UNIVERSIDAD DE SALAMANCA
MÁSTER UNIVERSITARIO EN MODELIZACIÓN MATEMÁTICA

Modelos causales para el análisis del consumo energético en edificios residenciales

AUTOR/AUTORA: Antonio José Moreno Soria

TUTOR/TUTORES: Cruz Enrique Borges
Diego Casado
Juan Manuel Rodríguez

Curso 2020-2021

UNIVERSIDAD DE SALAMANCA
MÁSTER UNIVERSITARIO EN MODELIZACIÓN MATEMÁTICA

Modelos causales para el análisis del consumo energético en edificios residenciales

AUTOR/AUTORA: Antonio José Moreno Soria

TUTOR/TUTORES: Cruz Enrique Borges
Diego Casado
Juan Manuel Rodríguez

Curso 2020-2021

Índice general

Índice de figuras

Consentimiento para la presentación del Trabajo Fin de Máster

Los tutores del Trabajo Fin de Máster titulado "Modelos causales para el análisis del consumo energético en edificios residenciales"llevado a cabo por Antonio José Moreno Soria, Cruz Enrique Borges y Juan Manuel Rodríguez, consideran que es apto para ser presentado y defendido.

Salamanca, a 12 de Julio de 2021

Firmado:

Cruz Enrique Borges

RODRIGUEZ DIAZ JUAN MANUEL - 07960523Q

Firmado digitalmente por RODRIGUEZ DIAZ JUAN MANUEL - 07960523Q
Fecha: 2021.07.12 14:03:24 +02'00'

Juan Manuel Rodríguez

Capítulo 1

Introducción

Horizonte 2020 es el instrumento financiero que pone en marcha la Unión Europea por la Innovación, situándolo en el centro del proyecto de la UE para fomentar crecimiento y empleo inteligentes, sostenibles e integradores, considerando que la investigación es una inversión de futuro.

H2020 integra todas las fases desde la generación del conocimiento hasta las actividades más próximas al mercado: investigación básica, desarrollo de tecnologías, proyectos de demostración, líneas piloto de fabricación, innovación social, transferencia de tecnología, pruebas de concepto, normalización, apoyo a las compras públicas pre-comerciales, capital riesgo y sistema de garantías.

Los objetivos estratégicos del programa H2020 son los siguientes:

- **Crear una ciencia de excelencia**, que permita reforzar la posición de la UE en el panorama científico mundial.
- **Desarrollar tecnologías y sus aplicaciones para mejorar la competitividad europea.**
- **Investigar en las grandes cuestiones que afectan a los ciudadanos europeos**

El programa está abierto a todo el mundo, fomentando una colaboración público-privada, con una estructura sencilla que garantiza que los nuevos proyectos se pongan en marcha rápidamente y consigan resultados con mayor rapidez.

Uno de los principales apartados dentro del programa es **H2020-EU.3.3.1. Reducir el consumo de energía y la huella de carbono mediante un uso inteligente y sostenible**. En este punto, las actividades se centran en la investigación y ensayo a escala real de nuevos conceptos, soluciones no tecnológicas, componentes tecnológicos más eficientes, socialmente aceptables y asequibles y sistemas con inteligencia incorporada, con el fin de poder gestionar la energía en tiempo real. De esta manear se busca lograr edificios con emisiones cercanas a cero o que generen más energía de la que consumen, infraestructuras modernizadas, calefacción y refrigeración renovables, industrias altamente eficientes y adopción masiva de soluciones y servicios de eficiencia energética y de ahorro de energía. En este marco es donde se encuentra el proyecto WHY.

1.1. WHY: Subiendo la Escalera de la Causalidad para Entender y Proyectar la Demanda Energética del Sector Residencial

Para mitigar los efectos del cambio climático, es necesario tomar medidas urgentes en todos los sectores de la economía para reducir significativamente las emisiones de gases de efecto invernadero (GEI). Los modelos de sistemas energéticos (ESM) son herramientas que ayudan a los analistas, planificadores y responsables políticos del sector de la energía a describir racionalmente los sistemas energéticos y a evaluar sistemáticamente los impactos de los diferentes escenarios a largo plazo. Por el lado de la oferta, los ESM han proporcionado resultados útiles, pero sin embargo, por el lado de la demanda, carecen del grado de precisión necesario para caracterizar adecuadamente, entre otros, el uso de la energía en los hogares. Una de las dificultades intrínsecas es que la demanda de energía en el sector residencial está influenciada por un sinfín de factores (como la gran diversidad de viviendas, las condiciones socio-económicas de las unidades sociales y los patrones de consumo relacionados con el comportamiento) que no pueden contabilizarse fácilmente en los ESM tradicionales.

Para superar este reto, se utilizará la novedosa modelización causal para analizar cuantitativamente la toma de decisiones humanas en el consumo de energía y sus reacciones a las intervenciones, como por ejemplo, los cambios en las políticas. Esto se combinará con un enfoque innovador de FFORMA que permite clasificar múltiples perfiles de carga diferentes mediante un conjunto de vectores que los describen. Por lo tanto, WHY creará metodologías innovadoras para la previsión de la carga a corto y largo plazo. La modelización de WHY permitirá evaluar directamente el impacto de una multitud de políticas en el sistema energético, así como realizar una evaluación ex ante y ex post de las medidas políticas. Por lo tanto, WHY contribuirá a una comprensión holística del consumo energético de los hogares y a una mejor modelización de la demanda.

El conjunto de herramientas WHY se utilizará para evaluar varios escenarios que simulan diferentes medidas políticas. Se integrará con ESM ampliamente utilizados (PRIMES, TIMES) y se analizarán los resultados. Todos los resultados serán de código abierto para maximizar su aceptación y se difundirán ampliamente entre diversos públicos.

El presente trabajo se enmarca en la primera fase del proyecto WHY, es por ello que en el mismo se elabora un extenso estado del arte que recoge la toda la información relevante para poder llevar a cabo el proyecto. En particular, se centra en la comprensión de la causalidad y el estudio de los diferentes modelos causales. En el Capítulo 2 se presentan los principales objetivos tanto del proyecto como del trabajo, y la metodología seguida para su elaboración. En el Capítulo 3 se recoge la teoría relativa a la causalidad y los modelos causales extraída de la literatura, además, se presenta una taxonomía de las tareas que estos modelos deben realizar y se evalúan las diferentes librerías disponibles para su implementación. En los Capítulos 4 y 5 se presentan las herramientas técnicas que se utilizarán para el estudio causal junto a una colección de ejemplos resueltos que permiten comprender como operan los diferentes modelos causales y las conclusiones al respecto. Por último, en el Capítulo 6 se detallan los siguientes pasos a seguir en el proyecto WHY. La información necesaria para la creación de los diferentes escenarios de interés y las bases sobre las que se sentará la modelización del comportamiento humano se expone en los Apéndices B y C, ya que esta información queda fuera del objetivo principal del trabajo.

Capítulo 2

Objetivos y metodología

El proyecto WHY persigue el objetivo de desarrollar el WHY Toolkit, que facilita un Modelo Causal en combinación con otros tipos de modelos para proporcionar una mejor y más profunda comprensión de la demanda de electricidad en los hogares simulando el comportamiento de los consumidores. Este software permitirá a los usuarios:

- Realizar mejores previsiones sobre el consumo de electricidad en los hogares
- Analizar, evaluar y validar las decisiones políticas y cualquier otro tipo de intervenciones en el ámbito energético
- Examinar como sería el mundo actual si se hubieran tomado o no determinadas decisiones en política energética

El objetivo de este trabajo es sentar las bases para todos los desarrollos posteriores del proyecto WHY. En particular, los objetivos específicos del trabajo son:

- Objetivos principales: crear una librería en el Marco de Resultados Potenciales que permita comprender con facilidad los conceptos relacionados con el efecto causal; presentar una taxonomía de las tareas que los Modelos Causales deben realizar y se evaluar las diferentes librerías disponibles para su implementación, seleccionando la más conveniente; y por último, presentar una colección de ejemplos que ayuden a comprender el análisis causal y que muestren las funciones de las que dispone dicha librería.
- Como objetivos secundarios, se tienen explorar el estado del arte y modelado de los diferentes aspectos técnicos relacionados con el consumo energético, como por ejemplo los sistemas energéticos, medidas de eficiencia y suficiencia energética, sistemas de gestión, movilidad y aspectos políticos, y explorar el estado del arte de los conocimientos sobre el modelado del comportamiento humano.

El trabajo se ha llevado a cabo en equipo, manteniendo reuniones semanales para compartir los avances realizados.

Para cumplimentar el estado del arte de los diferentes aspectos técnicos, cada uno de los apartados del Apéndice B eran asignados a los miembros del equipo, recogiendo toda la información al respecto en una tabla y elaborando un esquema que resumiera las

principales características. Estas tablas y esquemas están disponibles En mi caso, elaboré el estado del arte de los Sistemas de Gestión de Energía, Sección B.3, de los Sistemas de Almacenamiento de Energía, Sección B.5 y de los Modelos de Negocio, Sección B.9.

Con respecto al estado del arte del modelado del comportamiento humano, se llevó a cabo la lectura de numerosos artículos, compartiendo los principales puntos con el resto del equipo. En mi caso, estudié el artículo de Klöckner y Nayum, 2016, cuyo análisis se presenta en el Apéndice C.

Una vez finalizados ambos estados del arte, se comenzó a estudiar la causalidad. En primer lugar se llevó a cabo un profundo estudio del Marco de Resultados Potenciales, Sección 3.1. Para una mejor comprensión de las medidas del efecto causal y de las relaciones entre las variables, se desarrolló una librería en este marco teórico. El trabajo se realizó en un equipo de dos, trabajando con un repositorio en remoto en GitHub donde se iba actualizando el código. Para gestionar correctamente los paquetes necesarios para el proyecto, se trabaja dentro de un entorno virtual de Python que se crea dentro del repositorio en local. En mi caso, trabajé utilizando el sistema operativo Ubuntu y el entorno de desarrollo Spyder. En esta parte, añadí varias funciones a la clase principal que permiten el análisis causal de conjuntos de datos como los de la Tabla 4.1c, la documentación pertinente y sus correspondientes pruebas unitarias. Al principio del Capítulo 4, se detalla la estructura del repositorio y en la Sección 4.1 se explica dicha librería junto con los ejemplos analizados.

Por último, se estudiaron los Modelos Causales Estructurales (Sección 3.2). Se realizó una taxonomía de las tareas que los Modelos Causales deben llevar a cabo y se evaluaron las diferentes librerías disponibles para su implementación, seleccionando la librería DoWhy para implementar el Modelo Causal en el proyecto (Sección 3.2.8). Trabajando sobre el mismo repositorio y en las mismas condiciones, elaboré un ejemplo sencillo que muestra cómo trabajar con la librería DoWhy y que aúna los resultados que se obtienen a partir del Marco de Resultados Potenciales y de los Modelos Causales Estructurales creados utilizando esta librería (Sección 4.2). Por último, trabajé sobre un Modelo Causal obtenido de la literatura (Ma, Chow y Xu, 2017), este sirve como ejemplo para los desarrollos futuros y para mostrar la cantidad de información es posible obtener a partir un Diagrama Causal y la potencia que tiene DoWhy a la hora de trabajar con modelos más complejos.

Capítulo 3

Estado del arte

3.1. Inferencia Causal

El espectacular éxito del machine learning, o aprendizaje automático, ha provocado una explosión de aplicaciones de Inteligencia Artificial (IA) y un aumento de las expectativas de sistemas autónomos que muestren una inteligencia de nivel humano. Sin embargo, estas expectativas se han topado con obstáculos fundamentales que afectan a muchos ámbitos de aplicación.

Uno de estos obstáculos es la falta de adaptabilidad o robustez, es decir, la falta de capacidad de reaccionar a circunstancias para las cuales no ha sido específicamente programado. Otro obstáculo es la explicabilidad, los modelos de aprendizaje automático son en su mayoría cajas negras incapaces de explicar las razones que hay detrás de sus predicciones. Un tercer obstáculo es la comprensión de las relaciones causa-efecto, sello distintivo de la cognición humana. Estos tres obstáculos requieren de equipar a las máquinas con herramientas de modelado causal, en particular, diagramas causales y su lógica asociada (Pearl, 2019).

A continuación, se presentan una serie de conceptos básicos que serán necesarios para la comprensión de las relaciones causa-efecto y los modelos causales.

3.1.1. Conceptos Previos

Dado que la estadística no se ocupa generalmente de los absolutos, el lenguaje de la probabilidad es extremadamente importante para ella. De igual manera, la probabilidad es importante para el estudio de la causalidad, ya que la mayoría de las afirmaciones causales son inciertas y la probabilidad es la forma de expresar la incertidumbre (Pearl, Glymour y Jewell, 2016).

Probabilidad Condicionada e Intervenciones

La probabilidad de que ocurra un suceso $X = x$, sabiendo que $Y = y$, es lo que se denomina **probabilidad condicional de** $X = x$ **dado** $Y = y$, denotado por $P(X = x|Y = y)$, o de forma abreviada, $P(x|y)$. A menudo, la probabilidad que se asigna al suceso x cambia drásticamente en función del conocimiento y al que se condiciona.

Las probabilidades condicionales desempeñan un papel importante en la investigación de cuestiones causales, ya que por lo general se quiere comparar cómo cambia la probabilidad (o, equivalentemente, el riesgo) de un resultado bajo diferentes condiciones de filtrado o exposición.

Una **intervención** en una variable consiste en fijar su valor. Para ello, es necesario un cambio en el sistema, lo que conlleva, a menudo, a un cambio en el valor de las otras variables. Para representar una intervención se utiliza el operador $do()$, de tal manera que la probabilidad de $Y = y$ cuando se interviene para hacer $X = x$ se denota por $P(Y = y|do(X = x))$. Esta última probabilidad representa la distribución de población de Y si toda la población tuviera fijo su valor de X en x. De igual manera, $P(Y = y|do(X = x), Z = z)$ representa la probabilidad condicionada de $Y = y$ a que $Z = z$ en la distribución creada por la intervención $do(X = x)$.

La diferencia entre una intervención en una variable y condicionar en esa variable es obvia, mientras que la intervención requiere de un cambio en el sistema, condicionar en una variable no requiere de ningún cambio, simplemente consiste en limitar el enfoque al subconjunto de casos en los que la variable toma el valor que interesa.

Independencia

Puede ocurrir que la probabilidad de un suceso se mantenga inalterada con la observación de otro. En estos casos, se dice que los dos sucesos son **independientes**. Formalmente, dos sucesos x e y se dice que son independientes si:

$$P(x|y) = P(x) \tag{3.1}$$

es decir, el conocimiento de que $Y = y$ no otorga información adicional sobre la probabilidad de que ocurra $X = x$. Si esta igualdad no se cumple, entonces se dice que x e y son **dependientes**. Las relaciones de dependencia e independencia son simétricas, es decir, si x es dependiente/independiente de y, y lo es de x. Dos variables, X e Y, son independientes si para todo valor x e y que pueden tomar se cumple la igualdad 3.1.

Dos sucesos x e y son **condicionalmente independientes** dado un tercer suceso z si:

$$P(x|y, z) = P(x|z) \tag{3.2}$$

es decir, x e y son condicionalmente independientes dado z si x e y son independientes tras filtrar por z. Si son independientes antes del filtrado, se dice que son **marginalmente independientes**.

Ley de la Probabilidad Total y Teorema de Bayes

Para cualquier conjunto de sucesos $y_1, y_2, \ldots, y_n$, tal que exactamente uno de ellos debe ser verdadero, se tiene que:

$$P(x) = P(x, y_1) + P(x, y_2) + \ldots + P(x, y_n) \tag{3.3}$$

Esta igualdad se conoce como la **ley de probabilidad total**, donde $P(x)$ es la probabilidad marginal de x.

Si se conoce la probabilidad de y y la probabilidad de x condicionado a y, se deduce la probabilidad de x e y:

$$P(x, y) = P(x|y)P(y) \tag{3.4}$$

Esta ecuación se conoce como la **regla del producto** e implica que la noción de independencia tiene una representación numérica en la distribución de probabilidad, es decir, para que los sucesos x e y sean independientes, se requiere que:

$$P(x, y) = P(x)P(y) \tag{3.5}$$

A partir de la ecuación 3.4 y teniendo en cuenta la simetría $P(x, y) = P(y, x)$, se obtiene inmediatamente una de las leyes más importantes en probabilidad, la **regla de Bayes**:

$$P(x|y) = \frac{P(y|x)P(x)}{P(y)} \tag{3.6}$$

Cuando se utiliza la regla de Bayes, se denomina al suceso x como la hipótesis y al suceso y como la evidencia. En muchos casos, se puede determinar fácilmente $P(y|x)$ (probabilidad de que se produzca la evidencia, dado que la hipótesis es correcta), pero es mucho más difícil averiguar $P(x|y)$ (probabilidad de que la hipótesis sea correcta, dado que se obtiene una evidencia). Esta última es la pregunta a la que a menudo se quiere dar respuesta, ya que, generalmente, se busca actualizar la creencia en una hipótesis, $P(x)$, después de que se haya producido alguna evidencia y, a $P(x|y)$. Para utilizar la regla de Bayes de esta manera, se debe tratar cada hipótesis como un suceso y asignar a todas ellas una distribución de probabilidad, llamada *a priori*.

Una vez comprendidos estos conceptos básicos, es posible adentrarse en el mundo de la causalidad. Para ello se presenta un brillante esquema desarrollado por Judea Pearl, en el cual muestra la capacidad del ser humano de ver el mundo en términos de causas y efectos, donde pueden diferenciar tres peldaños que implican distinta potencia predictiva. A su vez, indica qué tipo de preguntas corresponden a cada nivel y los datos necesarios para responderlas.

3.1.2. La Escalera de la Causalidad

Una idea útil revelada por la teoría de los modelos causales es la clasificación de la información causal en términos del tipo de preguntas que cada clase es capaz de responder. Esta clasificación forma una jerarquía de tres niveles de tal manera que las preguntas del nivel $i(i = 1, 2, 3)$ únicamente pueden responderse si se dispone de información del nivel $j(j \geq i)$. En la Figura 3.1 se muestra esta jerarquía.

Primer nivel: Asociación. Se dice que un evento está asociado con otro si la observación de uno cambia la probabilidad de observar el otro. En este nivel las relaciones son puramente estadísticas, definidas por los datos desnudos. Las preguntas de este nivel (¿Qué pasa si veo...?) se sitúan en el inferior de la jerarquía, ya que no requieren información causal. Este nivel se caracteriza por sentencias de probabilidad condicional, las cuales son fácilmente computables utilizando Redes Bayesianas o cualquier modelo de aprendizaje automático.

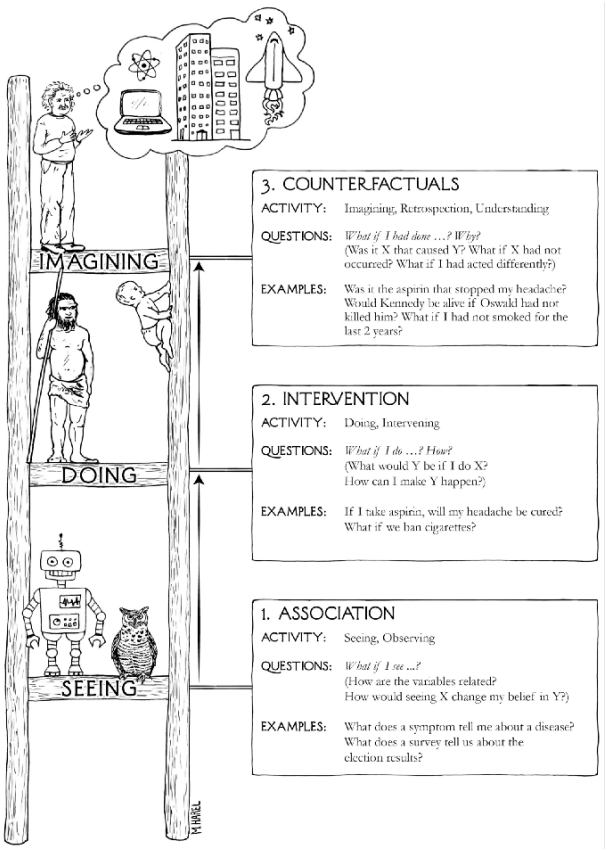

Figura 3.1: La Escalera de la Causalidad, con organismos representativos en cada nivel(Pearl y Mackenzie, 2018).

Para una mejor comprensión, se presenta un ejemplo. Se supone un director de marketing que se pregunta: ¿Qué probabilidad hay de que un cliente que ha comprado pasta de dientes compre también hilo dental? La respuesta viene dada por una probabilidad condicionada obtenida a partir del análisis de datos utilizando métodos asociativos, pertenecientes al primer escalón.

Segundo nivel: Intervención. Se sitúa por encima del nivel de asociación ya que implica no sólo ver, sino cambiar lo que es. Este nivel se caracteriza por sentencias del tipo: $P(y|do(x), z)$, que indica la probabilidad del evento $Y = y$ dada la intervención en X fijando su valor a x y observando el evento $Z = z$. Una pregunta típica en este nivel sería: ¿Qué pasaría si...? No se puede responder a este tipo de preguntas únicamente con datos recogidos pasivamente, sin importar lo grande que sea el conjunto de datos o la profundidad de la red neuronal. Un método directo para predecir el resultado de una intervención es la realización de experimentos bajo condiciones cuidadosamente controladas.

Siguiendo con el ejemplo, si el director se preguntara: ¿Qué pasará con las ventas de hilo dental si duplicamos el precio de la pasta de dientes? Este tipo de preguntas exige un nuevo tipo de conocimiento, carente de datos y perteneciente al segundo escalón, la intervención. La respuesta no puede obtenerse simplemente consultando los datos obtenidos en otras ocasiones que se doblara el precio, ya que este aumento podría

haberse ocasionado por diferentes razones, por ejemplo, por un suministro insuficiente que obligara a todas las tiendas a subir el precio. Es por ello que para dar respuesta a estas preguntas se necesitan de experimentos bajo condiciones controladas, o en su defecto, un modelo causal suficientemente sólido y preciso que nos permitiría utilizar datos del primer nivel para responder preguntas del segundo.

Tercer nivel: Contrafactuales. Están en la cima de la jerarquía ya que incluyen las preguntas de Intervención y Asociación. Una pregunta típica de este nivel sería: ¿Qué hubiera pasado si...? No es posible responder una pregunta contrafactual a partir de información puramente intervencionista, como por ejemplo la adquirida en experimentos controlados, por lo que la jerarquía es direccional, siendo el nivel superior el que más información contiene. Los contrafactuales son los elementos básicos del pensamiento científico, así como del razonamiento jurídico y moral y se caracterizan por expresiones del tipo: $P(y_x|x', y')$, que indica la probabilidad de que el suceso $Y = y$ se hubiera observado si X hubiera sido x cuando realmente se observó que $X = x'$ e $Y = y'$. Para dar respuesta a este tipo de preguntas, es necesario un modelo del proceso causal subyacente, a veces llamado teoría o incluso ley de la naturaleza.

Por último, volviendo al ejemplo, una pregunta perteneciente al tercer escalón sería: ¿Cuál es la probabilidad de que un cliente que compró pasta de dientes la siguiera comprando si hubiéramos duplicado el precio? En este caso, se compara el mundo real, donde se conoce que el cliente compró la pasta de dientes, con un mundo ficticio, donde el precio es el doble y donde es imposible realizar experimentos controlados como en el escalón anterior. Es por ello que para dar respuesta a preguntas contrafactuales es necesario un modelo causal preciso.

Para acabar de comprender el papel especial de la causalidad, se examinará uno de los rompecabezas más intrigantes de la literatura, donde se ilustra perfectamente por qué el lenguaje tradicional de la estadística debe enriquecerse con nuevos ingredientes para abordar las relaciones causa-efecto.

3.1.3. La Paradoja de Simpson

Llamada así por Edward Simpson, el estadístico que la popularizó por primera vez en 1951 (Simpson, 1951). La paradoja se refiere a la existencia de datos en los que una asociación estadística que se mantiene para toda una población se invierte en cada subpoblación, a continuación se presenta un ejemplo (Pearl, Glymour y Jewell, 2016).

Ejemplo 3.1.1. Se registraron las tasas de recuperación de 700 pacientes que tuvieron acceso a un medicamento. Un total de 350 pacientes optaron por tomar el fármaco y otros 350 no lo hicieron. Los resultados del estudio se muestran en la Tabla 3.1.

En los pacientes varones, los que tomaron el fármaco tuvieron una mejor tasa de recuperación que los que no lo tomaron (93 % frente a 87 %). En las mujeres, de nuevo, las que tomaron el fármaco tuvieron una mejor tasa de recuperación que las que no lo tomaron (73 % frente a 69 %). Sin embargo, en la población combinada, los que no tomaron

Tabla 3.1: Resultados del estudio de un fármaco teniendo en cuenta el género.

	Fármaco	No fármaco
Hombre	81/87 recuperados(93 %)	234/270 recuperados(87 %)
Mujeres	192/263 recuperados(73 %)	55/80 recuperados(69 %)
Total	273/350 recuperados(78 %)	289/350 recuperados(83 %)

el fármaco tuvieron una mejor tasa de recuperación que los que sí lo hicieron (83 % frente a 78 %).

