inferencia causal y se corresponde con los pasos 2-7 de la Figura 4.8.
- */compute-estimation-methods*: Esta ruta devuelve los métodos de estimación disponibles en el *backend*. Se activa en la secuencia de diálogos para realizar inferencia causal y se corresponde con los pasos 2-7 de de la Figura 4.8.
- */compute-effect-with-estimand-and-estimator*: Dado un método de identificación y un método de estimación, esta ruta devuelve un diccionario que especifica el tratamiento, el resultado y el efecto causal del tratamiento en el resultado. Esta ruta se

activa en la secuencia de diálogos para realizar inferencia causal y se corresponde con los pasos 2-7 de la Figura 4.8.

- *refutation-tests*: Esta ruta realiza unas pruebas de refutación y devuelve un diccionario con los resultados obtenidos. Esta ruta se activa en la secuencia de diálogos para realizar inferencia causal y se corresponde con los pasos 2-7 de la Figura 4.8.

Haciendo uso de estas rutas conseguimos las funcionalidades descritas en el apartado de diseño.

En el Capítulo 5 se mostrará la utilización de CAUSAINF a través de dos estudios.

4.2.1. Requisitos del sistema

Existen dos aspectos necesarios para poder ejecutar CAUSAINF:

- Tener *Python 3.8* o superior y *Poetry 1.1.7* instalados en nuestro sistema. *Poetry* es una herramienta que permite manejar las dependencias del proyecto de manera automática, y se puede instalar a través de *pip*.
- CAUSAINF, al igual que *DAGitty*, debe ejecutarse en un navegador que soporte *Javascript*, *HTML* y *SVG*.

CAPÍTULO 5

Estudios experimentales

En este capítulo se muestra la utilización de CAUSAINF en dos ámbitos diferentes. Particularmente, el primer estudio corresponde a un entorno médico y el segundo aborda relaciones en un ámbito psicológico. De esta manera ejemplificaremos el uso de la herramienta CAUSAINF en dos entornos de distinta naturaleza.

5.1 Estudio aplicado al ámbito médico

En esta sección se pretende ilustrar tanto el funcionamiento de la herramienta como los pasos de la inferencia causal a través de un ejemplo del ámbito médico.

LUCAS (*LUnG CAncer Simple set*) es un conjunto de datos cuyo objetivo es el diagnóstico y prevención del cáncer de pulmón. En este conjunto de datos disponemos de dos mil filas que representan las unidades o individuos de la población, y varias columnas que representan las variables o características de los individuos. Estos datos han sido generados artificialmente a partir de la red bayesiana causal de la Figura 5.1, dónde cada nodo representa una variable y los arcos representan relaciones causales entre las variables.

Las variables disponibles en este escenario son: *Ansiedad*, *Presión de grupo*, *Fumar*, *Dedos amarillos*, *Alergia*, *Cáncer de pulmón*, *Tos*, *Fatiga*, *Genética*, *Déficit de atención*, *Accidente de coche* y *Nacido un día par*. Todas ellas son variables binarias y tendrán valor 1 si se cumplen para el individuo, y 0 si no se cumplen. Por ejemplo, un individuo para el que la variable *Ansiedad* es 1 nos indica que ese individuo sufre ansiedad. Además, también tenemos las probabilidades condicionales asociadas a los diferentes nodos que se han aplicado para generar los datos. Algunos ejemplos de probabilidades dadas son:

- $P(\text{Ansiedad} = 1) = 0.64$
- $P(\text{Presión de grupo} = 1) = 0.33$
- $P(\text{Fumar} = 1 | \text{Presión de Grupo} = 1, \text{Ansiedad} = 1) = 0.91$

A partir de la información disponible (el conjunto de datos, las probabilidades condicionales y la red causal bayesiana), procedemos a aplicar los pasos de la inferencia causal para estudiar si realmente existe una relación causa-efecto entre las diferentes variables definidas o simplemente se identifica una correlación entre ellas. Específicamente, a través de CAUSAINF estudiaremos las relaciones entre:

- El efecto de *Fumar* en *Cáncer de Pulmón*.

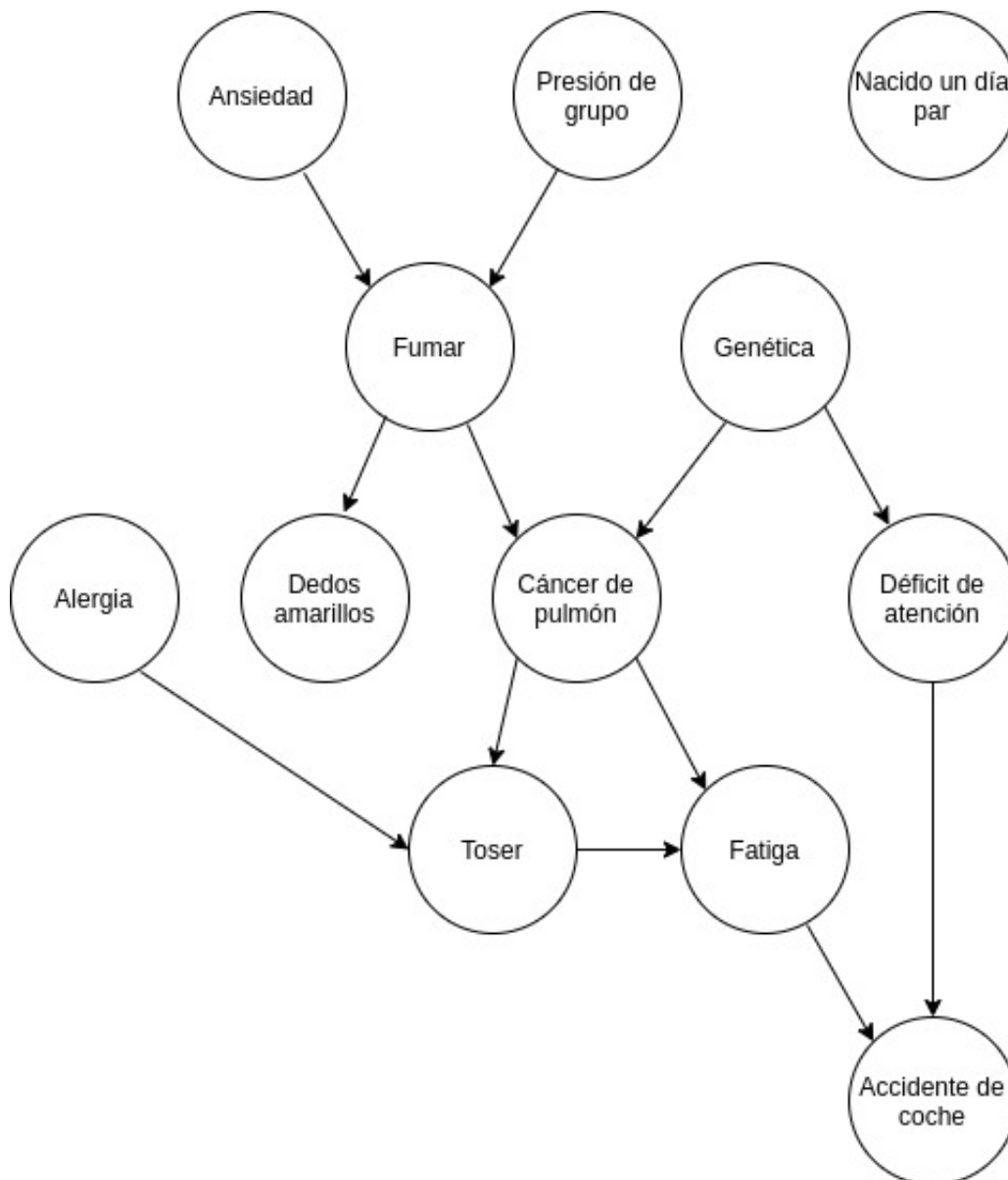


Figura 5.1: Red bayesiana causal de LUCAS.

