



Universidad de Valladolid

Facultad de Ciencias

Trabajo Fin de Grado

Grado en Estadística

**ALGUNAS APLICACIONES DEL ÍNDICE DE
PROPENSIÓN**

Autor:

D. Antonio Fraile Valderrey

Tutor:

D. José Antonio Menéndez Fernández

ÍNDICE GENERAL

1. Resumen. -----	1
2. Introducción. -----	2
3. Índice de Propensión. -----	5
3.1. Diseños experimentales y observacionales. -----	5
3.2. Definición Índice de Propensión. -----	8
3.3. modelo causal de Rubin. -----	18
3.4. Aplicación del Índice de Propensión. -----	22
3.4.1. Estimación del Índice de Propensión. -----	22
3.4.1.1. Regresión logística. -----	22
3.4.2. Técnicas de reducción del sesgo obtenido y estimación del efecto del tratamiento. -----	23
3.4.2.1. Estratificación. -----	24
3.4.2.2. Emparejamiento. -----	27
3.4.2.2.1. Emparejamiento mediante distancia de Mahalanobis. -----	28
3.4.2.2.2. Emparejamiento mediante Vecino más próximo. -----	28
3.4.2.2.3. Emparejamiento mediante Calipers. -----	29
3.4.2.2.4. Emparejamiento mediante Comparación N:N. -----	29
3.4.2.2.5. Emparejamiento mediante Kernel. -----	29
3.4.2.2.6. Emparejamiento mediante estratificación. -----	30
3.4.2.2.7. Comparación de los métodos de emparejamiento del Índice de Propensión. --	30
3.4.2.3. Ponderación. -----	31
3.4.3. Elección del mejor método de ajuste. -----	32
3.5. Usos, ventajas y limitaciones del Índice de Propensión. -----	32
4. Ejemplo de aplicación del Índice de Propensión. -----	34
4.1. Introducción y descripción del problema. -----	34
4.2. Cálculo del Índice de Propensión. -----	37
4.3. Corrección del sesgo mediante el método de emparejamiento. -----	39
4.4. Corrección del sesgo mediante el método de estratificación. -----	43
4.5. Estimación del efecto del tratamiento. -----	48
5. Conclusiones. -----	51

6.	Referencias y bibliografía.-----	53
7.	Anexos.-----	56
7.1.	ANEXO 1. Código R para resolver ejemplo con datos <i>pride</i> de la librería <i>nonrandom</i> .-----	56
7.2.	ANEXO 2. Código R para resolver ejemplo con datos propios sobre PBC. -----	58

ÍNDICE DE TABLAS

Tabla 1. Variables del conjunto de datos pride	9
Tabla 2. Emparejamiento del Índice de Propensión del ejemplo pride.	12
Tabla 3. Balance del emparejamiento del ejemplo pride.	13
Tabla 4. Estratos en la estratificación del ejemplo pride.	14
Tabla 5. Balance de la estratificación del ejemplo pride.	15
Tabla 6. Estimación del efecto del tratamiento por el emparejamiento ejemplo pride. --	16
Tabla 7. Estimación del efecto del tratamiento por la estratificación ejemplo pride. ---	17
Tabla 8. Tabla 2x2 para cada estrato para estimar efecto del tratamiento.	26
Tabla 9. Tabla 2x2 para estimar efecto del tratamiento en emparejamiento.	27
Tabla 10. Variables del conjunto de datos PBC.	35
Tabla 11. Análisis variables numéricas conjunto datos PBC.	36
Tabla 12. Análisis variables categóricas conjunto datos PBC.	36
Tabla 13. Estimadores por reg. Logística de los parámetros.	38
Tabla 14. Diferencias medias estandarizadas variables antes emparejamiento.	40
Tabla 15. Diferencias medias estandarizadas variables después emparejamiento.	41
Tabla 16. Resumen emparejamiento ejemplo PBC.	41
Tabla 17. Estratificación con 5 estratos ejemplo PBC.	43
Tabla 18. Resumen balance estratificación 5, ejemplo PBC.	44
Tabla 19. Balance diferencias estandarizadas estratificación 5, ejemplo PBC.	45
Tabla 20. Estratificación con 4 estratos ejemplo PBC.	46
Tabla 21. Resumen balance estratificación 4, ejemplo PBC.	46
Tabla 22. Balance diferencias estandarizadas estratificación 4, ejemplo PBC.	47
Tabla 23. Estimación del efecto del tratamiento por el emparejamiento ejemplo PBC. --	48

ÍNDICE DE GRÁFICOS

Gráfico 1. Distribución del Índice de Propensión del ejemplo pride. -----	11
Gráfico 2. Balance de las covariables antes y después del emparejamiento ejemplo pride. -----	13
Gráfico 3. Distribución del Índice de Propensión antes emparejamiento y estratificación ejemplo PBC. -----	38
Gráfico 4. Balance de las covariables antes y después del emparejamiento ejemplo PBC. -----	42
Gráfico 5. Distribución del Índice de Propensión tras emparejamiento ejemplo PBC. --	42

1. Resumen

Cuando se comparan dos tratamientos mediante la asignación de una muestra de individuos a cada uno de ellos, se denomina Índice de Propensión a la probabilidad de que un individuo sea asignado a uno de ambos tratamientos. La técnica estadística que hace uso del Índice de Propensión fue introducida por P. Rosenbaum y D. Rubin en estudios observacionales cuando se trata de estimar el efecto de un determinado tratamiento teniendo en cuenta una serie de covariables observadas. Mediante el Índice de Propensión se trata de controlar el sesgo de selección y de confusión inherentes a los estudios observacionales y que dificulta la valoración del efecto de cada tratamiento. En este trabajo presentamos varias técnicas asociadas, entre las que se destacan el emparejamiento y la estratificación. Las técnicas basadas en el Índice de Propensión están basadas en el modelo causal para datos observacionales desarrollado por D. Rubin.

2. INTRODUCCIÓN

En algunas situaciones puede resultar de interés conocer qué efecto tiene un determinado tratamiento sobre los individuos. Según el diseño mediante el cual fueron obtenidos los datos, habrá que buscar técnicas que permitan estimar dicho efecto.

Los conjuntos de datos se pueden obtener mediante diseños experimentales o diseños observacionales. Estos diseños se diferencian principalmente en la forma en la que los individuos han sido incluidos en el conjunto de observación y en cómo se les asigne a estos individuos el tratamiento de interés.

Si tanto la inclusión de los individuos en la muestra que conformará el conjunto de datos como la asignación del tratamiento a dichos individuos se realiza de forma aleatoria, entonces se trata de un diseño experimental. Las estimaciones que se deseen obtener bajo este tipo de diseño no se verán afectadas por covariables externas que den lugar a un sesgo de confusión. Tampoco estarán afectadas por el sesgo de selección, ya que la inclusión de los individuos se realizó, como hemos comentado anteriormente, de forma aleatoria. A pesar de las ventajas que nos ofrece un diseño experimental, también lleva asociado alguna que otra limitación, como puede ser en el ámbito económico, ya que resulta inviable la obtención de los datos, y además en ocasiones esto viola algún principio ético.

Por otro lado, si la inclusión de los individuos en la muestra y la asignación del tratamiento no se realizaron de forma aleatoria sino que el investigador recogió los datos según se observaron, entonces se dice que los conjuntos de datos proceden de diseños observacionales. Al contrario de lo que ocurre en un diseño experimental, las estimaciones obtenidas por los investigadores cuando se trabaja con este tipo de diseños sí pueden verse afectadas por covariables externas; dicho de otra forma, las estimaciones obtenidas de este modo pueden verse afectadas por sesgos tanto de confusión como de selección. Es por este motivo por el que se necesitan métodos y técnicas estadísticas que permitan al investigador controlar y reducir, en la medida de lo posible, dichos sesgos, obteniendo de ese modo estimaciones lo más precisas posibles.

Uno de los métodos que podemos encontrar para solventar los problemas de sesgo mencionados es el que se conoce como Índice de Propensión, que es una técnica que introdujeron D. Rubin y P. Rosenbaum y que trata de controlar los sesgos que aparecen a la hora de hacer estimaciones a partir de los datos obtenidos mediante diseños observacionales. Es una técnica basada principalmente en la teoría del modelo causal desarrollada por D. Rubin para datos observacionales {[Thomas A.DiPrete y otro (2004)] y [Peter C. Austin (2011)]}.

Para obtener las estimaciones del Índice de Propensión se disponen de varias técnicas propuestas por Rubin como pueden ser, entre otras, la regresión logística, la regresión logit multinomial, la estimación mediante árboles de clasificación y regresión (CARTS) o mediante el uso de Redes Neuronales.

Tras estimar el Índice de Propensión tanto para el grupo de individuos a los que se les ha asignado un tratamiento como para aquellos individuos que conforman el grupo control, se comprueba si ambas distribuciones son muy diferentes. En caso afirmativo, se tendría que realizar un paso intermedio antes de poder estimar el efecto del tratamiento, de modo que podamos controlar de alguna manera los sesgos existentes consiguiendo así que las distribuciones en ambos grupos sean lo más similares posibles. Para conseguirlo contamos con métodos como la estratificación o el emparejamiento, que a su vez, pueden ser complementados con el uso de ponderaciones que ayuden a marcar la importancia de unos individuos frente a otros.

La estratificación consiste en dividir a los individuos en un número determinado de estratos, de manera que en cada uno de ellos se encuentren los individuos que tienen un Índice de Propensión semejante. Según explica Rubin, cuando aplicamos este método a un conjunto de datos, dotándole de 5 estratos, conseguimos reducir el sesgo en más de un 90% [Peter C. Austin (2011)].

El método de emparejamiento consiste en emparejar a los individuos del grupo de los tratados con uno o varios individuos del grupo de los controles, utilizando para ello diferentes medidas como pueden ser la distancia de Mahalanobis o el vecino más próximo, por ejemplo. Se descartan aquellos individuos que no hayan resultado emparejados.

El trabajo aquí presentado incluye los siguientes contenidos: en primer lugar se introducen los diseños observacionales y experimentales, que se diferencian entre sí en la forma de incluir a los individuos en la muestra y en la manera en la que a éstos se les asigna un tratamiento. A continuación se introduce el concepto de Índice de Propensión y se describe el modelo causal de Rubin, que da base teórica a dicho índice. Una vez explicados estos conceptos, se especifican los distintos pasos que deben seguirse para poder estimar el efecto del tratamiento. Por último, se presenta un ejemplo en el que se aplica paso a paso la técnica estudiada.

3. EL ÍNDICE DE PROPENSIÓN

El efecto que puede producir un determinado tratamiento o suceso sobre un individuo depende de variables predictoras o covariables que pueden influir sobre este dependiendo del diseño mediante el cual se obtuvieron los datos. Por lo tanto, debido a este factor influyente, en primer lugar habrá que determinar el diseño a través del cual se estimará el efecto del tratamiento.

Según sea el diseño escogido para hacer tanto la asignación de los individuos a la muestra como la asignación del tratamiento a los individuos de la misma, el efecto del tratamiento será estimable o no. Además, los resultados obtenidos podrían no ser generalizables al resto de la población por lo que los resultados no reflejarían el verdadero efecto que el tratamiento genera en la población. Según el tipo de diseño empleado, la diversas covariables existentes podrían o no influir en los resultados por lo que sería otro factor a tener en cuenta.

Por lo tanto en un principio se tendría que realizar un estudio de los distintos diseños aplicables y, según las características y desventajas que estos posean, elegir aquel mediante el cual la estimación del efecto del tratamiento sea correcta y los resultados obtenidos generalizables al resto de la población.

3.1. Diseños experimentales y observacionales

Los diseños experimentales constituyen aquellos diseños donde la inclusión de los individuos en la muestra se realiza de forma aleatoria y además la asignación o no asignación del tratamiento a los individuos seleccionados se realiza también de forma aleatoria. De este modo, los individuos que conforman el grupo a los que se les ha asignado un tratamiento están muy igualados en todos los factores o covariables a los individuos que conforman el grupo que no se les asigna tratamiento.

Debido a la aleatoriedad con la que se realiza tanto la asignación del tratamiento a los individuos como la inclusión de estos en la muestra y al hecho de que por ello las covariables observadas y no observadas se encuentran distribuidas de igual forma, la estimación del efecto del tratamiento será insesgada, lo que ocasiona que este tipo de

diseños experimentales sean ideales para realizar inferencias causales. Pero los diseños experimentales presentan diversos inconvenientes que dificultan su uso. Su correcta aplicación puede resultar demasiado costosa; la aleatorización puede violar algún principio ético e incluso puede producir resultados poco realistas que impidan la generalización a la población.

Los diseños observacionales son aquellos en los que el investigador simplemente se dedica a observar, y por tanto no existe aleatoriedad en las asignaciones de tratamiento o de los individuos a la muestra. De este tipo son los estudios de cohorte, estudios de casos y controles y los estudios de “corte transversal”.

Los estudios de cohorte son aplicados en estudios donde sea necesario determinar la incidencia y la historia natural de una enfermedad (evolución del proceso patológico sin intervención médica). Estos estudios pueden ser prospectivos, donde el investigador mide unas condiciones en los individuos y tras realizar un seguimiento comprueba si desarrollan el suceso de interés, o retrospectivos, en los cuales la investigación se plantea después de haberse producido el suceso de interés.

Este método de cohortes es ideal en situaciones donde la exposición de los individuos al agente sea poco ético por lo que los diseños experimentales no pueden ser aplicables. En estos estudios de cohortes es posible medir las posibles causas antes de observar la respuesta, por lo que se evita el debate sobre cuál es la causa y cuál es el efecto.

Según [J. Canela Soler y otro] cuando el suceso de interés es raro, el estudio de cohortes resulta ineficiente. Otro problema de estos estudios es la pérdida de algún individuo durante el seguimiento, lo que puede afectar significativamente en los resultados.

Los estudios transversales son utilizados principalmente para determinar la prevalencia, que es la proporción de individuos de un grupo que presentan una característica o evento determinado en un momento en el tiempo. Los estudios transversales también se utilizan para inferir sobre la causalidad. En un momento en el

tiempo los individuos se evalúan para determinar si fueron expuestos al agente en cuestión y si tienen la respuesta de interés.

Los estudios transversales poseen la ventaja de que los datos sólo se recogen una vez y no hay que realizar un seguimiento de los individuos, por lo que aplicar este tipo de diseños puede ser relativamente barato. Por otro lado en este tipo de estudios es difícil diferenciar en la mayoría de los casos la causa y el efecto de la simple asociación.