Para decidir si el medicamento es beneficioso, hay que comprender la información que hay detrás de los datos, es decir, el mecanismo causal que generó los resultados obtenidos. Por ejemplo, suponiendo que el estrógeno tiene un efecto negativo en la recuperación, las mujeres tendrían menos probabilidades de recuperarse que los hombres, independientemente de si tomaron o no el fármaco. Además, como se puede ver en los datos, las mujeres son significativamente más propensas a tomar la droga que los hombres (más del 75 % de los que tomaron el fármaco son mujeres). Así pues, la razón por la que la droga parece ser perjudicial es que, si se selecciona al azar a un individuo que tomó el fármaco, es más probable que esa persona sea una mujer y por lo tanto, es menos probable que se recupere que otra persona seleccionada al azar que no lo tomara (más del 77 % de los que no tomaron el fármaco son hombres). Dicho de otro modo, ser mujer es una causa común tanto del consumo del fármaco como de la falta de recuperación. Por lo tanto, para evaluar la eficacia, hay que comparar sujetos del mismo sexo, garantizando así que cualquier diferencia en las tasas de recuperación entre los que toman el fármaco y los que no, no es atribuible al estrógeno. Esto significa que se debe consultar los datos segregados, los cuales son más específicos e informativos que los no segregados, y muestran inequívocamente que el fármaco es útil.

La paradoja quedará resuelta de una manera muy sencilla y visual en la Sección 4.2. Para ello, primero es necesario manejar una serie de medidas y condiciones que permiten medir, o mejor dicho estimar, el efecto causal.

3.1.4. Medida del Efecto Causal

Como seres humanos, estamos familiarizados de forma innata con los conceptos fundamentales de la inferencia causal. Por el mero hecho de existir, sabemos lo que es un efecto causal, entendemos la diferencia entre asociación y causalidad, y hemos utilizado este conocimiento de forma constante a lo largo de nuestras vidas. Sin embargo, esto es muy diferente para las máquinas (Hernán y Robins, 2020).

Efecto causal individual y promedio

El ser humano razona sobre los efectos causales comparando el resultado cuando se realiza una acción con el resultado cuando no se realiza dicha acción. Si los resultados difieren, se dice que la acción tiene un efecto causal sobre el resultado. En caso contrario,

si los resultados fueran iguales, la acción no tendría ningún efecto causal sobre el resultado. Esta acción también es llamada tratamiento.

Para simplificar, se supone un tratamiento dicótomo, A (1: tratado, 0: no tratado), y un resultado también dicótomo, Y (1: recuperado, 0: no recuperado). Por lo tanto, la variable $Y^{a=1}$ es el resultado que se habría observado bajo el valor de tratamiento $a = 1$, y la variable $Y^{a=0}$, el resultado que se habría observado en el caso $a = 0$. A estas dos variables se les llama resultados potenciales (*potential outcomes*) o resultados contrafactuales.

Bajo esta notación, es posible dar una definición formal de **efecto causal individual** (Hernán y Robins, 2020):

Definición 3.1.1. *El tratamiento A tiene un efecto causal en el resultado Y de un individuo si $Y^{a=1} \neq Y^{a=0}$ para ese individuo.*

Estos efectos causales individuales se definen en función de los resultados potenciales, pero únicamente se observa uno de estos resultados para cada individuo, el correspondiente al tratamiento que realmente recibe. Todos los demás resultados potenciales no se observan. Debido a la falta de datos, estos efectos individuales no pueden ser identificados, es decir, no pueden expresarse como una función de los datos observados.

Debido a que identificar efectos causales individuales es por lo general imposible, es necesario estudiar el **efecto causal medio en una población**:

Definición 3.1.2. *El tratamiento A tiene un efecto causal medio sobre el resultado Y en una población si $P(Y^{a=1} = 1) \neq P(Y^{a=0} = 1)$ en esa población.*

Hay que destacar que esta definición se ajusta al caso en el que las variables son discretas, que es el objeto de estudio en este trabajo. Si se trabajara con variables continuas, habría que precisar más en la definición.

El riesgo constituye una medida de probabilidad estadística de que en un futuro se produzca un acontecimiento. Debido a que en este caso el riesgo es igual a la media y a que la letra E suele emplearse para representar el promedio o esperanza, se puede reescribir esta definición como $E(Y^{a=1}) \neq E(Y^{a=0})$ de modo que la definición se aplica tanto a los resultados dicotómicos como a los no dicotómicos. Cuando el efecto causal medio en la población es nulo, se dice que la hipótesis nula de que no hay efecto causal medio es verdadera.

El efecto causal medio, o simplemente efecto causal en lo sucesivo, es igual a la media de los efectos causales individuales, por lo que la ausencia del primero, no implica necesariamente la ausencia de los segundos. Como se verá, el efecto causal puede, a veces, determinarse a partir de datos, incluso si es imposible determinar los efectos causales individuales. A continuación se describen las diferentes medidas de la magnitud de un efecto causal:

$$DRC = P(Y^{a=1} = 1) - P(Y^{a=0} = 1) \tag{3.7}$$

$$RRC = \frac{P(Y^{a=1} = 1)}{P(Y^{a=0} = 1)} \tag{3.8}$$

Estas medidas se denominan **diferencia de riesgo causal** y **ratio de riesgo causal**, respectivamente. La primera representa el número absoluto de casos favorables atribuibles

al tratamiento, mientras que la segunda calcula cuántas veces el tratamiento, en relación con la ausencia del mismo, aumenta el riesgo de obtener un resultado favorable. Teniendo en cuenta estas definiciones, es trivial ver que la hipótesis nula de ausencia de efecto causal es verdadera si:

$$P(Y^{a=1} = 1) - P(Y^{a=0} = 1) = 0 \tag{3.9}$$

$$\frac{P(Y^{a=1} = 1)}{P(Y^{a=0} = 1)} = 1 \tag{3.10}$$

En este momento, el dicho *asociación no es causalidad* es fácilmente explicable. El riesgo $P(Y = 1|A = a)$ es una probabilidad condicional, esto es, el riesgo de Y en el subconjunto de la población que cumple la condición $A = a$. En cambio, el riesgo $P(Y^a = 1)$ es una probabilidad incondicional o marginal, es decir, el riesgo de Y^a en toda la población. Por lo tanto, la asociación se define por un riesgo diferente en dos subconjuntos disjuntos de la población determinados por el valor de tratamiento real de los individuos ($A = 1$ o $A = 0$), mientras que la causalidad se define por un riesgo diferente en la misma población bajo dos valores de tratamiento diferentes ($a = 1$ o $a = 0$). En la Figura 3.2 se muestra de forma gráfica esta diferencia, donde el rombo representa una población entera.

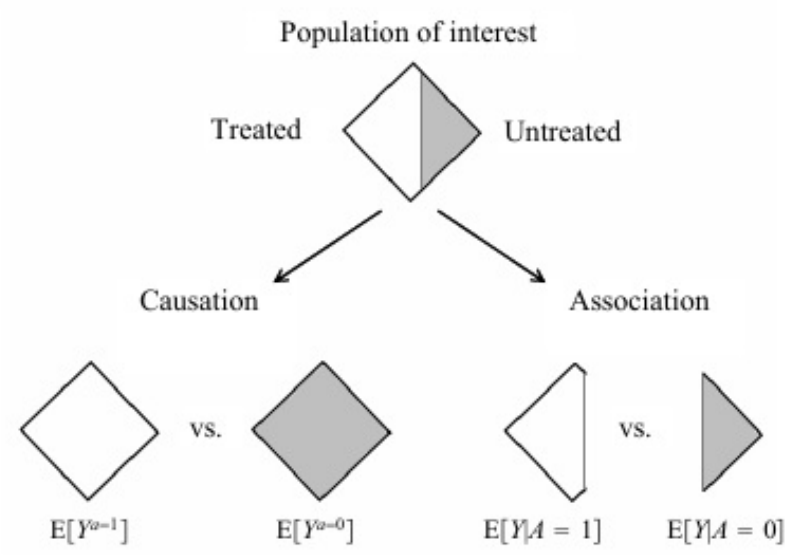

Figura 3.2: Diferencia entre asociación y causalidad (imagen extraída de Hernán y Robins, 2020)

.

La inferencia causal requiere, por lo tanto, datos hipotéticos en los que se conoce el resultado para cada valor del tratamiento en cada individuo de la población, pero lo único que podemos esperar es tener datos del mundo real, es decir, un único resultado para cada individuo correspondiente al valor del tratamiento que reciba. La cuestión es, entonces, en qué condiciones pueden utilizarse los datos del mundo real para la inferencia causal. El la próxima sección se ofrece la respuesta, un experimento aleatorio.

3.1.5. Experimentos Aleatorios

Como se ha mencionado en la sección anterior, en un estudio en el mundo real, sólo se conoce uno de todos los posibles resultados potenciales para cada individuo, el correspondiente al nivel de tratamiento que realmente recibió, resultando así una falta de datos que serían necesarios para calcular las medidas de efecto causal.

Los **experimentos aleatorios** generan datos con valores perdidos de los resultados contrafactuales, como cual otro tipo de estudio. Sin embargo, la aleatorización garantiza que esos valores perdidos se produzcan por azar, por lo tanto, las medidas de efecto causal pueden estimarse a partir de este tipo de experimentos (Hernán y Robins, 2020).

Se supone un **experimento aleatorio ideal** si se realiza sin pérdidas en el seguimiento de los individuos, se mantiene plena adherencia al tratamiento asignado durante toda la duración del estudio, se administra una única versión del tratamiento a cada individuo y se realiza la asignación a doble ciego, es decir, ni los individuos ni los investigadores conocen a quién se le administra cada tratamiento.

Un método para llevar a cabo un experimento aleatorio ideal es la aleatorización simple. Esta está basada en una única secuencia de asignaciones aleatorias. La forma más común y básica de este tipo de aleatorización es el lanzamiento de una moneda para distribuir a los individuos en los diferentes grupos de tratamiento. Este enfoque es simple y fácil de implementar, además de generar un número similar de sujetos entre los grupos si el experimento es de gran envergadura. Por supuesto, existen otros muchos métodos para realizar un experimento aleatorio descritos en la literatura (p.j. Suresh, 2011)

Llevar a cabo un experimento aleatorio ideal implica la condición de **intercambiabilidad**. Esto significa que si los tratamientos se aplicaran de una manera diferente a los grupos de individuos, el resultado que cabría esperar sería el mismo, es decir, el riesgo de obtener un resultado en un grupo hubiera sido el mismo que el riesgo de obtener el mismo resultado en otro grupo si los individuos del primer grupo hubieran recibido el tratamiento que recibieron los del segundo. Matemáticamente, el riesgo bajo el tratamiento potencial a cumple que:

$$P(Y^a|A=1) = P(Y^a|A=0) = P(Y^a) \tag{3.11}$$

con $a=0$ o $a=1$ y $Y^a=0$ o $Y^a=1$. La última igualdad sostiene una consecuencia obvia, los riesgos condicionales deben ser iguales al riesgo marginal bajo el tratamiento a en toda la población.

Debido a que el riesgo contrafactual bajo el valor del tratamiento a es el mismo en ambos grupos $A=1$ y $A=0$, se dice que el tratamiento real A no predice el resultado contrafactual Y^a. Equivalentemente, la intercambiabilidad implica que el resultado contrafactual y el tratamiento actual son independientes para todos los posibles valores de a, lo cual se puede representar como $Y^a \perp\!\!\!\perp A$.

Este razonamiento conduce a que, si existe esta intercambiabilidad, el riesgo contrafactual bajo un tratamiento en un grupo sería igual al riesgo contrafactual bajo ese mismo tratamiento en la población entera. Pero en realidad el riesgo en el grupo no es en absoluto contrafactual, ya que este grupo realmente fue, o no, tratado. Por lo tanto, un experimento aleatorio ideal permite calcular el riesgo contrafactual bajo un tratamiento en la población:

$$P(Y^{a=1}) = P(Y|A=1) \tag{3.12}$$

$$P(Y^{a=0}) = P(Y|A = 0) \tag{3.13}$$

Y por lo tanto, calcular las medidas de efecto causal, ya que en este caso, la asociación si es causalidad.

Hasta este momento, solo se ha tenido en cuenta una única probabilidad de aleatoriedad marginal, la cual es común a todos los individuos. Es por ello que a este tipo de experimentos también se les conoce como **experimentos marginalmente aleatorios**. En el caso de tener alguna otra variable que condicione el experimento, se tendrán varias probabilidades de aleatorización que dependen de los valores de esta variable, este es el caso de estudio en la siguiente sección.

3.1.6. Experimentos Condicionalmente Aleatorios

En un **experimento condicionalmente aleatorio** existe alguna variable, L, que condiciona el experimento y por lo tanto los resultados del mismo. En este tipo de experimentos no suele darse la condición de intercambiabilidad, ya que por diseño, cada grupo puede tener una proporción diferente de individuos con un pronóstico. Sin embargo, si se entienden como la combinación de dos experimentos marginalmente aleatorios, uno para cada valor de la variable condicional ($L = 1$ y $L = 0$), en cada uno de ellos los tratados pueden ser intercambiados por los no tratados:

$$P(Y^a|A = 1, L = l) = P(Y^a|A = 0, L = l) \tag{3.14}$$

con $l = 0$ o $l = 1$. Es decir, Y^a y A son independientes dado $L = l$, lo cual se puede denotar como $Y^a \perp\!\!\!\perp A|L = l$.

Por lo tanto, la aleatorización condicional no garantiza intercambiabilidad marginal, $Y^a \perp\!\!\!\perp A$, pero si que garantiza intercambiabilidad condicional, $Y^a \perp\!\!\!\perp A|L$ dentro de los niveles de la variable L.

En este caso, sería posible calcular el efecto causal en cada uno de los subconjuntos o estratos de la población, ya que en cada uno de ellos, la asociación si es causalidad y se cumple que:

$$P(Y^a|L = l) = P(Y|L = l, A = a) \tag{3.15}$$

con $l = 0$ o $l = 1$ y $a = 0$ o $a = 1$.

Este método de calcular los efectos causales específicos de cada estrato se denomina **estratificación**. Si la medida del efecto causal varía de un estrato al otro, se dice que existe una **modificación del efecto** por parte de L. Para calcular el efecto causal del tratamiento en toda la población en vez de únicamente en un estrato de ella, se presentan dos métodos.

Estandarización

El riesgo contrafactual marginal, $P(Y^a)$, es la media ponderada de los riesgos específicos de cada estrato de la población, $P(Y^a|L = 0)$ y $P(Y^a|L = 1)$, con pesos iguales a la proporción de individuos de la población con $L = 0$ y $L = 1$ respectivamente (Hernán

y Robins, 2020). Es decir, aplicando la ley de la probabilidad total, Ecuación 3.3, y la regla del producto, Ecuación 3.4, se tiene:

$$P(Y^a) = \sum_l P(Y^a|L = l)P(L = l) \tag{3.16}$$

Utilizando la intercambiabilidad condicional, se obtiene una expresión para el riesgo contrafactual que depende únicamente de cantidades observadas y permite calcular las medidas del efecto causal:

$$P(Y^a) = \sum_l P(Y|L = l, A = a)P(L = l) \tag{3.17}$$

Cuando una cantidad contrafactual puede expresarse en función de la distribución de los datos observados, se dice que la cantidad contrafactual es **identificable**.

Ponderación de la Probabilidad Inversa

Uno de los mayores problemas del análisis de datos es que los grupos estén desequilibrados. Esto puede crear una comparación sesgada entre ellos.

Una posible solución consiste en ponderar cada grupo de manera que refleje la distribución global de los tamaños en lugar de los individuos de cada grupo de tratamiento. Estas ponderaciones son precisamente las inversas de la puntuación de propensión, es decir, la inversa de la probabilidad de ser asignado a un determinado grupo de tratamiento, dados los atributos de los individuos (Villa, 2021):

$$W^a = \frac{1}{P(A|L)} \tag{3.18}$$

Por lo tanto, multiplicando la ecuación 3.17 por la puntuación de propensión se obtiene la fórmula de la Ponderación de la Probabilidad Inversa:

$$P(Y^a) = \sum_l \frac{P(Y|L = l, A = a)P(L = l)P(A = a|L = l)}{P(A = a|L = l)} \tag{3.19}$$

$$P(Y^a) = \sum_l \frac{P(Y, A = a, L = l)}{P(A = a|L = l)} = \sum_l P(Y, A = a, L = l)W^a \tag{3.20}$$

Donde la Ecuación 3.20 se ha obtenido aplicando la regla del producto, ver Ecuación 3.4.

Todo el análisis teórico-práctico presentado en esta sección ayuda a comprender cómo hay que enfocar el estudio de la causalidad, qué es el efecto causal y cómo, bajo unas ciertas condiciones, es posible calcularlo. Además, es la base teórica sobre la que se construye el Marco de Resultados Potenciales, uno de los modelos causales más utilizados. En la siguiente sección, se ofrece una visión global sobre los diferentes Modelos Causales y sus principales características, profundizando en el más práctico de ellos para el proyecto, los Modelos Causales Estructurales.

3.2. Modelos Causales

Un **Modelo Causal** es una abstracción matemática que cuantitativamente describe las relaciones causales entre variables. En primer lugar, los supuestos causales o el conocimiento causal previo puede representarse mediante un Modelo Causal incompleto. Luego, lo que falta puede aprenderse a partir de los datos (Pearl, 2019).

Los dos modelos causales más conocidos son el Marco de Resultados Potenciales (POF, por sus siglas en inglés) y los Modelos Causales Estructurales (SCM, por sus siglas en inglés). Ambos permiten una representación coherente de los conocimientos causales previos, las suposiciones y las estimaciones, por lo que se les considera los fundamentos del análisis causal. En particular, el POF toma como punto de partida los resultados potenciales y los relaciona mediante reglas de observación con los resultados observados. Por el contrario, la perspectiva de los SCM define un modelo basado en los resultados observados del que se pueden derivar los resultados potenciales.

Ambos marcos teóricos son lógicamente equivalentes, lo que significa que una suposición en uno siempre puede traducirse a su contraparte en el otro. Sin embargo, existen algunas diferencias entre ellos. En el POF, los efectos causales de las variables distintas del tratamiento, como las variables instrumentales, no se definen, mientras que en los SCM es posible estudiar el efecto causal de cualquier variable. Por lo tanto, si se conocen con certeza todas las relaciones causales entre un conjunto de variables, es preferible utilizar los SCM, mientras que si únicamente el objetivo es estimar un determinado efecto del tratamiento, el POF puede resultar mucho más sencillo. En el proyecto WHY, el objetivo es el análisis causal de un amplio rango de variables, por lo que se utilizarán SCM.

3.2.1. Modelos Causales Estructurales

Formalmente, un **Modelo Causal Estructural** consiste en dos conjuntos de variables, U y V, y una serie de funciones f, denominadas ecuaciones estructurales, que asignan a cada variable en V un valor, basado en los valores de las demás variables del modelo (Pearl, Glymour y Jewell, 2016).

Las variables en U se denominan **variables exógenas**, lo que significa que son externas al modelo, es decir, se elige, por la razón que sea, no explicar cómo son causadas. Las variables en V son **variables endógenas**, es decir, son parte del modelo. Toda variable endógena es descendiente de al menos una exógena, mientras que las exógenas no pueden ser descendientes de ninguna otra variable. Si se conoce el valor de todas las variables exógenas, se puede determinar a partir de las funciones en f el valor de todas las variables endógenas con certeza.

Todo SCM está asociado con un **Diagrama Causal**. Estos diagramas causales consisten en un conjunto de nodos que representan las variables en U y V y un conjunto de aristas que representan las funciones en f. Si la función f_X para una variables X en V contiene a la variable Y, es decir, X depende de Y para su valor, entonces existe una arista dirigida desde Y hasta X. Estos diagramas causales serán principalmente **Gráficos Acíclicos Dirigidos** (DAG, por sus siglas en inglés).

Los diagramas causales nos permiten dar una definición gráfica de causalidad: si, en un diagrama causal, una variable X es hija de otra variable Y, entonces Y es causa directa de X; si X es descendiente de Y, entonces Y es una causa potencial de X. De esta manera,

los diagramas causales codifican los supuestos causales.

Una ventaja de los diagramas causales es que permiten expresar las distribuciones conjuntas de una forma muy eficiente. Para cualquier DAG, la distribución conjunta de las variables viene dada por el producto de las distribuciones condicionales $P(hijo|padres)$ sobre todas las familias en el grafo, matemáticamente se expresa de la siguiente manera:

$$P(x_1, x_2, ..., x_n) = \prod_i P(x_i|pa_i) \tag{3.21}$$

donde pa_i representa los valores de los padres de la variable x_i e $i = 1, ..., n$. Esta regla permite ahorrar tiempo y memoria a la hora de calcular distribuciones conjuntas.

A pesar de que los SCM contengan algo más de información que los diagramas causales, son estos últimos los que proporcionan una comprensión más intuitiva de la causalidad. Normalmente el conocimiento que se tiene sobre las relaciones causales no es cuantitativo, como exige un SCM, sino cualitativo, como representa un diagrama causal. En la siguiente sección, se verá cómo los diagramas causales revelan mucha más información de la que resulta obvia a primera vista, se puede aprender mucho sobre un conjunto de datos e inferir sobre él utilizando únicamente el grafo de su historia causal.

3.2.2. Diagramas Causales y sus Aplicaciones

Otra forma de entender los modelos causales estructurales consiste en pensarlos como el mecanismo por el cual se generan los datos. Como se dijo en la sección anterior, normalmente se conoce qué variables son causadas por otras, pero se desconoce la fuerza o la naturaleza de estas relaciones. Incluso con una información tan limitada, se puede discernir mucho sobre el conjunto de datos generado por el modelo.

A partir del diagrama causal se puede determinar qué variables del conjunto de datos son independientes entre sí y cuáles son independientes entre sí condicionadas en otras variables. Estas independencias serán ciertas en cada conjunto de datos generado por cualquier SCM con una misma estructura gráfica, independientemente de las ecuaciones estructurales específicas que generen cada conjunto de datos.

A continuación se presentan los DAG más simples y relevantes con los que se puede construir cualquier diagrama causal y se detallan las relaciones de independencia que subyacen tras su estructura.

Cadena

La configuración que se muestra en la Figura 3.3, tres nodos y dos aristas, con una arista dirigida hacia la variable central y la otra saliendo de ella, se denomina **cadena**.

Teniendo en cuenta que en cualquier diagrama causal dos variables que estén conectadas a través de una arista son dependientes, simplemente observando la estructura de la Figura 3.3, se puede obtener el siguiente Teorema:

Teorema 3.2.1. *Relación entre variables en una cadena, $X \to Y \to Z$:*

- ***Z e Y son dependientes***
  Para algún z e y, $P(Z = z|Y = y) \neq P(Z = z)$

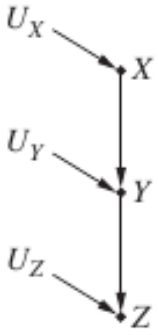

Figura 3.3: Diagrama causal de una cadena donde se muestran las variables exógenas que afectan a las variables endógenas X, Y y Z.

- ***Y y X son dependientes***
  Para algún y y x, $P(Y = y|X = x) \neq P(Y = y)$
- ***Z y X son probablemente dependientes***
  Para algún z y x, $P(Z = z|X = x) \neq P(Z = z)$
- ***Z y X son independientes condicionadas en Y***
  Para todos los x, y, z, $P(Z = z|X = x, Y = y) = P(Z = z|Y = y)$

El tercer punto viene dado a partir del hecho de que si Z depende de Y, e Y depende de X, entonces Z probablemente dependa de X. Sin embargo, condicionando en Y, es decir, filtrando los datos en grupos de tal manera que el valor de Y sea constante, el valor de Z permanecería inalterado, ya que únicamente depende de Y y U_Z. Por lo tanto, la selección de un valor diferente de X no cambia el valor de Z, llevando a la conclusión del cuarto punto.

Un razonamiento análogo al anterior, llevaría a que en cualquier modelo gráfico, dadas dos variables cualesquiera X e Y, si el único camino entre ellas se compone enteramente de cadenas, estas dos variables serían independientes condicionadas a cualquier variable intermedia en ese camino. Esta relación de independencia se mantiene sean cuales sean las funciones que conecten las variables, lo cual conlleva a una regla (Pearl, Glymour y Jewell, 2016):

Definición 3.2.1. Regla 1 (Independencia Condicional en Cadenas) *Dos variables, X e Y, son condicionalmente independientes dado Z, si sólo hay un camino unidireccional entre X e Y y Z es cualquier conjunto de variables que intercepta ese camino.*

Esta regla sólo se mantiene si se asume que los términos de error U_X, Y_Y y U_Z son independientes entre si.

Bifurcación

La configuración que se muestra en la Figura 3.4, tres nodos, con dos aristas saliendo de la variable central, se denomina **bifurcación**. La variable central en un bifurcación se denomina **causa común** de las otras dos variables y de sus descendientes.

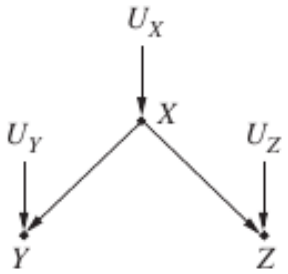

Figura 3.4: Diagrama causal de un bifurcaciónor donde se muestran las variables exógenas.

Asumiendo que los términos de error son independientes, examinando el diagrama causal de la Figura 3.4 se puede obtener el siguiente Teorema:

Teorema 3.2.2. *Relación entre variables en una bifurcación,* $Y \leftarrow X \rightarrow Z$*:*

- X ***e*** Y ***son dependientes***
  Para algún x *e* y, $P(X = x|Y = y) \neq P(X = x)$
- X ***y*** Z ***son dependientes***
  Para algún x *y* z, $P(X = x|Z = z) \neq P(X = x)$
- Z ***e*** Y ***son probablemente dependientes***
  Para algún z *e* y, $P(Z = z|Y = y) \neq P(Z = z)$
- Y ***y*** Z ***son independientes condicionadas en*** X
  Para todos los x, y *y* z, $P(Y = y|X = x, Z = z) = P(Y = y|X = x)$

El primer y segundo punto vienen, una vez más, del hecho de que Y y Z están directamente conectados con X, por lo que cuando el valor de X cambia, los valores de Y y Z cambian. Si Y cambia cuando X cambia, y Z cambia cuando X cambia, probablemente Y cambia junto con Z, y viceversa. Si se condiciona en X, se comparan casos en los que X es constante, por lo que los valores de Y y Z no cambian de acuerdo a X, haciéndolo únicamente respecto a U_Y y U_Z, respectivamente. Como estos últimos son independientes, se llega al tercer punto.