- El efecto de *Nacido un día par* sobre *Cáncer de Pulmón*.
- El efecto de *Fatiga* en *Accidente de Coche*.
- El efecto de *Presión de Grupo* en *Dedos Amarillos*.

5.1.1. Efecto de *Fumar* en *Cáncer de Pulmón*

Procedemos a comprobar si existe una relación causal entre *Fumar* y *Cáncer de Pulmón*. El primer paso de este proceso es el de modelización, mediante el cual se determinan las asunciones que tenemos sobre el dominio y se codifican a través del diagrama causal.

Para realizar estas asunciones, nos basamos en la Figura 5.1. Por lo tanto, en este estudio estamos asumiendo que tanto *Ansiedad* como *Presión de grupo* son una causa de *Fumar*.

Fumar, a su vez, provoca *Dedos amarillos* y *Cáncer de pulmón*. *Cáncer de pulmón*, aparte de *Fumar*, también puede tener como precursor *Genética*. *Toser* tiene como causa *Alergia* o *Cáncer de pulmón*, y tiene efecto en *Fatiga* (que también puede venir provocada por *Cáncer de pulmón*). Finalmente, *Genética*, a parte de *Cáncer de pulmón*, también puede preceder al *Déficit de atención*, y tanto este como *Fatiga* pueden ser el detonante de *Accidente de coche*.

Todas las relaciones causales entre variables que no hayan sido mencionadas en el anterior párrafo, se asume que no existen. Por ejemplo, asumimos que *Nacido un día par* no tiene absolutamente ningún efecto en el resto de variables del dominio y es completamente independiente de estas. También asumimos que *Ansiedad* no tiene ninguna relación con *Presión de grupo*, ni con *Genética*, ni con *Alergia*, etc.

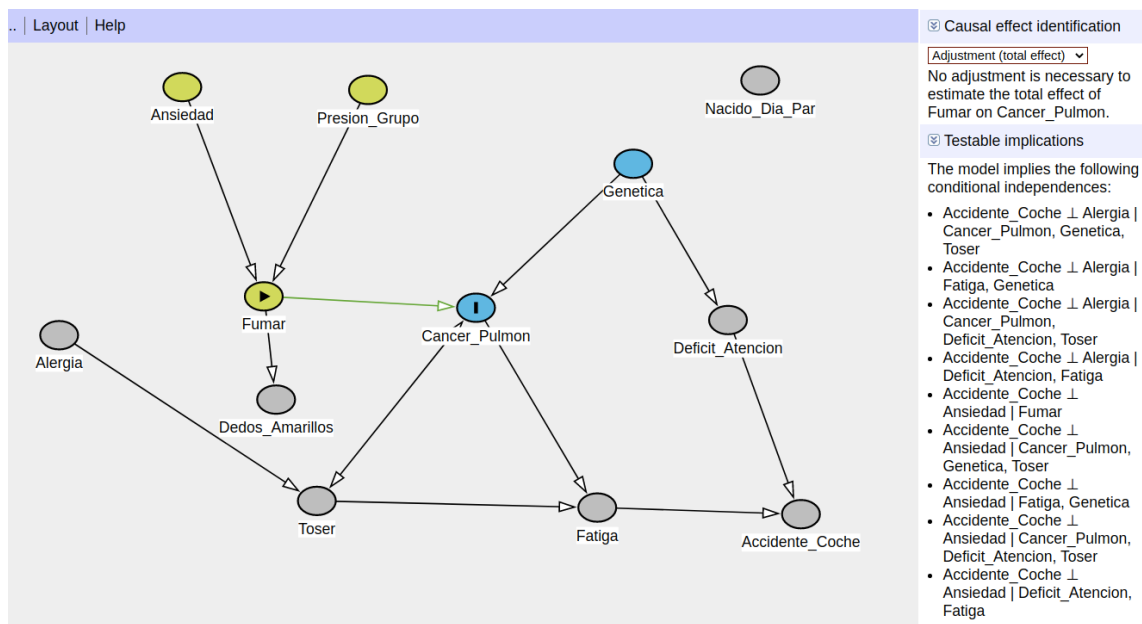


Figura 5.2: Diagrama causal representativo del efecto de *Fumar* en *Cáncer de Pulmón*

El primer paso es construir en CAUSALINF el diagrama causal que representará nuestras asunciones, especificando que *Fumar* es la variable tratamiento y *Cáncer de Pulmón* es la variable resultado (Figura 5.2). Una vez señaladas las variables tratamiento y resultado en el diagrama causal, CAUSALINF nos muestra información calculada exclusivamente a partir del diagrama causal: el mínimo conjunto de ajuste detectado y las implicaciones estadísticas entre variables:

- El mínimo conjunto de ajuste detectado para calcular el efecto de *Fumar* en *Cáncer de Pulmón* es el conjunto vacío. Esto es así porque la herramienta no encuentra ningún camino abierto no causal entre *Fumar* y *Cáncer de Pulmón*. Por lo tanto, en la parte superior del panel derecho de la Figura 5.2 se indica que no necesitamos ajustar en ninguna variable para estimar el efecto causal sin sesgos.
- Las implicaciones estadísticas entre variables simplemente son las implicaciones estadísticas deducibles a partir de nuestro diagrama, como por ejemplo que *Accidente de Coche* es independiente de *Alergia* dado *Cáncer de Pulmón*, *Genética* y *Toser*.

Una vez clara la información disponible, y antes de empezar con el proceso de identificación, deberemos subir el conjunto de datos LUCAS en la herramienta. Bastará con

hacer clic en el botón *Choose File...* (situado dentro del desplegable *Upload csv*) y seleccionar el archivo *csv* correspondiente. Apretando el botón *Compute ATE* procedemos con la fase de identificación. En esta fase decidimos el método de identificación de los presentados en 3.3.2 se va a utilizar para convertir el *ATE* en una expresión equivalente sin intervenciones, así como las variables que vamos a utilizar en el proceso. Recordemos que con esto queremos computar el efecto del tratamiento en la población anulando posibles influencias externas o sesgos con el fin de simular una intervención. Una vez apretado el botón *Compute ATE*, se despliega un diálogo con todas las opciones disponibles (Figura 5.3). Debemos seleccionar la que queramos (todas las opciones ofrecidas son correctas).



Figura 5.3: Conjuntos de ajuste disponibles para computar el efecto de *Fumar* en *Cáncer de Pulmón*

Previamente, la herramienta ha señalado que no es necesario ajustar en ninguna variable para computar el efecto medio de *Fumar* en *Cáncer Pulmón*. Esto quiere decir que asumimos que no existe ninguna causa común entre nuestras variables tratamiento y resultado, y que, por consiguiente, tanto el grupo fumador como el grupo no fumador tienen características similares. De esta manera, escogeremos no ajustar en ninguna variable, asumiendo así que el efecto es directamente computable sobre los datos. Una vez seleccionada y confirmada la opción deseada, seguiremos con el proceso de estimación. Escogiendo *Double Machine Learning*, nos aparece que el efecto medio de *Fumar* en *Cáncer de Pulmón* es 0.51, tal y como se puede ver en la Figura 5.4.



Computed Causal Effect

The causal effect Fumar on Cancer_Pulmon is 0.508.

Do you wanna refute your model?
(This process may take a few minutes).

Refute

Cancel

Figura 5.4: Efecto de *Fumar* en *Cáncer de Pulmón*

Antes de sacar conclusiones sobre este resultado, es recomendable realizar una serie de pruebas sobre el modelo construido para comprobar su robustez. Así, apretando en el botón *Refute*, realizamos las pruebas de refutación explicadas en 3.3.4. Los resultados se muestran en la Figura 5.5.

Original Causal effect : 0.508		
p-value : 0.001		
Refutation method	New effect	Procedure
Bootstrap Validation	0.505	Replaces the given dataset with a randomly selected samples of the dataset. The estimated effect should be close to the original effect.
Placebo treatment	-0.001	Substitutes the treatment values with values of a random independent variable. The estimated effect should be close to zero.
Random common cause	0.492	Inserts a random common cause and recomputes the causal effect. The estimated effect should be close to the original effect.
Subset validation	0.507	Replaces the given dataset with a randomly selected subset. The estimated effect should be close to the original effect.