Por último están los estudios de casos y controles. Estos estudios suelen ser de tipo retrospectivo. Los individuos con el suceso de interés se comparan con un grupo control que no están sometidos al suceso. El investigador determina que los individuos fueron expuestos al agente, tratamiento o la prevalencia de una variable en cada uno de los grupos de estudio. Cuando el suceso es poco frecuente, los estudios de casos y controles pueden ser el único enfoque viable. En estas situaciones los estudios de casos y controles generan una gran cantidad de información en comparación con el reducido número de individuos de los que se disponen.

Los principales problemas en los diseños observacionales son la existencia de las variables de confusión y el sesgo. El sesgo con el que un investigador se puede encontrar es un sesgo de selección, debido a la forma por la que los individuos fueron incluidos en la muestra, y un sesgo de confusión, debido a las variables de confusión.

Como en muchas situaciones el único diseño aplicable para realizar el estudio de interés es el diseño observacional donde las estimaciones que se deseen obtener se pueden ver afectadas por la existencia de un sesgo de selección y de confusión, el investigador precisará de métodos que le permitan controlar y reducir en la medida de lo posible el sesgo que estos estudios observacionales generan cuando se desea obtener una estimación del efecto del tratamiento. Algunas de las técnicas existentes que permiten el control del sesgo serían el uso de más grupos de controles en el estudio, ajustes analíticos, estratificación o emparejamiento entre los controles y los tratados. Una técnica relativamente reciente del control del sesgo para la estimación del efecto del tratamiento está basada en el uso del Índice de Propensión.

3.2. Definición Índice de Propensión

El Índice de Propensión es una técnica estadística mediante la cual se trata de estimar el efecto de un determinado tratamiento en función de una serie de covariables medidas y realizar inferencias causales cuando no es factible realizar la asignación aleatoria de ese tratamiento a los individuos y por lo tanto se tienen que emplear diseños observacionales para obtener la muestra que se va a analizar. Con la aplicación de esta técnica sobre conjuntos de datos obtenidos mediante un diseño observacional, también se intenta reducir el sesgo de confusión que se puede encontrar en las estimaciones del efecto del tratamiento.

Esta técnica del Índice de Propensión fue descrita y publicada en 1983 por Paul Rosenbaum y Donald Rubin y emplea, como base para la estimación del efecto del tratamiento, la teoría del modelo causal para datos observacionales descrito por Donald Rubin. Aunque generalmente se emplea para la comparación de los individuos de dos grupos, también puede generalizarse a estudios con más de dos grupos.

Una definición estadística del Índice de Propensión es la siguiente:

Dado un individuo i , con valores de covariables X_i , al que se asocia una variable Z_i indicadora de la asignación de i al tratamiento, se define el Índice de Propensión $\lambda(X_i)$ como la probabilidad de que el individuo i sea asignado al tratamiento:

$$\lambda(X_i) = P(Z_i = 1 / X_i) \quad (1)$$

Si se dispone de n individuos $i = 1, \dots, n$, entonces:

$$P(Z_1 \dots Z_n / X_1 \dots X_n) = \prod_{i=1}^n (\lambda(X_i)^{Z_i} * (1 - \lambda(X_i))^{(1-Z_i)}) \quad (2)$$

donde:

- $Z_i = 1$ si el individuo i pertenece al grupo de los tratados y $Z_i = 0$ si el individuo i pertenece al grupo de los controles.
- X_i es el vector de covariables observadas en el individuo i -ésimo.

En los diseños experimentales la covariables no tienen ningún efecto en la asignación de los individuos al tratamiento. Si se aplica el Índice de Propensión en un conjunto de datos obtenido mediante un diseño experimental para comparar dos grupos, el Índice de Propensión que se obtendrá para cada uno de los individuos será de 0.5. Esto es debido a que cada individuo es asignado al tratamiento con una probabilidad del 50%. Cuando el estudio se realiza en diseños donde la asignación de los individuos a los grupos no se realiza de una forma aleatorizada, como los estudios observacionales, el Índice de Propensión es desconocido y debe ser estimado. El valor del Índice de Propensión depende de un vector de covariables observadas las cuales están asociadas a la asignación del tratamiento.

Como ejemplo de aplicación del Índice de Propensión, podemos observar uno de los ejemplos de la librería nonrandom [Susanne Stampf (2014)], librería construida para la aplicación de este índice. El conjunto de datos que vamos a emplear es el llamado pride. Este ejemplo completo se puede encontrar en [Susanne Stampf (2014)]. El código correspondiente a los análisis realizados se encuentra en el anexo 1 del trabajo.

Este conjunto de datos recoge la información sobre las infecciones respiratorias en vías bajas (LTRI) en 3078 bebés y niños menores de 3 años. Se trata de un estudio observacional que fue realizado en Alemania por el PRI.DE (Pediatric Respiratory Infection, Deutschland). Lo que se desea es conocer el efecto de tener infección RSV con la gravedad del LRTI teniendo en cuenta las covariables observadas.

En este estudio se recogieron las 16 variables que se muestran a continuación en la Tabla 1:

Variable	Descripción	Codificación o unidad de medida
PCR_RSV	Infección por el virus sincitial respiratorio corriente (RSV)	1:si
		0:no
Sexo	Sexo del niño	1: hombre
		2: mujer
Etnia	Etnia del niño	1: alemán
		2: unión europea
		3: otros
FRUEHG	Nacimiento del niño por parto prematuro	1:si
		0:no
RSVINF	Alguna infección anterior de RSV	1:si

		0:no
HERZ	Defecto congénito del corazón	1:si
		0:no
Región	Región alemana donde reside el niño	1: sur
		2: medio
		3: oeste
		4:norte
Edad	Edad del niño	En años
VOLLSTIL	Actualmente en lactancia o más de dos meses	1:si
		0:no
EINZ	Tiene hermanos	1:si
		0:no
Tabaco	Exposición al humo de tabaco pasivo en el hogar	1:si
		0:no
EXT	Atención externa	1:si
		0:no
ELTATOP	Atopia de los padres	1:si
		0:no
KRAMKSUM	Número de LRTI diagnosticados al niños	
SEVERE	Nivel de LRTI	1:grave
		0:no grave

Tabla 1. Variables del conjunto de datos pride.

Mediante este conjunto de datos se desea conocer el efecto de la presencia del RSV recogido en la variable PCR_RSV sobre la gravedad del LRTI, recogido en la variable SEVERE, ajustado por el resto de covariables observadas.

Como se trata de un diseño observacional, las covariables pueden influir en las estimaciones del efecto de la RSV ocasionando un sesgo de confusión, se necesita aplicar alguna técnica para corregirlo, que en este caso es la técnica del Índice de Propensión que se está estudiando.

El primer paso consiste en estimar el Índice de Propensión para los individuos del estudio. Se empleó el método de la regresión logística para estimarlo. En el Gráfico 1 se representan los resultados de las estimaciones obtenidas del Índice de Propensión separándolos por los grupos formados de tratamiento (individuos que tienen infección RSV) y de control o no tratados (individuos que no tienen infección RSV).

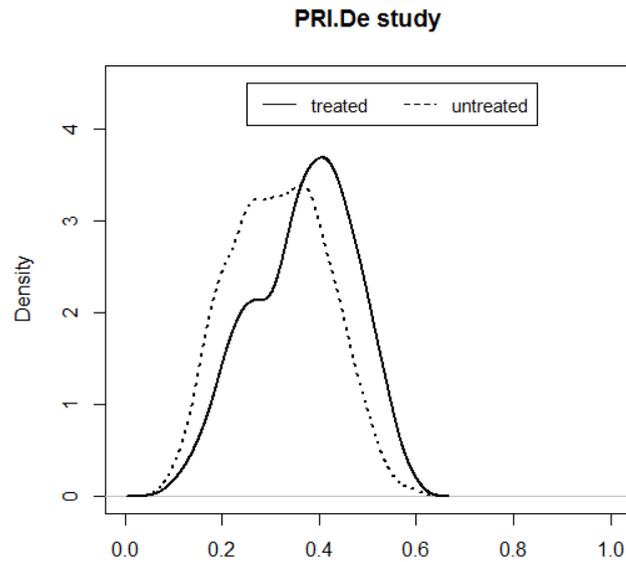


Gráfico 1. Distribuciones del índice de propensión del ejemplo pride.

Como se observa, hay bastante diferencia en el Índice de Propensión para los individuos ‘tratados’ y para los individuos ‘controles’. Esta diferencia, producida por el sesgo de confusión que introducen las covariables en el modelo, hace que no se puedan comparar ambos grupos.

Para reducir este sesgo de confusión el Índice de Propensión tiene varios métodos que se pueden seguir. El primero que se tratará es el camino del emparejamiento. Mediante esta técnica se emparejan individuos del grupo de los ‘tratados’ con individuos del grupo de los ‘controles’, descartando los individuos que queden desemparejados. Los resultados tras aplicar el emparejamiento son los que aparecen en la Tabla 2:

```

Matched by: ps

Matching parameter:

Caliper size:    0.102
Ratio:          1.000
Who is treated?: 1.000

Matching information:

Untreated to treated?: TRUE
Best match?:        TRUE

Matching data:

Number of treated obs.:          1031
Number of matched treated obs.: 1031
Number of untreated obs.:       2047
Number of matched untreated obs.: 1031
Number of total matched obs.:   2062
Number of not matched obs.:     1016
Number of matching sets:        1031
Number of incomplete matching sets: 0

```

Tabla 2. Emparejamiento del Índice de Propensión del ejemplo pride.

Existen varias técnicas de emparejamiento. En este ejemplo el emparejamiento que se empleó es un emparejamiento 1:1 con un caliper, valor considerado como valor máximo posible del valor absoluto de la diferencia de los índices de propensión en los individuos emparejados de cada grupo para mejorar la precisión del emparejamiento, de 0.102. Mediante el emparejamiento, se emparejaron los 1031 individuos del grupo de los ‘tratados’ con 1031 individuos de los ‘controles’ quedando sin emparejar y por lo tanto descartados los 1016 restantes individuos del grupo control.

Con este emparejamiento realizado, hay que comprobar si se consigue el equilibrio en las covariables entre los grupos de estudio. Al realizar el balance, los resultados son los que presenta la Tabla 3:

```

Summary of balance check:

                                before: no balance (0) before: balance (1)
after: no balance (0)           0                               0
after: balance (1)             3                               4

Covariates not completely tested: ---

Detailed balance check (overall):

                                SEX ETHNO FRUEHG RSVINF HERZ REGION AGE
table.before  1      1      1      0      1      0      0
table.after   1      1      1      1      1      1      1

Detailed balance check:
[Standardized differences (cut point: 10)]

                                SEX ETHNO FRUEHG RSVINF HERZ REGION AGE
Before      9.926 1.761  3.322 15.102 2.847 18.78 40.34
-----
After       0.39 4.472  1.123  1.665 4.998  4.602 4.572

-----
Scale      cat   cat   cat   cat   cat   cat   num

```

Tabla 3. Balance del emparejamiento del ejemplo pride.

Para comprobar el balance en las covariables, se emplea el método de las diferencias estandarizadas entre los grupos y se selecciona un punto límite de balance de 10. A este nivel, se puede observar que antes de aplicar la técnica de emparejamiento, 4 de las 7 covariables estaban balanceadas en ambos grupos. Tras aplicar el emparejamiento, se puede observar que todas las covariables han alcanzado el balance entre los grupos, como se destaca en el gráfico 2:

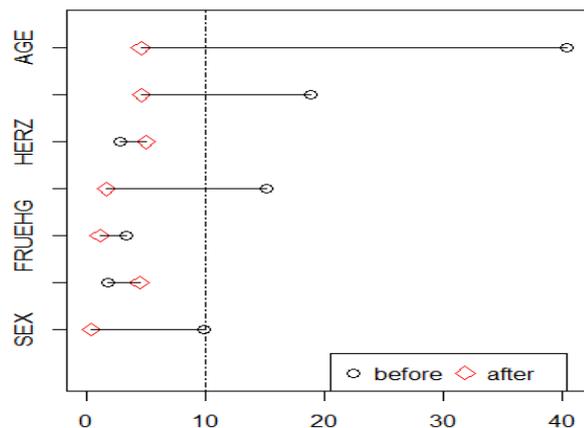


Gráfico 2. Diferencias estandarizadas de las covariables antes y después del emparejamiento ejemplo pride.

En el gráfico 2 se puede observar cómo tras aplicar el emparejamiento, las diferencias estandarizadas en cada una de las covariables se encuentran más próximas entre sí y más próximas al 0 que antes de aplicar el emparejamiento.

El ajuste de las covariables también se puede obtener mediante la técnica de estratificación en lugar de utilizar el emparejamiento. Si se emplea esta técnica el primer paso que hay que seguir es determinar el número de estratos que se van a construir y de qué forma. Se ha empleado una estratificación con 5 estratos. Estos estratos se construyen utilizando los cuantiles del Índice de Propensión de forma que en cada estrato haya un 20% de los individuos. La estratificación realizada es la que aparece en la Tabla 4:

```

Stratified by: ps

Strata information:

```

	Strata bounds	n	n (proportion)
1	[0.0624,0.236]	616	0.2
2	(0.236,0.306]	615	0.2
3	(0.306,0.369]	616	0.2
4	(0.369,0.431]	615	0.2
5	(0.431,0.594]	616	0.2

Tabla 4. Estratos en la estratificación del ejemplo pride.

Como se puede observar en cada uno de los estratos la proporción de individuos es del 20%, incluyendo entre 616 y 615 individuos cada uno.

Con la estratificación elegida y realizada, se puede proceder al balance de las covariables. Para comprobar el balance se emplea el método de las diferencias estandarizadas como se hizo en el balance mediante el emparejamiento. Para poder concluir que una covariable esta balanceada en los estratos, tiene que tener una diferencia estandarizada menor que el punto de corte límite marcado en todos los estratos. Los resultados del balance con un punto de corte de 20 se recogen en la Tabla 5:

Summary of balance check:

	before: no balance (0)	before: balance (1)
after: no balance (0)	1	2
after: balance (1)	0	3

Covariates not completely tested:
RSVINF

Detailed balance check (overall):

	SEX	ETHNO	FRUEHG	RSVINF	HERZ	REGION	AGE
table.before	1	1	1	1	1	1	0
table.after	1	0	1	NA	1	0	0

Detailed balance check (per stratum):
[Standardized differences (cut point: 20)]

	SEX	ETHNO	FRUEHG	RSVINF	HERZ	REGION	AGE
Before	9.926	1.761	3.322	15.102	2.847	18.78	40.34
-----	-----	-----	-----	-----	-----	-----	-----
Stratum 1	11.12	9.506	14.181	5.814	0.292	33.492	2.043
Stratum 2	13.906	36.838	1.596	9.599	6.792	27.774	16.367
Stratum 3	3.047	19.495	4.717	0.913	7.779	19.294	7.075
Stratum 4	13.512	13.614	3.836	NA	3.317	13.267	25.547
Stratum 5	1.073	4.336	7.723	NA	7.013	17.767	41.37
-----	-----	-----	-----	-----	-----	-----	-----
Scale	cat	cat	cat	cat	cat	cat	num

Tabla 5. Balance de la estratificación del ejemplo pride.