Siguiendo el mismo razonamiento que en el caso de la cadena, si se comparan los casos en los que el valor de X es constante, los valores de Y y Z no cambian en función de X, sólo cambian en respuesta a U_Y y U_Z, que se han asumido como independientes. Por lo tanto, cualquier cambio adicional en los valores de Y y Z debe ser independiente entre sí, llevando al cuarto punto.

De nuevo, este razonamiento conduce a una regla (Pearl, Glymour y Jewell, 2016):

Definición 3.2.2. Regla 2 (Independencia Condicional en Bifurcaciones) *Si una variable* X *es causa común de las variables* Y *y* Z*, y sólo existe un camino entre* Y *y* Z*, entonces* Y *y* Z *son independientes condicionadas en* X*.*

Colisionador

La última configuración, la cual se muestra en la Figura 3.5, consiste en tres nodos, con dos aristas llegando a la variable central y se denomina **colisionador**. La variable central recibe el nombre de nodo de colisión o colisionador y representa un efecto común de dos causas, las otras dos variables.

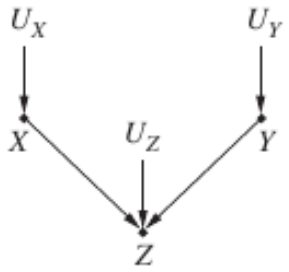

Figura 3.5: Diagrama causal de un colisionador donde se muestran las variables exógenas.

Al igual que en los otro casos, observando el diagrama causal de la Figura 3.5, se puede obtener el siguiente Teorema:

Teorema 3.2.3. *Relación entre variables en un colisionador,* $X \to Z \leftarrow Y$*:*

- X ***y*** Z ***son dependientes***
  Para algún x *y* z, $P(X = x|Z = z) \neq P(X = x)$
- Y ***y*** Z ***son dependientes***
  Para algún y *y* z, $P(Y = y|Z = z) \neq P(Y = y)$
- X ***e*** Y ***son independientes***
  Para todos los x *e* y, $P(X = x|Y = y) = P(X = x)$
- X ***e*** Y ***son dependientes condicionadas en*** Z
  Para algún x, y *y* z, $P(X = x|Y = y, Z = z) \neq P(X = x|Z = z)$

Los dos primeros puntos siguen el mismo razonamiento que en los casos anteriores. El tercer punto en este caso es una evidencia, ya que ni X ni Y son descendientes o ancestros el uno del otro ni dependen entre si para sus respectivos valores. Sin embargo, si se condiciona en Z, se limitan los casos a los que Z tiene un valor concreto. Pero Z depende de X e Y, por lo tanto, cualquier cambio en el valor de X ha de ser compensado por un cambio en el valor de Y, de otra manera, el valor de Z cambiaría.

Estas consideraciones llevan a una tercera regla (Pearl, Glymour y Jewell, 2016):

Definición 3.2.3. Regla 3 (Independencia Condicional en Colisionadores) *Si una variable* Z *es el nodo de colisión entre dos variables* X *e* Y*, y únicamente existe un camino entre* X *e* Y*, entonces* X *e* Y *son incondicionalmente independientes pero son dependientes condicionalmente de* Z *y de cualquier descendiente de* Z*.*

Los modelos causales no suelen ser tan sencillos como los casos que se han estudiado hasta ahora. En concreto, es raro que un diagrama causal consista en un único camino entre variables. En la mayoría, los pares de variables tendrán múltiples caminos posibles que los conecten, y cada uno de esos caminos atravesará una variedad de cadenas, bifurcaciones y colisionadores. En la siguiente sección se presenta un proceso que puede ser aplicado a cualquier diagrama causal, sea cual sea su complejidad, para predecir las dependencias entre las variables.

3.2.3. d-Separación

El criterio de la ***d*-separación** permite determinar, para cualquier par de nodos, si están ***d*-conectados**, lo que significa que existe un camino de conexión entre ellos, es decir, las variables son probablemente dependientes, o, por el contrario, si están ***d*-separados**, lo que significa que no existe tal camino y por lo tanto las variables que representan son definitivamente independientes. Es suficiente con que exista un único camino no bloqueado para que ambos nodos estén d-conectados.

Teniendo en cuenta las relaciones obtenidas en la sección anterior, se puede establecer una clasificación de los tipos de nodos que pueden bloquear un camino, dependiendo de si se realiza una d-separación incondicional o condicional:

- Si no se condiciona en ninguna variable, únicamente los nodos de colisión pueden bloquear un camino, ya que la dependencia incondicional no puede pasar a través de ellos.
- Si se condiciona en un conjunto de nodos Z, los siguientes tipos de nodos pueden bloquear un camino:
  - Un nodo de colisión que no está en Z y que no tiene descendientes en Z (ver Definición 3.2.1).
  - Una cadena o una bifurcación cuyo nodo central está en Z (ver Definiciones 3.2.2 y 3.2.3).

Teniendo en cuenta estas relaciones, es posible dar una definición formal de la d-separación:

Definición 3.2.4. *d*-separación *Un camino p está bloqueado por un conjunto de nodos Z si y solo si:*

1. *p contiene una cadena de nodos $A \to B \to C$ o una bifurcación $A \leftarrow B \to C$ tal que el nodo central B está en Z, es decir, B está condicionado, o*
2. *p contiene un colisionador $A \to B \leftarrow C$ tal que el nodo colisionador B no está en Z, y ningún descendiente de B está en Z*

Si Z bloquea todos los caminos entre dos nodos X e Y, entonces X e Y están d-separados, condicionado en Z, y por lo tanto son independientes condicionado en Z.

Con el criterio de la d-separación es posible analizar diagramas causales más complejos y determinar qué variables son independientes y dependientes, tanto marginalmente como condicionadas a otras variables. Esto es posible independientemente del modelo causal al que pertenezca el diagrama, es decir, sean cuales sean los tipos de variables y las relaciones que existan entre ellas. En este punto, es posible adentrarse en el objetivo final de muchos estudios estadísticos, la predicción de los efectos de las intervenciones.

3.2.4. Intervenciones

Cuando se interviene para fijar el valor de una variable, se restringe la tendencia natural de esa variable a variar en respuesta a otras variables. Esto equivale a realizar una especie de cirugía en el diagrama causal, eliminando todas las aristas dirigidas a esa variable.

En la Figura 3.6 se presenta cómo cambiaría el diagrama causal correspondiente al ejemplo presentado en la Sección 3.1.3 tras realizar una intervención en el tratamiento A. Se puede observar como al intervenir en una variable se da lugar a un patrón de dependencias totalmente diferente al patrón de dependencias que ocasiona el condicionamiento en la misma.

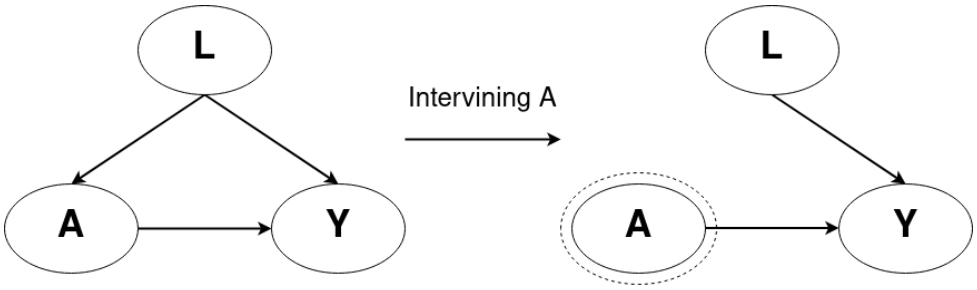

Figura 3.6: Diagrama causal del Ejemplo 3.1.1 tras una intervención en A.

Para conocer la eficacia de un tratamiento en la población, se imagina una hipotética intervención en la que se administra dicho tratamiento de manera uniforme a toda la población y el resultado se compara con el resultado que se obtendría con la intervención complementaria. Siguiendo con el Ejemplo 3.1.1, esto se denotaría como:

$$P(Y = y|do(A = 1)) - P(Y = y|do(A = 0)) \tag{3.22}$$

Esta diferencia se conoce como **diferencia de efecto causal** o **efecto causal medio**, que como se puede comprobar, se corresponde con la expresión 3.7.

Como bien se conoce hasta el momento, los efectos causales no se pueden estimar directamente de los datos sin un diagrama causal. Sin embargo, a partir de los diagramas causales de la Figura 3.6, se puede obtener una expresión para calcular dicho efecto a partir de la Ecuación 3.22.

La clave para calcularlo, consiste en tener en cuenta que **el efecto causal es igual a la probabilidad condicionada en el diagrama manipulado tras la intervención**, es decir:

$$P(Y = y|do(A = a)) = P_m(Y = y|A = a) \tag{3.23}$$

De tal manera que la probabilidad manipulada P_m comparte dos propiedades esenciales con la probabilidad P, correspondiente a la probabilidad original en el diagrama antes de ser intervenido:

- La probabilidad marginal $P(L = l)$ es invariante bajo la intervención (las proporciones de hombres y mujeres se mantienen constantes tras la intervención)
- La probabilidad condicional $P(Y = y|L = l, A = a)$ es invariante bajo la intervención, ya que el proceso por el que Y responde a A y L es el mismo, independientemente de si A cambia espontáneamente o por manipulación.

Además, se puede comprobar que L y A están d-separados en el diagrama modificado, ya que cumplen los criterios dados en la Definición 3.2.4, lo que significa que $P_m(L = l|A = a) = P_m(L = l) = P(L = l)$. Por lo tanto:

$$P(Y = y|do(A = a)) = P_m(Y = y|A = a) = \tag{3.24}$$

$$= \sum_l P_m(Y = y|A = a, L = l)P_m(L = l|A = a) = \tag{3.25}$$

$$= \sum_l P_m(Y = y|A = a, L = l)P_m(L = l) \tag{3.26}$$

Donde la Ecuación 3.24 viene dada por definición, es decir, aplicando directamente la Ecuación 3.23. La Ecuación 3.25 se ha obtenido a partir de la Regla de Bayes, Ecuación 3.6, condicionando en L y sumando sobre todos sus posibles valores. Por último, la Ecuación 3.26 se obtiene a partir de la independencia de L y A en el diagrama modificado. Por lo tanto, utilizando las relaciones de invarianza anteriormente descritas, se obtiene una expresión para el efecto causal en términos de las probabilidades antes de la intervención:

$$P(Y = y|do(A = a)) = \sum_l P(Y = y|A = a, L = l)P(L = l) \tag{3.27}$$

Conocida como la **fórmula de ajuste**, que como se puede ver, calcula la asociación entre A e Y para cada valor z de Z, haciendo la media sobre esos valores. Además, se puede comprobar que esta expresión se corresponde la Ecuación 3.17. Cabe destacar que en un experimento controlado aleatorio no es necesario ningún ajuste, ya que, en ese caso, los datos son generados por un modelo que ya posee la estructura modificada, por lo que $P_m = P$ independientemente de cualquier factor que afecte a Y. Esta derivación de la Ecuación 3.27 constituye, por tanto, una prueba formal de que la aleatorización permite obtener la cantidad que se busca estimar, $P(Y = y|do(A = a))$.

Esta fórmula de ajuste es fácilmente generalizable. El procedimiento que ha conducido a esta fórmula dicta que L debe coincidir con los padres de A, que se denotan por PA, ya que es la influencia de estos padres la que se neutraliza cuando se fija el valor de A mediante una manipulación externa. Por lo tanto, se puede establecer una regla (Pearl, Glymour y Jewell, 2016):

Definición 3.2.5. (La Regla del Efecto Causal) *Dado un grafo G en el cual un conjunto de variables PA son designadas como los padres de X, el efecto causal de X en Y viene dado por:*

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, PA = z)P(PA = z) \tag{3.28}$$

o de una manera más conveniente y resumida:

$$P(y|do(x)) = \sum_z \frac{P(x, y, z)}{P(x|z)} \tag{3.29}$$

donde z abarca todas las combinaciones de valores que pueden tomar las variables PA.

Como se puede comprobar, La Regla del Efecto Causal no es más que una aplicación del Teorema de la Probabilidad Total, Ecuaciones 3.3 y 3.4. Además, existe una gran similitud entre esta expresión y la Ecuación 3.20, y de igual manera, el factor $P(X = x|PA = z)$ se conoce como puntuación de propensión.

Este resultado es muy interesante, ya que utilizando los diagramas causales y sus supuestos subyacentes, es posible identificar las relaciones causales en datos puramente observacionales, es decir, aquellos que no provienen de experimentos controlados.

Sin embargo, en la mayoría de los casos prácticos, esto no es tan sencillo como en el ejemplo mostrado. Normalmente el conjunto de padres de X contendrá variables no observadas que impedirían calcular las probabilidades condicionales necesarias para utilizar la fórmula de ajuste. En las dos próximas secciones, se muestran dos criterios que permiten ajustar otras variables en el diagrama causal para sustituir los elementos no medidos de PA, el criterio de puerta trasera y el criterio de puerta delantera.

3.2.5. Criterio de Puerta Trasera

El **Criterio de Puerta Trasera** permite determinar, para cualesquiera dos variables X e Y en un modelo causal representado por un DAG, sobre qué conjunto de variables Z de ese modelo se debe condicionar cuando se busca la relación causal entre X e Y (Pearl, Glymour y Jewell, 2016).

Definición 3.2.6. (El Criterio de Puerta Trasera) *Dado un par de variables (X, Y) ordenadas en un gráfico acíclico dirigido G, un conjunto de variables Z satisface el criterio de puerta trasera relativo a (X, Y) si ningún nodo en Z es descendiente de X, y Z bloquea todos los caminos entre X e Y que contienen una arista hacia X.*

Si un conjunto de variables Z satisface este criterio para X e Y, entonces el efecto causal de X sobre Y viene dado por la siguiente fórmula:

$$P(Y = y|do(X = x)) = \sum_z P(Y = y|X = x, Z = z)P(Z = z) \tag{3.30}$$

Cabe destacar que los padres de X siempre satisfacen el criterio de puerta trasera, por lo que siempre se puede utilizar esta ecuación si $Z = PA(X)$.

La lógica detrás de este criterio es bastante sencilla. Por lo general, se busca condicionar a un conjunto de nodos Z tal que:

- Se bloqueen todos los caminos espurios entre X e Y, es decir, todos los caminos que no pertenezcan al camino causal o directo de X a Y
- Dejar todos los caminos directos de X a Y sin alterar
- No crear nuevos caminos espurios entre X e Y

Cuando se quiere encontrar el efecto causal de X sobre Y, se busca que los nodos sobre los que se condiciona bloqueen cualquier camino *de puerta trasera* en el que un extremo tenga una arista hacia X, ya que tales caminos pueden hacer que X e Y sean dependientes aunque, obviamente, no están transmitiendo influencias causales de X, y si no se bloquean, confundirán el efecto que X tiene sobre Y. Por lo tanto, se condiciona sobre los caminos de puerta trasera para cumplir el primer requisito.

Por otro lado, no se quiere condicionar ningún nodo que sea descendiente de X. Los descendientes de X se verían afectados por una intervención en X y podrían afectar ellos mismos a Y; condicionarlos bloquearía esas vías. Por lo tanto, no se condiciona a los descendientes de X, cumpliendo el segundo requisito.

Por último, para cumplir el tercer requisito, no se debe de condicionar sobre ningún colisionador que desbloquee un nuevo camino entre X e Y. El requisito de excluir a los descendientes de X también protege de condicionar a los hijos de los nodos intermedios entre X e Y. Tal condicionamiento distorsionaría el paso de la asociación causal entre X e Y, de forma similar a como lo haría el condicionamiento sobre sus padres.

3.2.6. Criterio de Puerta Delantera

El criterio de la puerta trasera proporciona un método sencillo para identificar los conjuntos de covariables, es decir, variables que posiblemente predicen el resultado bajo estudio, que deben ajustarse cuando se busca estimar los efectos causales a partir de datos observacionales. Sin embargo, no es la única manera de estimar dichos efectos. El operador $do()$ puede aplicarse a diagramas causales que no satisfacen el criterio de puerta trasera para identificar efectos que a primera vista parecen estar fuera de nuestro alcance. Uno de esos criterios es el **Criterio de Puerta Delantera** (Pearl, Glymour y Jewell, 2016).

Definición 3.2.7. (Criterio de Puerta Delantera) *Un conjunto de variables Z se dice que satisface el criterio de puerta delantera relativo a un par ordenado de variables (X, Y) si:*

1. *Z intercepta todos los caminos directos desde X hasta Y*
2. *No existe ningún camino sin bloquear desde X hasta Z*
3. Todos los caminos de puerta trasera de Z a Y están bloqueados por X

Teorema 3.2.4. ***(Ajuste de Puerta Delantera)*** *Si Z satisface el criterio de puerta delantera relativo a (X, Y) y si $P(x, z) > 0$, entonces el efecto causal de X sobre Y es identificable y viene dado por la fórmula de puerta delantera:*

$$P(Y = y|do(X = x)) = \sum_{z} P(Z = z|X = x) \sum_{x'} P(Y = y|X = x', Z = z)P(X = x') \tag{3.31}$$

donde z representa todos los posibles valores que pueden tomar las variables en Z y x' todos los posibles valores que puede tomar X.

Las condiciones establecidas en la Definición 3.2.7 son demasiado conservadoras, algunos de los caminos de puerta trasera excluidos por las dos últimas condiciones en realidad pueden estar permitidos siempre que estén bloqueados por algunas variables.

La combinación de la fórmula de ajuste, el criterio de puerta trasera y el criterio de puerta delantera cubre numerosos escenarios causales. Sin embargo, existe una poderosa maquinaria simbólica denominada *do-Calculus* que permite analizar todo tipo de diagramas causales sin importar su complejidad. De hecho, el *do-Calculus* descubre todos los efectos causales que pueden identificarse a partir de un gráfico dado.

3.2.7. do-Calculus

En esta sección se establece un conjunto de reglas de inferencia mediante las cuales las operaciones probabilísticas que implican intervenciones y observaciones pueden transformarse en otras sentencias de este tipo, proporcionando así un método sintáctico para derivar afirmaciones sobre las intervenciones. Cada regla de inferencia respetará la interpretación del operador $do()$ como una intervención que modifica un conjunto selecto de funciones en el modelo subyacente. El conjunto de reglas de inferencia que emergen de esta interpretación se llamará *do-Calculus* (Pearl, 2000 y Shpitser y Pearl, 2008).

Se supone conocida la estructura del diagrama causal en el que alguno de los nodos son observables mientras que otros permanecen inobservados. El objetivo es facilitar la reducción gradual de expresiones del tipo $P(y|do(x))$ a expresiones equivalentes que impliquen probabilidades estándar de las cantidades observadas, lo que se conoce como derivación del efecto causal. Siempre que dicha reducción sea factible, el efecto causal es identificable.

Sean X, Y y Z conjuntos disjuntos de nodos en un DAG causal G. Se denota por $G_{\bar{X}}$ el grafo obtenido al eliminar de G todas las aristas que apuntan a los nodos en X. De la misma manera, se denota por $G_{\underline{X}}$ el grafo obtenido al eliminar de G todas las aristas salientes de los nodos en X. Para representar la eliminación de las aristas tanto entrantes como salientes, se utiliza $G_{\bar{X}\underline{Z}}$, ver Figura 3.7.

Teorema 3.2.5. ***(Reglas del do-Calculus)*** *Sea G un gráfico acíclico dirigido asociado a un modelo causal, y sea $P(\dot{})$ la distribución de probabilidad inducida por dicho modelo. Para cualesquiera subconjuntos disjuntos de variables X, Y, Z y W, se tienen las siguientes reglas:*

Regla 1 *Inserción/supresión de observaciones*

$$P(y|do(x), z, w) = P(y|do(x), w) \text{ si } (Y \perp\!\!\!\perp Z|X; W)_{G_{\bar{X}}} \tag{3.32}$$

Regla 2 *Intercambio de acciones/observaciones*

$$P(y|do(x), do(z), w) = P(y|do(x), z, w) \text{ si } (Y \perp\!\!\!\perp Z|X; W)_{G_{\bar{X}\underline{Z}}} \tag{3.33}$$

Regla 3 *Inserción/supresión de acciones*

$$P(y|do(x), do(z), w) = P(y|do(x), w) \text{ si } (Y \perp\!\!\!\perp Z|X; W)_{G_{\bar{X}\bar{Z(W)}}} \tag{3.34}$$

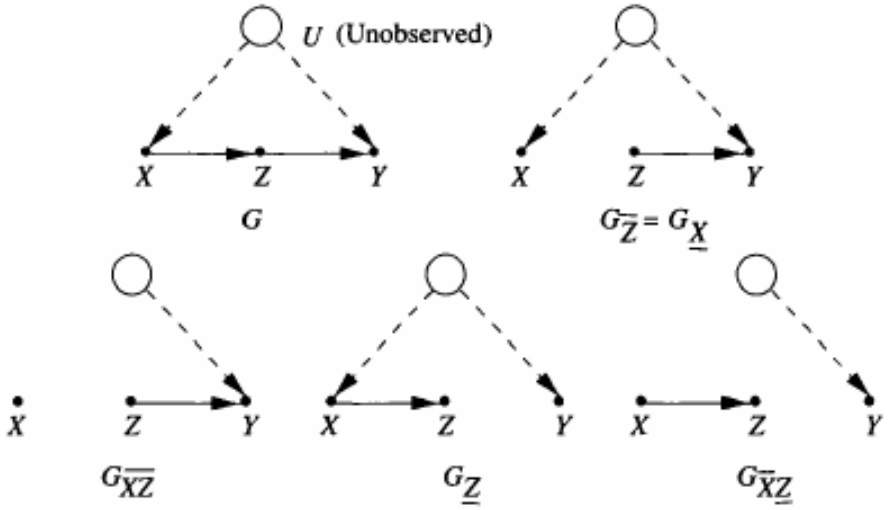

Figura 3.7: Subgrafos de G utilizados en la derivación de efectos causales (imagen extraída de Pearl, 2000).

> *donde $Z(W)$ es el conjunto de Z-nodos que no son ancestros de ningún W-nodo en $G_{\bar{X}}$*

La Regla 1 reafirma la *d*-separación como una prueba válida para la independencia condicional en la distribución resultante de la intervención $do(X = x)$, por ello el grafo $G_{\bar{X}}$. Esta regla se deriva del hecho de que la eliminación de las ecuaciones del sistema no introduce ninguna dependencia entre los términos de perturbación restantes.

La Regla 2 proporciona una condición para que una intervención externa $do(Z = z)$ tenga el mismo efecto sobre Y que la observación pasiva $Z = z$. La condición equivale a que $\{X \cup W\}$ bloquee todos los caminos de puerta trasera de Z a Y (en $G_{\bar{X}}$), ya que $G_{\bar{X}\underline{Z}}$ conserva todos y exclusivamente esos caminos.

La Regla 3 establece las condiciones para introducir, o eliminar, una intervención externa $do(Z = z)$ sin afectar a la probabilidad de $Y = y$. La validez de esta regla se deriva, de nuevo, de simular la intervención $do(Z = z)$ mediante la supresión de todas las ecuaciones correspondientes a las variables de Z, de ahí el grafo $G_{\bar{X}\bar{Z}}$.

Corolario 3.2.5.1. *Un efecto causal $q = P(y_l, ..., y_k)|do(x_1), ..., do(x_m)$ es identificable en un modelo caracterizado por un grafo G si existe una secuencia finita de transformaciones, cada una de ellas conforme a una de las reglas de inferencia del Teorema 3.2.5, que reduce q a una expresión de probabilidad estándar que implique cantidades observadas.*

Teorema 3.2.6. *Las Reglas 1-3 son completas.*

Que dichas reglas sean completas quiere decir que son suficientes para derivar todos los efectos causales identificables. La demostración del Teorema 3.2.6 se puede encontrar en la literatura (Shpitser y Pearl, 2006).

La tarea de decidir si existe una secuencia de reglas para reducir una expresión de efecto causal arbitraria puede ser un trabajo muy complejo. Es por ello que se utiliza

diferente software para el análisis causal. En la siguiente sección se presentan las tareas que deben proporcionar los modelos causales para ser útiles y se clasifican las librerías disponibles para construir estos modelos según estas tareas.

3.2.8. Las Siete Herramientas de Inferencia Causal

De acuerdo con Pearl, 2019, los modelos causales necesitan proporcionar siete tareas para ser útiles. A continuación se ofrece una visión general de estas tareas en el marco de los SCM, las herramientas utilizadas en cada una de ellas y la contribución única que cada herramienta aporta al arte del razonamiento automatizado.

Codificación de Supuestos Causales - Transparencia y Comprobabilidad. La transparencia permite a los analistas discernir si los supuestos codificados son plausibles o si se justifican supuestos adicionales. La comprobabilidad permite determinar si las hipótesis codificadas son compatibles con los datos disponibles y, en caso contrario, identificar las que necesitan ser reparadas. La comprobabilidad se facilita a través del criterio gráfico d-separación, que proporciona la conexión fundamental entre causas y probabilidades (Pearl, 1988).

El do-Calculus y el Control de la Confusión. Para los modelos en los que no se cumple el criterio de puerta trasera, se dispone del do-Calculus, que predice el efecto de las intervenciones siempre que sea posible (Shpitser y Pearl, 2008).

La Algoritmización de los Contrafactuales. Esta tarea formaliza el razonamiento contrafactual dentro de la representación gráfica. Cada modelo de ecuaciones estructurales determina el valor de verdad de cada frase contrafactual.