Figura 5.5: Refutación del modelo construido para el efecto de *Fumar* en *Cáncer de Pulmón*

En la Figura 5.5, no se observa ningún indicio de que nuestro modelo sea incorrecto y, por lo tanto, podemos interpretar los resultados obtenidos. Podríamos decir que si tuviéramos cien unidades de nuestra población a las que hemos obligado a fumar y cien unidades a las que hemos obligado a no fumar, existirían cincuenta y una personas más con cáncer de pulmón en el grupo al que hemos obligado a fumar que en el que no. Por lo tanto, *Fumar* y *Cáncer de Pulmón* efectivamente mantienen una relación causal en nuestra población.

5.1.2. Efecto de *Nacido un día par* en *Cáncer de Pulmón*

En este apartado, se quiere estudiar el efecto de *Nacido un día par* en *Cáncer de Pulmón*. Para ello, tal y como se ha hecho en la sección 5.1.1, se va a utilizar el diagrama causal definido en la Figura 5.1, con la excepción de que se añadirá una arista de *Nacido un día par* a *Cáncer de Pulmón*. Esta arista representa la relación causal que queremos corroborar o descartar. De esta manera, las asunciones son las mismas que en el anterior apartado con la única diferencia de que ahora, asumimos que *Nacido un día par* causa *Cáncer de Pulmón*. En la fase de modelización, simplemente bastará con construir el diagrama causal previamente descrito en CAUSALINF (Figura 5.6) y subir los datos.

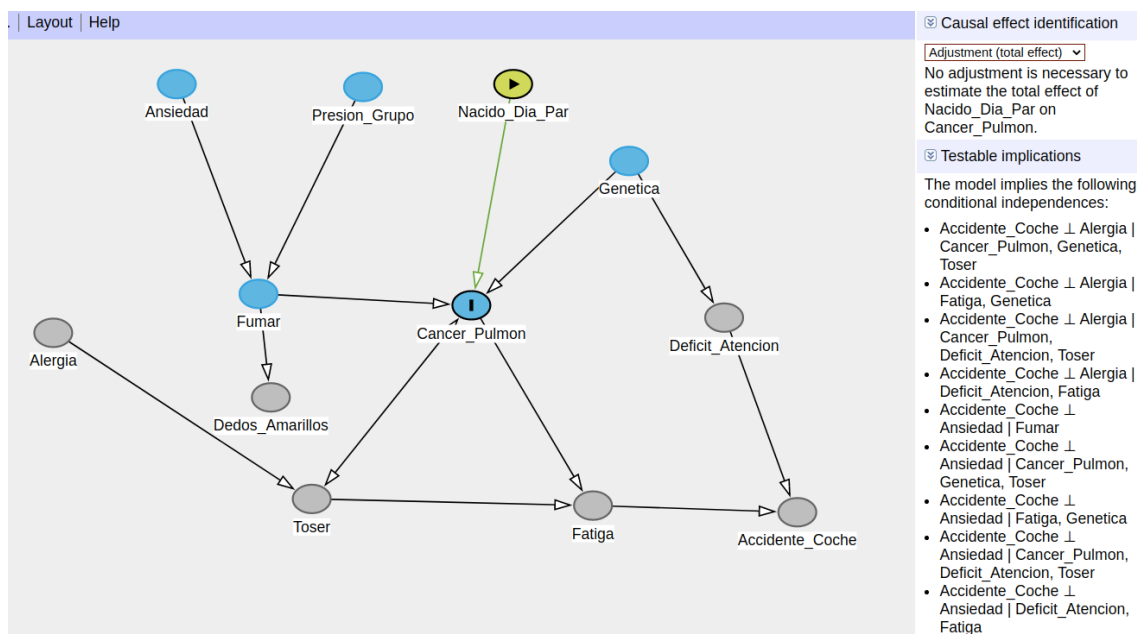


Figura 5.6: Diagrama causal representativo del efecto de *Nacido un día par* en *Cáncer de Pulmón*

En la fase de identificación, deberemos seleccionar nuevamente el método de identificación que queremos utilizar. Tal y como podemos observar en la Figura 5.6, no es necesario ajustar en ninguna variable para estimar el efecto deseado. Nuevamente no encontramos influencias que puedan introducir sesgos en nuestros resultados. De esta manera, seleccionamos la opción de no ajustar en ninguna variable y continuamos con el proceso de inferencia causal.

En la fase de estimación, seleccionamos regresión lineal. La herramienta nos indica que el efecto medio computado es de 0. En la Figura 5.7 podemos observar que las pruebas realizadas no indican que el modelo tenga ningún defecto.

Original Causal effect : -0.003		
p-value : 0.0		
Refutation method	New effect	Procedure
Bootstrap Validation	-0.004	Replaces the given dataset with a randomly selected samples of the dataset. The estimated effect should be close to the original effect.
Placebo treatment	0.002	Substitutes the treatment values with values of a random independent variable. The estimated effect should be close to zero.
Random common cause	-0.004	Inserts a random common cause and recomputes the causal effect. The estimated effect should be close to the original effect.
Subset validation	-0.003	Replaces the given dataset with a randomly selected subset. The estimated effect should be close to the original effect.

Figura 5.7: Refutación del modelo construido para el efecto de *Nacido un día par* en *Cáncer de Pulmón*

De esta manera, podemos concluir que en nuestra población, las variables *Nacido un día par* y *Cáncer de Pulmón* no mantienen ninguna relación causal, y deberíamos borrar la arista creada en la primera fase.

5.1.3. Efecto de *Fatiga* en *Accidente de Coche*

Seguidamente, aplicaremos inferencia causal para estudiar la relación existente entre *Fatiga* y *Accidente de Coche*. En el primer paso o paso de modelización, construiremos el diagrama causal en la herramienta. Vamos a asumir que el diagrama causal se corresponde con el diagrama causal de la Figura 5.2. Una vez construido, marcamos *Fatiga* como variable tratamiento y *Accidente de Coche* como resultado (Figura 5.8). CAUSALINF nos indica tres mínimos conjuntos de ajuste posibles: {*Cáncer de Pulmón*}, {*Déficit de Atención*} y {*Genética*}. Nótese en que existen dos caminos traseros diferentes entre *Fatiga* y *Accidente de Coche*:

- $Fatiga \leftarrow Toser \leftarrow Cáncer\ de\ Pulmón \leftarrow Genética \rightarrow Déficit\ de\ Atención \rightarrow Accidente\ de\ Coche$
- $Fatiga \leftarrow Cáncer\ de\ Pulmón \leftarrow Genética \rightarrow Déficit\ de\ Atención \rightarrow Accidente\ de\ Coche$

Ambos caminos traseros serían cerrados condicionando en cualquiera de los conjuntos mínimos de ajuste, y la utilización de cualquiera de ellos debería resultar en estimaciones similares.