En la Tabla 5 se observa que antes de aplicar la estratificación todas las variables excepto la edad estaban balanceadas. Tras aplicar la estratificación, se pierde el balance en la variable etnia por la diferencia en el segundo estrato, en la variable región por las diferencias en los dos primeros estratos y en la variable RSVINF porque no se dispone de información para los estratos cuatro y cinco. La variable edad sigue sin conseguir el balance.

Comparando los resultados que se han obtenido en los dos métodos de balance del Índice de Propensión, se puede observar que el método que nos proporciona un mejor balance es la técnica de emparejamiento dado que mediante su aplicación se consiguió equilibrar todas las covariables mientras que mediante la estratificación solo se consiguió el balance final en tres de las siete covariables.

El último paso en el proceso del Índice de Propensión es estimar el efecto del tratamiento. La primera estimación del efecto del tratamiento se obtendrá a partir del balance mediante el emparejamiento. En los resultados se mostrará el efecto del tratamiento antes de aplicar ninguna técnica y después de aplicar el emparejamiento.

Este efecto se medirá con la razón de ventajas u odds ratio. El efecto del tratamiento se estimará sobre los individuos emparejados. Este efecto puede ser diferente para los individuos descartados por el emparejamiento.

Para realizar el cálculo de las OR se emplea la tabla 2x2 que surge de cruzar la variable respuesta SEVERE con la variable PCR_RSV que recoge el “tratamiento”.

Las estimaciones del efecto del “tratamiento” a través del balance por el emparejamiento se puede observar en la tabla 6:

```

Effect estimation for treatment/exposure on outcome

Treatment/exposure: PCR_RSV
Outcome: SEVERE
Effect measure: odds ratio ('or')

Table of effect estimates:

      or  SE[log[or]]  [95%-CI[or]]
-----
Crude  1.677      0.0796  [1.435,1.96]
Matching
  Unadjusted  1.379      0.0916  [1.152,1.65]
  Adjusted                [,]
Regression
  Conditional                [,]
  Marginal                  [,]

```

Tabla 6. Estimación del efecto del tratamiento por el emparejamiento ejemplo pride.

El efecto del “tratamiento” cuando se estima no teniendo en cuenta las covariables es de 1.677, lo que indica que la “ventaja” de que un niño padezca LTRI severo es 1.677 veces mayor cuando el niño padece una infección RSV frente a no padecerla. El intervalo de confianza construido para esta estimación no incluye al valor 1 por lo que se puede concluir que, con una confianza del 95%, padecer una infección RSV aumenta la probabilidad de padecer un LRTI severo.

Cuando se aplica el emparejamiento, se reduce el sesgo de confusión y de selección y la estimación del efecto en esta situación será más precisa que la obtenida anteriormente. El efecto del “tratamiento” cuando se estima teniendo en cuenta el resto de covariables y tras el emparejamiento es de 1.379, lo que indica que la “ventaja” de que un niño padezca LTRI severo es 1.379 veces mayor cuando el niño padece una infección RSV frente a no padecerla. El intervalo de confianza construido para esta estimación no incluye al valor 1 por lo que se puede concluir que, con una

confianza del 95%, padecer una infección RSV aumenta la probabilidad de padecer un LRTI severo.

La segunda estimación del efecto del tratamiento se obtendrá a partir del balance mediante la estratificación. Las estimaciones del efecto del “tratamiento” a través del balance por la estratificación se puede observar en la tabla 7:

```

Effect estimation for treatment/exposure on outcome

Treatment/exposure: PCR_RSV
Outcome: SEVERE
Effect measure: odds ratio ('or')

Table of effect estimates:

              or  SE[log[or]]  [95%-CI[or]]
-----  -
Crude      1.677      0.0796  [1.435,1.96]
Stratification
  Outcome rates  1.362      0.0805  [1.163,1.595]
  MH            1.419      0.0823  [1.208,1.667]
  Adjusted     1.568      0.2017  [1.056,2.328]
Regression
  Conditional
  Marginal

```

Tabla 7. Estimación del efecto del tratamiento por la estratificación para el ejemplo pride.

Cuando se aplica la estratificación, se puede reducir el sesgo de confusión y de selección. Por ello, la estimación del efecto en esta situación será más precisa que la obtenida anteriormente. Para estimar el efecto del “tratamiento” cuando se estima teniendo en cuenta el resto de covariables y tras la estratificación se emplea el test de Mantel-Haetszen (test M-H) debido a que se disponen de 5 tablas 2x2 (1 por cada estrato) donde se cruzan la variable respuesta SEVERE con la variable “tratamiento” PCR_RSV. La estimación del efecto mediante el test de M-H es de 1.419, lo que indica que la “ventaja” de que un niño padezca LRTI severo es 1.419 veces mayor cuando el niño padece una infección RSV frente a no padecerla. El intervalo de confianza construido para esta estimación no incluye al valor 1 por lo que se puede concluir que, con una confianza del 95%, padecer una infección RSV aumenta la probabilidad de padecer un LRTI severo. En esta estimación se asume que todas las covariables se equilibraron tras aplicar la estratificación, pero como se pudo observar en la tabla 5,

las variables etnia, RSVINF, región y edad no se equilibraron por lo que habría que obtener la estimación teniendo en cuenta estas covariables no balanceadas.

Para obtener la estimación teniendo en cuenta estas covariables, se emplea también el test M-H. La estimación del efecto mediante el test de M-H es de 1.568, lo que indica que la “ventaja” de que un niño padezca LTRI severo es 1.568 veces mayor cuando el niño padece una infección RSV frente a no padecerla. El intervalo de confianza construido para esta estimación no incluye al valor 1 por lo que se puede concluir que, con una confianza del 95%, padecer una infección RSV aumenta la probabilidad de padecer un LRTI severo.

Como se puede observar en este ejemplo, las estimaciones del efecto obtenidas a partir del emparejamiento y las obtenidas a partir de la estratificación son bastante parecidas. A través de ambos métodos, se puede concluir que, con una confianza del 95%, padecer una infección RSV aumenta la probabilidad de padecer un LRTI severo.

Debido a que la técnica del Índice de Propensión trata de estimar el efecto de un tratamiento mediante inferencias causales basadas en el modelo causal de Rubin para datos observacionales, a continuación se introducirá dicho modelo.

3.3. Modelo causal de Rubin

El modelo causal de Rubin es una de las bases principales en la aplicación del Índice de Propensión sobre el conjunto de datos objeto de estudio. Donald Rubin presenta su modelo causal en 1983, basándose en el modelo construido por Jerzy Neyman en su tesis sobre experimentos aleatorios desde el punto de vista no paramétrico en el año 1923 [Jasjeet S. sekhon (2007)]. Por esta razón el modelo causal de Rubin también es conocido como el modelo causal de Neyman-Rubin. Según Neyman, cada individuo tiene dos resultados potenciales, uno si al individuo se le asignó el tratamiento y otro si pertenece a los controles. El efecto causal viene dado por las diferencias entre estos dos resultados potenciales, aunque solo uno de ellos es observado. A partir de esta idea, Rubin propone su modelo causal [Jasjeet S. sekhon (2007)]:

“Sea Y_{i1} el resultado potencial del individuo i cuando este individuo recibió el tratamiento cuyo efecto se quiere observar y sea Y_{i0} el resultado potencial del individuo i si fuera control, entonces el efecto del tratamiento en el individuo i viene definido por:

$$\tau_i = Y_{i1} - Y_{i0}.$$

Cuando existen datos faltantes la inferencia causal es un problema porque los resultados potenciales Y_{i1} o Y_{i0} no se observan.

Para poder avanzar es necesario desarrollar una serie de hipótesis. Las mejores hipótesis que se pueden realizar serían las elaboradas sobre un diseño experimental. Si Z_i es un indicador de tratamiento, valiendo 1 cuando i está en régimen de tratamiento y 0 en caso contrario, el resultado observado para el individuo i es entonces:

$$Y_i = Z_i * Y_{i1} + (1 - Z_i) * Y_{i0}$$

Rubin propuso su modelo causal para los dos diseños más comunes a partir de los cuales pueden extraerse los datos para el estudio, los diseños experimentales y observacionales.

Para el diseño experimental, realizar la inferencia causal sobre los individuos es sencillo debido a la inclusión aleatoria de los individuos en la muestra y la asignación aleatoria del tratamiento a estos individuos. La asignación del tratamiento es independiente de todas las variables por lo que el efecto del tratamiento no se ve afectado por la acción de otras covariables.

Para estimar el efecto del tratamiento en esta situación se usa generalmente el ATE (Average Treatment Effects o efecto del tratamiento promedio). El ATE lo podemos calcular por:

$$ATE = E(Y_i | Z_i = 1) - E(Y_i | Z_i = 0) = E(Y(1)) - E(Y(0)) \quad (3)$$

Para obtener (3) se calcula la diferencia del valor esperado del individuo i cuando este individuo está asignado al grupo del tratamiento ($Z_i = 1$) con el valor esperado del individuo cuando es asignado al grupo de los controles ($Z_i = 0$). Es útil para evaluar cuál es el efecto esperado del resultado si los individuos de la población son asignados aleatoriamente al tratamiento.

Uno de los problemas más comunes que se puede plantear en la inferencia causal es la “cuestión del conformidad” [Jasjeet S. sekhon (2007)]. Las personas a las que se asigna un tratamiento pueden rechazarlo y las personas asignadas al grupo control pueden encontrar la manera de recibir el tratamiento. Cuando este problema aparece, la ecuación (3) se denomina Estimador de la Intención de Tratamiento (ITT). Aplicar este nuevo estimador para estimar el efecto promedio del tratamiento puede ser difícil.

Por otro lado, cuando el diseño es observacional, donde no existe aleatoriedad en las asignaciones de tratamiento o de los individuos a la muestra, la aplicación del modelo causal presenta dificultades debido a que generalmente los individuos a los que se les ha aplicado el tratamiento no proceden de la misma población que los individuos que pertenecen al grupo control [Jasjeet S. sekhon (2007)]. Para calcular el efecto del tratamiento ya no se emplea el ATE sino que se emplea el ATT (Average Treatment Effect for the Treated o efecto promedio del tratamiento para el tratado):

$$ATT = E(Y_{i1} | Z_i = 1) - E(Y_{i0} | Z_i = 1) \quad (4)$$

La ecuación (4) no se puede calcular directamente dado que para los individuos tratados no se ha observado Y_{i0} . Además, para el proceso de estimación hay que tener en cuenta que la asignación del tratamiento a los individuos depende de una serie de covariables observadas, X . Al condicionar por X , se puede asumir que la asignación del tratamiento es independiente de Y_1 y Y_0 , y que hay un solapamiento (*para todo* $x \in X$; $0 < P(T = 1 | X = x) < 1$). Este solapamiento, junto a la independencia de Y_1 y Y_0 con la asignación del tratamiento, constituyen una propiedad necesaria para identificar el efecto del tratamiento conocido como “propiedad fuerte de ignorabilidad de la asignación”, necesaria para determinar el efecto del tratamiento.

Al condicionar por las covariables X_i observadas, los grupos de control y tratamiento estarían en equilibrio. En este caso ATT lo podemos calcular haciendo uso de la siguiente expresión:

$$ATT = E(E(Y_{i1}| X_i, Z_i = 1) - E(Y_{i0}|X_i, Z_i = 0) | Z_i = 1) \quad (5)$$

A partir de (5) se puede calcular el efecto del tratamiento, puesto que emplea para ello las variables de referencia del grupo de tratamiento. Además, comparando en esta ecuación los resultados por el estado del tratamiento en los individuos se obtiene una estimación sesgada del mismo. Este sesgo se produce como resultado de la diferencia que existe entre los tratados y los que juegan el papel de controles y se denomina sesgo de selección [Leonardo Grilli y otros (2011)]. Será:

$$\begin{aligned} ATT &= E(E(Y_{i1}| X_i, Z_i = 1) - E(Y_{i0}|X_i, Z_i = 0) | Z_i = 1) \\ &\Rightarrow E(Y_{i1}| X_i, Z_i = 1) - E(Y_{i0}| X_i, Z_i = 1) \\ &\quad + [E(Y_{i0}| X_i, Z_i = 1) - E(Y_{i0}| X_i, Z_i = 0)] = ATE + sesgo \quad (6) \end{aligned}$$

Hay que tener en cuenta que al calcular el valor del ATT cuando los individuos se ponderan, este se ve afectado. Si existen covariables que solo son observadas para individuos del grupo control, éstas serán irrelevantes en el cálculo del ATT y por tanto serán descartadas. La suposición del solapamiento para obtener el valor del ATT requiere que, el soporte de las covariables X para los individuos del grupo de tratamiento, sean un subconjunto del soporte de X para los individuos del grupo control [Jasjeet S. sekhon (2007)]. Podemos encontrarnos situaciones en las que sea necesario eliminar ciertos individuos del grupo al que se les aplica el tratamiento cuando estos poseen valores en las covariables para los cuales no exista solapamiento con los individuos del grupo control. Al eliminar los individuos que están fuera del soporte no solo se reduce el sesgo, sino también se puede reducir la varianza de las estimaciones.

3.4. Aplicación del Índice de Propensión

Existen varios métodos que permiten estimar el efecto de un determinado tratamiento teniendo en cuenta una serie de covariables observadas para cada uno de los individuos que conformen el conjunto de datos. Uno de estos métodos es el Índice de Propensión.

La aplicación del Índice de Propensión en un conjunto de datos consta de tres pasos. El primer paso consiste en estimar el Índice de Propensión, cuyas estimaciones obtenidas pueden estar afectadas por la existencia de un sesgo de selección o de confusión. El segundo paso consiste en reducir en la medida de lo posible el sesgo de selección y de confusión. El tercer paso es estimar del efecto del tratamiento.