Análisis de Mediación y Evaluación de los Efectos Directos e Indirectos. Esta tarea se refiere a los mecanismos que transmiten los cambios de una causa a sus efectos, lo cual es esencial para generar explicaciones, se debe recurrir al análisis contrafactual para facilitar la identificación de los mismos.

Adaptabilidad, Validez Externa y Sesgo de Selección de la Muestra. La robustez es reconocida por los investigadores de la IA como una falta de adaptabilidad que sale a la luz cuando las condiciones del entorno cambian. El do-Calculus ofrece una metodología completa para superar los sesgos debidos a los cambios del entorno. Puede utilizarse para reajustar las políticas aprendidas, para sortear los cambios del entorno y para controlar las disparidades entre las muestras no representativas y una población objetivo (Bareinboim y Pearl, 2016).

Recuperación de Datos Perdidos. El uso de modelos causales puede formalizar las condiciones bajo las cuales se pueden recuperar las relaciones causales y probabilísticas de los datos incompletos y, siempre que se cumplan las condiciones, producir una estimación consistente de la relación deseada.

Descubrimiento Causal. El criterio de la d-separación detecta y enumera las implicaciones comprobables de un modelo causal dado. Esto abre la posibilidad de inferir, con suposiciones leves, el conjunto de modelos que son compatibles con los datos, y

de representar este conjunto de forma compacta. En determinadas circunstancias, el conjunto de modelos compatibles puede podarse significativamente hasta el punto de que las preguntas causales pueden estimarse directamente a partir de ese conjunto (Jaber, Zhang y Bareinboim, 2018).

Además de estas siete tareas, se han tenido en cuenta ciertos aspectos respecto a la implementación para la selección de la librería:

Licencia. La licencia que la librería y el resto del entorno (IDE) tienen. Sólo se han encontrado modelos de código abierto.

Lenguaje de Programación. El lenguaje de programación utilizado para construir los modelos. Todas las librerías consideradas están implementadas en R o Python, alguna de ellas acepta ambos.

Documentación y Canales de Soporte. En todos los casos hay una buena documentación escrita (tutoriales, descripciones de la API, ejemplos y guías de preguntas y respuestas)

Herramientas para escribir Diagramas Causales Disponibilidad de herramientas de apoyo para escribir desde cero, cargar y modificar y convertir desde diferentes formatos de archivo el diagrama causal.

En la Tabla 3.2 se presentan las diferentes librerías consideradas, así como sus principales características basadas en las siete tareas y los requerimientos de implementación detallados anteriormente. Como se puede observar, DoWhy (Sharma, Kiciman y col., 2019 y Sharma y Kiciman, 2020) es la única librería que cubre todas las tareas. Además, tiene una licencia muy permisiva, tiene varios enlaces a diferentes lenguajes de programación y tiene una de las comunidades más activas. Por todas estas razones, DoWhy, implementada a través del lenguaje de programación Python, será la librería utilizada para implementar el Modelo Causal que se desarrollará en el proyecto. En la Sección 4.2 se ofrece una descripción detallada de cómo funciona DoWhy y se presenta algún ejemplo.

Tabla 3.2: Principales características de las diferentes librerías disponibles para implementar el Modelo Causal.

Aspectos	**DAGitty**	**DoWhy**	**Causal Graphical Models**	**Causality**	**Causal Inference**
Codificación de Supuestos Causales, Transparencia y Comprobabilidad	X	X	X	X	X
do-Calculus y Control de la Confusión	X	X	X	X	X
Algoritmización de Contrafactuales	X	X		X	X
Análisis de Mediación y Evaluación de los Efectos Directos e Indirectos	X	X	X	X	X
Adaptabilidad, Validez Externa y Sesgo de Selección de la Muestra	X	X			
Recuperación de Datos Perdidos		X			
Descubrimiento de la Causa	X	X	X	X	
Herramientas para Generar Diagramas Causales	X	X	X	X	X
Licencia	GNU	MIT	MIT	Open	BSD
Lenguaje de Programación	R	R/Python	Python	Python	Python
Documentación y Canales de Asistencia	X	X	X		X

Capítulo 4

Resultados

Para obtener una mejor comprensión de los conceptos relacionados con las medidas de efecto causal, se ha creado una librería basada en el Marco de Resultados Potenciales (POF) a partir de la teoría expuesta en la Sección 3.1. En la siguiente sección, se describe esta librería y se presenta una colección de ejemplos muy sencillos que permiten entender cómo está estructurada y los resultados que se pueden obtener a partir de ella.

Una vez afianzados estos conceptos, es posible comenzar a utilizar la librería DoWhy, que como se dijo en la Sección 3.2.8, es la librería seleccionada para realizar el análisis causal en el proyecto. Siguiendo el mismo esquema, se ofrece una descripción de este software y se presenta un ejemplo ilustrativo que muestra todos los resultados que se pueden obtener a partir de él. Por último, se realiza un análisis cualitativo de un Modelo Causal más complejo, viendo así la potencia que tiene DoWhy a la hora de trabajar con Diagramas Causales.

Todos los resultados presentados en este capítulo se encuentran en el directorio del proyecto, *why-scm*. Todos los módulos y scripts se encuentran en el directorio *src*, que será el directorio de trabajo, mientras que las tablas de datos están en el directorio *data*. En *doc* hay una serie de imágenes utilizadas para documentación del proyecto y en *docs* la documentación de las diferentes funciones. Hay que destacar que como el proyecto se encuentra activo, dicha documentación no está completa y ésta es más fácil de hallar dentro de cada una de las funciones. El directorio *dowhy* es el entorno virtual creado con todos los paquetes necesarios para trabajar con DoWhy. Por último, en *tests* se encuentran una serie de pruebas unitarias para comprobar la funcionalidad de la librería creada en el POF.

4.1. Análisis causal en el POF

La librería creada en el POF está basada en Programación Orientada a Objetos (POO), donde las funciones principales se encuentran en la clase *WhatIf* dentro del módulo *WhatIfClass.py* en el directorio *dataModel*.

La clase tiene como parámetro de entrada un diccionario de Python, este contiene los diferentes parámetros que caracterizan el conjunto de datos y que son necesarios para llevar a cabo el análisis:

'wd': Indica el directorio de trabajo (*str*)

'dataPath': Indica la ruta donde se encuentra el fichero *csv* que contiene los datos (*str*)

'csvFilename': Indica el nombre del fichero *csv* (*str*)

'popuLevel': Indica si los datos corresponden a una población o a una muestra (*int*)

'num_potOutcomes': Indica el número de resultados potenciales conocidos (*int*)

'num_treatment': Indica el número de tratamientos bajo estudio (*int*)

'num_conditional': Indica el número de variables condicionales que se contemplan (*int*)

Las diferentes funciones de la clase llevan a cabo un profundo análisis probabilístico del conjunto de datos y permiten obtener las medidas del efecto causal. Estas funciones tienen como parámetro de entrada el resultado de interés, el cual para todo el análisis que prosigue será $Y = 1$. Los resultados se exponen a través de uno o varios diccionarios de Python.

Para comprender cómo funciona esta librería, se han utilizado tres conjuntos de datos diferentes. Estos han sido obtenidos de Hernán y Robins, 2020 y se muestran en la Tabla 4.1. El análisis causal se lleva a cabo en el script *CE_measures_App.py*, alojado en el directorio *example_causalEffect*. En este script es posible elegir qué conjunto de datos se quieren analizar, seleccionando el diccionario que se le pasa a la función *main()*.

Resultados de la Tabla 4.1a

En la Tabla 4.1a se presentan 20 individuos que representan una población completa, $popuLevel = 1$, y para los cuales se conocen los dos resultados potenciales posibles, cuando un mismo individuo recibe el tratamiento ($a = 1$) y cuando no ($a = 0$), $num\_potOutcomes = 2$. Por lo tanto, como se conocen ambos resultados potenciales para cada individuo, no hay ninguna variable que indique si el individuo recibió o no el tratamiento, $num\_treatment = 0$, y tampoco hay ninguna variable que condicione el experimento, $num\_conditional = 0$. El análisis se lleva a cabo utilizando la función *setACE()*, a continuación se muestran los datos más interesantes que se ofrecen como salida:

```
{... ,
'potOutcome_0'    : ['Ya=0', 10, 20, 0.5],
'potOutcome_1'    : ['Ya=1', 10, 20, 0.5],
'causalRiskDiffe': 0.0,
'causalRiskRatio': 1.0,
'ace'             : False,
'nullHypothesis' : True}
```

Por un lado, las claves *'potOutcome_0'* y *'potOutcome_1'* contienen una lista con el número de medidas de interés ($Y = 1$) para cada resultado potencial, el número de medidas totales y la probabilidad $P(Y^a = 1)$. Por otro lado, *'causalRiskDiffe'* y *'causalRiskRatio'* muestran las medidas del efecto causal, obtenidas directamente a partir de las Ecuaciones 3.7 y 3.8, y por último, *'ace'* y *'nullHypothesis'* indican si existe o no un efecto causal del tratamiento en el resultado en la población.

(a)

	$Y^{a=0}$	$Y^{a=1}$
Rheia	0	1
Kronos	1	0
Demeter	0	0
Hades	0	0
Hestia	0	0
Poseidon	1	0
Hera	0	0
Zeus	0	1
Artemis	1	1
Apollo	1	0
Leto	0	1
Ares	1	1
Athena	1	1
Hephaestus	0	1
Aphrodite	0	1
Cyclope	0	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

(b)

	A	Y
Rheia	0	0
Kronos	0	1
Demeter	0	0
Hades	0	0
Hestia	1	0
Poseidon	1	0
Hera	1	0
Zeus	1	1
Artemis	0	1
Apollo	0	1
Leto	0	0
Ares	1	1
Athena	1	1
Hephaestus	1	1
Aphrodite	1	1
Cyclope	1	1
Persephone	1	1
Hermes	1	0
Hebe	1	0
Dionysus	1	0

(c)

	L	A	Y
Rheia	0	0	0
Kronos	0	0	1
Demeter	0	0	0
Hades	0	0	0
Hestia	0	1	0
Poseidon	0	1	0
Hera	0	1	0
Zeus	0	1	1
Artemis	1	0	1
Apollo	1	0	1
Leto	1	0	0
Ares	1	1	1
Athena	1	1	1
Hephaestus	1	1	1
Aphrodite	1	1	1
Cyclope	1	1	1
Persephone	1	1	1
Hermes	1	1	0
Hebe	1	1	0
Dionysus	1	1	0

Tabla 4.1: Conjuntos de datos básicos utilizados para un primer análisis causal en el Marco de Resultados Potenciales.

Como se puede observar, el efecto causal medio en la población es nulo, siendo la hipótesis nula cierta. Esto se debe a que, como se muestra en los resultados, $P(Y^{a=1} = 1) = P(Y^{a=0} = 1) = 0.5$, por lo que no se cumple la Definición 3.1.2.

Resultados de la Tabla 4.1b

En la Tabla 4.1b se presentan de nuevo 20 individuos, que en este caso representan una muestra de una población, $popuLevel = 0$. Únicamente se conoce un resultado potencial, el correspondiente al nivel de tratamiento que ha recibido, $num\_potOutcomes = 1$ y $num\_treatment = 1$. Y de nuevo, no hay ninguna variable que condicione el experimento, $num\_conditional = 0$. En este caso, para el análisis se utiliza la función *setAssociaRiskAndIndependence()*, a continuación se muestran los datos más interesantes que se ofrecen como salida:

```
{... ,
'treatment_v'     : [{'A=1': 13, '[Y=1|A=1]': 7,
                      'Pr[Y=1|A=1]': 0.54}],
'untreatment_v'   : [{'A=0': 7, '[Y=1|A=0]': 3,
                      'Pr[Y=1|A=0]': 0.43}],
'associaMeasures': {'associaRiskDiffe': 0.1099,
                    'associaRiskRatio': 1.2564,
                    'independence'    : False}}
```

Por un lado, las claves *'treatment_v'* y *'untreatment_v'* ofrecen las medidas probabilísticas del conjunto de datos, mostrando el número de individuos que recibieron y no el tratamiento, el número de resultados de interés condicionado a ese nivel de tratamiento y la probabilidad de obtener dicho resultado de interés condicionado a recibir o no el tratamiento. Por otro lado, se ofrecen las mismas medidas del efecto causal que en el caso anterior, dentro de la clave *'associaMeasures'*.

Como se puede observar, las medidas del efecto causal indican que la Definición 3.1.2 se cumple y que por lo tanto $Y \not\perp\!\!\!\perp A$, es decir, el resultado Y y el tratamiento A son dependientes, y por lo tanto existe un efecto causal del tratamiento en el resultado.

Cabe destacar que para que este análisis sea correcto, estos datos han de provenir de un experimento marginalmente aleatorio, ver Sección 3.1.5, y por lo tanto presentar la condición de intercambiabilidad $Y^a \perp\!\!\!\perp A$ (nótese la diferencia con $Y \perp\!\!\!\perp A$), es decir, si el experimento se hubiera realizado al revés y hubieran recibido el tratamiento los que no lo han recibido y viceversa, el resultado tendría que haber sido el mismo. Si esto no se cumpliera y los datos no provinieran de un experimento aleatorio, no sería posible calcular las medidas del efecto causal a partir del riesgo bajo un tratamiento en una muestra de la población ya que no se cumplirían las Ecuaciones 3.12 y 3.13.

Para una mejor comprensión de qué significa la condición de intercambiabilidad se presenta un ejemplo numérico. Suponiendo que se dispone de los datos contrafactuales de la Tabla 4.1a y estos se corresponden al mismo estudio que la Tabla 4.1b, se puede calcular el riesgo de $Y = 1$ sin tratamiento en los individuos que si han sido tratados (hecho contrafactual obtenido de la Tabla 4.1a ya que en realidad no se ha observado), y el riesgo de $Y = 1$ sin tratamiento en los individuos que no han sido tratados(hecho observado y medido en la Tabla 4.1b):

$$Pr[Y^{a=0} = 1|A = 1] = 7/13$$
$$Pr[Y^{a=0} = 1|A = 0] = 3/7$$

La intercambiabilidad se daría si ambos riesgos fuesen iguales, es decir, si el resultado hubiera sido el mismo en caso de que el experimento se hubiese llevado a cabo al revés, tratando a los que no fueron tratados y viceversa. Sin embargo, $7/13 > 3/7$, por lo tanto los tratados y los no tratados no son intercambiables y la intercambiabilidad $Y^a \perp\!\!\!\perp A$ no se sostiene para $a = 0$ y por lo tanto no se sostiene en general. En este caso, se podría afirmar la asociación entre el tratamiento A y el resultado Y pero esta asociación no sería causalidad.

Resultados de la Tabla 4.1c

La Tabla 4.1c es la misma que la Tabla 4.1b pero con una variable condicional L a partir de la cual se puede estratificar el conjunto de datos, $num\_conditional = 1$. En

este caso se utilizan dos funciones para llevar a cabo el análisis, *setConditionalRandomizationIndependence()* y *setInverseProbabilityWeighting()*, la primera de ellas calcula el efecto causal por Estandarización, ver Sección 3.1.6, y la segunda, por Ponderación de la Probabilidad Inversa, ver Sección 3.1.6, la salida, por lo tanto, son dos diccionarios diferentes:

```
{... ,
'treatment_v'     : [{'A=1 & L=1': 9, 'A=1 & L=0': 4,
                      '[Y=1|A=1,L=1]': 6, '[Y=1|A=1,L=0]': 1,
                      'Pr[Y=1|A=1,L=1]': 0.66,
                      'Pr[Y=1|A=1,L=0]': 0.25}],
'untreatment_v'   : [{'A=0 & L=1': 3, 'A=0 & L=0': 4,
                      '[Y=1|A=0,L=1]': 2, '[Y=1|A=0,L=0]': 1,
                      'Pr[Y=1|A=0,L=1]': 0.66,
                      'Pr[Y=1|A=0,L=0]': 0.25}],
'associaMeasures': {'strataRat_L1': 1.0, 'strataRat_L0': 1.0,
                    'strataDiff_L1': 0.0, 'strataDiff_L0': 0.0,
                    'A modified by L': False,
                    'causalRiskDiff': 0.0,
                    'causalRiskRatio': 1.0,
                    'independence': True}}


{... ,
'probabilities'   : [{'Pr[Y=1,A=0,L=0]': 0.05,
                      'Pr[Y=1,A=0,L=1]': 0.1,
                      'Pr[Y=1,A=1,L=0]': 0.05,
                      'Pr[Y=1,A=1,L=1]': 0.3}],
'weights'         : [{'w(A=0|L=0)': 2.0, 'w(A=0|L=1)': 4.0,
                      'w(A=1|L=0)': 2.0, 'w(A=1|L=1)': 1.33}],
'associaMeasures': {'causalRiskDiff': 0.0,
                    'causalRiskRatio': 1.0,
                    'independence': True}}
```

En el primero de ellos, de nuevo, las claves *'treatment_v'* y *'untreatment_v'* muestran las medidas probabilísticas del resultado de interés de los individuos que recibieron el tratamiento y los que no, condicionados a los diferentes valores de la variable L. De la misma manera, dentro de la clave *'associaMeasures'* se presentan todas las medidas de asociación. Las primeras, *'strataRat_L1'* y *'strataRat_L0'*, y *'strataDiff_L1'* y *'strataDiff_L0'*, muestran el ratio y la diferencia de riesgo causal para cada estrato del conjunto de datos en función del valor de L, $L = 1$ y $L = 0$, respectivamente. Este se calcula sustituyendo en las Ecuaciones 3.7 y 3.8 la Ecuación 3.15. Como en este caso los ratios causales calculados por estratificación son iguales, se dice que el efecto del tratamiento no es modificado por la variable L. Por último, en *'causalRiskDiff'* y *'causalRiskRatio'* se muestran la diferencia y el ratio de riesgo causal del conjunto de datos completo, obtenidos por estandarización, sustituyendo en las Ecuaciones 3.7 y 3.8 la Ecuación 3.17. En este caso, como la Definición 3.1.2 no se cumple, el tratamiento A y el resultado Y son independientes dado L, $Y \perp\!\!\!\perp A|L$.

En el segundo de ellos, las medidas de asociación se han obtenido a partir del método de la Ponderación de la Probabilidad Inversa, sustituyendo en las Ecuaciones 3.7 y 3.8 la Ecuación 3.20. Es por ello que en la clave *'probabilities'* se muestran las probabilidades conjuntas y en la clave *'weights'* las ponderaciones de los diferentes grupos, es decir,

las inversas de la puntuación de propensión. Obviamente el resultado obtenido para las medidas de asociación es el mismo que utilizando el método anterior.

Una vez más, cabe destacar que para que este análisis sea correcto, los datos de la Tabla 4.1c han de provenir de un experimento condicionalmente aleatorio, ver Sección 3.1.6, y cumplir la condición de intercambiabilidad condicional $Y^a \perp\!\!\!\perp A|L$. De esta manera, tanto los individuos con $L = 0$ como los individuos con $L = 1$ son realmente los representantes medios de toda la población.

Como se puede observar, el resultado obtenido para las medidas del efecto causal a partir de las Tablas 4.1b y 4.1c es diferente a pesar de que los valores del tratamiento A y el resultado Y son idénticos para todos los individuos en ambas tablas. Esta diferencia se deriva del hecho de contemplar la variable L en la Tabla 4.1c, ya que se realiza una media ponderada en la que cada grupo formado a partir del valor de L recibe un peso proporcional a su tamaño. Este ejemplo muestra la importancia de tener en cuenta todas las variables relevantes a la hora de realizar cualquier experimento y análisis causal.

Una vez comprendido el análisis causal en el Marco de Resultados Potenciales es posible dar un salto a los Diagramas Causales. En la siguiente sección se ve con más detalle como opera el software DoWhy y se presenta un ejemplo comparativo entre los resultados que ofrece la librería que se acaba de estudiar y los que ofrece DoWhy.

4.2. Análisis Causal utilizando DoWhy

Como se ha mencionado anteriormente, dar respuesta a cuestiones del tipo ¿funcionará?, ¿qué deberíamos hacer?, implica la estimación de un contrafactual. De la misma manera, para llevar a cabo intervenciones, es necesario estimar el efecto de cambiar una entrada respecto a su valor actual para el que no existen datos. Este tipo de análisis requiere de un razonamiento causal.

Aunque existen muchos métodos para la inferencia causal, es difícil comparar sus supuestos y la solidez de los resultados, para ello, DoWhy aporta tres diferencias clave en comparación con el resto de software disponible (Sharma, Kiciman y col., 2019):

Supuestos explícitos de identificación. Los supuestos son una pieza clave en DoWhy. Cada análisis comienza con la construcción de un modelo causal. Estas suposiciones pueden verse gráficamente o en términos de independencia condicionada. Siempre que sea posible, DoWhy también puede comprobar automáticamente los supuestos establecidos utilizando los datos observados.

Separación entre identificación y estimación. La identificación es el problema causal, mientras que la estimación es simplemente un problema estadístico. DoWhy respeta este límite y los trata por separado. Esto centra el esfuerzo de inferencia causal en la identificación, llevando a cabo la estimación utilizando cualquier estimador estadístico disponible para el estimando objetivo. Además, se pueden utilizar múltiples métodos de estimación para un único estimando identificado y viceversa.

Controles de solidez automatizados. La parte más crítica, y a menudo omitida, del análisis causal es la comprobación de la solidez de una estimación frente a supuestos no verificados. DoWhy facilita la ejecución automática de comprobaciones de sensibilidad y solidez de la estimación obtenida.

Como se puede deducir a partir del segundo punto clave, DoWhy se basa en dos de los marcos más potentes para la inferencia causal, los diagramas causales y los resultados potenciales. Utiliza los criterios basados en grafos y el do-Calculus para modelar los supuestos e identificar un efecto causal. Para la estimación, cambia a métodos basados principalmente en resultados potenciales. A pesar de que este procedimiento pueda parecer complicado, DoWhy lleva a cabo el análisis siguiendo cuatro pasos clave:

I. Modelar un problema causal. DoWhy crea un diagrama causal subyacente para cada problema, que sirve para hacer explícita cada suposición causal. Este grafo no tiene por qué ser completo, se puede proporcionar un grafo parcial, que represente el conocimiento previo sobre algunas de las variables, y el resto de las variables se consideran como posibles factores de confusión. DoWhy admite dos formatos de entrada de gráficos, *gml* (recomendado) y *dot*. Aunque no se recomienda, también es posible especificar las diferentes variables presentes en el modelo en lugar de proporcionar un gráfico.

II. Identificar un estimando objetivo bajo el modelo. Basándose en el diagrama causal, DoWhy encuentra todas las formas posibles de identificar un efecto causal deseado. Utiliza diferentes criterios y el do-Calculus para encontrar posibles expresiones que puedan identificar el efecto causal. Actualmente DoWhy soporta el Criterio de Puerta Trasera, el Criterio de Puerta Delantera, Variables Instrumentales y Mediación.

III. Estimar el efecto causal a partir del estimando identificado. DoWhy admite métodos basados tanto en el Criterio de Puerta Trasera como en las Variables Instrumentales. También proporciona intervalos de confianza no paramétricos y una prueba de permutación para comprobar la significación estadística de la estimación obtenida. También es posible llamar a métodos de estimación externos utilizando DoWhy, como por ejemplo los paquetes EconML y CausalML.

IV. Refutar la estimación obtenida. DoWhy tiene acceso a múltiples métodos de refutación para validar una estimación del efecto de un estimador causal, como por ejemplo añadir una causa común aleatoria, tratamiento con placebo...

Para la instalación, DoWhy requiere Python 3.6+ y una serie de paquetes adicionales. Por propia experiencia, se recomienda la instalación del software en el sistema operativo Ubuntu. En el Apéndice A se ofrece una guía detallada para llevar a cabo la instalación de DoWhy en Ubuntu utilizando el administrador de paquetes pip.

A continuación, se vuelve a recurrir al Ejemplo 3.1.1. Se lleva a cabo un análisis utilizando ambas librerías, la que se ha creado en el POF y DoWhy, con el fin de presentar un análisis comparativo entre ambas y, a su vez, resolver por completo la paradoja.

4.2.1. La Paradoja de Simpson

Se ha generado un conjunto de datos que se corresponde con los resultados mostrados en la Tabla 3.1, *gender_simpson_paradox.csv*. Se tiene una variable condicional, el género, que se denota por L, el tratamiento, denotado por A y el resultado potencial debido al nivel de tratamiento recibido, Y, donde el resultado de interés son los recuperados ($Y = 1$).

El fichero que resuelve este problema es *dowhy_simpson_paradox.py*, que se encuentra en el directorio *example_CausalEffect*, en el directorio de trabajo. Este script resuelve el ejemplo de la Paradoja de Simpson tanto en el Marco de Resultados Potenciales como en los Modelos Causales utilizando la librería DoWhy.