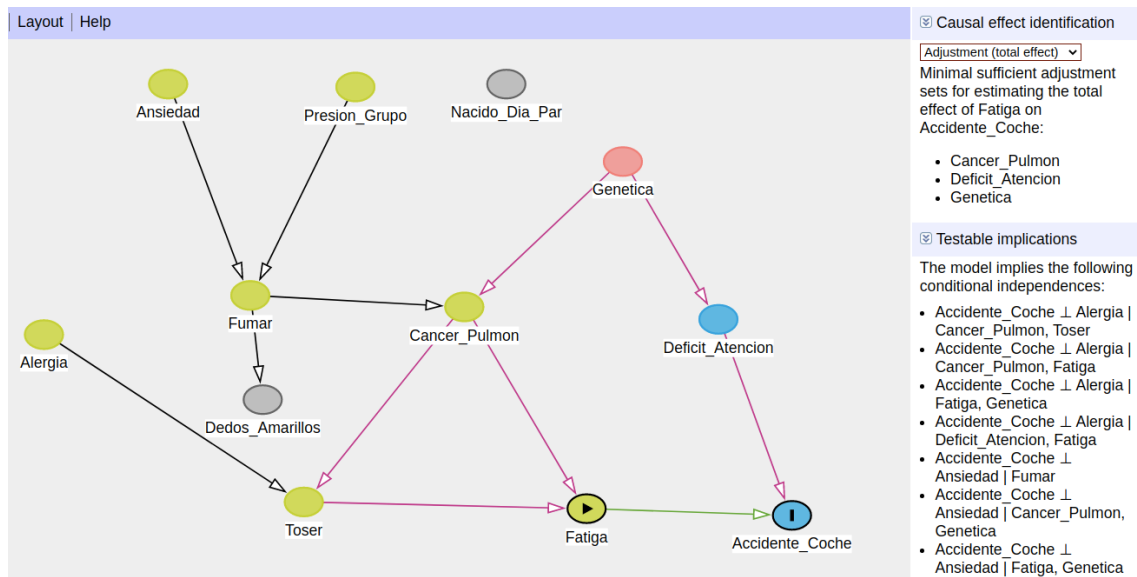


Figura 5.8: Diagrama causal representativo del efecto de *Fatiga* y *Accidente de Coche*

Subiendo los datos y apretando el botón *Compute ATE* iniciamos la fase de identificación. En este estudio, vamos a probar a computar el efecto causal ajustando en todos los conjuntos mínimos de ajuste por separado, utilizando todos los métodos de estimación disponibles en la herramienta. De esta manera, realizaremos varios estudios probando con diferentes conjuntos de ajuste y métodos de estimación. Los resultados son mostrados en la Tabla 5.1, dónde *Matching* se corresponde a *Propensity Score Matching* y *Stratification* se corresponde a *Propensity Score Stratification*.

Conjunto de ajuste aplicado	<i>Double Machine Learning</i>	<i>Matching</i>	<i>Stratification</i>
{Cáncer de Pulmón }	0.46	0.788	0.462
{Déficit de Atención}	0.452	0.359	0.454
{Genética}	0.45	0.788	0.451

Tabla 5.1: Efecto causal de *Fatiga* en *Accidente de Coche* aplicando diferentes conjuntos de ajuste y métodos de estimación

Para todos los experimentos realizados, se han realizado pruebas de refutación. Estas señalan un error en el modelo de *Matching*.

Original Causal effect : 0.359		
p-value : 0.111		
Refutation method	New effect	Procedure
Bootstrap Validation	0.473	Replaces the given dataset with a randomly selected samples of the dataset. The estimated effect should be close to the original effect.
Placebo treatment	-0.044	Substitutes the treatment values with values of a random independent variable. The estimated effect should be close to zero.
Random common cause	0.451	Inserts a random common cause and recomputes the causal effect. The estimated effect should be close to the original effect.
Subset validation	0.458	Replaces the given dataset with a randomly selected subset. The estimated effect should be close to the original effect.

Figura 5.9: Refutación del modelo construido para el efecto de *Fatiga* en *Accidente de Coche* ajustando en *Cáncer de Pulmón*

Para empezar, debemos señalar que una de las tres pruebas realizadas mediante *Propensity Score Matching* obtiene un *p-value* mayor a 0.05, por lo que rechazaríamos los resultados obtenidos. Además, tanto *Bootstrap Validation* como *Subset validation* indican que utilizando un subconjunto de los datos, en vez de todos ellos, conseguimos acercarnos más al resultado obtenido por los otros métodos de estimación. Esto se puede deber a que existen algunos valores atípicos que están introduciendo bastante error en el resultado. También podría ser porque para determinar la similitud entre dos unidades, *Propensity Score Matching* necesitaría utilizar más variables para mejorar el emparejamiento.

Dejando de lado *Matching*, el resto de métodos nos dan a entender que el efecto causal de *Fatiga* en *Accidente de Coche* es 0.45. Es decir, si tenemos un grupo de cien personas de nuestra población a las que fatigamos y un grupo de cien personas de nuestra población a las que obligamos a no fatigarse, existirán cuarenta y cinco personas más que tendrán un accidente de coche en el grupo que hemos fatigado. Como conclusión, podemos afirmar que en nuestra población *Fatiga* y *Accidente de Coche* mantienen una relación causal.

5.1.4. Efecto de *Presión de Grupo* y *Dedos Amarillos*

A continuación, centramos nuestra atención en la relación existente entre *Presión de Grupo* y *Dedos Amarillos*. Para ello, una vez más, construiremos el diagrama causal de la Figura 5.2 en *DAGitty*, marcando *Presión de Grupo* como tratamiento y *Dedos Amarillos* como resultado (Figura 5.10). La herramienta nos indica que no es necesario ajustar en

ninguna variable para estimar el efecto deseado. Esto es así porque no existe ningún camino trasero (abierto y no causal) entre *Presión de Grupo* y *Dedos Amarillos*.

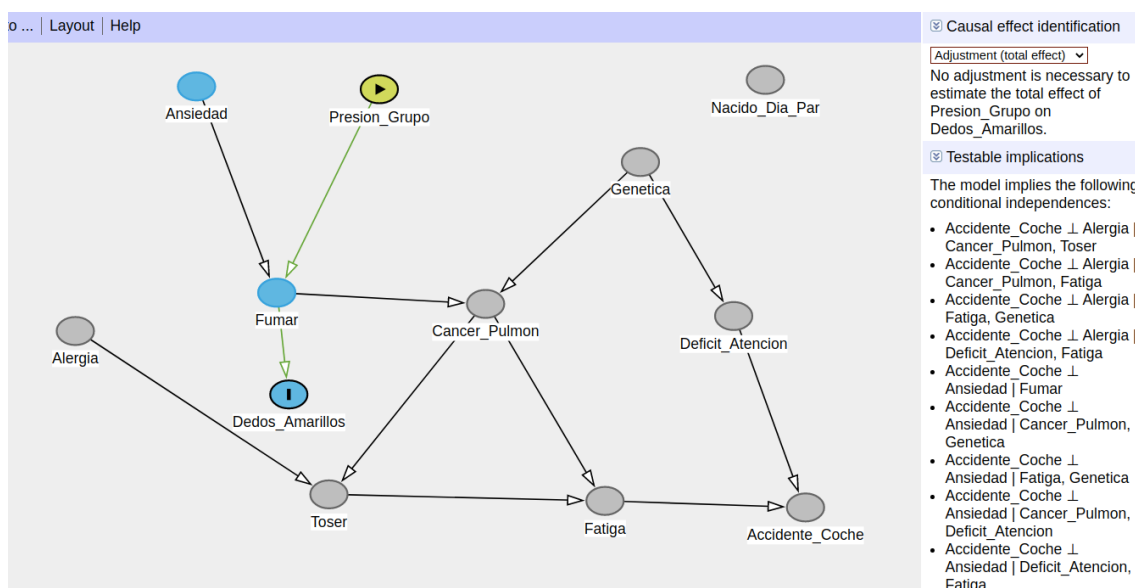


Figura 5.10: Diagrama causal representativo del efecto de *Presión de Grupo* en *Dedos Amarillos*

Apretando en *Compute ATE*, la herramienta muestra todas las posibilidades para realizar la fase de identificación. Podemos o bien no ajustar en nada, o computar el efecto aplicando el ajuste de caminos directos o *frontdoor criterion*. Vemos que $\{Fumar\}$ cumple con los criterios establecidos por este método puesto que:

1. No existe ningún camino directo entre *Presión de Grupo* y *Dedos Amarillos* que no pase por $\{Fumar\}$.
2. No existe ningún camino abierto entre *Presión de Grupo* y *Fumar*.
3. No existe ningún camino trasero entre *Fumar* y *Dedos Amarillos*.