3.4.1. Estimación del Índice de Propensión

Se dispone de varios métodos de estimación del Índice de Propensión, como son el método de regresión logística, el método de Árboles de Clasificación y Regresión, el método de la regresión de Cox o el método de Redes Neuronales. El método más empleado para estimar el Índice de Propensión es la regresión logística, debido a su simplicidad. En este trabajo nos centraremos en este método de la regresión logística.

3.4.1.1. Regresión logística

El método de regresión logística se emplea para estimar la probabilidad de que un individuo i se encuentre en el grupo de los tratados (individuos que reciben tratamiento), teniendo en cuenta las covariables observadas. Cuando la variable respuesta consta de dos grupos, que son grupo de tratamiento y grupo de controles (individuos que no reciben tratamiento), se utiliza lo que se conoce como regresión logística binomial. Si por el contrario, la variable tiene k categorías con $k > 2$, siendo una de ellas el grupo de tratados y las $k - 1$ restantes grupos de controles, se utiliza una regresión logística multinomial.

Para estimar la probabilidad de que un individuo i reciba el tratamiento se empleará la función enlace “logit” que, permitirá construir una relación lineal, y estimar a partir de ella dicha probabilidad. A continuación mostramos la relación de ecuaciones (9) y (10).

Partiendo de la función enlace “logit” linealizamos el valor esperado de la variable respuesta.

$$\lambda_i = E(Z_i = 1/X_i) = \frac{e^{(\alpha + \beta^t * X_i)}}{1 + e^{(\alpha + \beta^t * X_i)}} \quad (9)$$

$$\text{logit}(\lambda_i) = \ln\left(\frac{\lambda_i(X_i)}{1 - \lambda_i(X_i)}\right) = \ln\left(\frac{P(Z_i=1/X_i)}{1 - P(Z_i=1/X_i)}\right) = \alpha + \beta_k^t X_i \quad (10)$$

α Es el intercept o termino independiente.

β_k Son los coeficientes de regresión

$Z_i = 1$ si el individuo i pertenece al grupo de los tratados y $Z_i = 0$ si el individuo i pertenece al grupo control.

X_i es el vector de covariables observadas en el individuo i -ésimo.

$\lambda_i(X_i)$ es el Índice de Propensión para el individuo i -ésimo.

Por lo tanto de la siguiente forma obtendremos los estimadores del Índice de Propensión:

$$\hat{\lambda}(X_i) = \frac{e^{(\hat{\alpha} + \hat{\beta}^t * X_i)}}{1 + e^{(\hat{\alpha} + \hat{\beta}^t * X_i)}} \quad (11)$$

3.4.2. Técnicas de reducción de sesgo y estimación del efecto del tratamiento

Tras la estimación del Índice de Propensión, mediante la regresión logística u otro de los métodos alternativos nombrados, el siguiente paso es reducir en la medida de lo posible los sesgos de confusión y selección que puedan tener las estimaciones del Índice de Propensión mediante alguno de los métodos disponibles.

Algunos de estos métodos son la estratificación, en la cual se construyen un número de subgrupos a partir de las estimaciones del Índice de Propensión para clasificar a los individuos con valores del Índice de Propensión similares; el emparejamiento, donde se emparejará individuos del grupo de tratamiento con individuos del grupo control a partir de los valores estimados del Índice de Propensión; y la ponderación, donde cada individuo se pondera de una forma u otra según

pertenezca al grupo control o al grupo tratamiento teniendo en cuenta el valor estimado del Índice de Propensión. Este último método siempre se emplea junto con los otros dos métodos. Tras aplicarlos, se puede estimar el efecto del tratamiento.

3.4.2.1. Estratificación

El método conocido como estratificación es uno de los métodos utilizados generalmente para intentar reducir el sesgo que obtenemos cuando nos vemos obligados a estimar el efecto de que produce un determinado tratamiento cuando el diseño es observacional. Mediante el uso de este método se trata de clasificar en un número determinado de subclases o estratos homogéneos los individuos por el valor del Índice de Propensión obtenido con anterioridad para cada uno de ellos.

Según [Ismini Savvala (2008)] es importante en primer lugar cuantificar el sesgo inicial de las variables de confusión. Los métodos populares para ello son el test t, ANOVA, test de diferencias de proporciones en dos muestras, regresión logística,... Para observar el desequilibrio inicial que puedan presentar los grupos de tratamiento y de control, se pueden utilizar el test B de la diferencia o el test R de la relación de la varianza. El test de la diferencia es el siguiente:

$$B = \frac{\mu_t - \mu_c}{\sqrt{\frac{\sigma_t^2 - \sigma_c^2}{2}}} \quad (12)$$

Donde μ_t y μ_c son las medias de cada factor de confusión para el grupo del tratamiento y grupo de los controles respectivamente. De igual forma σ_t y σ_c con las desviaciones típicas de cada factor para estos grupos. Cuanto más cercanas son las medias de los grupos, más equilibrados serán estos respecto a las covariables de interés. La otra medida utilizada, la relación entre las varianzas de los grupos de tratamiento y de control, se calcula de la siguiente manera:

$$R = \frac{\sigma_t^2}{\sigma_c^2} \quad (13)$$

Cuanto más próximo a 1 sea el valor que se obtiene a través de este estadístico, mayor equilibrio habrá entre los dos grupos del estudio.

Tras estimar el Índice de Propensión se comprueba que existe un solapamiento adecuado entre las estimaciones de los grupos. Los individuos fuera del rango de

solapamiento son descartados. Los individuos restantes se dividen en grupos de igual tamaño, denominados estratos, basados en la distribución de las estimaciones del Índice de Propensión. Según Cochran [Peter C. Austin (2011)], el número óptimo de estratos en los que clasificar estos individuos restantes sería el de cinco estratos. Si estos cinco estratos están bien contruidos, el sesgo que obtenemos por las covariables que influyen en el efecto del tratamiento se vería reducido en un 90%. Aunque la estratificación óptima es de cinco estratos, en algunos casos los datos precisan la utilización de un número menor de estratos.

El siguiente paso consiste en medir el desequilibrio en cada uno de los estratos contruidos. Los métodos que se emplearán para medir el desequilibrio serán tanto el estadístico B como el estadístico R dentro de cada estrato y el valor medio de estos estadísticos en cada uno de los estratos será comparado con los valores iniciales de estos estadísticos.

La estratificación combina la información en todos los grupos para proporcionar una estimación precisa del efecto del tratamiento para toda la población. En este sentido la estandarización es un procedimiento estratificado donde la población estándar provee la base para combinar la información de todos los estratos.

Cuando se logra el equilibrio, el investigador calcula dentro de cada estrato el efecto promedio del tratamiento restando el resultado promedio de los individuos tratados de los individuos de control. La estimación general del efecto del tratamiento se calcula promediando estos números en los estratos:

$$\widehat{ATT} = \sum_{k=1}^s \frac{n_k}{N} (\hat{Y}_{tk} - \hat{Y}_{ck}) \quad (14)$$

Donde n_k es el número de individuos de los grupos de tratamiento y control en el estrato k, N es el número total de individuos, \hat{Y}_{tk} es el resultado promedio de los índices de propensión para los individuos del grupo de tratamiento en el estrato k y \hat{Y}_{ck} es el resultado promedio de los índices de propensión para los individuos del grupo de control en el estrato k.

El error estándar estimado del efecto del tratamiento mediante la diferencia de medias es:

$$\hat{s}(\widehat{ATT}) = \sqrt{\sum_{k=1}^s \frac{n_k^2}{N^2} \left(\frac{s_{tk}^2}{n_{tk}} + \frac{s_{ck}^2}{n_{ck}} \right)} \quad (15)$$

En el caso de que el resultado fuera ordinal y el tratamiento tenga dos niveles, el investigador tiene dos formas de estimar el efecto del tratamiento, diferencia de proporciones y odds ratio. En el caso de las diferencia de proporciones dentro de cada estrato, se resta las proporciones de los individuos tratados de los de control. Una estimación global del efecto del tratamiento se calcula promediando las diferencias en los estratos:

$$\widehat{ATT} = \sum_{k=1}^S \frac{n_k}{N} (\hat{p}_{tk} - \hat{p}_{ck}) \quad (16)$$

El error estándar estimado de la diferencia estimada en proporciones es:

$$\hat{s}(\widehat{ATT}) = \sqrt{\sum_{k=1}^S \frac{n_k^2}{N^2} \left(\frac{\hat{p}_{tk}\hat{q}_{tk}}{n_{tk}-1} + \frac{\hat{p}_{ck}\hat{q}_{ck}}{n_{ck}-1} \right)} \quad (17)$$

Donde $q_{jk} = 1 - p_{jk}$ para $j = c, t$.

Para estimar el efecto del tratamiento a través de las odds ratio o razón de ventajas se aplicará el estimador de mantel-Haenszel (M-H). La notación general aplicada será que para cada uno de los k niveles de los factores de confusión ($k=1,2,\dots, K$) y para cada uno de los estratos es la recogida en la Tabla 8:

Grupos	Expuesto	No expuesto	Total
tratamiento	a_k	b_k	$a_k + b_k$
Control	c_k	d_k	$c_k + d_k$
Total	n_k	m_k	N_k

Tabla 8. Tabla 2x2 para cada estrato en el estimador M-H

De la anterior tabla podremos calcular el estimador del efecto del tratamiento:

$$\widehat{ATT} = \frac{\sum_{k=1}^S (a_k d_k) / N_k}{\sum_{k=1}^S (b_k c_k) / N_k} \quad (18)$$

Este método de la estratificación posee como todos los métodos, algunas limitaciones. La mayor limitación que presenta dicho método es que cuando el número de covariables existentes en el modelo aumenta, el número de estratos a formar aumenta de forma exponencial. Esto conlleva a que resulte más difícil crear las capas que nos permitan la comparación entre los individuos del grupo de tratamiento con los individuos del grupo de los controles.

3.4.2.2. Emparejamiento

Este método de emparejamiento consiste en emparejar uno o varios individuos pertenecientes al grupo de tratamiento con uno o varios individuos del grupo control según alguna característica específica.

El emparejamiento se aplica sobre los valores estimados del Índice de Propensión. Existen diferentes tipos de emparejamiento [Barbara Sianesi (2001)]. Alguno de estos serían el emparejamiento mediante el uso de la distancia de Mahalanobis, el emparejamiento por el vecino más cercano, emparejamiento estratificado,...

Tras obtener el equilibrio mediante el emparejamiento, el investigador podrá calcular el efecto promedio del tratamiento restando el resultado promedio de los individuos tratados de los individuos de control:

$$\widehat{ATT} = \frac{1}{n} \sum_{i=1}^n Z_i * Y_i - \frac{1}{n} \sum_{i=1}^n (1 - Z_i) * Y_i \quad (19)$$

Donde Z_i es una variable indicador que vale 1 si el individuo pertenece al grupo de los tratados y vale 0 si pertenece al grupo de los controles, n es el número de individuos emparejados y las Y_i son los individuos emparejados, tanto tratados como controles.

En el caso de que el resultado fuera ordinal y el tratamiento tenga dos niveles, el investigador puede emplear para estimar el efecto del tratamiento las odds ratio.

Para estimar el efecto del tratamiento a través de las odds ratio o razón de ventajas se construirá una tabla 2x2 como la tabla 9 tras el emparejamiento.

Grupos	Expuesto	No expuesto	Total
tratamiento	a	b	$a + b$
Control	c	d	$c + d$
Total	n	m	N

Tabla 9. Tabla 2x2 para estimar efecto del tratamiento en emparejamiento

Donde a, b, c, d son el número de individuos con esas condiciones. El estimador será por lo tanto:

$$\widehat{ATT} = \frac{(a*d)}{(b*c)} \quad (20)$$

3.4.2.2.1. Emparejamiento empleando la distancia Mahalanobis

Este método de emparejamiento aplica la medida de la distancia de Mahalanobis para medir las distancias entre los individuos del grupo de tratamiento con los individuos del grupo de los controles basándose en el Índice de Propensión estimado [Propensity score matching in observational data]. La distancia de Mahalanobis se calcula de la siguiente forma:

$$D_{ij} = \sqrt{(\hat{\lambda}(X_i) - \hat{\lambda}(X_j))^t * S^{-1} * (\hat{\lambda}(X_i) - \hat{\lambda}(X_j))} \quad (21)$$

donde $\hat{\lambda}(X_i)$ es el valor estimado del Índice de Propensión de los individuos del grupo de los tratamientos, $\hat{\lambda}(X_j)$ es el valor estimado del Índice de Propensión de los individuos del grupo control y S es la matriz de covarianza muestral de X [Jasjeet S. Sekhon (2007)].

Mediante esta distancia de Mahalanobis emparejamos a los individuos de los grupos de tratamiento y control de modo que su distancia de Mahalanobis sea menor. Se repite el procedimiento para todos los individuos del grupo de tratamiento. Si alguno de los individuos de ambos grupos no han sido emparejados serán, como decía, eliminados del estudio.

3.4.2.2.2. Emparejamiento por vecino más próximo

Este método de emparejamiento calcula la distancia mínima entre los valores estimados del Índice de Propensión de los grupos de tratamiento y control. Este método trata de emparejar los individuos del grupo de tratamiento y de control de forma que la diferencia absoluta entre los estimadores del Índice de Propensión sea la mínima que se pueda obtener. Esta sería la siguiente:

$$C(\hat{\lambda}(X_i)) = \min_j |\hat{\lambda}(X_i) - \hat{\lambda}(X_j)| \quad (22)$$

Donde $\hat{\lambda}(X_i)$ es el valor estimado del Índice de Propensión de los individuos del grupo de los tratamientos, $\hat{\lambda}(X_j)$ es el valor estimado del Índice de Propensión de los individuos del grupo control y $C(\hat{\lambda}(X_i))$ es el valor de la mínima distancia absoluta que se obtiene para el individuo i del grupo de los tratados.

En un primer lugar se ordenan de forma aleatoria los individuos. A continuación, al primer individuo del grupo de los tratados se le aplica la ecuación (22) para encontrar el individuo del grupo de los controles con el cual la diferencia absoluta de los valores del Índice de Propensión sea la mínima y con el cual quedará emparejado. Se repite el procedimiento para todos los individuos del grupo de tratamiento, comparándolos con los individuos del grupo control que todavía no han sido

emparejados. Si alguno de los individuos de ambos grupos no han sido emparejados serán, como decía, eliminados del estudio.