En primer lugar, se lleva a cabo el análisis en el POF. Para ello, se sigue la misma metodología indicada en la Sección 4.1, pasándole a la función *main()* los parámetros necesarios. Debido a que el género es una variable condicional, se utiliza la función *setConditionalRandomizationIndependence()* para obtener las medidas del efecto causal, estos son los resultados obtenidos:

```
{'treatment_v'    : [{'A=1 & L=1': 87, 'A=1 & L=0': 263,
                      '[Y=1|A=1,L=1]': 81, '[Y=1|A=1,L=0]': 192,
                      'Pr[Y=1|A=1,L=1]': 0.93,
                      'Pr[Y=1|A=1,L=0]': 0.73}],
'untreatment_v'   : [{'A=0 & L=1': 270, 'A=0 & L=0': 80,
                      '[Y=1|A=0,L=1]': 234, '[Y=1|A=0,L=0]': 55,
                      'Pr[Y=1|A=0,L=1]': 0.87,
                      'Pr[Y=1|A=0,L=0]': 0.69}],
'associaMeasures': {'strataACE_L1': 1.07, 'strataACE_L0': 1.06,
                    'strataDiff_L1': 0.0644, 'strataDiff_L0': 0.0425,
                    'A modified by L': True,
                    'causalRiskDiff': 0.053,
                    'causalRiskRatio': 1.07,
                    'independence': False}}
```

Como se puede observar, el tratamiento A es modificado por la variable L. Esto se debe a que las diferencias y ratios causales son diferentes para cada conjunto de datos estratificado por L. Esta diferencia de riesgo causal es casi un 51 % mayor para el conjunto estratificado por $L = 1$ que para el conjunto estratificado por $L = 0$, es decir, el efecto causal del tratamiento en el resultado de interés es mayor para los hombres, $L = 1$, que para las mujeres, $L = 0$.

Por otro lado, el resultado Y y el tratamiento A no son independientes dado L, $Y \not\perp\!\!\!\perp A|L$, es decir, existe un efecto causal del tratamiento en el resultado ya que se cumple la Definición 3.1.2. Este efecto causal es positivo, $'causalRiskDiff' = 0.053$, es decir, el efecto que tiene el tratamiento $A = 1$ en el resultado de interés $Y = 1$ es positivo, lo cual da una clara ventaja a la toma del tratamiento resolviendo por completo la paradoja.

Una vez realizado el análisis en el POF, se procede a llevar a cabo el análisis utilizando la librería DoWhy. Lo primero es generar el Diagrama Causal en formato *gml* haciendo uso de la librería *networkx* y modificar los valores del tratamiento A en el dataframe de *pandas* cambiando el valor 1 por $True$ y el valor 0 por $False$, evitando así problemas que surgen a la hora de utilizar los diferentes estimadores de puntuación de propensión. A continuación se muestran tanto las entradas como las salidas para cada uno de los pasos descritos al principio de esta sección.

I. Crear un modelo causal a partir de los datos y del grafo dado.

En la Figura 4.1 se indica cómo generar el Modelo Causal utilizando la librería DoWhy. En este caso, se ha creado introduciendo el grafo causal en formato *gml*, ya que es la opción recomendada. Es necesario indicar qué variables son el tratamiento y la salida, en

este caso se ha hecho aprovechando el diccionario que da como salida la librería creada en el POF, $CondJSON['treatment'_k] =' A'$ y $CondJSON['outcome'] =' Y'$.

```
# I. Create a causal model from the data and given graph.
model=CausalModel(
    data      = data,
    treatment = CondJSON["treatment_k"],
    outcome   = CondJSON["outcome"],
    graph     = "simpson_causal_DAG.gml"
    )
# Causal Model DAG plot
model.view_model()
display(Image(filename="causal_model.png"))
```

Figura 4.1: Entrada para la creación de un Modelo Causal a partir de los datos y un grafo en formato *gml* utilizando DoWhy.

Cabe destacar que también sería posible generar el mismo modelo indicando las variables de cada tipo. En este caso, L es una causa común de A y L por lo que habría que añadir al modelo el parámetro $common\_causes = ['L']$. También es posible añadir variables instrumentales($instruments$), que solo afectan al tratamiento y modificadores de efecto ($effect\_modifiers$), que solo afectan al resultado. Sin embargo, este método es limitado en cuanto a la variedad de tipos de variables, introduciendo el grafo el Modelo Causal puede ser mucho más complejo.

En la Figura 4.2 se muestra el diagrama causal y por tanto los supuestos codificados en el Modelo Causal. A continuación, se utiliza este grafo para identificar el efecto causal, es decir, pasar de un estimando a una expresión de probabilidad, y por último, estimar el efecto causal.

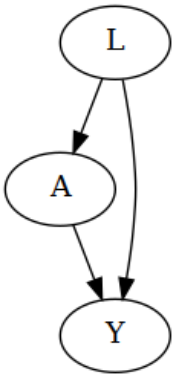

Figura 4.2: DAG generado por DoWhy para el Ejemplo 3.1.1 donde L es el género, A el tratamiento e Y el resultado.

II. Identificar el efecto causal y devolver los estimandos objetivo

Como se dijo al principio de la sección, la identificación y la estimación son dos procesos ortogonales. La identificación puede realizarse sin acceder a los datos, únicamente a través del grafo. El resultado de la identificación es una expresión que se puede calcular. Esta expresión puede evaluarse utilizando los datos disponibles en la etapa de estimación.

En la Figura 4.3 se muestra la entrada para identificar el efecto causal. El parámetro $proceed\_when\_unidentifiable = True$ transmite la suposición de que se ignora cualquier factor de confusión no observado, en caso de no indicarlo, DoWhy pedirá al usuario que compruebe que los factores de confusión no observados pueden ser ignorados para continuar con el análisis.

```
# II. Identify causal effect and return target estimands
identified_estimand = model.identify_effect(proceed_when_unidentifiable=True)
print(identified_estimand)
```

Figura 4.3: Entrada para identificar el efecto causal y obtener los estimandos objetivo utilizando DoWhy.

En la Figura 4.4 se muestra la identificación del efecto causal. Esta ha sido realizada utilizando el Criterio de Puerta Trasera, ver Sección 3.2.5, y la expresión obtenida para estimar el efecto causal de A sobre Y nos indica que es necesario condicionar sobre la variable L. Esto se debe a que el conjunto de variables que cumple la Definición 3.2.6 es $Z = \{L\}$. Por último, DoWhy indica que dicha identificación se sostiene bajo el supuesto de que no existen variables de confusión U, y si existen, se cumple $P(Y|A, L, U) = P(Y|A, L)$.

El fenómeno de confusión aparece cuando la relación observada entre el tratamiento y el resultado puede ser total o parcialmente explicada por otra variable, o por el contrario, cuando una relación real queda enmascarada por esta. A estas variables se las conoce como variables o factores de confusión y son causa común del tratamiento y del resultado. En este ejemplo, si se asumiera la estructura causal de la Figura 4.7, L sería un factor de confusión, ya que como se verá a continuación, la relación real estaría enmascarada por esta.

```
Estimand type: nonparametric-ate

### Estimand : 1
Estimand name: backdoor
Estimand expression:
 d
────(Expectation(Y|L))
d[A]
Estimand assumption 1, Unconfoundedness: If U→{A} and U→Y then P(Y|A,L,U) = P(Y|A,L)

### Estimand : 2
Estimand name: iv
No such variable found!

### Estimand : 3
Estimand name: frontdoor
No such variable found!
```

Figura 4.4: Estimandos objetivo obtenidos utilizando DoWhy.

III. Estimar el estimando objetivo utilizando un método estadístico.

A continuación, se estima el efecto causal basado en la estimando obtenido a partir del Criterio de la Puerta Trasera. En la Figura 4.5 se muestra la entrada necesaria para estimar el efecto causal utilizando puntuaciones de propensión inversas o IPS, ver Definición 3.2.5. DoWhy admite diferentes esquemas de ponderación de pronpensión,

como la ponderación IPS autonormalizada, o estimador Hajek (*'ips_normalized_weight'*) (Swaminathan y Joachims, 2015) y la ponderación IPS estabilizada (*'ips_stabilized_weight'*) (Xu y col., 2009).

Hay que destacar que el parámetro $target\_units =' ate'$ por defecto, esto significa que se está estimando el efecto medio del tratamiento, es decir, se mide el efecto medio de imponer el tratamiento frente a no imponerlo sobre toda la población. Sin embargo, este parámetro puede fijarse como $'att'$, para obtener el efecto medio del tratamiento sobre los tratados, es decir, medir efecto medio de negar el tratamiento a los que lo recibieron. O por el contrario, como $'atc'$, para obtener el efecto medio sobre los no tratados, es decir, medir el efecto medio de ampliar el tratamiento a los que no lo recibieron (Greifer y Stuart, 2021).

```
# III Estimate the target estimand using a statistical method.
estimate = model.estimate_effect(identified_estimand,
                                 method_name="backdoor.propensity_score_weighting",
                                 method_params={"weighting_scheme":"ips_weight"})
print(estimate)
```

Figura 4.5: Entrada para estimar el estimando objetivo utilizando el método de la Ponderación de la Probabilidad Inversa.

En la Figura 4.6 se muestra la estimación realizada para el estimando objetivo. Como se puede observar, el efecto causal medio es de 0.05029, siendo este ligeramente diferente al obtenido utilizando la librería creada en el POF (0.053). Cabe destacar que esta pequeña diferencia proviene del redondeo que se haga al obtener $P(Y = 1|A = a, L = l)$, ya que en este caso la Ecuación 3.29 es exactamente igual que la Ecuación 3.20. Se puede comprobar que la diferencia proviene del redondeo operando con dicha ecuación.

```
*** Causal Estimate ***

## Identified estimand
Estimand type: nonparametric-ate

### Estimand : 1
Estimand name: backdoor
Estimand expression:
 d
────(Expectation(Y|L))
d[A]
Estimand assumption 1, Unconfoundedness: If U→{A} and U→Y then P(Y|A,L,U) = P(Y|A,L)

## Realized estimand
b: Y~A+L
Target units: ate

## Estimate
Mean value: 0.050285349449815664
```

Figura 4.6: Estimación del estimando objetivo utilizando DoWhy.

Además, se ha utilizado DoWhy para obtener el efecto causal en el conjunto de datos estratificado por L. El proceso es el mismo cambiando únicamente el conjunto de datos que se introduce a la hora de crear el Modelo Causal, los resultados son los siguientes:

```
{'Causal estimate in full dataset': 0.05029,
 'Causal estimate in stratified dataset by L=0': 0.04254,
 'Causal estimate in stratified dataset by L=1': 0.06437}
```

Hay que notar que los resultados coinciden con los obtenidos en el POF. También se ha obtenido el efecto causal del tratamiento ignorando la variable L, como se representa en la Figura 4.7, a partir de las Ecuaciones 3.7, 3.12 y 3.13. El resultado es el siguiente:

```
Causal Effect only taking into account treatment 'A',
ignoring 'L': -0.04571
```

Por lo que si se ignora el género, el efecto que tiene el tratamiento $A = 1$ en el resultado de interés $Y = 1$ es negativo, lo cual da una clara ventaja a no tomar el tratamiento.

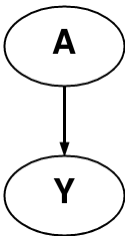

Figura 4.7: Diagrama Causal donde únicamente el tratamiento A tiene un efecto sobre el resultado Y.

Una vez más, se comprueba que asumir una estructura causal resuelve por completo la paradoja. Si se asumiera la estructura de la Figura 4.7, el efecto del tratamiento sobre el resultado sería claramente negativo, por el contrario, asumiendo la estructura de la Figura 4.2, el efecto del tratamiento es claramente positivo. En ninguno de los dos casos existiría la paradoja, ya que se está asumiendo dicha estructura, sin embargo, aquí se refleja la importancia de conocer con precisión la estructura causal y no dejar fuera del modelo ninguna variable relevante para el análisis, ya que por el contrario se podría llegar a conclusiones erróneas.

IV. Refutar la estimación obtenida mediante varias comprobaciones de robustez

Una vez obtenida la estimación del efecto causal, se procede a refutarla. En la Figura 4.8 se presenta la entrada necesaria para refutar dicha estimación utilizando dos métodos diferentes:

```
# IV.I Refute the obtained estimate using multiple robustness checks
# Method: Radom Common Cause

refute_results1 = model.refute_estimate(identified_estimand, estimate,
                                        method_name="random_common_cause")
print(refute_results1)

# IV.II Refute the obtained estimate using multiple robustness checks
# Method: Placebo Treatment Refuter

refute_results2 = model.refute_estimate(identified_estimand, estimate,
                                        method_name="placebo_treatment_refuter")
print(refute_results2)
```

Figura 4.8: Entrada para refutar la estimación obtenida añadiendo una Causa Común Aleatoria y utilizando un Tratamiento Placebo.

- **Añadir una causa común aleatoria:** Se añade una variable aleatoria independiente como causa común al conjunto de datos y se comprueba si utilizando el mismo método cambia la estimación obtenida. En la Figura 4.9 se muestra el diagrama causal modificado para llevar a cabo la refutación a partir de este método.
- **Tratamiento con placebo:** Se sustituye la verdadera variable de tratamiento por una variable aleatoria independiente y se vuelve a estimar el efecto causal. En la Figura 4.10 se muestra el diagrama causal modificado para llevar a cabo la refutación a partir de este método.

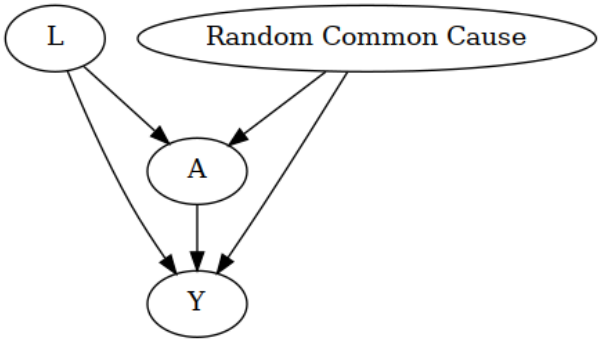

Figura 4.9: Diagrama Causal para refutar la estimación obtenida añadiendo una causa común aleatoria.

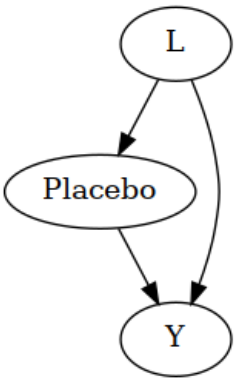

Figura 4.10: Diagrama Causal para refutar la estimación obtenida sustituyendo el tratamiento A por un placebo.

En la Figura 4.11 se muestra el resultado utilizando el primer método. El hecho de añadir una variable aleatoria como causa común no debería de afectar al efecto causal que tiene el tratamiento sobre el resultado, y efectivamente, el nuevo efecto causal obtenido al añadir dicha variable prácticamente es el mismo.

En la Figura 4.12 se muestra el resultado al utilizar el segundo método. En este caso, si se cambia el tratamiento por uno placebo, el efecto causal que cabría esperar de este nuevo tratamiento sería cero. Como se puede ver, el nuevo efecto causal es prácticamente nulo con un p-valor suficientemente alto para no rechazar este resultado.

Por lo tanto, se puede apreciar que el método de estimación utilizando la Ponderación de la Probabilidad Inversa es suficientemente robusto y que la medida del efecto causal obtenida es buena. Con este análisis quedaría completo el Modelo Causal, sin embargo, DoWhy permite ir más allá y obtener el resultado que se obtendría al realizar una intervención.

```
Refute: Add a Random Common Cause
Estimated effect:0.050285349449815664
New effect:0.050347679042794136
```

Figura 4.11: Resultado de la refutación de la estimación al añadir una Causa Común Aleatoria.

```
Refute: Use a Placebo Treatment
Estimated effect:0.050285349449815664
New effect:-0.0011808043203290063
p value:0.47
```

Figura 4.12: Resultado de la refutación de la estimación al utilizar un Tratamiento Placebo.

Intervenciones

DoWhy permite representar las distribuciones de probabilidad intervenidas mediante el muestreo de las mismas con un objeto llamado $do - sampler$. Con estas muestras, es posible calcular las estadísticas de muestras finitas de los datos intervenidos.

El $do - sampler$ toma muestras de la distribución de resultados, por lo que los resultados variarán significativamente de una muestra a otra. Para utilizarlos para calcular las medidas de efecto causal, se recomienda generar varias muestras de este tipo para tener una idea de la varianza posterior de la estadística de interés.

Siguiendo la noción de una intervención en un Modelo Causal según Pearl, Glymour y Jewell, 2016 y Pearl, 2000, el procedimiento a través del cual el $do - sampler$ toma dichas muestras se implementa en una secuencia de tres pasos:

1. **Interrumpir las causas:** Se eliminan todas las aristas de entrada a todas las variables sobre las que se interviene.

2. **Hacer efectivas:** Se fija el valor de esas variables a sus cantidades intervenidas.

3. **Propagar y muestrear:** Se propaga ese valor a través del modelo para calcular los resultados de la intervención con un procedimiento de muestreo.

En la práctica, hay muchas formas de implementar estos pasos. En el caso del $do - sampler$ MCMC, se ajusta una red bayesiana a los datos, y luego se construye una nueva red que representa la red de intervención. Las ecuaciones estructurales se establecen con los parámetros ajustados en la red inicial, y se muestrea a partir de esa nueva red para obtener las muestras de intervención utilizando los métodos de cadenas de Markov Monte Carlo. En el caso del $do - sampler$ de ponderación, las puntuaciones de propensión contienen la información que se utiliza para bloquear los caminos traseros,

y por lo tanto tienen el mismo efecto estadístico que eliminar las aristas en el diagrama causal. Se hace efectivo el tratamiento seleccionando el subconjunto del conjunto de datos inicial con el valor de intervención correcto. Por último, utilizando la ponderación de propensión inversa, se genera una muestra aleatoria ponderada.

En la Figura 4.13 se muestra la entrada para obtener una muestra aleatoria utilizando el $do - sampler$ de ponderación para diferentes intervenciones de interés. El parámetro $keep\_original\_treatment$ indica si se mantiene o no el tratamiento original. En el caso de fijarlo como *False*, se debe especificar la intervención para la cual se desea obtener la muestra al utilizar el $do - sampler$. En el caso contrario, si se fija como *True*, se mantiene el nivel de tratamiento como estaba eliminando así el sesgo de confusión al estimar el efecto causal. En este caso, se saltará la segunda etapa del proceso descrito anteriormente.

```
# Different interventios we are interested in
intervention = [{'A':False},{'A':True},{'A':True,'L':0},{'A':True,'L':1}]

for interv in intervention:
    sampler = WeightingSampler(df,
                               causal_model = model,
                               keep_original_treatment = False,
                               variable_types = {'L': 'd', 'A': 'd', 'Y': 'd'})
    interventional_df = sampler.do_sample(interv)
```

Figura 4.13: Entrada para obtener una muestra de datos para diferentes intervenciones de interés utilizando DoWhy.

La salida consiste en un DataFrame que contiene la muestra de intervención, $interventional\_df$. A partir de dicha muestra, se han obtenido las diferentes probabilidades de interés para cada intervención:

- **do(A=1):** Toda la población recibe el tratamiento.

$$\begin{aligned} P(Y|do(A=1)) &: 0.8171 \\ P(Y|do(A=1),\ L=1) &: 0.9504 \\ P(Y|do(A=1),\ L=0) &: 0.6891 \end{aligned}$$

- **do(A=0):** Nadie recibe el tratamiento.

$$\begin{aligned} P(Y|do(A=0)) &: 0.7814 \\ P(Y|do(A=0),L=1) &: 0.8483 \\ P(Y|do(A=0),L=0) &: 0.7122 \end{aligned}$$

- **do(A=1) y do(L=0):** Toda la población recibe el tratamiento y todos los individuos son mujeres.

$$P(Y|do(A=1),do(L=0)) : 0.7257$$

- **do(A=1) y do(L=1):**Toda la población recibe el tratamiento y todos los individuos son hombres.

$$P(Y|do(A=1),do(L=1)) : 0.9371$$

Los resultados obtenidos son los esperados, obteniendo tasas de recuperación más altas para los hombres. En este ejemplo se puede apreciar claramente la diferencia entre condicionar e intervenir un valor.

Si se calcula el efecto causal para esta muestra utilizando la Ecuación 3.22, el valor que se obtiene es 0.0357, siendo ligeramente menor que la estimación realizada a partir del Modelo Causal (0.05029). Para obtener un mejor resultado, es necesario obtener un gran número de muestras para cada intervención, $do(A = 1)$ y $do(A = 0)$, y promediar las probabilidades $P(Y|do(A = 0))$ y $P(Y|do(A = 0))$. Para ello, es necesario fijar el parámetro de entrada $'computeACE' : 1$. De esta, se generan 2000 muestras para cada intervención, devolviendo el valor del efecto causal y la media y varianza de cada probabilidad:

```
P(Y=1|do(A=1)) - P(Y=1|do(A=1)) = 0.0508
Mean and variance of P(Y=1|do(A=1)): 0.8307 and 0.000206
Mean and variance of P(Y=1|do(A=0)): 0.7799 and 0.000250
```

Como se puede observar, la medida del efecto causal obtenida en este caso es prácticamente la misma que la estimación obtenida a partir del Modelo Causal, comprobando así que es necesaria la generación de varias muestras para calcular las medidas del efecto causal.

En la siguiente sección, se lleva a cabo el análisis cualitativo de un Modelo Causal más complejo, mostrando así la potencia de los Modelos Causales y en particular de la librería DoWhy.

4.3. Modelo Causal para la Elección del Modo de Transporte

La elección del modo de transporte es un factor clave a la hora de reducir la huella de carbono en el mundo. Es por ello que numerosas acciones políticas se centran en la movilidad sostenible buscando diferentes estrategias para fomentar el transporte público.

A continuación, se presenta la metodología seguida para la creación del diagrama causal que rige la toma de decisiones a la hora de elegir un modo de transporte (Ma, Chow y Xu, 2017). Además, se llevará a cabo un análisis cualitativo utilizando la librería DoWhy ya que no se dispone de los datos necesarios para realizar estimaciones, de esta forma se muestra la cantidad de información que es posible obtener a partir de los diagramas causales por ellos mismos.

4.3.1. Creación del Diagrama Causal

Son numerosos los factores que influyen directa o indirectamente en la elección del modo de transporte. Según Ma, Chow y Xu, 2017, se pueden resumir los resultados empíricos de la influencia de los diferentes determinantes en la decisión al elegir el modo de transporte ($Mode\_choice$) de la siguiente manera:

- **Factores socio-demográficos:** género ($Gender$), edad (Age), vida en pareja ($Couple$), número de hijos ($Child$), ingresos (HH_income), número de coches (N_car), permiso de conducir ($D\_license$), horario de trabajo ($WH\_flexible$) y subsidios para el transporte público ($Subsidy$).

- **Factores espaciales:** parking en el trabajo ($Parking\_work$), parking en casa ($Parking\_res$), tipo de urbanización ($Urbanization$), condiciones meteorológicas ($Weather$) y accesibilidad al trabajo en transporte público ($Accessibility\_PT$)
- **Características del viaje:** número de paradas en transporte público (T_tour), hora de salida (D_time), distancia del viaje ($T\_distance$), tiempo de viaje en coche (TT_car), tiempo de viaje en transporte público (TT_pt), coste del viaje en coche ($Cost\_car$) y coste del viaje en transporte público ($Cost\_pt$).

Se supone que pueden existir relaciones a priori entre pares de variables. Estas relaciones se pueden obtener de la literatura y del conocimiento de diversos expertos en la materia y se utilizan para identificar las restricciones más relevantes a la hora de crear el diagrama causal. Dichas restricciones son usadas para limitar el espacio de búsqueda de la estructura del diagrama causal y para comprobar la consistencia y el grado de violación que presentan las diferentes relaciones causales obtenidas a partir de los experimentos llevados a cabo. Según Campos y Castellano, 2007 existen tres tipos de restricciones en este tipo de estructuras:

- **Restricción de existencia:** Existencia de ciertas aristas dirigidas o no dirigidas en el grafo.
- **Restricción de ausencia:** Ausencia de ciertas aristas.
- **Restricción de ordenación:** Existe un camino dirigido ordenado entre dos nodos, presentan una relación temporal o funcional entre variables.

Para el presente estudio, se asumen las siguientes restricciones basadas en esta clasificación:

- **Restricción de existencia:**
  - Arista entre Tiempo de viaje y Decisión en la elección del modo
  - Arista entre Distancia del viaje y Tiempo de viaje en coche
  - Arista entre Tiempo de viaje en coche y Coste del viaje en coche
  - Arista entre Tiempo de viaje en transporte público y Coste del viaje en transporte público
- **Restricción de ausencia:**
  - Prohibidas las aristas dirigidas desde la variable objetivo (elección del modo) a sus variables explicativas.
  - Prohibidas las aristas dirigidas de las características del viaje a las características socio-demográficas y espaciales

Sin embargo, es imposible obtener las restricciones de orden a partir de la literatura. La presencia de estas restricciones más relevantes mantendrá las aristas esenciales y evitará las aristas irracionales a la hora de crear el diagrama causal.

Para la creación del diagrama causal, Ma, Chow y Xu, 2017 llevaron a cabo una encuesta de movilidad entre los trabajadores transfronterizos en la Región de Luxemburgo durante 2010 y 2011. Los encuestados detallaron el diario de viaje de un día laboral, la información espacial y socio-demográfica y dieron su opinión sobre las satisfacciones de su movilidad diaria, la percepción de los diferentes modos de transporte y los factores que pueden promover el uso del transporte público. Las variables continuas se discretizan en variables categóricas, de tal manera que hay una variable objetivo o resultado de interés (resultado de la elección del modo de transporte) y 21 variables explicativas que toman diferentes valores discretos en función de la magnitud que representan (para más información acerca de los valores que puede tomar cada variable, acudir a Ma, Chow y Xu, 2017).

Para la validación del modelo, Ma, Chow y Xu, 2017 utilizan el Método de Validación Cruzada con $k = 5$. Es decir, el conjunto de datos se divide aleatoriamente en k subconjuntos donde cada uno de ellos se utiliza para la validación cruzada utilizando el diagrama causal ajustado a partir de los $k - 1$ subconjuntos restantes.