Si marcamos la opción de realizar *frontdoor* con $\{Fumar\}$, y realizamos la estimación a través de *Double Machine Learning* aparece que el efecto causal es 0.095. Las pruebas de refutación realizadas sobre el modelo no muestran ningún indicio de que el modelo sea incorrecto de alguna manera. Por lo tanto, aunque pequeña, existe ninguna relación causal entre *Presión de Grupo* y *Dedos Amarillos*.

5.2 Estudio aplicado al ámbito psicológico

En esta sección, al igual que en la sección 5.1, se ilustrará el funcionamiento de CAUSALINF así como los pasos de la inferencia causal a través de un escenario del ámbito psicológico.

En esta ocasión los datos utilizados proceden de una encuesta de investigación pagada con el objetivo de explorar el vínculo entre las enfermedades mentales y el desempleo. En esta encuesta participaron 334 personas (80 de ellas con enfermedades mentales), las cuáles respondieron preguntas acerca de su género, edad, si tenían alguna enfermedad

mental, máximo grado de educación, etc. Los resultados de esta han sido publicados en [Kaggle](#) como un conjunto de datos de libre uso. Particularmente, podemos encontrar el conjunto de datos en [21].

Por lo tanto, el conjunto de datos disponible consiste en varias filas que representan las unidades o individuos de la población, y varias columnas que representan las variables o características de los individuos. Existe un total de treinta y una variables en los datos, pero con el fin de simplificar el estudio, reduciremos las variables utilizadas a nueve, las cuales se muestran en la Tabla 5.2 junto con sus descripciones y posibles valores. Las variables categóricas se han transformado en variables numéricas para poder aplicar todos los métodos de la inferencia causal (como la regresión lineal). Estas variables no podrán ser utilizadas ni como tratamiento ni como resultado puesto que en este proyecto sólo se manejan como variables binarias. Sin embargo, sí que podrán ser ajustadas.

Cabe destacar que en este estudio los datos no son sintéticos y se desconoce cómo se han generado. Es decir, no disponemos de un grafo que describa el proceso de generación de los datos en el que podamos apoyarnos en todos los casos. De esta manera, deberemos tomar asunciones basándonos en nuestro conocimiento del dominio y apoyándonos en la medida de lo posible en fuentes fiables de información. Así, procedemos a aplicar los pasos de la inferencia causal para estudiar si realmente existe una relación causa-efecto entre las diferentes variables definidas o simplemente existe correlación entre ellas. Específicamente, a través de la herramienta estudiaremos las relaciones entre:

- El efecto de *Enfermedad Mental* en *Empleo*.
- El efecto de *Discapacidad* en *Empleo*.
- El efecto de *Enfermedad Mental* en *Hospitalización*.
- El efecto de *Género* en *Vivir padres*.

5.2.1. Efecto de *Enfermedad Mental* en *Empleo*

Procedemos a estudiar el efecto medio de *Enfermedad Mental* en *Empleo*. Para ello, en contraste al estudio previamente realizado, deberemos plantearnos qué relaciones existen entre nuestras variables, puesto que no sabemos cómo se han generado los datos. Esto lo haremos en la fase de modelización, con el objetivo de construir un diagrama causal que codifique nuestras asunciones. De esta manera, a continuación se van a enumerar las asunciones consideradas entre las variables del dominio:

1. Queremos corroborar o descartar la relación existente entre *Enfermedad Mental* y *Empleo*, por lo que vamos a asumir que *Enfermedad Mental* causa *Empleo*.
2. Se considera que el grado de educación recibido puede generar más presión por las expectativas, y esto podría desembocar en enfermedades mentales, por lo que asumimos que *Educación* causa *Enfermedad Mental*.
3. También consideramos que *Educación* es una causa directa de *Empleo*, puesto que el grado de educación recibido suele estar relacionado con la facilidad a la hora de encontrar empleo.
4. Además, por la anterior afirmación también asumimos que *Educación* causa *Periodo Sin Trabajar*.

Variable	Descripción	Posibles valores
<i>Educación</i>	Grado de educación que el individuo ha recibido	<ul style="list-style-type: none"> ■ 1(Sin estudios) ■ 2(Estudios sin finalizar) ■ 3(Instituto sin finalizar) ■ 4(Instituto) ■ 5(Máster sin finalizar) ■ 6(Máster) ■ 7(Doctorado sin finalizar) ■ 8(Doctorado)
<i>Periodo Sin Trabajar</i>	Existe algún periodo en el que el individuo no ha trabajado	0 / 1
<i>Género</i>	Género del individuo	0(Hombre) / 1(Mujer)
<i>Empleo</i>	El individuo tiene un empleo al menos de a tiempo parcial	0 / 1
<i>Cobrar Desempleo</i>	El individuo recibe beneficios monetarios por desempleo. Cabe destacar que un individuo también puede cobrar desempleo si tiene un empleo a tiempo parcial.	0 / 1
<i>Enfermedad Mental</i>	El individuo tiene una enfermedad mental (esquizofrenia, anorexia, depresión, etc.)	0 / 1
<i>Discapacidad</i>	El individuo tiene una discapacidad (daño cerebral, ceguera, etc.)	0 / 1
<i>Hospitalización</i>	El individuo ha sido hospitalizado por su enfermedad mental	0 / 1
<i>Vivir Padres</i>	El individuo vive con sus padres	0 / 1

Tabla 5.2: Descripción variables utilizadas en este estudio

5. Por otro lado, se asume que tener periodos sin trabajar en el currículum dificulta el encontrar empleo, por lo que *Periodo Sin Trabajar* es una causa directa de *Empleo*.
6. Se considera que muchas veces, el género femenino afronta más dificultades a la hora de encontrar trabajo y, por lo tanto, asumimos que *Género* es una causa directa de *Empleo*.
7. Muchas veces, las personas con enfermedades mentales toman medicación que les afecta de manera física, pudiendo llegar a causa una discapacidad. Así, asumimos que *Enfermedad Mental* puede ser una causa de *Discapacidad*.
8. Asumimos que *Enfermedad Mental* causa *Hospitalización*.

9. También asumimos que *Discapacidad* es una causa directa de *Hospitalización*.
10. También asumimos que *Discapacidad* es una causa directa de *Empleo*, puesto que suelen tener dificultades físicas para trabajar.
11. Se asume que la falta de empleo puede llevar a falta de dinero, que a su vez puede causar vivir con tus padres, por lo que *Empleo* constituye una causa directa de *Vivir Padres*.
12. El hecho de no tener empleo, o tener un empleo que no te da lo suficiente para vivir es una causa directa de cobrar del estado, por lo que asumimos que *Empleo* causa *Cobrar Desempleo*.

Cualquier relación causal no comentada, se asumirá que no existe. Teniendo en cuenta todas estas asunciones, construimos el diagrama causal objetivo en CAUSAINF marcando *Enfermedad Mental* como tratamiento y *Empleo* como resultado, tal y como se puede ver en la Figura 5.11.