3.4.2.2.3. Emparejamiento mediante calipers

El método de emparejamiento mediante calipers es un método que se puede aplicar tanto por sí mismo para emparejar los individuos de los grupos del tratamiento con los individuos del grupo de los controles, como junto a otros métodos de emparejamiento para que la calidad de dicho emparejamiento entre los grupos mejore considerablemente.

Para poder aplicar este método, se establece un rango de valores e , (generalmente se considera como un cuarto del error estándar de la estimación de la propensión obtenida) y se considera que el valor absoluto de la diferencia de los Índices de propensión estimados de los individuos emparejados tiene que ser menor que este valor e :

$$|\lambda(X_i) - \lambda(X_j)| < e \quad (23)$$

Donde $\hat{\lambda}(X_i)$ es el valor estimado del Índice de Propensión de los individuos del grupo de los tratamientos, $\hat{\lambda}(X_j)$ es el valor estimado del Índice de Propensión de los individuos del grupo control.

Si la diferencia absoluta para algún individuo no cumple la condición (23) y por lo tanto estén fuera del rango considerado, estos individuos se eliminan del modelo.

3.4.2.2.4. Emparejamiento por emparejamiento N:N

Este tipo de emparejamiento consiste en emparejar más de un individuo por cada grupo. Los emparejamientos más habituales son los del tipo 1:1, 1:N, N:1, N:N.

El emparejamiento del tipo 1:1 es el emparejamiento clásico donde un individuo del grupo control es emparejado con un individuo del grupo de tratamiento. El emparejamiento del tipo 1:N es un emparejamiento especial donde 1 individuo del grupo control se empareja con N individuos del grupo de los tratados. El emparejamiento N:1 empareja N individuos del grupo control con 1 individuo del grupo de los tratados. Por último, el emparejamiento N:N empareja N individuos del grupo de los controles con N individuos del grupo de los tratados.

3.4.2.2.5. Emparejamiento por Kernel

El emparejamiento por kernel, [Marco Caliendo y otros (2005)] y [Propensity score matching in observational data], es un método no paramétrico en donde se

utilizan los promedios ponderados de todos los individuos del grupo de los controles para construir el emparejamiento.

Una de las ventajas que se tiene al aplicar este método de emparejamiento en lugar de otros métodos paramétricos, es que la varianza obtenida es más baja que la de otros métodos.

Con este método, los individuos del grupo de los tratados se emparejan con una media ponderada de los individuos del grupo control. Los pesos que se apliquen serán inversamente proporcionales a las distancias de los índices de propensión entre los individuos del grupo de los tratados con los individuos del grupo de los controles.

Cuando se aplica el emparejamiento por kernel, se tiene que elegir la función núcleo y el parámetro de ancho de banda [Marco Caliendo y otros (2005)]. La elección del parámetro de ancho de banda es muy importante. Un parámetro de ancho de banda con un alto rendimiento produce una estimación suave de la función de densidad, consiguiendo un mejor ajuste y un decrecimiento de la varianza entre la verdadera función de densidad subyacente y la densidad estimada. Por otro lado, puede suavizar demasiado las características subyacentes de la función de densidad, conduciendo a una estimación sesgada. El parámetro de ancho de banda elegido tiene que mantener un equilibrio entre el decrecimiento de la varianza y el sesgo estimado con la función de densidad real.

3.4.2.2.6. Emparejamiento por estratificación

Es una técnica de emparejamiento mediante la cual los individuos, antes de ser emparejados, son distribuidos según el valor estimado del Índice de Propensión en diferentes estratos. En cada uno de estos estratos construidos se aplican técnicas de emparejamiento entre los individuos de dichos estratos para emparejar individuos tratados con individuos del grupo control.

3.4.2.2.7. Comparación de los métodos de emparejamiento del Índice de Propensión.

En general no existe ningún método que sea más eficaz en todas las situaciones posibles frente al resto de métodos. Cada método indicado en los puntos anteriores funciona de forma más eficiente cuando estos son aplicados bajo ciertas circunstancias.

Según [Propensity score matching in observational data], en situaciones en donde el número de individuos del grupo control es pequeño, el método más adecuado y más eficaz en dar resultados al aplicarlo es el método de emparejamiento con reemplazamiento. Cuando se dispone de un gran conjunto de individuos en el

grupo de los controles en comparación con el número de individuos en el grupo de los tratados, el emparejamiento más apropiado para aplicar en estas situaciones sería el emparejamiento 2:1 (donde se emparejan 2 individuos del grupo control con 1 individuos del grupo de tratamiento) o $N:1$ (donde se emparejan N individuos del grupo control con 1 individuos del grupo de tratamiento). En situaciones donde el número de individuos control es grande y la distribución es asimétrica, los emparejamientos más adecuados son el emparejamiento por kernel y el emparejamiento por la distancia de Mahalanobis.

3.4.2.3. Ponderación del Índice de Propensión.

La ponderación en el Índice de Propensión se emplea en situaciones donde se desea que los individuos sean más representativos de la población. La ponderación considerada para los individuos del grupo de tratamiento es el inverso del valor estimado del Índice de Propensión para ese individuo. La ponderación para los individuos del grupo control es el inverso de uno menos el valor estimado del Índice de propensión para dicho individuo. Las ponderaciones serán:

$$w_{t;i} = \frac{1}{\hat{\lambda}(X_i)} \quad (24)$$

$$w_{c;i} = \frac{1}{1-\hat{\lambda}(X_i)} \quad (25)$$

Donde $w_{t;i}$ es la ponderación para el individuo i perteneciente al grupo de tratamiento, $w_{c;i}$ es el peso para el individuo i perteneciente al grupo control y $\hat{\lambda}(X_i)$ es el valor estimado del Índice de Propensión.

La ponderación del Índice de Propensión se emplea junto a uno de los otros dos métodos. Como en los otros métodos, tras realizar el ajuste, se procede a la estimación del efecto del tratamiento. Según el método complementario empleado junto a la ponderación, el efecto del tratamiento se calculará de una forma u otra. Un ejemplo de estimación será la estimación del efecto del tratamiento en una situación donde se emplee la ponderación junto al emparejamiento, cuando la respuesta no es ordinal. La estimación del efecto del tratamiento resultante será:

$$\widehat{ATT} = \frac{1}{n} \sum_{i=1}^n \frac{Z_i * Y_i}{\hat{\lambda}(X_i)} - \frac{1}{n} \sum_{i=1}^n \frac{(1-Z_i) * Y_i}{1-\hat{\lambda}(X_i)} \quad (26)$$

Donde Z_i es una variable indicador que vale 1 si el individuo pertenece al grupo de los tratados y vale 0 si pertenece al grupo de los controles y las Y_i son los individuos, tanto tratados como controles.

Según [Propensity score matching in observational data], este método de ponderación tiene un inconveniente de que si el valor de los Índices de Propensión

estimados están próximos a cero o a uno, entonces las ponderaciones son a menudo poco realistas para estos individuos.

3.4.3. Elección del mejor método de ajuste.

Cada uno de los métodos descritos posee ventajas y limitaciones según el tipo de datos sobre los que se apliquen. No existe un consenso general sobre cuál de estos métodos es el más eficaz y preferible para aplicar.

El emparejamiento basado en el Índice de Propensión se ha considerado como el método más eficaz [Propensity score matching in observational data]. La estratificación y el emparejamiento sobre las estimaciones del Índice de Propensión proporcionan el equilibrio buscado en las covariables y entre los individuos de los grupos de tratamiento y de control. Sin embargo, el emparejamiento es más eficaz en la reducción del desequilibrio existente en los individuos de los grupos de tratamiento y de control, así como en la reducción del sesgo de selección que aparece en la aplicación del tratamiento. Además, mediante el emparejamiento se consigue que las distribuciones de los grupos de tratamiento y control se asemejen entre sí.

3.5. Usos, ventajas y limitaciones del Índice de Propensión.

El Índice de Propensión es una técnica empleada para controlar los factores de confusión. Una ventaja que presenta el Índice de Propensión frente a las técnicas tradicionales (por ejemplo, la regresión lineal múltiple) es que permite comparar las covariables después del emparejamiento o en cada uno de los estratos después de la estratificación con los valores que se obtuvieron antes de aplicar estos métodos [Wolfgang C. Wilkenmayer y otros (2004)]. También permite equilibrar los grupos de tratamiento y control en un gran número de covariables. Para no recaer en un problema de dimensionalidad cuando se equilibran un gran número de covariables, se necesitarán un gran número de individuos.

Una de las limitaciones está en que el Índice de Propensión solo puede equilibrar las covariables observadas entre los grupos de tratamiento y control, las covariables no observadas no las puede equilibrar. Estas covariables ocasionan un sesgo oculto en el modelo. Al equilibrar las covariables observadas, el sesgo oculto puede verse aumentado.

Otra limitación que este método posee es la necesidad de que exista una superposición sustancial entre los grupos de tratamiento y control en las estimaciones de los Índices de Propensión para que esta técnica sea más eficaz. Además, hay que tener en cuenta que no se pueden utilizar covariables que solamente afecten a uno de los grupos.

En algunos casos, [Wolfgang C. Wilkenmayer y otros (2004)], los resultados que se obtienen mediante la aplicación del Índice de Propensión sobre un conjunto de datos obtenidos a partir de un diseño observacional no se diferencian a los resultados de los métodos tradicionales como puede ser la regresión lineal múltiple. En situaciones donde el número de covariables de confusión es moderado, la utilidad del Índice de Propensión como herramienta analítica es bastante incierta. Cuando el tamaño de la muestra es bastante grande, el uso del método tradicional de la regresión lineal múltiple es preferible al uso del Índice de Propensión. También se prefiere la utilización de la regresión lineal múltiple cuando el suceso de interés no es un suceso de los denominados sucesos raros. El Índice de Propensión se debe aplicar en las situaciones donde el suceso de interés es un suceso poco común y existe un gran número de factores de confusión.

4. Ejemplo Índice de Propensión.

4.1. Introducción y descripción del problema

La cirrosis biliar primaria, (PBC, Primary Biliary Cirrhosis), es una enfermedad autoinmune del hígado a través de la cual se produce una destrucción lenta y progresiva de los conductos biliares del hígado, los conductos interlobulares afectados en el principio de la enfermedad. Esto produce que la bilis se acumule en hígado y a lo largo del tiempo produzca daños en el tejido, provocando cicatrices, fibrosis o cirrosis.

El objetivo de este análisis es determinar mediante el Índice de Propensión cuál es el posible efecto de la administración de D-penicillamine en los individuos en la probabilidad de que un individuo presente o no edema en el hígado y señalar los diferentes pasos que se deben seguir en la aplicación de esta técnica esta técnica del Índice de Propensión.

El conjunto de datos se obtuvo de la siguiente página web:

<http://biostat.mc.vanderbilt.edu/wiki/Main/DataSets>

Como el conjunto de datos empleado en el ejemplo no es exactamente el localizado en la página web anterior, este se encuentra en el documento de texto anexo a este trabajo.

En este conjunto de datos se dispone de información de 418 individuos de la clínica Mayo con cirrosis biliar primaria. Unos de estos pacientes que cumplían con una serie de condiciones recibieron un placebo mientras que los otros individuos fueron tratados mediante D-penicillamine. Se recogieron una serie de variables para cada uno de estos individuos.

Si se tratara de estimar el efecto del tratamiento de forma directa, estas estimaciones podrían verse afectadas por la existencia de un sesgo de selección y un sesgo de confusión. En estas situaciones las distribuciones del Índice de Propensión estimado en tratados y controles pueden ser muy diferentes y la estimación del efecto del tratamiento puede no ser adecuada.

Las variables que intervienen en este estudio están recogidas en la tabla 10:

Variable	Descripción de la Variable	Codificación o unidad de medida
Bili	Bilirrubina sérica	En mg/dl
Medicamento	Indica el medicamento utilizado en los individuos	Placebo
		D-penicillamine
Edad	Edad de los individuos	En años
Spiders	Malformaciones en los vasos sanguíneos de la piel	Presente
		ausente
Hepatom	Presencia de hígado agrandado o hepatomegalia	Presente
		ausente
Colesterol	Colesterol sérico	En mg/dl
Triglicéridos	Triglicéridos	En mg/dl
Edema	Presencia de edema en el individuo	No edema
		Si edema

Tabla 10. Variables conjunto PBC

La variable que recoge el tratamiento en este estudio es la variable medicamento. La variable que recoge la respuesta para la cual se desea conocer el efecto del tratamiento es la variable edema. Ambas variables son dicotómicas. El resto de las variables observadas actuarán como covariables que ayudarán a la estimación del efecto del tratamiento sobre la variable respuesta, pero produciendo con su inclusión un sesgo de confusión que hay que tener en cuenta.

En primer lugar realizamos un análisis descriptivo de estas variables recogido en la tabla 11:

Para total de los individuos				
Variable	Media	sd	Mínimo	Máximo
Bili	3.221	4.4075	0.3	28
Edad	50.74	10.4472	26.28	78.44
Trig	150.51	77.7023	33	598
Colesterol	329.5	204.32	115	1775
Para individuos tratados con placebo				
Variable	Media	sd	Mínimo	Máximo
Bili	2.971	3.46155	0.3	20
Edad	52	10.5378	26.28	78.44
Trig	165	84.1938	33	598
Colesterol	307.4	172.2491	115	1712
Para individuos tratados con D-penicillamine				
Variable	Media	sd	Mínimo	Máximo
Bili	3.649	5.28195	0.3	28
Edad	48.58	9.9578	30.57	74.52
Trig	125.67	57.33049	44	432
Colesterol	367.4	245.95577	120	1775

Tabla 11. Análisis variables numéricas conjunto datos PBC

En la tabla 11 se recogen tanto las medias como las desviaciones estándar de las variables numéricas para los individuos del conjunto de datos según sean tratados con la D-penicilline o si pertenecen al grupo de los controles. Se puede ver que las diferencias en las variables trig, colesterol y bili están en las diferencias existentes en las desviaciones estándar, mientras que para la variable edad, no se observan grandes diferencias entre los dos grupos.

En la Tabla 12 se recogen para las variables categóricas del conjunto de datos, el número de individuos según la categoría de la variable y según que fueran tratados con D-penicillamine o placebo.