Llevaron a cabo diferentes experimentos para probar los efectos de las restricciones estructurales, para comprobar qué algoritmo de aprendizaje es más eficaz y para comparar los distintos diagramas ajustados mediante el método de validación cruzada, obteniendo así varios diagramas causales similares entre ellos. Como el objetivo de esta sección es el análisis cualitativo de un Diagrama Causal, se ha seleccionado el diagrama de la Figura 4.14. Este incluye todas las restricciones citadas y muestra una estructura consistente, ya que la decisión de elegir el modo de transporte está directamente influenciada por el tiempo de viaje y la facilidad para encontrar aparcamiento en el lugar de trabajo. El coste del viaje está influido por el tiempo y la distancia del viaje y no tiene un efecto directo o indirecto en la elección del modo de transporte. Además, este diagrama es coherente con otro estudio sobre una población diferente también de Luxemburgo (Ma y col., 2015). En la siguiente sección se lleva a cabo el análisis cualitativo de este diagrama utilizando DoWhy.

4.3.2. Análisis del Diagrama Causal

Debido a la falta de datos, es imposible realizar ningún tipo de estimación del efecto causal que tienen las diferentes variables en la decisión en la elección del modo de transporte. Por la misma razón, es imposible obtener las muestras que se producirían al realizar algún tipo de intervención sobre las variables. Es por ello que resulta imposible llevar a cabo un análisis cuantitativo del modelo causal semejante al realizado en la Sección 4.2.

Sin embargo, como se ha venido diciendo a lo largo del trabajo, los modelos causales permiten obtener gran cantidad de información únicamente a partir del diagrama causal a partir de los diferentes criterios que se han expuesto.

Para poder llevar a cabo un análisis cualitativo utilizando DoWhy, es necesario introducir un conjunto de datos a la hora de crear el modelo. Para ello, se ha definido el parámetro $newData$, el cual es el único parámetro de entrada de la función $main()$. Si se fija $newData = 1$, se genera un nuevo conjunto de datos aleatorio de acuerdo con los valores que pueden tomar cada una de las variables y se guarda en el fichero $'mobility\_random\_data.csv'$. Si por el contrario $newData = 0$, se utiliza el conjunto de datos que previamente ha sido guardado en ese mismo fichero.

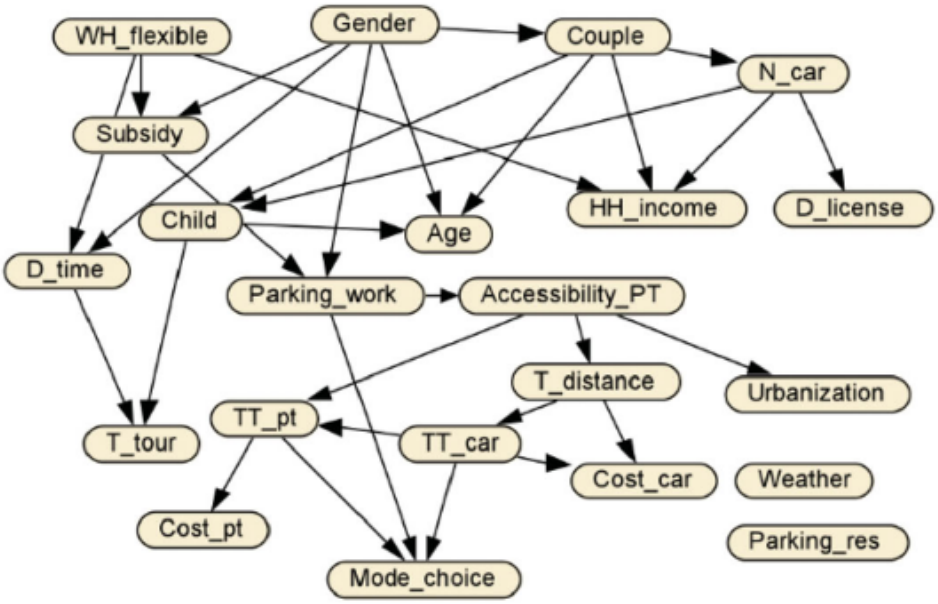

Figura 4.14: Diagrama del Modelo Causal para la Elección del Modo de Transporte (Ma, Chow y Xu, 2017).

De la misma manera, a la hora de crear el modelo, es necesario especificar qué variable es el tratamiento, es decir, la variable de interés para la cual se quiere estimar el efecto causal que tiene sobre el resultado. Como en este caso se carece de datos y este efecto no puede ser estimado, simplemente se va a comprobar si es posible identificarlo, obteniendo las suposiciones necesarias para ello y determinando los conjuntos de variables que cumplen los criterios de puerta trasera (Definición 3.2.6) y puerta delantera (Definición 3.2.7) que permitirían su estimación.

En la Figura 4.15 se puede observar la entrada necesaria para identificar el efecto causal de varias variables mientra que en la Figura 4.16 el diagrama causal generado por DoWhy.

```
treatments = ['Accessibility_PT', 'Subsidy']

for treatment in treatments:
    # I. Create a causal model from the data and given graph.
    model = CausalModel(
        data                    = df,
        treatment               = treatment,
        outcome                 = 'Mode_choice',
        graph                   = 'mobility_DAG.gml')
    model.view_model()
    display(Image(filename="causal_model.png"))

    # II. Identify causal effect and return target estimands
    identified_estimand = model.identify_effect(proceed_when_unidentifiable=True)
    print(identified_estimand)
```

Figura 4.15: Entrada para identificar del efecto causal de varias variables utilizando DoWhy.

En este caso, las variables que se han considerado más relevantes, y por tanto, sobre las que se busca identificar el efecto causal, son la accesibilidad al trabajo en transporte público, *Accessibility_PT*, y los subsidios disponibles para el transporte público, *Subsidy*,

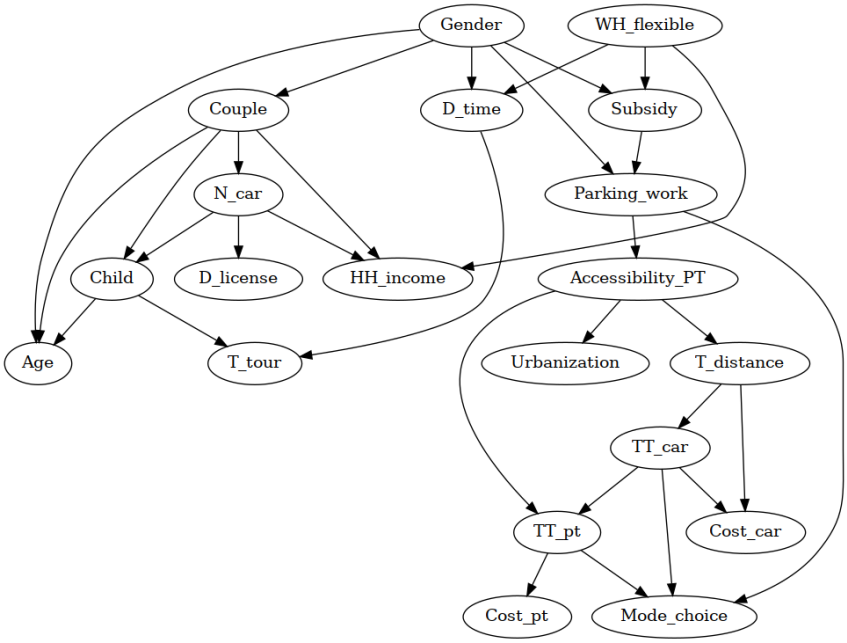

Figura 4.16: Diagrama del Modelo Causal para la Elección del Modo de Transporte generado por DoWhy.

ya que estas pueden tener un gran interés a la hora de tomar acciones políticas para fomentar el uso del transporte público y así reducir la huella de carbono. A continuación se presentan los resultados que ofrece la librería DoWhy al llevar a cabo el análisis de este modelo causal:

Accessibility_PT

Es posible identificar el efecto causal que tiene *Accessibility_PT* sobre *Mode_choice* únicamente a partir del **Criterio de Puerta Trasera**. El conjunto de variables Z que cumple con la Definición 3.2.6 es el siguiente:

```
Z = {Gender, Parking_work, T_tour, HH_income, Child, D_license, N_car,
Couple, WH_flexible, D_time, Subsidy, Age}
```

Por lo tanto, es necesario condicionar sobre todas las variables en Z para poder obtener una estimación del efecto causal a partir de la Ecuación 3.30. Para ello se asume la inconfundibilidad, es decir, no existen causas comunes U sin identificar y en caso de que existieran, estas no serían relevantes, siendo $P(Mode\_choice|Z, U) = P(Mode\_choice|Z)$.

Subsidy

En este caso, es posible identificar el efecto causal a partir de tres criterios diferentes:

- **Criterio de Puerta Trasera.** El conjunto de variables Z que cumple con la Definición 3.2.6 es el siguiente:

```
Z = {Gender, T_tour, HH_income, Child, D_license, N_car,
Couple, D_time, Age}
```

Al igual que para la variable $Accessibility\_PT$, el efecto causal se puede estimar a partir de la Ecuación 3.30 condicionando sobre todas las variables en Z asumiendo la inconfundibilidad.

- **Criterio de Puerta Delantera.** El conjunto de variables Z que satisfacen la Definición 3.2.7 es el siguiente:

$$\mathrm{Z} = \{\mathrm{Parking\_work}\}$$

El efecto causal se podría estimar a partir de la Ecuación 3.31. Para ello, se asume la condición de mediación completa, es decir, Z bloquea todos los caminos directos desde $Subsidy$ hasta $Mode\_choice$, lo cual ocurre si el diagrama causal es el asumido. También se asumen dos etapas de inconfundibilidad, si U es causa común de $Subsidy$ y Z, se tiene que $P(Z|Subsidy, U) = P(Z|Subsidy)$, mientra que si U es causa común de Z y $Mode\_choice$, se cumple que $P(Mode\_choice|Z, Subsidy, U) = P(Mode\_choice|Z, Subsidy)$.

- **Variables Instrumentales.** Las variables instrumentales son aquellas que únicamente tienen un efecto sobre el tratamiento, en este caso la única variable instrumental es $WH\_flexible$. El criterio que permite estimar el efecto causal a partir de dichas variables queda fuera del objeto de estudio de este trabajo, sin embargo, es interesante tener en cuenta las suposiciones necesarias. Por un lado, si existe una arista entre una variable U y la variable $Mode\_choice$, entonces no existirá arista entre U y $WH\_flexible$. Por otro lado, si se elimina la arista entre $WH\_flexible$ y $Subsidy$, entonces $WH\_flexible$ no tendrá efecto sobre $Mode\_choice$. Estas suposiciones se pueden obtener directamente de la propia definición de variable instrumental.

Los resultados obtenidos no manifiestan si existe o no un efecto causal de las variables sobre el resultado, sin embargo, a partir de diferentes criterios, es posible conocer que dicho efecto causal puede ser identificado, es decir, se puede encontrar una expresión para obtener dicha magnitud en función de la distribución de los datos observados. En función del criterio utilizado, se obtienen las diferentes expresiones que permiten estimar el efecto causal, indicando, por lo tanto, sobre que variables es necesario condicionar para poder calcularlo. De esta manera, al tener que condicionar sobre ciertas variables, es necesario disponer de datos suficientes acerca de las mismas para poder ejecutar dicha operación.

Este conocimiento resulta especialmente interesante a la hora de llevar a cabo un experimento, ya que si se tiene claro sobre qué variables se quiere medir el efecto causal y se conoce las relaciones causales entre las variables, es decir, el diagrama causal, es posible orientar el experimento para obtener los datos necesarios para realizar el análisis y estimar el efecto causal.

Capítulo 5

Análisis y Conclusiones

La correlación se da cuando dos variables están relacionadas, pero una no es necesariamente la causa de la otra, mientras que la causalidad se da cuando una variable causa la otra, por lo tanto, correlación no implica causalidad.

A menudo es fácil encontrar pruebas de correlación entre dos variables, sin embargo, es difícil encontrar pruebas de que una causa realmente la otra. Cuando los cambios en una variable provocan el cambio de otra, se habla de una relación causal. Incluso en este caso, existiendo una relación causal entre las variables, puede ser difícil saber la dirección de la relación.

Algunos tipos de investigación pueden proporcionar pruebas de las relaciones causales entre variables, mientras que a través de otros tipos de experimentos, únicamente se pueden encontrar correlaciones. Por un lado, los experimentos aleatorios pueden aportar buenas pruebas de relaciones causales y por otro lado, los experimentos observacionales, al no ser aleatorios, no pueden controlar todos los demás factores inevitables, a menudo no medibles, que pueden estar causando los resultados. Por lo tanto, si es posible diseñar y realizar experimentos, normalmente con pruebas A/B (como se suponen generados los datos de los ejemplos de la Sección 4.1), es siempre mejor que utilizar técnicas de inferencia causal, ya que no es necesario modelar el mecanismo a través del cual se generan los datos. Sin embargo, hay muchas situaciones en las que esto no es del todo posible. Para este tipo de situaciones, en las que únicamente se dispone de datos observacionales, tomando como verdaderas ciertas suposiciones, es posible llevar a cabo un análisis causal, para ello se presentan dos marcos teóricos a lo largo del trabajo.

En el Marco de los Resultados Potenciales, los problemas se definen algebraicamente, como supuestos de independencia contrafactual, también conocidos como supuestos de ignorabilidad. Este tipo de suposiciones pueden resultar demasiado complicadas para interpretarlas o verificarlas a simple vista. En el Marco de los Modelos Causales Estructurales, en cambio, los problemas se definen en forma de diagramas causales, a partir de los cuales las dependencias de los contrafactuales pueden derivarse mecánicamente siguiendo criterios como el de la puerta trasera.

El POF difiere del relato estructural únicamente en el lenguaje en el que se definen los problemas, y por tanto, en las herramientas matemáticas disponibles para su solución. Sin embargo, ambos marcos son lógicamente equivalentes, es decir, un problema resuelto en un marco daría solución en otro, al igual que un teorema en uno, es un teorema en el otro, y cada suposición en uno puede traducirse en una suposición equivalente en el otro.

Este hecho se ve reflejado en el ejemplo analizado en la Sección 4.2, donde se obtienen los mismos resultados en ambos marcos teóricos. En este ejemplo se resuelve fácilmente la Paradoja de Simpson, lo cual sería imposible llevando a cabo un análisis puramente estadístico, demostrando así la importancia de la inferencia causal.

Para implementar los diferentes Modelos Causales en el proyecto WHY se han elegido los Modelos Causales Estructurales, y en particular, la librería DoWhy. Esto se debe al hecho de que las suposiciones son explicitas a través del Diagrama Causal. A partir de la literatura y del conocimiento de los expertos, es posible generar un Diagrama Causal que represente el mecanismo que rige ciertas acciones y derivar a partir de él ciertos resultados incluso si no se dispone de datos, como se ha mostrado en la Sección 4.3.

Como conclusión, se puede afirmar que la inferencia causal es una herramienta muy potente en la ciencia de datos y el aprendizaje automático ya que, no sólo permite estimar la magnitud del efecto que tiene una variable en otra, si no que también permite evaluar el efecto que tendrían ciertas intervenciones e incluso examinar como sería el mundo si se hubieran tomado o no determinadas decisiones. En concreto, los Modelos Causales en el proyecto WHY permitirán realizar mejores previsiones sobre el consumo energético y analizar y validar las decisiones políticas que se llevarán y se llevaron a cabo en el ámbito energético, contribuyendo así a la reducción de la huella de carbono en el mundo.

Capítulo 6

Desarrollos futuros

Una vez sentadas las bases para llevar a cabo el proyecto, es posible comenzar a construir el modelo causal de la demanda de energía en el sector doméstico con el que se pretende comprender plenamente el proceso de toma de decisiones que hay detrás de las acciones y el comportamiento de los usuarios que dan lugar al consumo de electricidad.

Para ello, hay que llevar a cabo dos tareas principales: construir el diagrama causal, para el consumo de energía y sus factores subyacentes, y proporcionar datos compatibles con él para construir el modelo causal.

Para la primera de ellas, se ha contactado con un amplio grupo de expertos en cada uno de los aspectos de la transición energética mencionados al inicio del Apéndice B. Serán estos los que indiquen los factores que llevan a los individuos a realizar inversiones para la mejora del consumo energético, tanto económicas como en tiempo y esfuerzo, diferenciando entre cada uno de los escenarios posibles plasmados al final del Apéndice B.

Para la segunda de ellas, se llevará a cabo un conjunto de experimentos para obtener los datos necesarios para desarrollar el Modelo Causal. Estos experimentos se basarán en una serie de encuestas para cada uno de los aspectos de la transición energética que se distribuirán entre consumidores abarcando todos los escenarios descritos.

De acuerdo con el conocimiento de los diferentes expertos y a partir de los datos que se recojan en los experimentos, se desarrollará y ajustará un Diagrama Causal que proporcione un mapa completo de las causas y efectos de las decisiones y el comportamiento del usuario sobre en el consumo de energía. El procedimiento que se seguirá será similar al descrito en la Sección 4.3. Una vez se disponga del Diagrama Causal, este será traducido en un Modelo Causal utilizando la librería DoWhy para su posterior uso en el WHY Toolkit.

A continuación, se adaptarán los diferentes modelos energéticos y de comportamiento disponibles para que puedan ser utilizados en el WHY Toolkit. Del mismo modo, se desarrollará e implementará un modelo multiagente para simular la interacción de los habitantes y los diferentes dispositivos. De esta manera se podrá evaluar las medidas políticas a diferentes niveles facilitando al mismo tiempo las microsimulaciones.

Por último, se traducirán las macroconsultas pertinentes al lenguaje de los modelos causales, es decir, se expresarán las posibles preguntas políticas y técnicas como asociaciones, intervenciones y contrafactuales. Se validará el WHY Toolkit mediante la simulación en diversos escenarios de los impactos de las intervenciones en la demanda energética, perfiles de carga y costes energéticos y se crearán los plugins necesarios para que todos los participantes a niveles nacionales, europeos y mundiales puedan acceder al WHY Toolkit.

Bibliografía

Ajzen, Icek (1991). «The theory of planned behavior». En: *Organizational Behavior and Human Decision Processes* 50.2. Theories of Cognitive Self-Regulation, págs. 179-211. ISSN: 0749-5978. DOI: https://doi.org/10.1016/0749-5978(91)90020-T.

Ayaz, Ramazan, Ismail Nakir y Mugdesem Tanrioven (ago. de 2014). «An Improved Matlab-Simulink Model of PV Module considering Ambient Conditions». En: *International Journal of Photoenergy* 2014. DOI: 10.1155/2014/315893.

Bamberg, Sebastian (jun. de 2013a). «Changing environmentally harmful behaviors: A stage model of self-regulated behavioral change». En: *Journal of Environmental Psychology* 34, págs. 151-159. DOI: 10.1016/j.jenvp.2013.01.002.

— (2013b). «Environmental psychology: An introduction». En: BPS textbooks in psychology. Cap. Processes of change, págs. 267-279.

Bareinboim, Elias y Judea Pearl (2016). «Causal inference and the data-fusion problem». En: *Proc Natl Acad Sci* 113(27). DOI: 10.1073/pnas.1510507113.

Bouwma, I.M. y col. (dic. de 2015). *Policy instruments and modes of governance in environmental policies of the European Union*. Inf. téc. Wageningen.

Campos, Luis M. de y Javier G. Castellano (2007). «Bayesian network learning algorithms using structural restrictions». En: *International Journal of Approximate Reasoning* 45, págs. 233-254. DOI: 10.1016/j.ijar.2006.06.009.

Greifer, Noah y Elizabeth A. Stuart (2021). *Choosing the Estimand When Matching or Weighting in Observational Studies*. arXiv: 2106.10577 [stat.ME].

Hernán, Miguel A. y James M. Robins (2020). *Causal Inference: What If*. CRC Press/-Chapman & Hall. ISBN: 978-1420076165.

Horni, Andreas, Nagel Kai y Kay W. Axhausen (2016). *The Multi-Agent Transport Simulation MATSim*. Ubiquity Press. DOI: https://doi.org/10.5334/baw..

IEC (sep. de 2019). *Electrical Energy Storage*. Inf. téc. IEC Market Strategy Board y Fraunhofer Institut für Solare Energiesysteme ISE.

IRENA (2020). *Electrical Energy Storage*. Inf. téc. International Renewable Energy Agency, Abu Dhabi.

Jaber, A., J.J. Zhang y E. Bareinboim (2018). «Causal Identification under Markov Equivalence». En: *Proceedings of the Thirty Fourtth Conference on Uncertainty in Artificial Intelligence*, 978–-987.

Klöckner, Christian A. y Alim Nayum (sep. de 2016). «Specific Barriers and Drivers in Different Stages of Decision Making about Energy Efficiency Upgrades in Private Homes». En: *Frontiers in Psychology* 7. DOI: 10.3389/fpsyg.2016.01362.

Krajzewicz, Daniel y col. (dic. de 2012). «Recent Development and Applications of SUMO - Simulation of Urban MObility». En: *International Journal On Advances in Systems and Measurements* 3-4.

Ma, Tai-Yu, Joseph Y.J. Chow y Jia Xu (2017). «Causal structure learning for travel mode choice using structural restrictions and model averaging algorithm». En: *Transportmetrica A: Transport Science* 13:4, págs. 299-325. DOI: 10.1080/23249935.2016.1265019.

Ma, Tai-yu y col. (ene. de 2015). «Mode choice with latent preference heterogeneity: a case study for employees of the EU Institutions in Luxembourg». En: *Transportmetrica*. DOI: 10.1080/23249935.2015.1007175.

Notter, Benedikt y col. (ago. de 2019). «Development Report 4.1. HBEFA». En: *Infras. Research and Consulting*.

Onwezen, Marleen C., Gerrit Antonides y Jos Bartels (2013). «The Norm Activation Model: An exploration of the functions of anticipated pride and guilt in pro-environmental behaviour». En: *Journal of Economic Psychology* 39, págs. 141-153. ISSN: 0167-4870. DOI: https://doi.org/10.1016/j.joep.2013.07.005.

Pearl, Judea (1988). *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann. ISBN: 978-0-934613-73-6.

— (2000). *Causality: Models, Reasoning and Inference*. Cambridge University Press. ISBN: 978-0-521-89560-6.

— (feb. de 2019). «The seven tools of causal inference, with reflections on machine learning». En: *Communications of the ACM* 62, págs. 54-60. DOI: 10.1145/3241036.

Pearl, Judea, Madelyn Glymour y Nicholas P. Jewell (2016). *Causal Inference in Statistics: A Primer*. John Wiley & Sons, Ltd. ISBN: 978-1-119-18684-7.

Pearl, Judea y Dana Mackenzie (2018). *The Book of Why*. Penguin Books. ISBN: 978-0-141-98241-0.

Pflugradt, Noah (ago. de 2016). «Modellierung von Wasser und Energieverbräuchen in Haushalten». Tesis doct.

Plangklang, Boonyang y Pornchai Pornharuthai (ene. de 2013). «Mathematical Model and Experiment of Temperature Effect on Discharge of Lead-Acid Battery for PV Systems in Tropical Area». En: *Energy and Power Engineering* 05. DOI: 10.4236/epe.2013.51006.

Prochaska, James y Carlo Diclemente (ene. de 1982). «Trans-Theoretical Therapy - Toward A More Integrative Model of Change». En: *Psychotherapy: Theory, Research and Practice* 19, págs. 276-288. DOI: 10.1037/h0088437.

Prochaska, James, Colleen Redding y Kerry Evers (2008). «Health behavior and health education». En: Jossey-Bass. Cap. Five: The transtheoretical model and stages of change, págs. 97-121.

Prochaska, James y Wayne Velicer (sep. de 1997). «The Transtheoretical Model of Health Behavior Change». En: *American journal of health promotion : AJHP* 12, págs. 38-48. DOI: 10.4278/0890-1171-12.1.38.

Sharma, Amit y Emre Kiciman (2020). «DoWhy: An End-to-End Library for Causal Inference». En: *arXiv preprint arXiv:2011.04216*.

Sharma, Amit, Emre Kiciman y col. (2019). *DoWhy: A Python package for causal inference*. https://github.com/microsoft/dowhy.

Shpitser, Ilya y Judea Pearl (2006). «Identification of Joint Interventional Distributionsin Recursive Semi-Markovian Causal Models». En: *Proceedings of the Twenty-Second Conference on Uncertainty in Artificial Intelligence*, págs. 437-444.

— (2008). «Complete Identification Methods for the Causal Hierarchy». En: *Journal of Machine learning Research* 9.9, págs. 1941-1979. URL: https://www.jmlr.org/papers/volume9/shpitser08a/shpitser08a.pdf.

Simpson, E. H. (1951). «The Interpretation of Interaction in Contingency Tables». En: *Journal of the Royal Statistical Society. Series B (Methodological)* 13.2, págs. 238-241. ISSN: 00359246. URL: http://www.jstor.org/stable/2984065.

Suresh, KP (2011). «An overview of randomization techniques: An unbiased assessment of outcome in clinical research». En: *Journal of Human Reproductive Sciences* 4.

Swaminathan, Adith y Thorsten Joachims (2015). «The self-normalized estimator for counterfactual learning». En: *Advances in Neural Information Processing Systems* 28.

Uddin, Mostafa y Tamer Nadeem (ago. de 2012). «EnergySniffer: Home energy monitoring system using smart phones». En: págs. 159-164. ISBN: 978-1-4577-1378-1. DOI: 10.1109/IWCMC.2012.6314195.

USDepartmentOfEnergy (2020). *EnergyPlus*. URL: https://energyplus.net/.

Villa, Aleix Ruiz de (feb. de 2021). *Propensity Scores and Inverse Probability Weighting in Causal Inference. A global overview.* DOI: https://towardsdatascience.com/propensity-scores-and-inverse-probability-weighting-in-causal-inference-97aa53f3b6ce.

Xu, Stanley y col. (2009). «Use of stabilized inverse propensity scores as weights to directly estimate relative risk and its confidence intervals». En: *Value Health*. DOI: 10.1111/j.1524-4733.2009.00671.x.