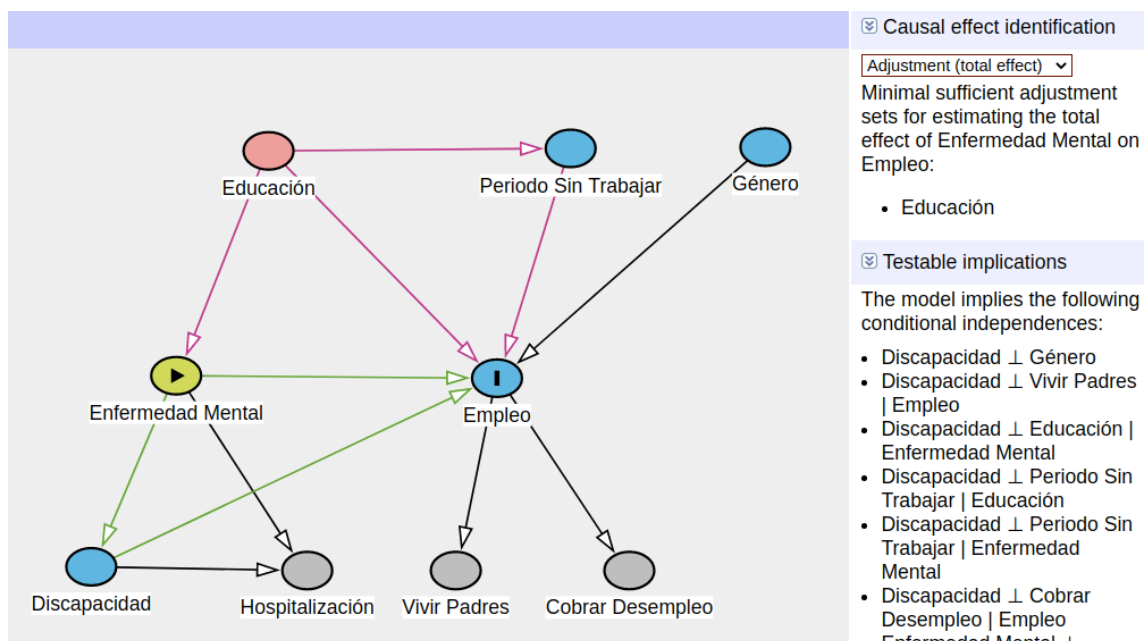


Figura 5.11: Diagrama causal representativo del efecto de *Enfermedad Mental* en *Empleo*

El mínimo conjunto de ajuste detectado por CAUSAINF corresponde con $\{Educación\}$. Esto tiene sentido porque existen dos caminos traseros entre *Enfermedad Mental* y *Empleo*, y ambos son cerrados ajustando en *Educación*.

Una vez finalizada la fase de modelización y subidos los datos, hacemos clic en *Compute ATE* para empezar el proceso de inferencia causal. Indicando a la herramienta que ajuste en $\{Educación\}$ con el fin de identificar el ATE, y posteriormente marcando *Double Machine Learning* como método de estimación, CAUSAINF nos indica que el efecto medio de *Enfermedad Mental* en *Empleo* es 0.047. Las pruebas de refutación no indican ningún problema en el modelo, pero obtenemos un *p-value* de 0.145, por lo que automáticamente descartamos la relación causal entre *Enfermedad Mental* en *Empleo*.

5.2.2. Efecto de *Discapacidad* sobre *Empleo*

Vamos a estudiar el efecto medio de *Discapacidad* sobre *Empleo*. Para ello, utilizaremos el diagrama causal de la Figura 5.12, el cual se construye sobre la base del diagrama 5.11, con la única diferencia de que suprimiremos la arista de *Enfermedad Mental* a *Empleo*, puesto que ya hemos descartado su relación en el anterior apartado. De esta manera, construimos el diagrama en CAUSALINF, marcando *Discapacidad* como tratamiento y *Empleo* como resultado. La herramienta nos indica que existen dos posibles conjuntos mínimos de ajuste: {*Educación*} o {*Enfermedad Mental*} (Figura 5.12). Esto es así porque existen dos caminos traseros entre *Discapacidad* y *Empleo*:

- $Discapacidad \leftarrow Enfermedad\ Mental \leftarrow Educación \rightarrow Empleo$
- $Discapacidad \leftarrow Enfermedad\ Mental \leftarrow Educación \rightarrow Periodo\ Sin\ Trabajar \rightarrow Empleo$

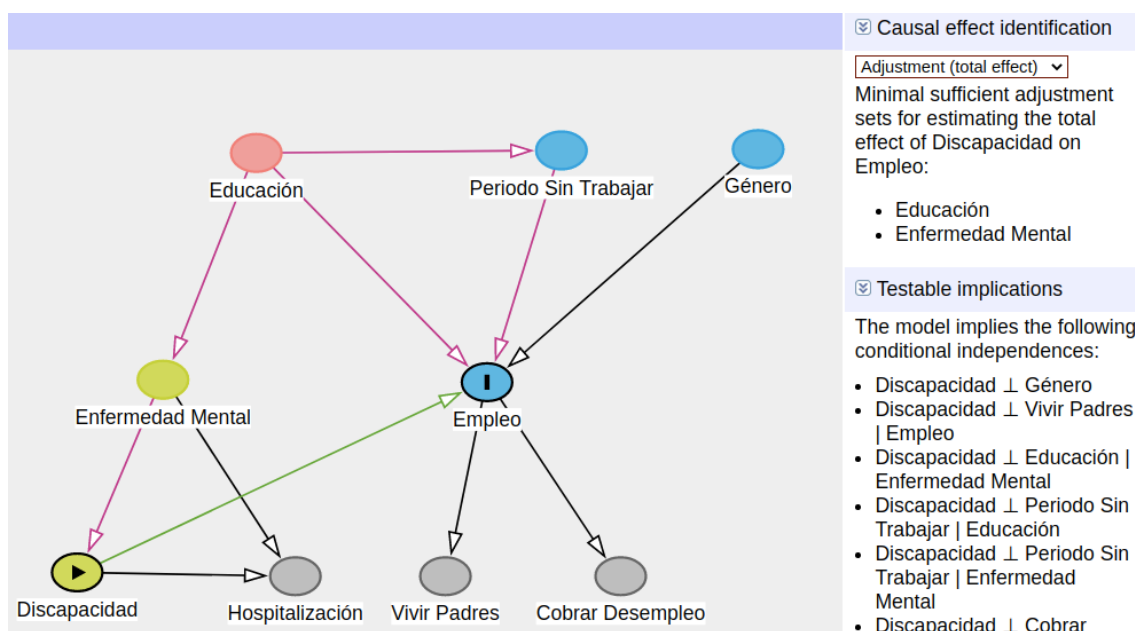


Figura 5.12: Diagrama causal representativo del efecto de *Discapacidad* en *Empleo*

De esta manera, tal y como nos indica la herramienta, debemos ajustar al menos en *Educación* o en *Enfermedad Mental* para cerrar los caminos no causales y poder estimar de manera correcta el efecto medio de *Discapacidad* en *Empleo*. A continuación seleccionamos el botón *Compute ATE* para empezar la fase de identificación. Por simplicidad vamos a indicarle a la herramienta que ajuste las variables *Educación* y *Enfermedad Mental* para poder convertir el ATE en una expresión equivalente pero sin intervenciones. Tras confirmar la selección, pasamos a la fase de estimación, en la que seleccionaremos *Double Machine Learning* por precisión. Confirmando nuevamente, obtenemos que el efecto medio de *Discapacidad* en *Empleo* es del -0.465 . Las pruebas de refutación muestran algo de sesgo en prácticamente todas las pruebas. Puede deberse a que existe algún problema con los datos o que existe algún problema con el diagrama.

Asumiendo el diagrama causal y los datos disponibles, e interpretando esta información en nuestro dominio, si tenemos cien personas de nuestra población a las que infligimos una discapacidad, y cien personas a las que obligamos a no tener ninguna discapacidad, existirán al menos cuarenta y seis personas más con trabajo en el grupo sin

discapacidad. Por lo tanto, podemos concluir que existe una relación causal entre *Discapacidad* y *Empleo* (si nuestras asunciones son correctas y los datos son representativos del mundo observable).

5.2.3. Efecto de *Enfermedad Mental* en *Hospitalización*

En este apartado, vamos a estudiar el efecto medio de *Enfermedad Mental* en *Hospitalización*. Para ello, construiremos el diagrama causal de la Figura 5.13, que se construye a partir del diagrama visto en la Figura 5.12, marcando *Enfermedad Mental* como tratamiento y *Hospitalización* como resultado.