		Total = 418	D-penicillamine = 154	Placebo=264
Spiders	Presente	160	45	115
	Ausente	258	109	149
hepatom	Presente	204	87	117
	Ausente	214	67	147
edema	No edema	354	131	223
	edema	64	23	41

Tabla 12. Análisis variables categóricas conjunto datos PBC

Para la estimación del efecto del tratamiento, el primer paso que seguiremos es la estimación del Índice de Propensión.

Tras la estimación del Índice de Propensión y la comprobación de que existen diferencias en las estimaciones del índice entre los grupos de tratados y controles, se puede proceder al segundo paso, en el que se aplicará alguna de las técnicas de control de sesgo descritas en el documento para reducir las diferencias entre ambos grupos producidas por el sesgo de confusión y de selección.

El último paso será el de estimar el efecto del tratamiento propiamente dicho.

Para resolver este problema aplicando la técnica del Índice de Propensión, emplearemos el software estadístico libre R (<http://cran.r-project.org/>) y las librerías “nonrandom” ([Susanne Stampf (2014)]) y “MatchIt” ([Daniel Ho y otros (2013)]). El código correspondiente a los análisis realizados se encuentra en el anexo 2 del trabajo.

4.2. Cálculo del Índice de Propensión

Para el primer paso de la estimación del Índice de Propensión, emplearemos la regresión logística. Lo primero que hay que determinar antes de aplicar la técnica es determinar que covariables intervendrán en el modelo, si se tendrán en cuenta alguna o todas las interacciones entre las variables,... Para este ejemplo se ha decidido seleccionar como variables que intervendrán en el modelo las covariables medidas y alguna de las interacciones de segundo orden entre las covariables seleccionadas mediante stepwise.

El ajuste del modelo logístico (27) permite estimar el Índice de Propensión (28):

$$\eta_i = \text{logit}(\lambda(X_i)) = \alpha + \beta_1 * \text{hepatom} + \beta_2 * \text{spiders} + \beta_3 * \text{edad} + \beta_4 * \text{bili} + \beta_5 * \text{colesterol} + \beta_6 * \text{trig} + \beta_7 * \text{spiders:trig} + \beta_8 * \text{edad:trig} + \beta_9 * \text{colesterol:trig} \quad (27)$$

Mediante esta ecuación se puede estimar para cada uno de los individuos su Índice de Propensión aplicando (28):

$$\lambda(X_i) = \frac{e^{\eta_i(X_i)}}{1+e^{\eta_i(X_i)}} \quad (28)$$

Donde $\lambda(X_i)$ es el Índice de Propensión para el individuo con covariables X_i .

Los estimadores de los parámetros del modelo se recogen en la tabla 13:

	Estimación	Error estándar	Z	p-valor
Término independiente	0.6975	1.445	0.483	0.62937
Hepatom	0.6	0.2374	2.528	0.01148
Spiders	0.4876	0.5859	0.832	0.40527
edad	0.00618	0.0263	0.235	0.81427
Bili	0.0719	0.0311	2.309	0.0209
Colesterol	-0.00299	0.001439	-2.08	0.03748
trig	0.00135	0.01051	0.128	0.898
Spiders:trig	-0.009165	0.004109	-2.23	0.02572
Edad:trig	-0.0003126	0.0001965	-1.591	0.11165
Colesterol:trig	0.0000528	0.0000095	2.7	0.00693
AIC:	493.61			

Tabla 13. Estimadores por reg. Logística de los parámetros

Con el modelo de regresión logística ajustado se calculan los índices de propensión para cada uno de los individuos empleando la ecuación (28).

En el Gráfico 3 se pueden observar las densidades del Índice de Propensión para ambos grupos tratados y placebo.

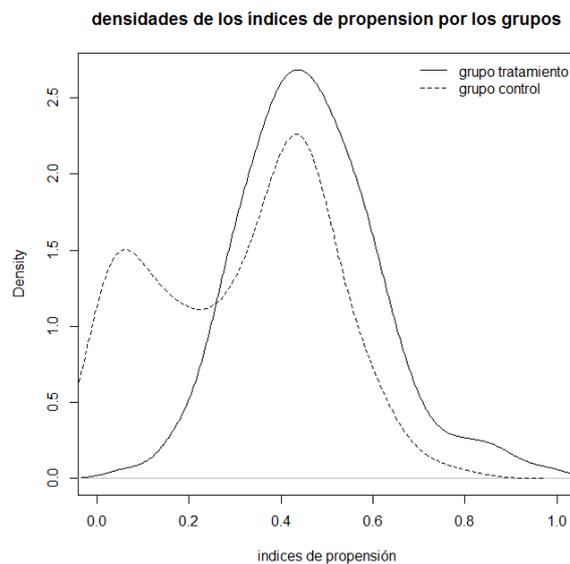


Gráfico 3. Distribución del Índice de Propensión antes emparejamiento y estratificación ejemplo PBC

A la vista del Gráfico 3, se puede observar que la distribución del Índice de Propensión estimado por la regresión logística para los individuos tratados mediante la D-penicillamine es diferente a la distribución en los individuos controles. Estas

diferencias pueden ser debidas a un sesgo de selección. Para controlar o reducir en la medida de lo posible este sesgo la técnica de emparejamiento y la de estratificación.

4.3. Corrección del sesgo mediante el método de emparejamiento

El primer método que se aplicará será el método de emparejamiento, mediante el cual se emparejarán individuos del grupo de tratamiento con individuos del grupo de control. El método de emparejamiento que se utilizará será el emparejamiento 1:1 por el vecino más cercano.

Para determinar si el emparejamiento realizado es útil existen diversas técnicas. Una técnica que se puede emplear es el ASAM (Average Standardized Absolute Mean difference o diferencia absoluta de medias estandarizadas) mediante la cual se desea que el valor obtenido tras el emparejamiento sea menor que antes de aplicarlo [Brian Lee (2013)]. El ASAM se calculará mediante la diferencia entre la media absoluta de los tratados con la media absoluta de los controles:

$$ASAM = media\ absoluta_{tratados} - media\ absoluta_{controles} \quad (29)$$

Además en esta técnica de ASAM se tiene que comprobar las diferencias de medias estandarizadas para cada covariable sean menores que un punto límite de equilibrio. Con esta técnica se emparejaran individuos del grupo de los tratados con un individuo del grupo de control, descartando del modelo los individuos que no han sido emparejados. Para calcular las diferencias de medias estandarizadas para cada covariable se aplicará la fórmula (30):

$$d_i \cong \frac{\bar{X}_{tratamiento(i)} - \bar{X}_{control(i)}}{\sqrt{\frac{var_{control(i)} + var_{tratamiento(i)}}{2}}} \quad (30)$$

Cuando las variables no son numéricas, la fórmula (30) se varía modificada de la siguiente manera en la fórmula (31):

$$d_i \cong \frac{p_{tratamiento(i)} - p_{control(i)}}{\sqrt{\frac{p_{tratamiento(i)}(1-p_{tratamiento(i)}) + p_{control(i)}(1-p_{control(i)})}{2}}} \quad (31)$$

Los primeros resultados que se calculan son las diferencias medias estandarizadas para cada covariable y las medias absolutas de los tratados y de los controles antes de aplicar el emparejamiento. Estos resultados se recogen en la Tabla 14.

ASAM	0.1523
-------------	--------

	Medias en tratados	Medias en controles	SD controles	Diferencia media estandarizada
absoluta	0.4646	0.3123		
Hepatom	0.4351	0.5568	0.4977	-0.2448
Spiders	0.2922	0.4356	0.4968	-0.3143
Edad	48.5825	52.0008	10.5378	-0.3433
bili	3.6487	2.9712	3.7949	0.1283
Colesterol	367.3896	307.4318	172.2491	0.2438
trig	125.6688	165.0076	84.1938	-0.6862
Spiders:trig	36.8831	84.6667	114.567	-0.6836
Edad:trig	6080.9122	8632.4089	4807.1568	-0.8625
Colesterol:trig	50101.4286	50098.7159	40718.1492	0.0001

Tabla 14. Diferencias medias estandarizadas variables antes emparejamiento

Podemos comprobar los resultados obtenidos aplicando las fórmulas 29, 30 y 31. Por ejemplo el ASAM para los datos antes del emparejamiento sería:

$$ASAM = 0.4646 - 0.3123 = 0.1523 \quad (32)$$

Si queremos calcular las diferencias medias estandarizadas para por ejemplo heptom se aplicaría la fórmula (31) y los resultados serían:

$$d_{heptom} = \frac{0.4351 - 0.5567}{\sqrt{\frac{0.4351(1-0.4351) + 0.5567(1-0.5567)}{2}}} = -0.245 \cong -0.2448 \quad (33)$$

En estos resultados se puede observar las diferencias entre las medias del grupo de los tratados y las medias del grupo de los controles. Con los emparejamientos se espera disminuir las diferencias. Los resultados tras el emparejamiento se recogen en la Tabla 15.

ASAM	0.0293
------	--------

	Medias en tratados	Medias en controles	SD controles	Diferencia media estandarizada	% reducción diferencia
absoluta	0.4646	0.4353			
Hepatom	0.4351	0.5065	0.5016	-0.1436	41.3333
Spiders	0.2922	0.2922	0.4563	0	100
Edad	48.5825	49.7214	10.3040	-0.1144	66.6838
bili	3.6487	3.0104	3.8018	0.1208	5.7827
Colesterol	367.3896	337.2208	205.9056	0.1227	49.6832
trig	125.6688	124.2208	62.5691	0.0253	96.319
Spiders:trig	36.8831	35.9091	71.5532	0.0139	97.9616
Edad:trig	6080.9122	6092.5009	2974.2706	-0.0039	99.5458
Colesterol:trig	50101.4286	44064.5649	45741.3114	0.1134	incrementa

Tabla 15. Diferencias medias estandarizadas variables después emparejamiento

	Control	Tratamiento
Todos	264	154
Emparejados	154	154
No emparejados	110	0
Descartados	0	0

Tabla 16. Resumen emparejamiento ejemplo PBC

Se han emparejado los 154 individuos pertenecientes al grupo del tratamiento con 154 individuos del grupo de los controles, quedando 110 individuos no emparejados del grupo control, que son descartados del análisis.

Para comprobar el balance en las covariables, se emplea el método de las diferencias estandarizadas entre los grupos. Se selecciona un punto límite de balance de 15. A este nivel, se puede observar que antes de aplicar el emparejamiento, la variable bili y la interacción entre colesterol y trig estaban equilibradas en ambos grupos. Tras el emparejamiento, se obtiene que se ha alcanzado el balance en todas las covariables e interacciones entre los grupos, aunque la diferencia estandarizada de la interacción de trig con colesterol ha aumentado de forma considerable. El ASAM se ha visto reducido de 0.1523 a 0.0293. La mejora en el equilibrio que presentan las diferencias medias estandarizadas antes y después del emparejamiento en las covariables lo podemos observar en el Gráfico 4,

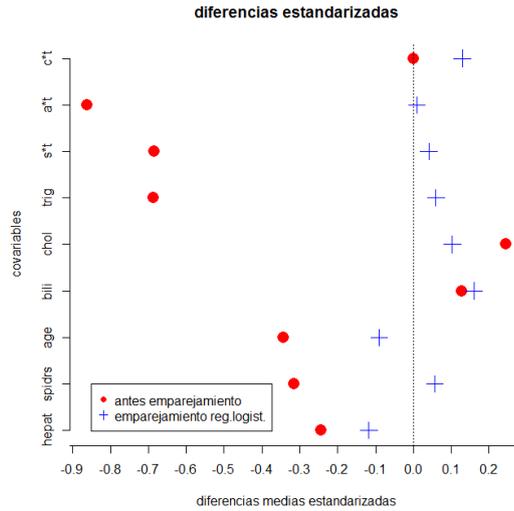


Gráfico 4. Balance de las covariables antes y después del emparejamiento ejemplo PBC

Se ve que las diferencias estandarizadas en casi todas las variables excepto en la interacción entre trig y colesterol se han reducido y se encuentran más próximas al 0 que antes de aplicar el emparejamiento.

Con los Índices de Propensión ajustados a través del emparejamiento por el vecino más próximo, representamos las distribuciones de los índices de propensión para los dos grupos.

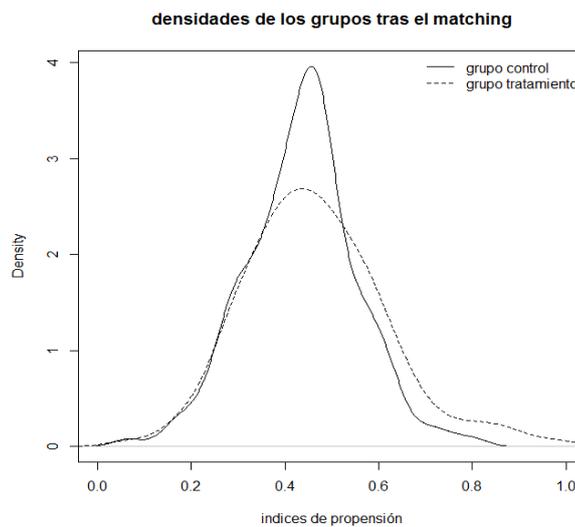


Gráfico 5. Distribución del Índice de Propensión tras emparejamiento ejemplo PBC

En este gráfico 5 se puede observar que la diferencia que existía entre las distribuciones de los grupos de tratamiento y de control se ha reducido.

4.4. Corrección del sesgo mediante el método de estratificación

El segundo método que utilizaremos como técnica de control del sesgo en el Índice de Propensión es el método de estratificación. En cada uno de los estratos contruidos se agruparán los individuos que posean un Índice de Propensión estimado similar.

Existen varias formas de construir los estratos. La forma que emplearemos en este ejemplo será empelando los percentiles de la distribución del Índice de Propensión.

En un principio, el número de estratos que se van a construir es de 5, ya que según [Peter C. Austin (2011)], Cochran demostró que este es el número óptimo de estratos para clasificar los individuos y si los estratos están bien contruidos entonces el sesgo de confusión que se tiene en el modelo se verá reducido en un 90%. Se aplicará el método de estratificar por los percentiles de la distribución del Índice de propensión, calculando los percentiles 20%,40%, 60% y 80%, de modelo que cada estrato contendrá al 20% de los individuos. Los estratos resultantes se presentan en la Tabla 17:

Estrato	Límite del estrato	Número de individuos	Proporción de individuos
1	[0.0201,0.196]	84	0.201
2	(0.196,0.327]	83	0.199
3	(0.327,0.416]	84	0.201
4	(0.416,0.524]	83	0.199
5	(0.524,0.913]	84	0.201

Tabla 17. Estratificación con 5 estratos ejemplo PBC

Con los estratos contruidos y los individuos clasificados en cada uno de estos, se procede a calcular el balance de las covariables. Para comprobar el balance emplearemos el método de las diferencias estandarizadas, con un punto de corte límite marcado con anterioridad. En la tabla 18 se recoge el equilibrio las covariables antes y después de aplicar la estratificación con un punto de corte límite de 35. El punto de corte límite se escoge de forma arbitraria. En este ejemplo se decide emplear este en concreto.