Zandi, Helia y col. (ene. de 2018). «Home Energy Management Systems: An Overview». En: URL: https://www.osti.gov/biblio/1423114.

Apéndice A

Instalación de DoWhy en Ubuntu

- Python 3.6+ is required, you can check your version using:

```
user@computer:~$ python3 --version
```

- Make sure that *pip* and *apt* are up-to-date by running:

```
user@computer:~$ sudo apt update
user@computer:~$ python3 -m pip install --user --upgrade pip
```

  Note: In some versions it may be necessary to use *apt-get* instead of *apt*.

- For better-looking graphs, *pygraphviz* should be install, so *graphviz* is required:

```
user@computer:~$ sudo apt install graphviz libgraphviz-dev graphv
```

- *venv* allows you to manage separate package installations for different projects. It essentially allows you to create a "virtual" isolated Python installation and install packages into that virtual installation. It is always recommended to use a virtual environment while developing Python applications. If *venv* is not install run:

```
user@computer:~$ sudo apt install python3-venv
```

- Go to your project's directory and create a virtual environment named *env*:

```
user@computer:~/project$ python3 -m venv env
```

- Before start installing or using packages in your virtual environment you need to activate it:

```
user@computer:~/project$ source env/bin/activate
```

- **DoWhy requires the following packages:**

numpy≥1.15
scipy
pandas≥0.24
networkx≥2.0
matplotlib
sympy≥1.4
scikit-learn
statsmodels
pydot≥1.4
seaborn
pygraphviz
ipython
jupyter
notebook
lightgbm
simplegeneric
dowhy
econml

- Now that you are in your virtual environment you can install packages running:

```
(env) user@computer:~/project$ python3 -m pip install package_name=
```

you can also use >= instead of using ==.

- Instead of installing packages individually, *pip* allows you to declare all dependencies in a requirements text file and tell *pip* to install all of the packages in this file using the *-r* flag:

```
(env) user@computer:~/project$ python3 -m pip install -r requiremen
```

- To leave the virtual environment simply run:

```
(env) user@computer:~/project$ deactivate
```

Note: This guide has been created to install DoWhy using Ubuntu 20.04.2 LTS.

Apéndice B

Modelos Técnicos

Durante el proyecto WHY se desarrollarán varios modelos causales. En concreto, grupos de expertos debatirán y construirán un Modelo Causal para las decisiones de inversión sobre cada uno de los siguientes aspectos de la transición energética:

- **Edificios**: este grupo abarcará el sistema de climatización y el aislamiento del edificio, incluyendo aspectos de suficiencia como la temperatura del termostato.
- **Electrodomésticos**: este grupo incluirá el uso y la renovación de todos los electrodomésticos de una casa.
- **Flexibilidades**: este grupo evaluará la instalación de generación distribuida, el almacenamiento de energía y la participación en los diferentes mercados de flexibilidad existentes. Además, la estrategia de control para operar estos componentes también es incluida.
- **Transporte**: este grupo discutirá la necesidad de desplazamiento, el modo de transporte propio y, finalmente, el utilizado.

Para la construcción de dichos modelos causales, es necesario un análisis exhaustivo y preciso del consumo energético actual. Para ello, es necesario contemplar los diferentes escenarios que se pueden dar a nivel residencial en los que se pueden dividir estos aspectos de la transición energética. A continuación se presentan las diferentes tecnologías y modelos que permiten la clasificación de los diferentes escenarios que se estudian en el proyecto.

B.1. Modelos para Diferentes Escalas de Sistemas Energéticos

En primer lugar, se ha realizado un análisis de los **modelos para diferentes escalas de sistemas energéticos**, prestando especial interés en los modelos a pequeña escala, ya que el enfoque del proyecto son los edificios residenciales. Estos modelos proporcionan una comprensión más profunda del consumo de energía, tanto térmica como eléctrica, teniendo en cuenta los diferentes electrodomésticos que puede haber en un hogar. Desde el punto de vista de la modelización, los modelos de sistemas energéticos son una colección de

diferentes modelos que trabajan juntos en un marco común, cuyo detalle varía en función de la escala del sistema. Un ejemplo de estos modelos son los siguientes:

- **Load Profile Generator**: compleja y potente herramienta de modelización del consumo energético residencial centrada en los hogares individuales. Realiza una simulación completa del comportamiento de las personas de un hogar y la utiliza para generar curvas de carga(Pflugradt, 2016).
- **Energy Plus**: es un programa de análisis energético y simulación de cargas térmicas. A partir de la descripción que haga el usuario de la geometría, los materiales de construcción, el uso y los sistemas de un edificio, el software calculará las cargas de calefacción y refrigeración necesarias para mantener los puntos de control(USDepartmentOfEnergy, 2020).
- **MATLAB & Simulink:** proporcionan una plataforma integrada con análisis de datos y diseño basado en modelos. Se crean modelos predictivos de la demanda y modelos de optimización para minimizar el coste computacional en MATLAB y se combinan con un modelo de sistema construido con Simulink y Simscape que integre los dispositivos electrónicos y los sistemas de control.

B.2. Medidas de Eficiencia y Suficiencia Energética

Una opción importante que afecta al consumo de energía y por tanto a los modelos de simulación, son las **medidas de eficiencia energética**. Se han identificado dos enfoques principales para aumentar la eficiencia energética en el sector residencial:

- **Medidas de renovación**: principalmente son medidas que afectan al aislamiento de diversas partes de la vivienda(paredes, tejado, suelo, ventanas, tuberías...)
- **Etiquetas energéticas y certificados de dispositivos**: utilizar aparatos que sean lo más sostenibles posible. Se han contemplado diferentes sistemas de etiquetado presentes en diferentes países, principalmente el sistema de etiquetas reescalado de la UE.

Mientras que la eficiencia energética suele conseguirse en los hogares a través de equipos modernos y a menudo costosos, la **suficiencia energética** se consigue mediante intervenciones de bajo o nulo coste, como cambios de comportamiento o la adaptación adecuada de los equipos ya existentes en el hogar. Estas las medidas de suficiencia energética también afectan considerablemente al consumo de energía. Alguna de las acciones de suficiencia energética más relevantes encontradas en la literatura son las siguientes: fijar la temperatura del termostato a 18^{o}C, no utilizar secadoras, desenchufar aparatos electrónicos cuando no son usados, darse duchas más cortas...

B.3. Sistemas de Gestión de la Energía

Para obtener un mayor control sobre el consumo de energía y el comportamiento de los dispositivos, se puede hacer uso de los **sistemas de gestión de la energía (EMS)**.

Además de aportar información, estos sistemas permiten la activación de la flexibilidad energética, es decir, utilizan datos medidos, previsiones y algoritmos de autoaprendizaje para desplazar las cargas flexibles a los momentos en los que es más económico, ecológico o conveniente utilizarlas.

En el mercado existen multitud de opciones de EMS, que difieren en precio, aplicabilidad, requerimientos para la instalación y opciones de gestión ofrecidas al usuario. Se pueden clasificar en tres categorías diferentes(Zandi y col., 2018):

- **EMS de código abierto**: software disponible para cualquier usuario, escritos en diferentes lenguajes de programación y muy versátiles, ya que se pueden añadir nuevos dispositivos y funcionalidades. Un ejemplo es OpenEMS, escrito principalmente en Java y HTML.
- **EMS comerciales**: distribuidos por diferentes compañías, de código cerrado y por lo general más intuitivos y fáciles de utilizar que los de código abierto. Home iOs y Energy Elephant son dos ejemplos.
- **EMS de investigación**: proyectos de investigación que buscan mejorar y desarrollar nueva tecnología para los EMS. Un ejemplo es EnergySniffer(Uddin y Nadeem, 2012).

B.4. Tecnologías de Generación

Uno de los cambios clave en el sistema energético de los últimos años ha sido el avance tecnológico en el sector de las tecnologías de generación descentralizada. Éstas proporcionan a los consumidores residenciales los medios para generar su propia energía para el autoconsumo y otros fines. Las inversiones en **tecnologías de generación** son un factor de influencia clave cuando se trata de comportamientos de consumo y por lo tanto un punto importante del Modelo Causal.

Las tecnología de generación que se más relevantes para el sector residencial es la generación fotovoltaica/solar, aunque también son relevantes la generación micro eólica y la generación combinada de calor y electricidad (CHP). Para esta última se consideraron diferentes tecnologías, como la cogeneración con gas o la cogeneración con hidrógeno. En la figura B.1 se muestra un esquema con las tecnologías disponibles en el mercado y las principales variables que afectan a la toma de decisión en las instalaciones.

En función de la tecnología de generación que se quiera considerar, hay una gran variedad de modelos para su representación. Por ejemplo, en (Ayaz, Nakir y Tanrioven, 2014) se presenta un modelo en Matlab/Simulink que simula paneles fotovoltaicos mediante un circuito equivalente.

B.5. Sistemas de Almacenamiento de Energía

Una de las opciones clave para flexibilizar el consumo residencial y mitigar los efectos secundarios negativos de una cantidad creciente de fuentes de energía renovables es el uso de **sistemas de almacenamiento de energía**. Se han identificado una multitud de diferentes tecnologías(IEC, 2019), desde sistemas de almacenamiento de baterías (y subtipos), pasando por almacenamientos térmicos hasta los almacenamientos mecánicos.

En la Figura B.2 se muestra un esquema con los sistemas disponibles en el mercado y las principales variables que afectan a la toma de decisión en las instalaciones.

El comportamiento técnico de los sistemas de almacenamiento se modela mediante ecuaciones matemáticas. Hay una multitud de enfoques diferentes para las distintas tecnologías disponibles, que difieren en el grado de detalle y en el tiempo necesario para resolver las ecuaciones subyacentes. Un ejemplo de modelado de una batería de ácido sólido como un circuito equivalente con varias resistencias e impedancias conectadas en serie se presenta en (Plangklang y Pornharuthai, 2013). Este modelo de circuito abierto aproxima los procesos electroquímicos dentro de la batería y proporciona una representación realista de la misma. Por otro lado, la estrategia de control para la carga y descarga de los diferentes sistemas se modeliza a partir de los EMS.

B.6. Sistemas HVAC

Los **sistemas de calefacción, ventilación y aire acondicionado (HVAC)** suelen tener un potencial considerable para aportar flexibilidad al sistema energético y contribuir a la eficiencia energética de un edificio. En la Figura B.3 se muestra un esquema con los sistemas disponibles en el mercado y las principales variables que afectan a la toma de decisión en las instalaciones.

En términos de modelización y simulación matemática, hay que tener en cuenta tres partes:

- Los componentes: describen las partes individuales del sistema que contribuyen a la calefacción, la refrigeración o la ventilación.
- El control: es el mismo caso que para los sistemas de almacenamiento, se realiza a partir de EMS.
- El sistema en general: describe cómo están vinculados los diferentes componentes dentro del edificio.

Para su modelización hay una gran variedad de software disponible. Una opción muy utilizada es Simscape, un entorno de simulación de Matlab donde se pueden construir modelos de componentes basados en conexiones que son integrados directamente en diagramas de bloques. Permite modelar los diferentes dispositivos HVAC para ser ensamblados en un sistema de trabajo.

B.7. Movilidad

La **movilidad** o **transporte** es uno de los factores más importantes a la hora de reducir la huella de carbono en el mundo. Es por ello que numerosas acciones políticas están centradas en la movilidad sostenible, así como en la movilidad peatonal y las estrategias de transporte público. La modelización del transporte ha sido un área activa de investigación en las últimas tres décadas, es por ello que en la literatura existe una gran cantidad de estudios sobre la simulación del tráfico mediante sistemas basados en agentes. Alguno de los proyectos más relevantes son:

- **Sumo** (Krajzewicz y col., 2012): Un paquete de simulación de tráfico multimodal, microscópico y de código abierto diseñado para manejar grandes redes de carreteras y establecer un banco de pruebas común para los algoritmos y modelos de la investigación del tráfico. Dispone de un mecanismo para importar material cartográfico y generar automáticamente una entrada para la simulación del tráfico tomando datos directamente de OpenStreetMap. El principal inconveniente es tener que definir explícitamente a mano pasos de la ruta multimodal para cada ciudadano en lugar de que la simulación sea capaz de calcularlos en función de los vehículos de transporte público disponibles.

- **MATSim** (Horni, Kai y Axhausen, 2016): Un simulador de transporte basado en agentes y de código abierto, capaz de simular escenarios a gran escala. Ofrece un marco para la modelización de la demanda, la movilidad basada en agentes (simulación del flujo de tráfico), la replanificación, un controlador para ejecutar simulaciones de forma iterativa y métodos para analizar los resultados generados por los módulos. Incluye vehículos privados, transporte público, peatones y ciclistas.

Estas metodologías se centran principalmente en el cálculo y la estimación de rutas y evitar la congestión, pero no tienen en cuenta el comportamiento social que influye en la decisión de utilizar un método de transporte u otro. Las simulaciones deben tener en cuenta este paso previo para incluir cómo el comportamiento social afecta al resultado de cualquier escenario. Esto es de gran importancia ya que el consumo de energía de un viaje depende principalmente del modo de transporte y de la distancia recorrida, por lo tanto, es necesario modelar las preferencias sociales a la hora de viajar, como el coste del tiempo, la comodidad, el coste monetario o la conciencia ecológica.

El la Figura B.4 se muestra la taxonomía de los medios de transporte que se utilizarán en este proyecto. Además de su clasificación, se indica el tipo (o tipos) de energía que utiliza cada vehículo y el modelo equivalente del Manual de Factores de Emisión para el Transporte por Carretera (HBEFA, por sus siglas en inglés) (Notter y col., 2019) que se utilizará.

B.8. Intervenciones

Uno de los principales objetivos del proyecto WHY es proporcionar a los responsables políticos una herramienta para evaluar y analizar mejor los resultados de sus políticas sobre el comportamiento y las decisiones de los ciudadanos y, por tanto, sobre el consumo energético de los hogares. Esto se hace a través de **intervenciones** con las que se alimenta el Modelo Causal que simula la reacción de los residentes. Para identificar estas intervenciones, hay que entender la ciencia general que subyace a los instrumentos políticos. Según (Bouwma y col., 2015) los siguientes instrumentos políticos son relevantes:

- **Instrumentos legislativos y reglamentarios**: se describen mediante requisitos vinculantes definidos por una autoridad pública que que serán seguidos de sanciones en caso de incumplimiento. Se fuerza un determinado comportamiento.

- **Instrumentos económicos y fiscales**: se centran en proporcionar ayudas financieras fuera del mercado, como préstamos, subvenciones, impuestos, etc. Se aplican

bajo el marco de la gobernanza del mercado, por lo que se puede considerar que tienen un carácter voluntario.

- **Instrumentos basados en acuerdos o de cooperación**: se basa en gran medida en una cooperación voluntaria entre el gobierno y los actores individuales, como las asociaciones público-privadas. Los pactos o acuerdos definen las reglas a cumplir.
- **Instrumentos de información y comunicación**: trata de influir en el comportamiento del actor proporcionándole información sobre el tema en cuestión (comunicación unidireccional, desde información hasta material didáctico, etc.).
- **Instrumentos de conocimiento e innovación**: se centran en la adquisición conjunta de conocimientos a través del aprendizaje social. Los conocimientos adquiridos actúan como información y como capacidad para actuar en consecuencia.

Estos instrumentos políticos construyen los cimientos sobre los que se basarán las intervenciones del Modelo Causal. Para crear una base concluyente para el Modelo Causal se ha elaborado una ampliación de la taxonomía de medidas reglamentarias y legislativas expuesta para cubrir los aspectos más relevantes para el proyecto, identificando las diferentes acciones en cada categoría de instrumento político.

Esta extensión abarca también las iniciativas industriales, las también pueden fomentar un cambio de comportamiento, pero con un enfoque diferente. En lugar de tratar de fomentar el cambio de comportamiento energético desde una "perspectiva de autoridad", las iniciativas empresariales ven un modelo de negocio que proporciona beneficios a los clientes cuando cambian o adaptan su comportamiento. Las compañías de servicios energéticos son un claro ejemplo de un tipo de empresa que fomenta el cambio de comportamiento a través de iniciativas empresariales.

B.9. Modelos de Negocio

Hasta hace poco los consumidores domésticos de energía desempeñaban un papel muy pasivo en el sistema energético, quedando reducidos a únicamente consumir energía y pagar por ella. Con la apertura del mercado energético, debido a las nuevas leyes y reglamentos y a las nuevas tecnologías de generación descentralizada y cargas flexibles controlables(almacenamientos, bombas de calor, etc.), los antiguos consumidores pasivos están recibiendo más atención y están adquiriendo un papel más activo(prosumidores).

Para sacar provecho de todas estas nuevas posibilidades, se han desarrollado nuevos **modelos de negocio** que permiten a los usuarios obtener un mayor beneficio de sus activos y recompensar determinados comportamientos. Estos modelos de negocio son relevantes para el proyecto WHY, ya que afectan directamente al comportamiento de los consumidores, influyendo en su decisión de invertir o no en nuevas tecnologías. A continuación se muestran los modelos de negocio más relevantes(IRENA, 2020):

- **Energía como un servicio(EaaS)**: un modelo innovador en el que el proveedor ofrece varios servicios relacionados con la energía en lugar de limitarse a suministrar electricidad. Alguno de estos servicios son asesoramiento en materia de energía, esquemas de financiación de activos, tecnologías de gestión de energía, etc.

- **Comercio peer-to-peer(P2P)**: un concepto muy novedoso en el sistema eléctrico, ya que permite intercambiar electricidad entre consumidores (excedente de la propia generación) sin un intermediario (proveedor de energía). Existen diferentes opciones para realizar el comercio P2P, como los acuerdos directos de compra de energía o el uso de una plataforma interconectada.

- **Agregador**: es un proveedor de servicios que representa a un grupo de agentes (consumidores, productores, prosumidores, etc.) para que actúen como una sola entidad con el fin de alcanzar los umbrales necesarios para entrar en determinados segmentos del mercado del sistema eléctrico. Los agregadores suelen trabajar con las denominadas centrales eléctricas virtuales (VPP), que son una agregación de activos dispersos. A través de la VPP se procesan todos los datos(meteorológicos, precios, tendencias, etc.) para optimizar el uso de estos activos. Como reembolso por su servicio, los participantes obtienen una parte de los beneficios.

- **Comunidades energéticas**: tienen como objetivo la propiedad, gestión y uso colectivo de cualquier tipo de activo relacionado con la energía. Estos modelos reducen las barreras para que los particulares inviertan en tecnologías renovables y les permite poseer partes de activos con niveles de inversión más bajos. Hay diferentes modelos de propiedad comunitaria disponibles, como los sistemas de autoconsumo colectivo o los sistemas de activos comunitarios centrados en el reparto de los beneficios económicos.

- **Pago por uso (PAYG)**: son un nuevo enfoque para abordar la pobreza energética. Este modelo requiere que los clientes paguen por adelantado. Esto les da más control sobre su factura de electricidad y una mejor percepción de su consumo energético en general. En zonas en las que no hay conexión con la red eléctrica, se combina la generación de energía descentralizada y aislada a partir de fuentes de energía renovables con este modelo de negocio.

Teniendo en cuenta toda esta información, se han creado los siguientes escenarios especulativos en los que se puede definir cuáles son los factores que inducen a tomar una decisión de inversión. Cada escenario describe una realidad diferente para cada uno de los aspectos de la transición energética descritos al principio de la sección.

Línea de base. Estos escenarios tratan de reflejar el estado actual de los diferentes aspectos de la transición energética. Intentan dar una descripción general de una casa normal en una ciudad de Europa en la actualidad.

Mínimo. Estos escenarios incluyen el esfuerzo mínimo necesario para mejorar el escenario base hacia la descarbonización para cada campo de aplicación. Normalmente, estos escenarios se basan en aspectos de comportamiento más que en acciones monetarias.

Probables. Estos escenarios son una proyección de la toma de decisiones más probable que los ciudadanos de cualquier ciudad europea llevarían a cabo a partir del escenario base para avanzar hacia los objetivos de descarbonización.

Plausible. Estos conjuntos de escenarios son menos probables que los anteriores, pero su ocurrencia no sería demasiado extraña para ser observada en algunas ciudades y unidades familiares de la UE para alcanzar los objetivos de descarbonización.

Ideal. Estos son los escenarios ideales para alcanzar los objetivos de descarbonización, sin embargo, es muy poco probable que se produzcan debido a la innovación social masiva que debe llevarse a cabo o el cambio cultural que conlleva.

Según la bibliografía existente, parece que factores internos como las creencias, los valores o la preocupación por el medio ambiente son los principales impulsores de un cambio en el comportamiento de las personas para invertir tiempo y esfuerzo en la reducción de la huella de carbono y energía en su vivienda personal. Además, los factores externos dominantes para que se produzca este cambio de comportamiento proambiental parecen ser la presión de los compañeros, la comparación social y las normas sociales. En la siguiente sección se detalla cómo se produce este cambio de comportamiento indicando los factores que llevan a él.

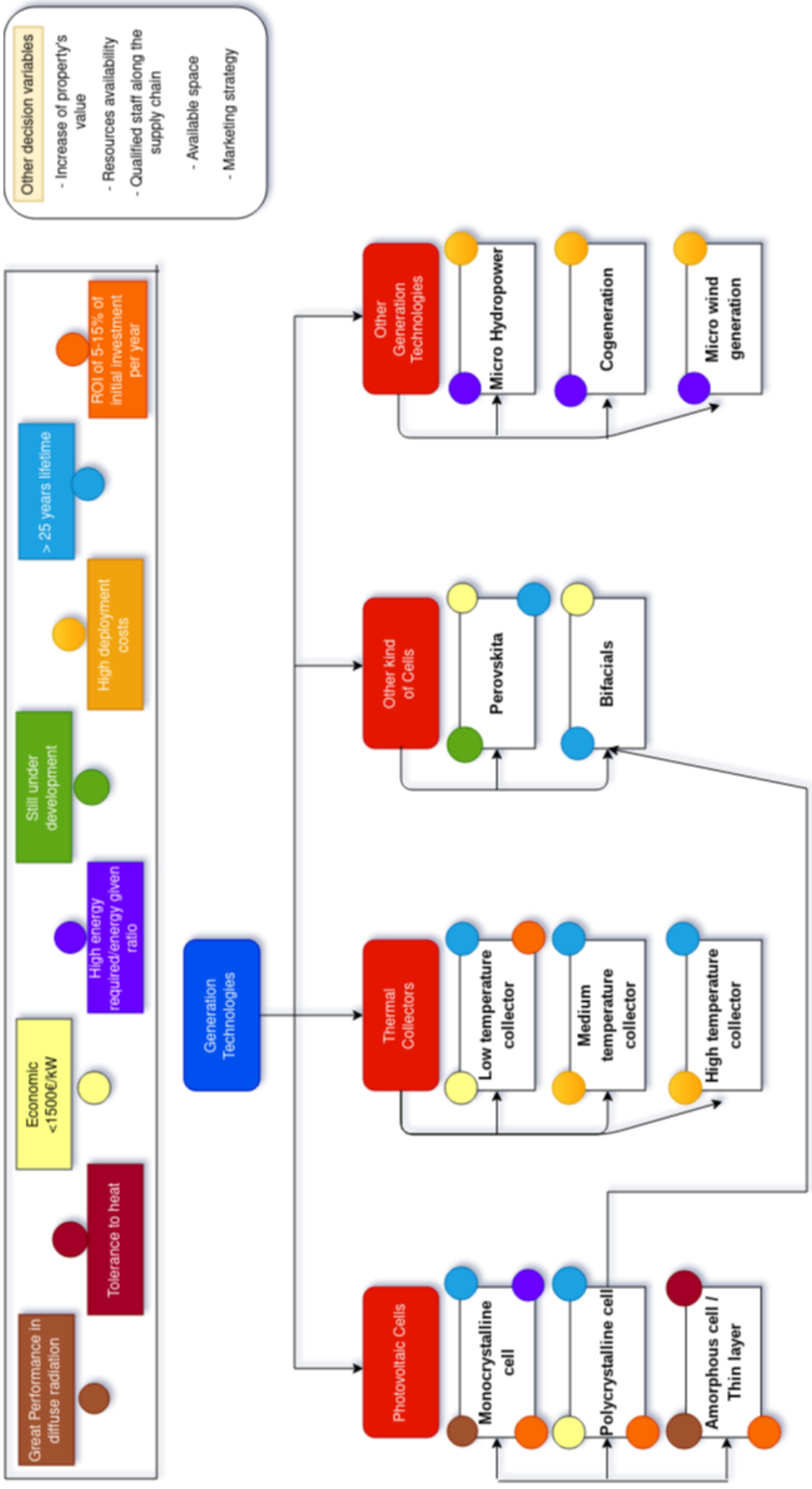

Figura B.1: Tecnologías de generación disponibles en el mercado y las principales variables de decisión.