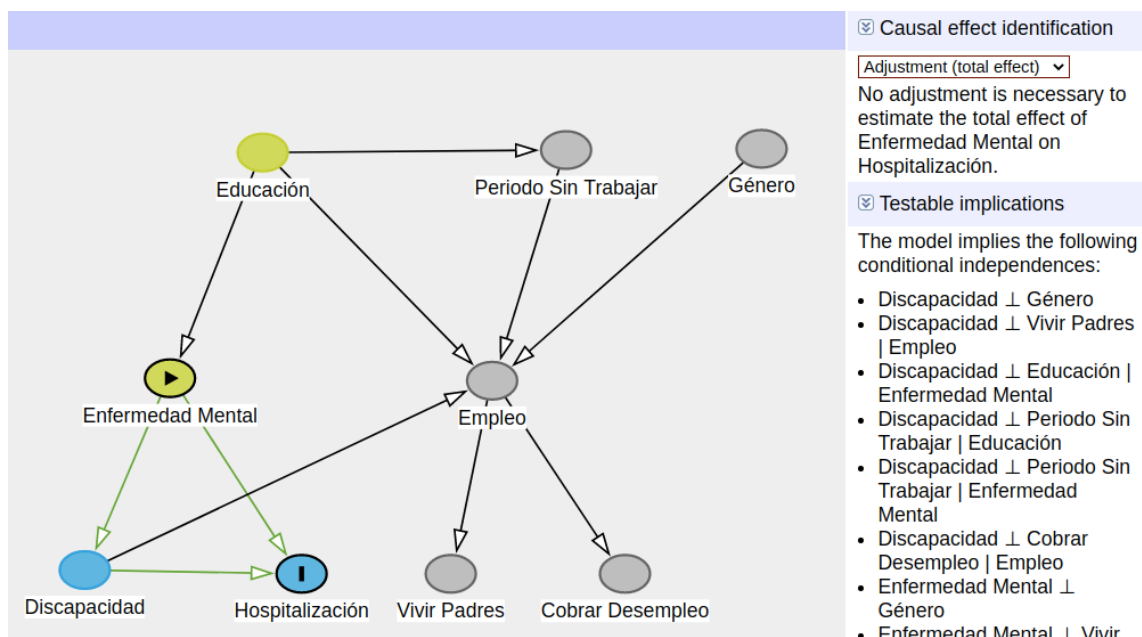


Figura 5.13: Diagrama causal representativo del efecto de *Enfermedad Mental* en *Hospitalización*

De esta manera, como se puede ver en la Figura 5.13, CAUSAINF no encuentra ningún camino trasero entre *Enfermedad Mental* y *Hospitalización*, por lo que nos indica que no necesitamos ajustar en ninguna variable para calcular el efecto medio de *Enfermedad Mental* en *Hospitalización*.

Empezando el proceso de inferencia causal, le indicamos a CAUSAINF que no ajuste en ninguna variable, convirtiendo así el ATE en una expresión equivalente pero sin intervenciones. Marcando *Double Machine Learning* como método de estimación y confirmando, nos aparece que el efecto medio de *Enfermedad Mental* en *Hospitalización* es de 0.274. Las pruebas de refutación no indican ningún problema en el modelo construido. Interpretando estos resultados en nuestro dominio, si tenemos cien personas de nuestra población a las que les provocamos una enfermedad mental y cien personas a las que no provocamos una enfermedad mental, existirán veintisiete personas más hospitalizadas en el grupo en el que provocamos enfermedades mentales. Por lo tanto, apoyándonos en nuestras asunciones y en los datos disponibles, podemos concluir que existe una relación causal entre *Enfermedad Mental* y *Hospitalización*.

5.2.4. Efecto de Género en Vivir Padres

A continuación, pasamos a estudiar el efecto medio de *Género* en *Vivir Padres*. Es decir, queremos estudiar si existe una relación causal entre estas dos variables o solamente correlación. Para ello, utilizaremos el diagrama causal de la Figura 5.14, que se construye apoyándose en el diagrama causal visto en la Figura 5.12, marcando *Género* como tratamiento y *Vivir Padres* como resultado.

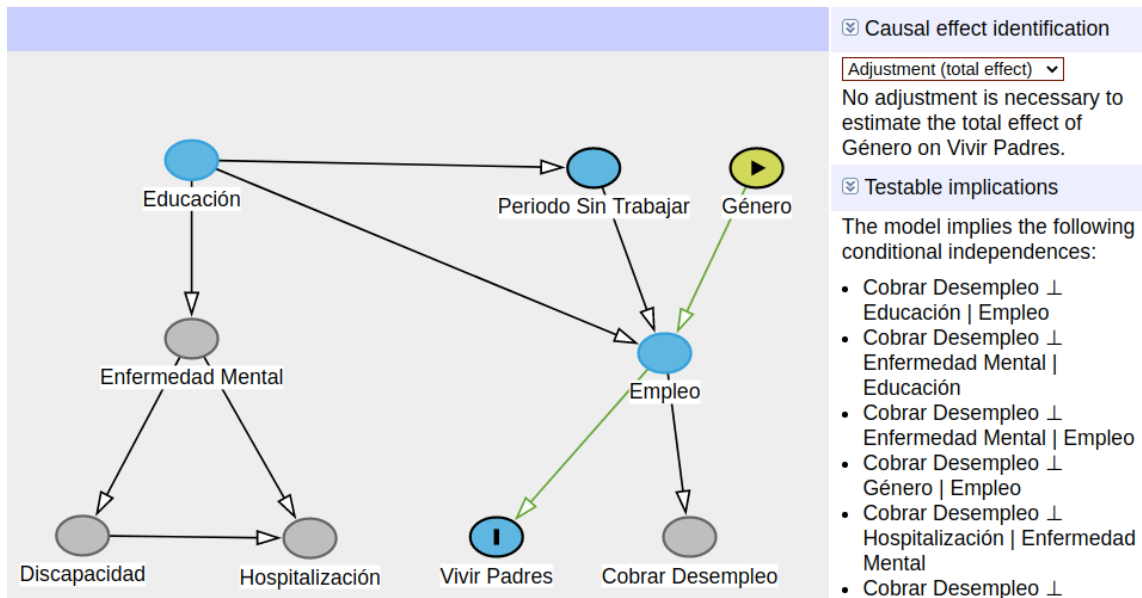


Figura 5.14: Diagrama causal representativo del efecto de *Género* en *Vivir Padres*

De esta manera, como se puede ver en la Figura 5.14, CAUSAINF no encuentra ningún camino trasero entre *Género* y *Vivir Padres*, lo que nos indica que no necesitamos ajustar en ninguna variable para calcular el efecto medio.

Empezando el proceso de inferencia causal, le indicamos a la herramienta que no ajuste en ninguna variable, convirtiendo así el ATE en una expresión equivalente pero sin intervenciones. Tras confirmar *Double Machine Learning* como método de estimación, nos aparece que el efecto medio de *Género* en *Vivir Padres* es de 0. Las pruebas de refutación no indican ningún problema en el modelo construido, pero obtenemos un *p-value* del 0.602. Por lo tanto, podemos concluir que no existe una relación causal entre *Género* y *Vivir Padres*.

CAPÍTULO 6

Conclusiones, relación con asignaturas y trabajos futuros

Finalmente, en este capítulo se presentarán las conclusiones finales del proyecto, la relación de este con asignaturas cursadas en el Grado de Ingeniería Informática y una propuesta de trabajos futuros.

6.1 Conclusiones

Al inicio de esta memoria, se han planteado una serie de objetivos iniciales. En este apartado, se procede a evaluar el cumplimiento de estos.

El primer objetivo establecido inicialmente consiste en ser capaces de interpretar los conceptos teóricos subyacentes de la inferencia causal. Consideramos que este objetivo está cubierto en cierta medida, puesto que la inferencia causal es un campo muy amplio y se ha experimentado con conceptos básicos de esta. Sin embargo, estos conceptos, aunque básicos, nos han ayudado tanto en la creación de la herramienta como en el análisis de los estudios experimentales. Incluso hemos conseguido detectar un error en la librería *DoWhy* y avisar a sus creadores.

El segundo objetivo establecido consiste en desarrollar una herramienta que permita aplicar diferentes métodos de la inferencia causal. Consideramos que CAUSAINF cumple con este objetivo.

Los últimos objetivos establecidos pretenden realizar una búsqueda de al menos dos conjuntos de datos de diferentes ámbitos y realizar dos estudios sobre estos. Consideramos que este objetivo se ha cumplido tal y como se expone en el Capítulo 5.