En la tabla 19 se recoge, de forma detallada los resultados de la tabla 18. Se recogen las diferencias estandarizadas en cada estrato y antes de la estratificación. Para determinar si la covariable está equilibrada en un estrato, la diferencia estandarizada obtenida tiene que ser menor que este punto de corte límite de 35. Para que se diga que una variable está balanceada después de la estratificación, entonces todas las diferencias estandarizadas de los estratos de esa variable tienen que ser menores que el punto de corte límite.

		balance de las variables					
		Hepatom	Spiders	Edad	Bili	Colesterol	Trig.
Antes estratificación		1	1	1	1	1	0
Después estratificación		0	1	1	0	0	1

Tabla 18. Resumen balance estratificación 5, ejemplo PBC

Covariables	Antes de la estratificación		Después de la estratificación													
	Diferencia estandarizada	Punto de corte límite=35														
		Estrato 1 Diferencia estandarizada	Estrato 2 Diferencia estandarizada	Estrato 3 Diferencia estandarizada	Estrato 4 Diferencia estandarizada	Estrato 5 Diferencia estandarizada	Estrato 1		Estrato 2		Estrato 3		Estrato 4		Estrato 5	
Hepatom	24.534	62.092	31.359	45.611	3.845	10.318										
Spiders	30.142	28.028	2.197	24.254	16.839	4.192										
Edad	33.343	31.751	34.33	0.17	2.173	6.399										
Bili	14.731	41.541	12.028	11.27	24.23	19.065										
Colesterol	28.239	54.498	3.314	28.238	4.034	8.151										
Trig.	54.618	29.884	18.98	27.531	12.993	1.1										
	Placebo	D- penicilamine	Placebo	D- penicilamine	Placebo	D- penicilamine	Placebo	D- penicilamine	Placebo	D- penicilamine	Placebo	D- penicilamine	Placebo	D- penicilamine	Placebo	D- penicilamine
N	264	154	81	3	52	31	49	35	49	34	33	51				

Tabla 19. Balance diferencias estandarizadas estratificación 5, ejemplo PBC

En estas tablas 18 y 19 se puede observar que antes de aplicar la estratificación, las variables bili, heptom, spiders, edad y colesterol están equilibradas en los grupos de tratamiento y control. Tras aplicar la estratificación, se obtiene que las covariables spiders, edad y trig logran un equilibrio entre ambos grupos, perdiendo el balance inicial en tres covariables. En la tabla 19 también se recogen los tamaños de los dos grupos antes y después de la estratificación para cada estrato. Observando estos tamaños muestrales, se puede pensar que las grandes diferencias estandarizadas en las covariables no equilibradas en el estrato 1 se deben a la gran diferencia entre el número de individuos tratados y controles.

Con los resultados de la estratificación anterior, nos planteamos una segunda estratificación para ver si se obtienen mejores resultados. En esta segunda estratificación se construirán, por el mismo método, cuatro estratos. Para aplicar el método de estratificación de los individuos se necesitará calcular los percentiles 25%, 50% y 75%. Los estratos se recogen en la Tabla 20:

Estrato	Límite del estrato	Número de individuos	Proporción de individuos
1	[0.00128,0.24]	105	0.251
2	(0.246,0.401]	104	0.149
3	(0.401,0.491]	104	0.241
4	(0.491,0.981]	105	0.251

Tabla 20. Estratificación con 4 estratos ejemplo PBC

Con los estratos contruidos y los individuos clasificados en cada uno de estos, se procede a calcular el balance de las covariables. Para comprobar el balance emplearemos de nuevo el método de las diferencias estandarizadas con un punto de corte límite de 35. En la tabla 21 se recoge un resumen del equilibrio las covariables antes y después de aplicar la estratificación.

En la tabla 22 se recoge, de forma detallada los resultados de la tabla 21. Se recogen las diferencias estandarizadas en cada estrato y antes de la estratificación. Para determinar si la covariable está equilibrada en un estrato, la diferencia estandarizada obtenida tiene que ser menor que este punto de corte límite de 35. Para que se diga que una variable está balanceada después de la estratificación, entonces todas las diferencias estandarizadas de los estratos de esa variable tienen que ser menores que el punto de corte límite.

	balance de las variables					
	Hepatom	Spiders	Edad	Bili	Colesterol	Trig.
Antes estratificación	1	1	1	1	1	0
Después estratificación	1	1	0	1	0	0

Tabla 21. Resumen balance estratificación 4, ejemplo PBC

	Antes de la estratificación		Después de la estratificación Punto de corte límite=35							
Covariables	Diferencia estandarizada		Estrato 1		Estrato 2		Estrato 3		Estrato 4	
			Diferencia estandarizada		Diferencia estandarizada		Diferencia estandarizada		Diferencia estandarizada	
Hepatom	24.534		1.329		0.32		21.455		1.913	
Spiders	30.142		27.82		4.495		14.246		6.214	
Edad	33.343		43.613		18.404		7.236		1.123	
Bili	14.731		19.974		1.714		32.683		18.776	
Colesterol	28.239		41.568		19.891		14.28		8.262	
Trig.	54.618		42.77		4.462		18.708		0.671	
	Placebo	D-penicilamine	Placebo	D-penicilamine	Placebo	D-penicilamine	Placebo	D-penicilamine	Placebo	D-penicilamine
N	264	154	97	8	58	46	66	38	43	62

Tabla 22. Balance diferencias estandarizadas estratificación 4, ejemplo PBC

En las tablas 21 y 22 se puede observar que antes de aplicar la estratificación, las variables hepatom, spiders, bili, edad y colesterol están equilibradas en los grupos de tratamiento y control. Tras aplicar la estratificación, se obtiene que las covariables hepatom, spiders y bili logran un equilibrio entre ambos grupos, perdiendo el balance inicial en dos covariables. En la tabla 22 también se recogen los tamaños de los dos grupos antes y después de la estratificación para cada estrato. Observando estos tamaños muestrales, se puede pensar que las grandes diferencias estandarizadas en las covariables no equilibradas en el estrato 1 se deben a la gran diferencia entre el número de individuos tratados y controles.

Al comparar ambas estratificaciones construidas, decidimos que la que nos proporciona mejores resultados y sobre la que estimaremos el efecto de la D-penicillamine sobre la presencia de edema en el hígado es la estratificación con 4 estratos, debido a que solo se pierde el equilibrio inicial en dos covariables mientras que en la estratificación con 5 estratos el equilibrio antes de estratificar se pierde en 3 covariables.

Comparando los resultados del emparejamiento y la estratificación, se puede decir que la técnica de emparejamiento proporciona mejores resultados debido a que se consigue que todas las covariables estén equilibradas tras su aplicación.

4.5. Estimación del efecto del tratamiento

Con los resultados obtenidos a partir del emparejamiento y la estratificación del Índice de Propensión, se procede a estimar el efecto del tratamiento sobre la presencia de edema en el hígado. Los resultados que se obtengan por ambos métodos deberían ser similares.

En primer lugar se estimará el efecto del tratamiento para los datos tratados a partir del método de emparejamiento. En los resultados se mostrará el efecto del tratamiento antes de aplicar ninguna técnica y después de aplicarla. Este efecto se medirá a partir de la razón de ventajas u odds ratio. Los resultados obtenidos están recogidos en la Tabla 23.

Tabla de la estimación del efecto				
	OR	Se(OR)	Intervalo confianza	
			Límite inferior	Límite superior
Antes de aplicar la técnica	0.955	0.2828	0.549	1.662
Tras el emparejamiento	1.126	0.3637	0.552	2.297

Tabla 23. Estimación del efecto del tratamiento por el emparejamiento ejemplo PBC

El efecto del tratamiento cuando se estima no teniendo en cuenta las covariables es de 0.955, que indica que la ventaja de que un individuo presente edema en el hígado es 0.955 veces cuando se le administra la D-penicillamine al individuo frente a administrarle placebo. Al observar el intervalo de confianza, se observa que este intervalo incluye al valor 1 por lo que se puede concluir que, con una confianza del 95%, administrar D-penicillamine a un individuo no afecta a la probabilidad de presentar edema en el hígado.

Cuando se aplica el emparejamiento, se reduce el sesgo de confusión y de selección y la estimación del efecto en esta situación será más precisa. El efecto del tratamiento cuando se estima teniendo en cuenta el resto de covariables y tras el emparejamiento es de 1.126, lo que indica que la ventaja de que un individuo presente edema en el hígado es 1.126 veces mayor cuando se le administra la D-penicillamine al individuos frente a administrarle placebo. El intervalo de confianza construido para esta estimación incluye al 1, por lo que se puede concluir que, con una confianza del 95%, administrar D-penicillamine a un individuo no afecta a la probabilidad de presentar edema en el hígado.

La conclusión extraída en la estimación del efecto de la D-penicillamine con la presencia de edema en el hígado es similar antes y después de aplicar el emparejamiento, pero la “orientación” de la OR es diferente y el intervalo de confianza es más amplio.

En segundo lugar se estimará el efecto del tratamiento para los datos tratados a partir del método de estratificación con 4 estratos. Los resultados que se obtienen son los recogidos en la Tabla 24.

	Tabla de la estimación del efecto			
	OR	Se(OR)	Intervalo confianza	
			Límite	Límite
Antes de aplicar la técnica	0.955	0.2828	0.549	1.662
Tras la estratificación (MH)	1.123	0.3152	0.605	4.104
Ajustado por las covariables no equilibradas	1.305	0.7642	0.292	5.838

Tabla 24. Estimación del efecto del tratamiento por estratificación 4, ejemplo PBC

Para obtener los resultados después de la estratificación se emplea el estadístico M-H. El efecto del tratamiento cuando se estima teniendo en cuenta el resto de covariables y tras la estratificación es de 1.123, lo que indica que la ventaja de que un individuo presente edema en el hígado es 1.123 veces mayor cuando se le administra

la D-penicillamine al individuos frente a administrarle placebo. El intervalo de confianza construido para esta estimación incluye al 1, por lo que se puede concluir que, con una confianza del 95%, administrar D-penicillamine a un individuo no afecta a la probabilidad de presentar edema en el hígado.

Como no todas las covariables se equilibraron cuando se aplicó la estratificación, es conveniente estimar el efecto teniendo en cuenta estas variables no equilibradas. Para estimar el efecto del tratamiento se ajusta un modelo de regresión logística donde las covariables no equilibradas se incorporarán como variables predictoras en el modelo [Brian Lee (2013)]. El efecto del tratamiento cuando se estima teniendo en cuenta estas covariables y tras la estratificación es de 1.305, lo que indica que la ventaja de que un individuo presente edema en el hígado es 1.305 veces mayor cuando se le administra la D-penicillamine al individuos frente a administrarle placebo. El intervalo de confianza construido para esta estimación incluye al 1, por lo que se puede concluir que, con una confianza del 95%, administrar D-penicillamine a un individuo no afecta a la probabilidad de presentar edema en el hígado.

La conclusión extraída en la estimación del efecto de la D-penicillamine con la presencia de edema en el hígado es similar antes y después de aplicar la estratificación, pero la “orientación” de la OR es diferente y los intervalos de confianza son más amplios.

Con los resultados obtenidos, se llega a la conclusión de que administrar D-penicillamine no afecta a la probabilidad de presentar edema en el hígado.

5. Conclusiones.

En este trabajo se ha presentado una técnica relativamente nueva que permite estimar el efecto de un determinado tratamiento en situaciones donde el conjunto de datos fue extraído mediante un diseño observacional cuando existen una serie de covariables que afectan a dicho efecto.

La primera conclusión que se puede extraer de la aplicación de la técnica del Índice de Propensión es que esta permite comparar las covariables entre los grupos tanto en cada uno de los estratos como después del emparejamiento. La técnica permite equilibrar los grupos de tratamiento y control en un gran número de covariables pero, para no recaer en un problema de dimensionalidad cuando se equilibran un gran número de covariables, se requieren un gran número de individuos.

Una segunda conclusión que se obtiene en la aplicación del Índice de Propensión es que se necesita una superposición sustancial entre los grupos de tratamiento y control en las estimaciones del Índice de propensión. Además, esta técnica solo puede equilibrar las covariables observadas, por lo que el sesgo que procede de las covariables no observadas no se puede reducir.

La tercera conclusión la encontramos en que los resultados que proporciona esta técnica en muchas situaciones no se diferencian en gran medida de los resultados que proporcionan otras técnicas tradicionales, como puede ser la regresión lineal múltiple. El Índice de Propensión se empleará en situaciones donde el suceso de interés es de los denominados “suceso raro”.

En el trabajo se han presentado dos ejemplos donde se aplica el Índice de Propensión paso a paso. La estimación del Índice de Propensión se obtuvo en ambos casos mediante la regresión logística. Sobre estas estimaciones se aplicaron los métodos de emparejamiento y estratificación para reducir el sesgo. Tras aplicar los métodos se estimó el efecto del tratamiento en cuestión. Para resolver estos ejemplos se empleó el software estadístico R.

En el ejemplo con el conjunto de datos pride se deseaba conocer el efecto que tiene que un niño sufra un LRTI severo cuando padece una infección RSV. Se concluyó que, con una confianza del 95%, padecer una infección RSV aumenta la probabilidad de sufrir un LRTI severo.

En el ejemplo con el conjunto de datos PBC se deseaba conocer el efecto que la administración de D-penicillamine tenía sobre la presencia de edema en el hígado en pacientes con cirrosis biliar primaria. Se concluyó que, con una confianza del 95%, administrar D-penicillamine a un individuo no afecta a la probabilidad de presentar edema en el hígado.

6. REFERENCIAS Y BIBLIOGRAFÍA

Barbara Sianesi (2001) Implementing propensity score matching estimators with STATA

<http://fmwww.bc.edu/RePEc/usug2001/psmatch.pdf>

Brian Lee (2013) appendix: Ps matching in R

http://www.pages.drexel.edu/~bkl29/pres/SPER_methods_R_matching.pdf

Cassandra W. Pattanayak, Donald B. Rubin, Elizabeth R. Zell (2011) Métodos de puntuación de propensión para crear una distribución equilibrada de las covariables en los estudios observacionales.