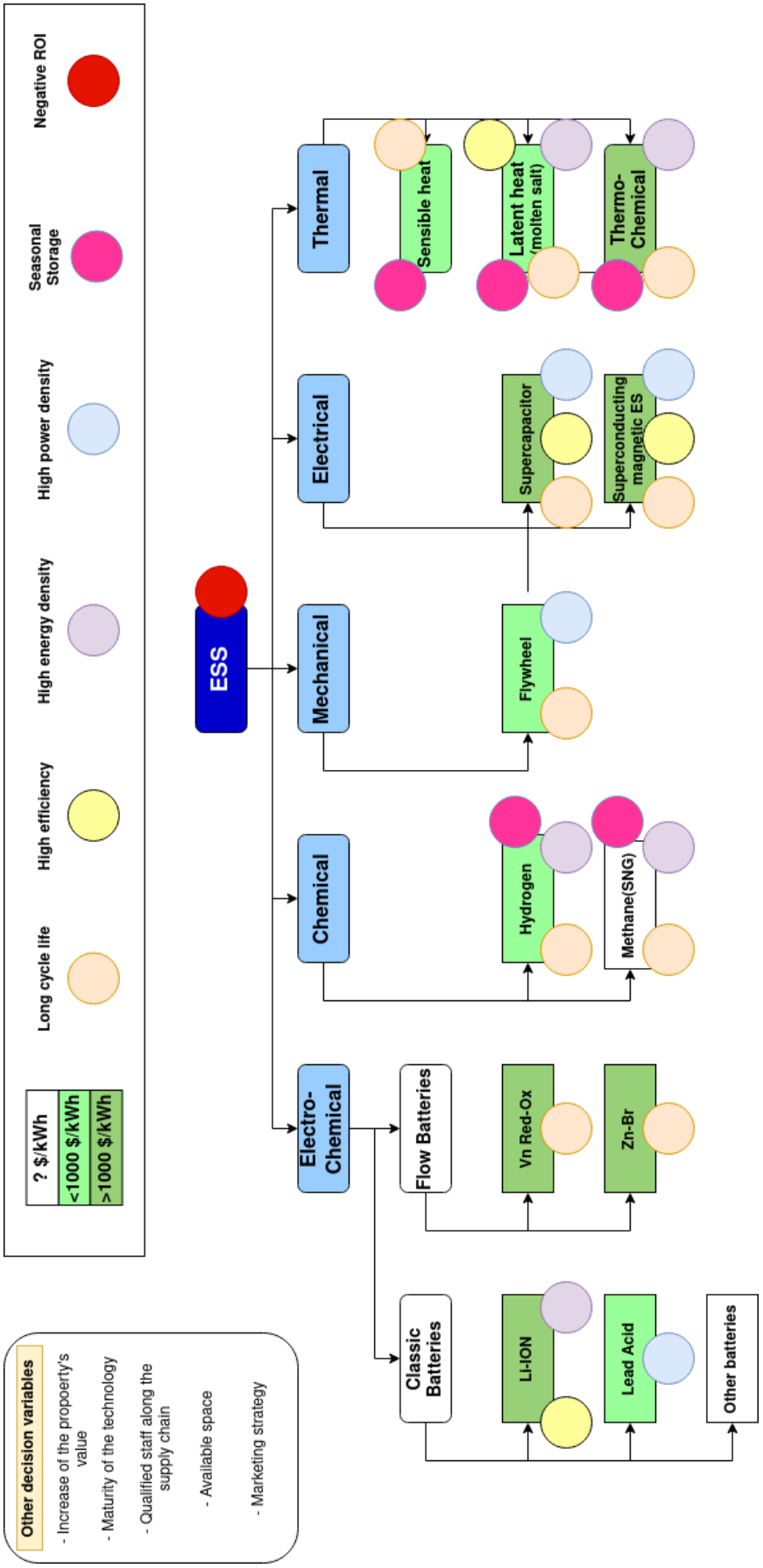

Figura B.2: Sistemas de almacenamiento disponibles en el mercado y las principales variables de decisión.

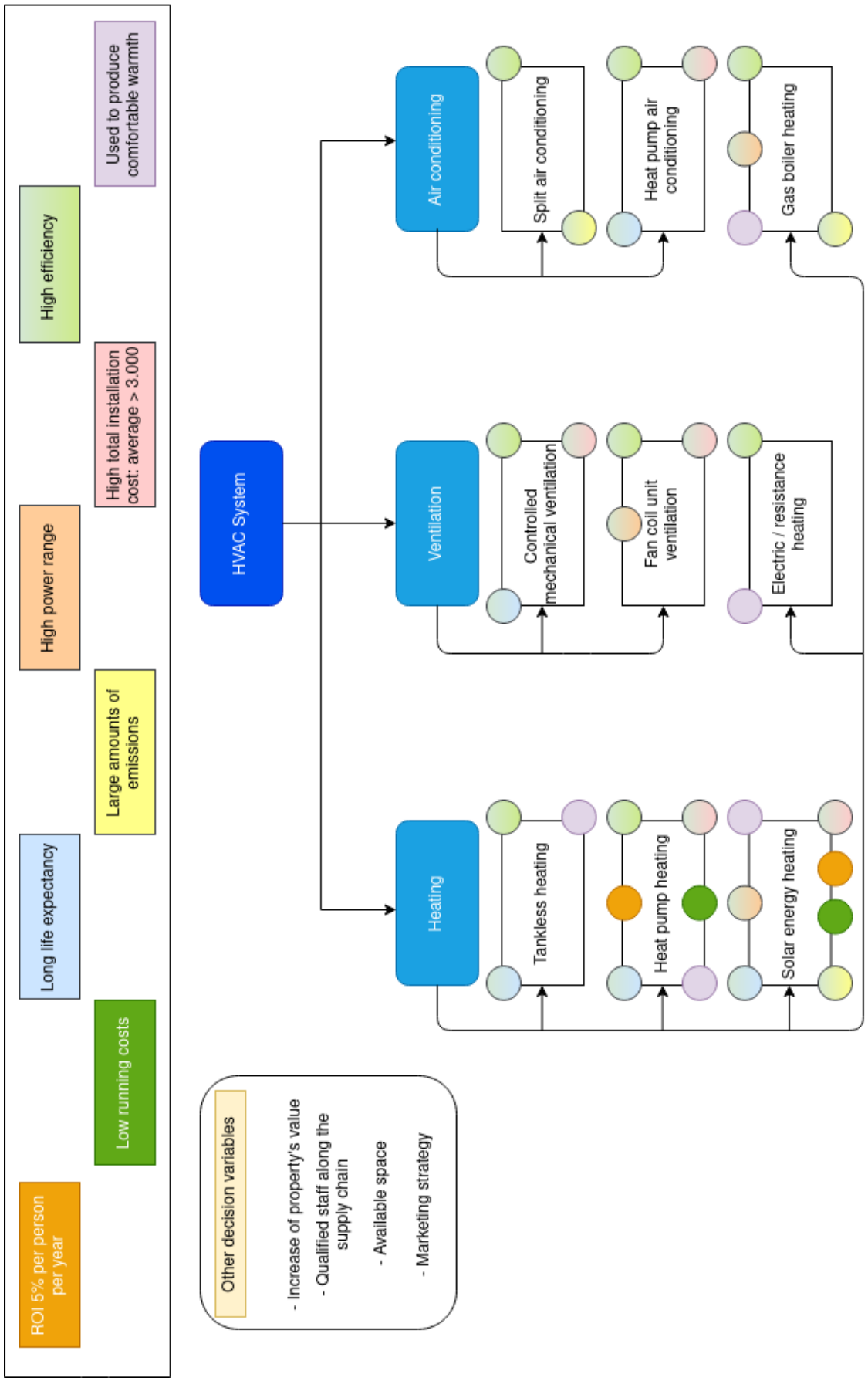

Figura B.3: Sistemas HVAC disponibles en el mercado y las principales variables de decisión.

Clasificación					Type of energy used and equivalent model of HBEFA			
					Animal traction	Electric	Hybrids	Combustion
Individual	Public	Saddle / Animal-drawn vehicles			Ø	--	--	--
		Bicycles / Scooters / Segways / etc.			Ø	eBike	--	--
		Mopeds (<50 km/h) / cars without a license			--	eScooter	--	Moped <=50cc Euro-5
		Motorcycles (> 50 km / h)			--	MC BEV	--	MC 4S <=250cc Euro-6
	Private	On foot / Saddle / Animal drawn vehicles			Ø	--	--	--
		Bicycles / Scooters / Mobility Vehicles			Ø	eBike	--	--
		Mopeds			--	eScooter	--	Moped <=50cc Euro-5
		Motorcycles			--	MC BEV	--	MC 4S <=250cc Euro-6
Collective	Public	Fixed route	Predictable	Metro / Tram / Train	--	Train	--	--
				Funicular / Elevator	--	Elevator	--	--
			Unpredictable	City bus	--	UBus Electric Std >15-18t	UBus Std >15-18t HEV Euro-VI	UBus Std >15-18t Euro-VI
				Coach	--	Coach BEV Std <=18t	--	Coach Std <=18t Euro-VI
		Variable route	Taxis		--	PC BEV	PC PHEV diesel Euro-6ab (EI)	PC diesel Euro-6ab PC petrol Euro-6ab
			Rental cars					
	Private	Passenger cars	Micro cars		--	PC BEV	PC PHEV diesel Euro-6ab (EI))	PC diesel Euro-6ab PC petrol Euro-6ab
			A segment (mini cars)					
			B segment (small cars)					
			C segment (medium cars)					
			D segment (family cars)					
			E segment (executive)					
			F segment (luxury)					
			S segment (sport)					
			M segment (monovolumen)					
			J segment (SUV)					
		Light commercial vehicles			--	LCV BEV N1-II	LCV PHEV petrol N1-II Euro-6	LCV diesel N1-II Euro-6c

Figura B.4: Taxonomía de los medios de transporte.

Apéndice C

Modelo Transteórico

El Modelo Transteórico (TTM, por sus siglas en inglés) utiliza una dimensión temporal, las etapas de cambio, para integrar los procesos y principios de cambio de las principales teorías de intervención. Este modelo emergió de un análisis comparativo de las principales teorías de la psicoterapia y del cambio de comportamiento, en un esfuerzo para integrar un campo que se había visto fragmentado en más de 300 teorías (Prochaska y Velicer, 1997).

Este ímpetu por el modelo surgió cuando Prochaska llevó a cabo un análisis comparativo entre personas que dejan de fumar de manera autónoma con las que llevan a cabo tratamientos profesionales (Prochaska y Diclemente, 1982). Se identificaron diez procesos de cambio que precedían al éxito del abandono del tabaco, incluyendo la toma de conciencia de la teoría de Freud, la gestión de contingencias de la teoría de Skinner, y las relaciones de ayuda de la teoría de Rogers(Prochaska, Redding y Evers, 2008):

Toma de conciencia Encontrar y aprender nuevos hechos, ideas y consejos que apoyen el cambio de comportamiento

Alivio significativo Experimentar las consecuencias negativas que conlleva adoptar un cierto comportamiento

Auto-reevaluación Darse cuenta de que el cambio de comportamiento es una parte importante de la propia identidad como persona

Reevaluación medioambiental Darse cuenta del impacto del cambio de comportamiento en el entorno próximo

Autoliberación Comprometerse firmemente con el cambio

Relaciones de ayuda Buscar y utilizar el apoyo social para el cambio de comportamiento

Contracondicionamiento Sustitución de la conducta por una alternativa más apropiada

Gestión del refuerzo Aumentar las recompensas por el cambio de comportamiento positivo y disminuirlas por el negativo

Control de estímulos Añadir señales o recordatorios para realizar la conducta más apropiada

Liberación social Darse cuenta de que las normas sociales están cambiando en la dirección de apoyar el cambio de comportamiento apropiado

Los participantes utilizaban diferentes procesos en diferentes momentos en su lucha contra la adicción. Estos individuos revelaron a los investigadores que el cambio de comportamiento se desarrolla a través de una serie de etapas, un fenómeno que no estaba recogido en ninguna teoría terapéutica y que condujo al desarrollo del TTM.

C.1. Etapas de Cambio

En el pasado, el cambio de comportamiento se entendía como un evento discreto, como por ejemplo, dejar de fumar. El TTM caracteriza el cambio de comportamiento como un proceso que se desdobla a lo largo del tiempo, una transición no lineal a través de una secuencia de etapas cualitativamente distintas. En cada etapa existen fuentes individuales de resistencia al cambio, específicas de cada una de ellas, que pueden mantener a las personas atascadas en una fase temprana durante un largo periodo de tiempo, por lo que son necesarias habilidades y estrategias específicas. El éxito de la transición a través de las etapas se refleja en la creciente disposición al cambio. El TTM parte de seis etapas de cambio:

Precontemplación: etapa en la que las personas no tienen intención de actuar en un corto periodo de tiempo, normalmente medido como los próximos seis meses. Las personas pueden estar en esta etapa debido a que no están bien informadas sobre las consecuencias de su comportamiento o puede que hayan intentado cambiar varias veces y se hayan desmoralizado.

Contemplación: etapa en la que las personas intentan cambiar sus comportamientos en los próximos seis meses. Son más conscientes de los pros del cambio, pero también son muy conscientes de los contras. Este balance entre pros y contras puede producir una profunda ambivalencia y mantiene a las personas atrapadas en la contemplación durante largos periodos de tiempo.

Preparación: etapa en la que las personas tienen intención de actuar pronto, generalmente en los próximos seis meses. Normalmente, ya han dado algún paso significativo hacia el comportamiento en el último año y tienen un plan de acción.

Acción: etapa en la que las personas han realizado modificaciones concretas y manifiestas en su estilo de vida en los últimos seis meses. No todos los cambios en la conducta cuentan como acción en este modelo, se tienen que alcanzar ciertos criterios que los científicos y los profesionales consideren suficientes.

Mantenimiento: etapa en la que las personas han realizado modificaciones específicas y manifiestas en su estilo de vida y trabajan para evitar las recaídas, no aplican los procesos de cambio con tanta frecuencia como las personas en acción. Cada vez tienen más confianza en que pueden continuar con los cambios.

Terminación: etapa en la que las personas tienen cero tentaciones de recaída y 100 % de autoeficacia. Es como si nunca hubiera existido la conducta en primer lugar o su nuevo comportamiento se hubiera convertido en algo automático. Este criterio puede ser demasiado estricto, siendo un objetivo ideal para la mayoría de las personas.

C.2. Evolución del Modelo Transteórico

Originalmente el TTM se desarrolló para entender la adopción de conductas en cuanto a salud personal, sin embargo, este modelo ha sido transferido a otro campos. Bamberg propuso el *Modelo de Etapas de Cambio de Comportamiento Autorregulado* (Bamberg, 2013a y Bamberg, 2013b), modelo que asume que el cambio de comportamiento medioambiental se puede caracterizar como una transición a través de una secuencia ordenada en el tiempo a través de cuatro etapas cualitativamente diferentes, cuyos nombre toma prestados del Modelo Transteórico.

El TTM ha sido criticado por ser un modelo principalmente descriptivo, que no ofrece respuestas claras a importantes preguntas como son la separación de las etapas y los procesos que desencadenan la transición entre ellas. Por esta razón, el Modelo de Etapas de Cambio de Comportamiento Autorregulado integra el concepto de etapa con constructos bien establecidos tomados de la TPB[1] y la NAM[2], dando una descripción detallada de las tareas que una persona tiene que resolver en las cuatro etapas que se consideran, así como la mentalidad cognitiva específica que las personas adoptan para resolverlas. El modelo asume que la transición a través de las etapas está marcada por tres puntos de transición críticos, la intención de meta, la intención de conducta y la intención de ejecución, los cuales reflejan la solución exitosa de las tareas específicas de cada etapa:

Precontemplación: el proceso de cambio comienza cuando las personas empiezan a reflexionar conscientemente sobre su comportamiento actual, dándose cuenta de que causa efectos sociales y medioambientales negativos. Si la persona acepta su responsabilidad individual, experimentará efectos negativos como el sentimiento de culpa y sentirá preocupación por lo que los demás esperan que haga (normas sociales), sintiéndose obligada a cambiar este comportamiento (norma personal). Si las normas personales y las emociones positivas por cumplirlas son fuertes y la viabilidad percibida de cambiar el comportamiento actual es alta, se forma una **intención de meta** que indica la transición a la segunda etapa.

Contemplación: la persona valora las consecuencias personales asociadas a las opciones de comportamiento alternativas (reflejadas en su actitud), así como la de llevarlas a cabo (reflejada en el control de comportamiento percibido). La **intención de**

[1] *Teoría del comportamiento planeado*(Ajzen, 1991) asume que la actitud, el control de comportamiento percibido(PCB) y las normas subjetivas son factores socio-cognitivos centrales que promueven la formación de una intención de comportamiento

[2] *Modelo de activación de la norma*(Onwezen, Antonides y Bartels, 2013) asume que cuando un individuo toma conciencia de que su comportamiento actual tiene consecuencias perjudiciales para otras personas y/o el medio ambiente (conciencia de las consecuencias) y acepta la responsabilidad personal por haber causado este daño (atribución de la responsabilidad), esto puede provocar sentimientos negativos como la culpa. También asume que la activación de una norma personal conduce a la anticipación de emociones positivas (orgullo, satisfacción) asociadas a un comportamiento más acorde con la norma personal

comportamiento, que marca la transición entre la etapa de contemplación y la de la etapa de preparación/acción, es el resultado de la ponderación por parte de la persona de los pros y los contras de de las diferentes opciones de comportamiento para alcanzar el objetivo, así como de la dificultad percibida de llevar a cabo estas opciones.

Preparación/acción: La principal tarea de esta etapa consiste en iniciar las acciones necesarias para poner en práctica las nuevas intenciones de comportamiento. Para ello, se planifica cuándo y dónde actuar para alcanzar el objetivo previsto mediante la realización de la conducta intencionada. En este caso, las capacidades de planificación cognitiva y las habilidades para hacer frente a los problemas de ejecución reales o previstos son variables importantes. Al final de esta fase, se establece una **intención de implementación** y la conducta se lleva a cabo.

Mantenimiento: comprende la estabilización de la conducta modificada y la implantación de nuevas rutinas o hábitos de comportamiento basados en la conducta modificada. La tarea en esta etapa consiste en hacer frente a las experiencias desagradables con el nuevo comportamiento y a la consiguiente tentación de volver a la conducta anterior. Para ello, son necesarias las habilidades para resistir esta tentación y, si se produce una recaída, las habilidades para recuperar y restablecer el nuevo comportamiento.

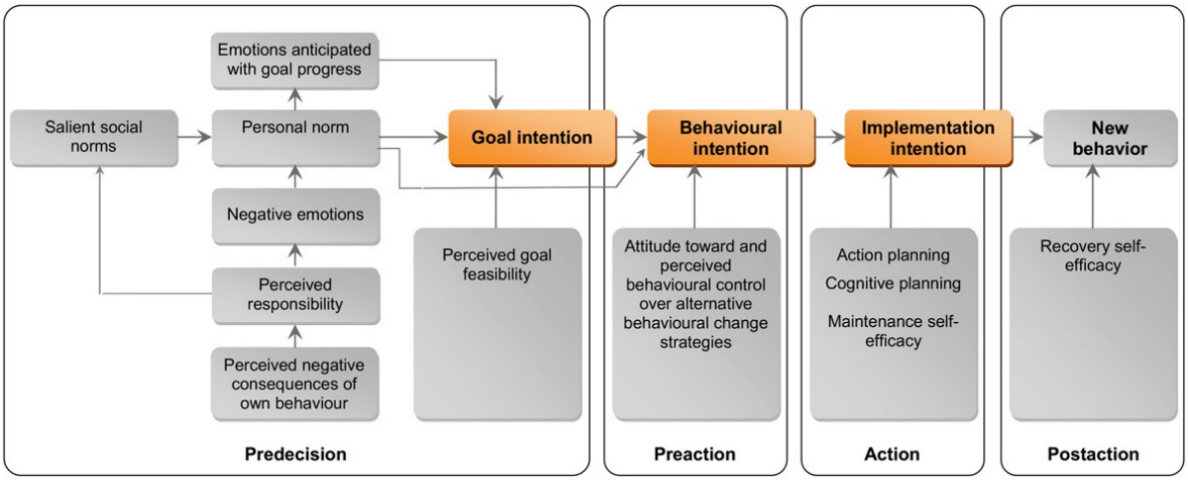

Figura C.1: El Modelo de Etapas de Cambio de Comportamiento Autorregulado (Bamberg, 2013a)

Este modelo se utiliza en el desarrollo sistemático de intervenciones destinadas a promover el cambio de comportamientos relevantes para el medio ambiente. Christian Klöckner, utilizando una variación del Modelo de Etapas de Cambio de Comportamiento Autorregulado, comprobó el efecto específico de 24 barreras e impulsores en las diferentes etapas de la toma de decisiones sobre la mejora de la eficiencia energética en las viviendas particulares (Klöckner y Nayum, 2016). En su estudio, propone el uso de cuatro etapas diferenciadas. La primera etapa es "no estar en modo de decisión", la cual corresponde a la mentalidad de la etapa de *precontemplación* en el modelo de Bamberg. La segunda la denominó "decidiendo qué hacer", que tiene un claro vínculo con la etapa *contemplación*.

La tercera, "decidiendo cómo hacerlo" estaría posicionada entre *contemplación* y *acción*. Y por último, la cuarta etapa, "decidiendo cómo implementarlo" que corresponde con la etapa *acción* de Bamberg. En la Figura C.2 se muestran las barreras y conductores más relevantes para la progresión a través de las diferentes etapas.

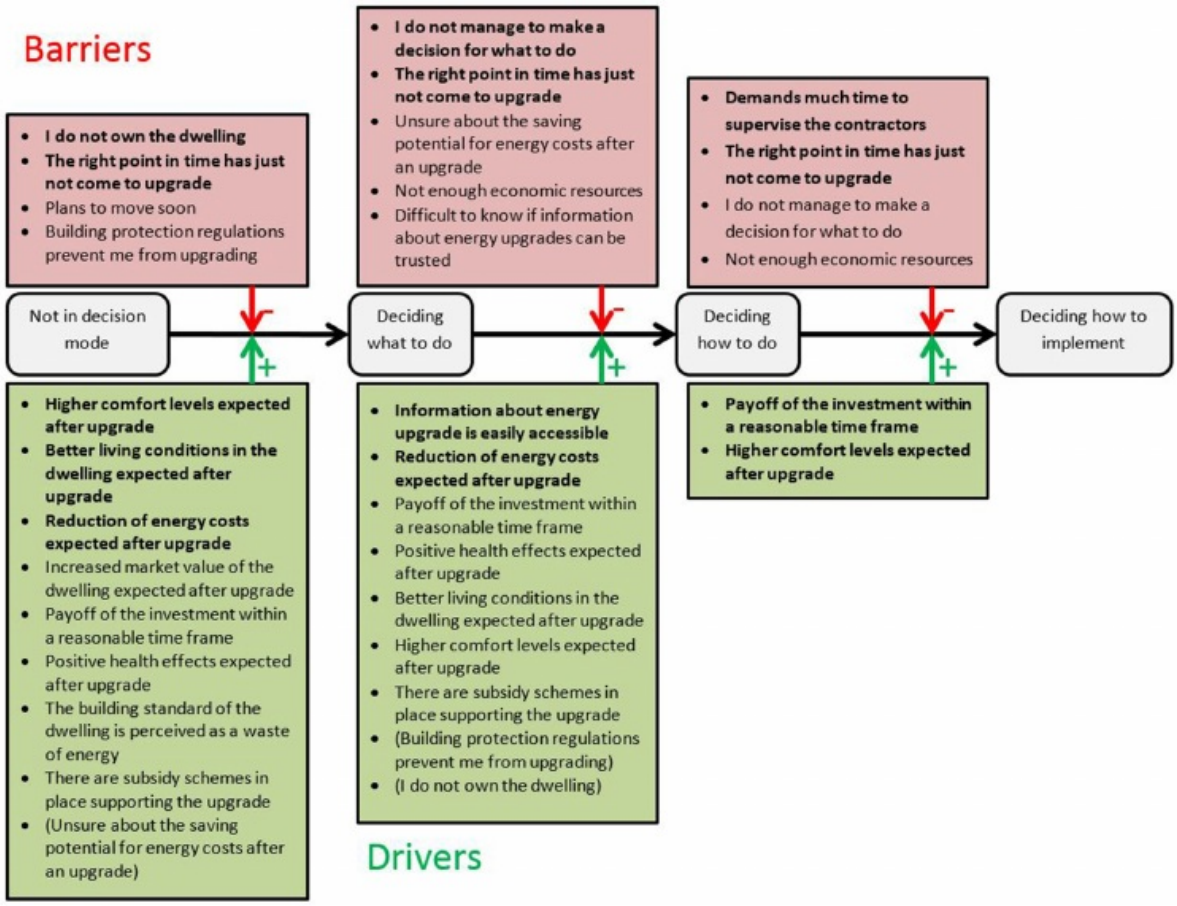

Figura C.2: Resumen de la relevancia de las barreras y conductores en la transición entre etapas (Klöckner y Nayum, 2016)

El análisis de los obstáculos y los factores que impulsan la toma de decisiones sobre la mejora de la eficiencia energética en cada etapa proporciona información que se perdería si se analizaran los mismos sin tener en cuenta hasta dónde ha llegado la persona en la toma de decisiones. En este estudio, la no linealidad del proceso no está suficientemente representada y no se tuvo en cuenta la última etapa del proceso de decisión, es decir, si las medidas decididas se aplicaron realmente en algún momento. A su vez, las correlaciones entre las etapas 2-4 son altas, lo que indica que probablemente no son lo suficientemente distintas o que el instrumento de medición no fue lo suficientemente bueno.

A pesar de estos inconvenientes, estos resultados podrían utilizarse para encontrar formas de promover la adopción de medidas de eficiencia energética en edificios residenciales y otras inversiones proambientales de alto coste. También permiten adaptar los planes de comunicación y apoyo, tanto públicos como privados, para conseguir que las personas interesadas superen el umbral para implementar esta tecnología.

Sin duda, este es uno de los objetivos principales del proyecto WHY, encontrar las

causas que llevan a la adopción de medidas para la descarbonización del sector residencial para poder intervenir en ellas mediante diferentes políticas. Es por ello que los diferentes Modelos Causales que se construyen para los diferentes aspectos de la transición energética tienen como base el Modelo Transteórico descrito en esta sección, y en particular, el modelo descrito en Klöckner y Nayum, 2016. Con esto se pretende dotar a los Modelos Causales con una fiel representación del comportamiento humano, y con los principales conductores que llevan a un cambio positivo en él.