De esta manera, se ha conseguido cumplir con todos los objetivos inicialmente planteados.

6.2 Relación del trabajo desarrollado con los estudios cursados

En este apartado se procede a relacionar los conocimientos aplicados en el proyecto con las asignaturas cursadas por el alumno en el grado de Ingeniería Informática.

Se considera que las asignaturas más relevantes en la interpretación y entendimiento del dominio han sido “Estadística” y “Matemáticas discretas”. Por una parte, la asigna-

tura de “Estadística” ha proporcionado los fundamentos básicos necesarios para poder realmente entender los cálculos que se estaban realizando en la inferencia causal e incluso el porqué. Por otro lado, la introducción de los grafos en la asignatura de “Matemáticas discretas” ha facilitado el entendimiento de diversos conceptos básicos para la fase de modelización.

Aunque no tan relevante, además se considera que los conocimientos aportados sobre el uso de *JavaScript* y su asincronía en la asignatura de “Tecnología de sistemas de información en la red” han sido útiles para adaptar el código de *DAGitty* de la mejor manera posible.

En cuanto a competencias transversales, se considera que las competencias que más se han tenido que poner en práctica en este proyecto han sido:

- **Comprensión e Integración:** se ha tenido que entender todo el proceso de inferencia causal a partir de la información sacada de artículos, libros e incluso cursos. Además, todos esos conocimientos se han aplicado en la creación de una herramienta.
- **Comunicación Efectiva:** a lo largo de toda la memoria, se ha procurado dejar todos los conceptos masticados con el fin de que sirva como una guía rápida o una base a futuros trabajos relacionados.
- **Planificación y Gestión del tiempo:** A lo largo del proyecto, se ha gestionado los hitos que deberían cumplirse cada cierto tiempo, para poder entregar el trabajo con el mayor tiempo posible.

6.3 Trabajos futuros

Finalmente, en esta sección se pretende plantear de manera breve ciertas líneas de trabajo futuras:

- Incluir variables instrumentales y mediación como métodos de identificación y estimación.
- Manejar variables no observadas.
- Integrar la funcionalidad de tratamiento múltiple.
- Gestionar el manejo de variables no binarias.
- Tratar filas con valores desconocidos en vez de ignorarlas.
- Además del *Average Treatment Effect*, agregar dos botones para calcular el *Natural Direct Effect* y *Natural Indirect Effect*.

Bibliografía

- [1] Jing Ma y col. «Assessing the Causal Impact of COVID-19 Related Policies on Outbreak Dynamics: A Case Study in the US». En: (mayo de 2021).
- [2] Matteo Bonvini y col. «Causal Inference in the Time of Covid-19». En: (mar. de 2021).
- [3] Peterson Ozili. «COVID-19 in Africa: socio-economic impact, policy response and opportunities». En: *International Journal of Sociology and Social Policy* ahead-of-print (mayo de 2020). DOI: [10.1108/IJSSP-05-2020-0171](https://doi.org/10.1108/IJSSP-05-2020-0171).
- [4] Oguzhan Gencoglu y Mathias Gruber. «Causal Modeling of Twitter Activity during COVID-19». En: *Computation* 8 (sep. de 2020), pág. 85. DOI: [10.3390/computation8040085](https://doi.org/10.3390/computation8040085).
- [5] Shuxi Zeng, Fan Li y Peng Ding. «Is being an only child harmful to psychological health?: evidence from an instrumental variable analysis of China's one-child policy». En: *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 183 (ago. de 2020). DOI: [10.1111/rssa.12595](https://doi.org/10.1111/rssa.12595).
- [6] Bernhard Schölkopf y col. «Towards Causal Representation Learning». En: (feb. de 2021).
- [7] Wenjie Wang y col. «Clicks Not Equal to "Like": Counterfactual Recommendation for Mitigating Clickbait Issue». En: (sep. de 2020).
- [8] Judea Pearl y Dana Mackenzie. *The Book of Why. The New Science of Cause and Effect*. Inglés americano. New York: Basic Books, 2018. ISBN: 978-0-465-09760-9.
- [9] Robins J.M. Hernan M.A. *Causal Inference: What If*. Boca Raton: Chapman & Hall/CRC. 2020. URL: https://cdn1.sph.harvard.edu/wp-content/uploads/sites/1268/2021/03/ciwhatif_hernanrobins_30mar21.pdf.
- [10] Brady Neal. *Introduction to Causal Inference*. 2020. URL: https://www.bradyneal.com/Introduction_to_Causal_Inference-Dec17_2020-Neal.pdf.
- [11] Bruno Gonçalves. *Interventions*. <https://medium.data4sci.com/causal-inference-part-ix-interventions-c3f94190191d>.
- [12] Sander Greenland y Judea Pearl. «Causal Diagrams». En: *International Encyclopedia of Statistical Science*. Ed. por Miodrag Lovric. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, págs. 208-216. ISBN: 978-3-642-04898-2. DOI: [10.1007/978-3-642-04898-2_162](https://doi.org/10.1007/978-3-642-04898-2_162). URL: https://doi.org/10.1007/978-3-642-04898-2_162.
- [13] Natalie Staplin y col. «Use of Causal Diagrams to Inform the Design and Interpretation of Observational Studies: An Example from the Study of Heart and Renal Protection (SHARP)». En: *Clinical journal of the American Society of Nephrology : CJASN* 12 (ago. de 2016). DOI: [10.2215/CJN.02430316](https://doi.org/10.2215/CJN.02430316).

- [14] Miguel Hernán. *Causal Diagrams: Draw Your Assumptions Before Your Conclusions*. URL: https://www.edx.org/es/course/causal-diagrams-draw-your-assumptions-before-your?source=aw&awc=6798_1622651045_d24326ce64d89bd2fb6dcf07587a834e&utm_source=aw&utm_medium=affiliate_partner&utm_content=text-link&utm_term=422873_Edflex.
- [15] Felix Elwert. «Graphical Causal Models». En: *Handbook of Causal Analysis for Social Research*. Ed. por Stephen L. Morgan. Dordrecht: Springer Netherlands, 2013, págs. 245-273. ISBN: 978-94-007-6094-3. DOI: [10.1007/978-94-007-6094-3_13](https://doi.org/10.1007/978-94-007-6094-3_13). URL: https://doi.org/10.1007/978-94-007-6094-3_13.
- [16] Marco Caliendo y Sabine Scheel-Kopeinig. «Some Practical Guidance For the Implementation of Propensity Score Matching». En: *Journal of Economic Surveys* 22 (feb. de 2008), págs. 31-72. DOI: [10.1111/j.1467-6419.2007.00527.x](https://doi.org/10.1111/j.1467-6419.2007.00527.x).
- [17] Peter Austin. «An Introduction to Propensity Score Methods for Reducing the Effects of Confounding in Observational Studies». En: *Multivariate behavioral research* 46 (mayo de 2011), págs. 399-424. DOI: [10.1080/00273171.2011.568786](https://doi.org/10.1080/00273171.2011.568786).
- [18] Johannes Textor y col. «Robust causal inference using directed acyclic graphs: the R package 'dagitty'». En: *International Journal of Epidemiology* 45.6 (2016), págs. 1887-1894. DOI: [10.1093/ije/dyw341](https://doi.org/10.1093/ije/dyw341).
- [19] Johannes Textor. «Drawing and Analyzing Causal DAGs with DAGitty». En: *CoRR* abs/1508.04633 (2015). arXiv: [1508.04633](https://arxiv.org/abs/1508.04633). URL: <http://arxiv.org/abs/1508.04633>.
- [20] Amit Sharma, Emre Kiciman y col. *DoWhy: A Python package for causal inference*. <https://github.com/microsoft/dowhy>. 2019.
- [21] Michael Corley. *Unemployment and mental illness survey: Exploring the causation of high unemployment among the mentally ill*. <https://www.kaggle.com/michaelacorley/unemployment-and-mental-illness-survey>. 2019.