<http://www.revespcardiol.org/es/metodos-puntuacion-propension-crear-una/articulo/90027210/>

Christopher F. Baum (2013) propensity score matching: Regression discontinuity limited dependent variables

<http://fmwww.bc.edu/EC-C/S2013/823/EC823.S2013.nn12.slides.pdf>

Daniel Ho, Kosuke Imai, Gary King, Elizabeth Stuart (2013) – Package “MatchIt”

<http://cran.r-project.org/web/packages/MatchIt/MatchIt.pdf>

Daniel Westreich, Justin Lessler, Michele Jonsson Funk (2011) – Propensity score estimation: machine learning and classification methods as alternatives to logistic regression

<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2907172/>

Fabiola Rivera Ramírez. Metodología para la evaluación del impacto: Propensity score matching

http://precesam.colmex.mx/images/secundaria/seminarios/Taller_ENHRUM/4impacto-PSM.pdf

Ismini Savvala (2008) - Quintile Stratification on Propensity Scores and Combining Strata for

Different Tests

<http://dspace.library.uu.nl/handle/1874/27716>

J. Canela Soler, X. Castellsagué Piqué. (). Estudios de cohortes.

<http://lbe.uab.es/vm/sp/materiales/bloque-4/Guardia-Rodes/cohortes-Guardia-Rodes.pdf>

Jasjeet S. Sekhon. (2007). The Neyman-Rubin model of causal inference and estimation via matching methods

<http://sekhon.berkeley.edu/papers/SekhonOxfordHandbook.pdf>

José I. Pijoán, Lorea Martínez Indart, Arantza Urkaregi (2011) Ensayos clínicos vs estudios observacionales para el estudio de asociaciones causales
ftp://ftp.hrc.es/pub/bioest/jubc/USO_DEL_PROPENSITY_SCORE_JUBC.pdf

Leonardo Grilli and Carla Rampichini (2011) - Propensity scores for the estimation of average treatment effects in observational studies
<http://www.bristol.ac.uk/cmm/software/support/workshops/materials/causal-inference/prop-scores.pdf>

Lieu n. hazelwood, Traci L. Mach, John D. Wolken (2007-10) Alternative methods of unit nonresponse weighting adjustments: an application from the 2003 survey of small business finances
<http://www.federalreserve.gov/pubs/feds/2007/200710/200710pap.pdf>

Lisa kaltenbach (2008) How to use propensity score analysis
http://www.mc.vanderbilt.edu/gcrc/workshop_files/2008-04-11.pdf

Martin Ravallion (2005) Evaluación de programas contra la pobreza
http://siteresources.worldbank.org/INTLACREGTOPLABSOCPRO/Resources/Handbook_chapter_on_evaluation-sp.pdf

Marco Caliendo, Sabine Kopeinig (2005) Some Practical Guidance For the Implementation of Propensity Score Matching
<http://webcache.googleusercontent.com/search?q=cache:rquPdXKVvngJ:ftp.iza.org/dp1588.pdf+&cd=3&hl=es&ct=clnk&gl=es>

Melissa D.A. Carlson y R.Sean Morrison – Study Design, Precision, and Validity in Observational Studies.
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2920077/>

Paul R. Rosenbaum (2002) – Observational Studies, New York: Springer-Verlag 2nd Edition

Peter C. Austin (2011) An introduction to propensity score methods for reducing the effects of confounding in observational studies
<http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3144483/>

Rajeev H. Dehejia, Sadek Wahba (1998) Propensity score matching methods for non-experimental causal studies
http://www.nber.org/papers/w6829.pdf?new_window=1

Ralph B. D'Agostino (1998) tutorial in biostatistics. Propensity score methods for bias reduction in the comparison of a treatment to a non-randomized control group

<http://www.stat.ubc.ca/~john/papers/DAgostinoSIM1998.pdf>

Sandra Carolina Serna Pineda (2009) – Comparación de árboles de regresión y clasificación y regresión logística.

http://www.bdigital.unal.edu.co/671/1/42694070_2009.pdf

Susanne Stampf (2014) Propensity score based data analysis

<http://cran.r-project.org/web/packages/nonrandom/vignettes/nonrandom.pdf>

Thomas A. DiPrete, Markus Gangl. (2004). Assessing bias in the estimation of causal effects: Rosenbaum bounds on Matching estimators and instrumental variables estimation with imperfect instruments.

http://www.wjh.harvard.edu/~cwinship/cfa_papers/HBprop_021204.pdf

Thomas E. Love, Ph. D., Instructor. (2004) Using Propensity Score: Methods Effectively

http://www.chrp.org/love/asa_love_oct112004.pdf

Thomas E. Love, Ph. D (2003) Propensity Score: What Do They Do, how Should I Use Them, and Why Should I Care?

<http://www.chrp.org/love/asacleveland2003propensity.pdf>

Wolfgang C. Wilkenmayer y Tobias Kurth (2004) Propensity scores: help or hype?

<http://ndt.oxfordjournals.org/content/19/7/1671.full>

. (). Propensity score matching in observational data

http://umanitoba.ca/faculties/medicine/units/community_health_sciences/departamental_units/mchp/protocol/media/propensity_score_matching.pdf

7. ANEXOS

ANEXO 1. Código R para resolver ejemplo con datos *pride* de la librería "nonrandom"

```
# cargamos la librería nonrandom que contiene el conjunto de datos y
funciones que se van a emplear

library(nonrandom)

# cargamos el conjunto de datos

data(pride)

# calculamos el índice de propensión

pride.ps <- pscore(data = pride, formula =
PCR_RSV~SEX+RSVINP+REGION+AGE+ELTATOP+EINZ+EXT, name.pscore = "ps")

# representamos la distribución del índice de propensión para cada
grupo

plot.pscore(pride.ps, main = "PRI.De study", with.legend = TRUE, par.1
= list(lty=1, lwd=2),

  par.0 = list(lty=3, lwd=2), xlab = "", ylim = c(0,4.5))

# construimos los estratos para la estratificación

pride.str5 <- ps.makestrata(object = pride.ps, breaks =
quantile(pride.ps$pscore,

  seq(0,1,0.2)), name.stratum.index = "stratum")

# generamos el emparejamiento que se va a utilizar

pride.m1 <- ps.match(object = pride.ps, ratio = 1, x = 0.2, caliper =
"logit",

  matched.by = "ps", setseed = 38902)

summary(pride.m1)

## balance de las covariables en la estratificación

pride.str5.bal <- ps.balance(object = pride.str5, sel = c(2:8),
cat.levels = 4,

  method = "stand.diff", alpha = 20)

pride.str5.bal

## balance de las covariables en el emparejamiento

pride.m1.bal2 <- ps.balance(object = pride.m1, sel = c(2:8),
cat.levels = 4,

  method = "stand.diff", alpha = 10)
```

```
## gráfico que recoge las diff. Estandarizadas antes y después del
emparejamiento para las covariables

plot.std(x= pride.m1.bal2)

# estimación del efecto del tratamiento tratado por el emparejamiento

pride.estimate.m <- ps.estimate(object = pride.m1, resp = "SEVERE",
family = "binomial")

pride.estimate.m

# estimación del efecto del tratamiento tratado por la estratificación
y ajustado por las covariables no balanceadas

pride.estimate.e <- ps.estimate(object = pride.str5, resp = "SEVERE",
family = "binomial",

adj=c('RSVINF', 'REGION', 'AGE', 'ETHNO'))

pride.estimate.e
```

ANEXO 2. Código R para resolver ejemplo con datos propios sobre *PBC*

```
# cargamos los datos y agregamos al conjunto de datos las
codificaciones de las variables de la droga y del edema.

a<-read.csv2('aaa.csv',header=T)

drug_cod<-rep(0,length(a$drug))

drug_cod[a$drug=="D-penicillamine"]<-1

a<-cbind(a,drug_cod)

edema_cod<-rep(1,length(a$edema))

edema_cod[a$edema=='no edema']<-0

a<-cbind(a,edema_cod)

## estimación del índice de propensión

# mediante un procedimiento stepwise seleccionamos las interacciones
que intervengan para reducir el modelo. Estimamos el índice de
propensión por la función glm

ps.model12 <-
glm(drug_cod~hepatom+spiders+age+bili+chol+trig+spiders*trig+age*trig+
chol*trig,data=a, family=binomial(link="logit"))

summary(ps.model12)

a$PS12 <- predict( ps.model12, type='response')

# mediante la función pscore de la librería nonrandom

library(nonrandom)

pbc.ps2 <- pscore(data = a,formula =

drug_cod~hepatom+spiders+age+bili+chol+trig+spiders*trig+age*trig+chol
*trig, name.pscore = "ps")

# Gráfico de la distribución del índice de propensión para cada grupo

plot(density(a$PS12[a$drug=='D-penicillamine']),xlim=c(0,1),

main='densidades de los índices de propension por los
grupos',xlab='indices de propensión')

lines(density(a$PS12[a$drug=='placebo']),lty=2)

legend('topright',lty=c(1,2),bty='n',legend=c('grupo
tratamiento','grupo control'))

## emparejamiento del índice de propensión

# mediante la función matchit de la librería MatchIt

library(MatchIt)
```

```

nn1 <-
matchit(drug_cod~hepatom+spiders+age+bili+chol+trig+spiders*trig+age*trig+chol*trig,data=a, distance=a$PS12,method="nearest")

nn1.data <- match.data(nn1)

summary(nn1, standardize=T)

# gráfico del balance de las covariables antes y después del emparejamiento.

nombres<-
c('hepat','spidrs','age','bili','chol','trig','s*t','a*t','c*t')

antes<-c(-0.2448,-0.3143,-0.3433,0.1283,0.2438,-0.6862,-0.6836,-0.8625,0.0001)

log<-c(-0.1175,0.0569,-0.0911,0.1593,0.1026,0.0597,0.0404,0.0095,0.1291)

plot(antes,1:9, col=2,axes=F,xlab='diferencias medias estandarizadas',ylab='covariables',cex=2,pch=16)

points(log,1:9,pch=3,col=4,cex=2)

axis(2, at=1:9, lab=nombres)

axis(1,seq(-1,0.5,0.1))

title(main='diferencias estandarizadas')

abline(v=0,lty=3)

legend(-0.85,2,pch=c(16,3),legend=c('antes emparejamiento','emparejamiento reg.logist.'),col=c('red','blue'))

## emparejamiento por las funciones de la librería nonrandom

pbc_mat<-ps.match(pbc.ps2,setseed = 38902)

# balance del emparejamiento

balance_mat<-ps.balance(object = pbc_mat, sel = c('hepatom','spiders','age','bili','chol','trig'),treat='drug_cod',cat.levels = 3, method = "stand.diff",alpha = 30)

balance_mat

# distribución del índice de propensión tras el emparejamiento de los individuos

plot(density(nn1.data$PS12[nn1.data$drug=='placebo']),main='densidades de los grupos tras el matching',xlim=c(0,1),xlab='indices de propensión')

lines(density(nn1.data$PS12[nn1.data$drug=='D-penicillamine']),lty=2)

```

```

legend('topright',lty=c(1,2),bty='n',legend=c('grupo control','grupo
tratamiento'))

## estratificación del índice de propensión

# estratificación con 5 estratos

pbc.str52 <- ps.makestrata(object = pbc.ps2, breaks =
quantile(pbc.ps2$pscore, seq(0,1,0.2)),

  name.stratum.index = "stratum")

pbc.str52

# balance de la estratificación

pbc.str5.bal2 <- ps.balance(object = pbc.str52, sel =
c('hepatom','spiders','age','bili','chol','trig'),

  treat='drug_cod', cat.levels = 3, method = "stand.diff",alpha = 35)

pbc.str5.bal2

# número de individuos tratados y controles en cada estrato

q1<-pbc.str52$data[pbc.str52$stratum==1,]

length(q1$drug[q1$drug=='placebo'])

length(q1$drug[q1$drug=='D-penicillamine'])

q2<-pbc.str52$data[pbc.str52$stratum==2,]

length(q2$drug[q2$drug=='placebo'])

length(q2$drug[q2$drug=='D-penicillamine'])

q3<-pbc.str52$data[pbc.str52$stratum==3,]

length(q3$drug[q3$drug=='placebo'])

length(q3$drug[q3$drug=='D-penicillamine'])

q4<-pbc.str52$data[pbc.str52$stratum==4,]

length(q4$drug[q4$drug=='placebo'])

length(q4$drug[q4$drug=='D-penicillamine'])

q5<-pbc.str52$data[pbc.str52$stratum==5,]

length(q5$drug[q5$drug=='placebo'])

length(q5$drug[q5$drug=='D-penicillamine'])

# estratificación con 4 estratos

pbc.str42 <- ps.makestrata(object = pbc.ps2,
breaks=quantile(pbc.ps2$pscore, seq(0,1,0.25)),name.stratum.index=
"stratum")

```

```

pbc.str42

# balance de la estratificación

pbc.str4.bal2 <- ps.balance(object = pbc.str42, sel =
c('hepatom','spiders','age','bili','chol','trig'),

  treat='drug_cod', cat.levels = 3, method = "stand.diff",alpha = 35)

pbc.str4.bal2

# número de individuos tratados y controles en cada estrato

q1<-pbc.str42$data[pbc.str42$stratum==1,]

length(q1$drug[q1$drug=='placebo'])

length(q1$drug[q1$drug=='D-penicillamine'])

q2<-pbc.str42$data[pbc.str42$stratum==2,]

length(q2$drug[q2$drug=='placebo'])

length(q2$drug[q2$drug=='D-penicillamine'])

q3<-pbc.str42$data[pbc.str42$stratum==3,]

length(q3$drug[q3$drug=='placebo'])

length(q3$drug[q3$drug=='D-penicillamine'])

q4<-pbc.str42$data[pbc.str42$stratum==4,]

length(q4$drug[q4$drug=='placebo'])

length(q4$drug[q4$drug=='D-penicillamine'])

### estimación del efecto del tratamiento

# estimación efecto del tratamiento a partir de la estratificación con
4 estratos

bb<-ps.estimate(object = pbc.str42,family='binomial', resp =
'edema_cod',treat='drug_cod',

adj=c('age','chol','trig'))

bb

# estimación efecto del tratamiento a partir del emparejamiento

vvv<-ps.estimate(object =pbc_mat,family='binomial',resp =
'edema_cod',treat='drug_cod')

vvv

```