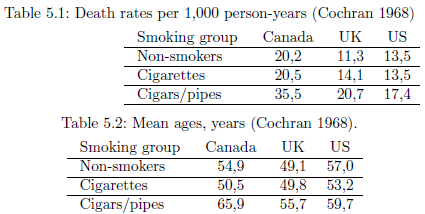
**Trabajo Práctico N° 2:**

**Matching.**

**Ejercicio 1: Exact Matching.**

*(Basado en Cunningham) Este ejercicio propone utilizar el procedimiento de matching sobre una variable para pensar en términos simples de dónde provienen los estimadores de Matching. Hoy en día, es conocimiento común que fumar aumenta la tasa de mortalidad. Sin embargo, esta afirmación no proviene de datos experimentales.*



**(a)** *Considerar los datos de mortalidad y condición de fumador de la Tabla 1. ¿Qué resultado da la diferencia de medias entre grupos de fumadores en términos de tasa de mortalidad? ¿Qué pide el supuesto de independencia? En particular, comentar qué espera que pase con otras variables observables.*

El resultado que da la diferencia de medias entre grupos de fumadores en términos de tasa de mortalidad es 15, 6,6 y 3,9 para Canadá, UK, US, respectivamente. El supuesto de independencia pide que la media de la tasa de mortalidad entre grupos sea igual. En particular, se espera que las variables observables entre fumadores no deberían ser, en promedio, diferentes, es decir, los grupos deberían ser balanceados.

**(b)** *Considerar la edad de las personas. ¿Qué se puede decir sobre el supuesto de independencia?*

Considerando la edad de las personas, lo que se puede decir sobre el supuesto de independencia es que no se cumpliría, ya que esta característica no es independiente del resultado potencial.

**(c)** *Estimar el efecto correcto.*

**Ejercicio 2: Teorema de Rosenbaum y Rubin.**

*Demostrar el siguiente teorema. Sean Y (0), Y (1) resultados potenciales, X un tratamiento binario, W un vector de características observables. Suponiendo que vale el supuesto de unconfoundedness, se define e (W) como la probabilidad de recibir el tratamiento en función de variables observables. Luego (Y (0), Y (1) X | e (W)). Como corolario:*

*ATE= E [E [ | X= 1, e (W)] - E [ | X= 0, e (W)]].*

P (X= 1 | Y (0), Y (1), e (W))= E (X | Y (0), Y (1), e (W))

P (X= 1 | Y (0), Y (1), e (W))= E (E (X | Y (0), Y (1), e (W), W) | Y (0), Y (1), e (W))

P (X= 1 | Y (0), Y (1), e (W))= E (E (X | W) | Y (0), Y (1), e (W))

P (X= 1 | Y (0), Y (1), e (W))= E (e (W) | Y (0), Y (1), e (W))

P (X= 1 | Y (0), Y (1), e (W))= e (W).

El resultado implica que los resultados potenciales son independientes del tratamiento habiendo condicionado en la probabilidad de recibir el tratamiento, e (W) (el llamado *propensity score*). Como se tiene el resultado demostrado al nivel de la probabilidad, esto implica el resultado deseado para las esperanzas y se puede calcular el ATE habiendo condicionado en el *propensity score* en vez de en todas las características W.

El teorema de Rosenbaum y Rubin sugiere un procedimiento aplicado para estimar los efectos promedios de tratamiento: calcular la probabilidad e (W) de recibir el tratamiento en función de las características W y hacer un *matching* en esta variable que resume las anteriores. Esto, además, simplifica ca el problema de dimensionalidad de emparejar muchas variables con algunas, potencialmente, continuas. El supuesto de *unconfoundedness* es necesario para el teorema de Rosenbaum y Rubin. No se puede reemplazar fallas en este supuesto y esperar que se solucione haciendo *Propensity Score Matching*.

**Ejercicio 3: Propensity Score Matching.**

*Para solucionar el problema de la dimensionalidad utilizando matching sobre múltiples variables, se puede reducir el conjunto de variables en una sola, la llamada propensity score, puntaje de propensión o probabilidad de recibir el tratamiento. En este ejercicio, se propone realizar una evaluación del impacto de fumar durante el embarazo sobre el peso de los bebés en base a datos observacionales. Utilizando la base de datos “cattaneo2.dta” que utiliza Cattaneo (XX), responder las siguientes preguntas.*

**(a)** *Comprobar si los grupos de control y de tratamiento están balanceados.*

Stata.

**(b)** *¿Cuál es la diferencia de medias simple?*

Stata.

**(c)** *Estimar la probabilidad de recibir el tratamiento con un modelo Logit. Utilizar como variables explicativas mmarried, deadkids, nprenatal, monthslb, prenatal, fbaby y alcohol. Definir una sección de soporte común.*

Stata.

**(d)** *Estimar el efecto promedio de tratamiento sobre los tratados utilizando Propensity Score Matching implementando nearest-neighbor matching, radius matching, kernel matching y stratification matching.*

Nearest-Neighbor Matching:

ATT estimation with Nearest Neighbor Matching method

(random draw version)

Bootstrapped standard errors

---------------------------------------------------------

n. treat. n. contr. ATT Std. Err. t

---------------------------------------------------------

864 659 -235.279 37.700 -6.241

---------------------------------------------------------

Note: the numbers of treated and controls refer to actual

nearest neighbour matches

Radius Matching:

ATT estimation with the Radius Matching method

Bootstrapped standard errors

---------------------------------------------------------

n. treat. n. contr. ATT Std. Err. t

---------------------------------------------------------

863 864 -246.371 34.574 -7.126

---------------------------------------------------------

Note: the numbers of treated and controls refer to actual

matches within radius

Kernel Matching:

ATT estimation with the Kernel Matching method

Bootstrapped standard errors

---------------------------------------------------------

n. treat. n. contr. ATT Std. Err. t

---------------------------------------------------------

864 864 -229.593 31.817 -7.216

---------------------------------------------------------

Stratification Matching:

ATT estimation with the Stratification method

Bootstrapped standard errors

---------------------------------------------------------

n. treat. n. contr. ATT Std. Err. t

---------------------------------------------------------

864 864 -227.836 25.709 -8.862

---------------------------------------------------------

**(e)** *Estimar el efecto promedio de tratamiento utilizando nearest-neighbour matching.*

Matching estimator: Average Treatment Effect

Number of obs = 1728

Number of matches (m) = 1

------------------------------------------------------------------------------

bweight | Coefficient Std. err. z P>|z| [95% conf. interval]

-------------+----------------------------------------------------------------

SATE | -209.5832 28.82159 -7.27 0.000 -266.0725 -153.0939

------------------------------------------------------------------------------

Matching variables: e\_x

Bias-adj variables: mmarried alcohol deadkids nprenatal monthslb prenatal fbaby

**(f)** *Estimar el efecto promedio de tratamiento utilizando ajuste por regresión.*

Treatment-effects estimation Number of obs = 1,728

Estimator : regression adjustment

Outcome model : linear

Treatment model: none

----------------------------------------------------------------------------------------

| Robust

bweight | Coefficient std. err. z P>|z| [95% conf. interval]

-----------------------+----------------------------------------------------------------

ATE |

mbsmoke |

(smoker vs nonsmoker) | -203.0403 26.27955 -7.73 0.000 -254.5473 -151.5334

-----------------------+----------------------------------------------------------------

POmean |

mbsmoke |

nonsmoker | 3343.881 19.0128 175.88 0.000 3306.616 3381.145

----------------------------------------------------------------------------------------

**(g)** *Estimar el efecto promedio de tratamiento utilizando inverse probability weighting.*

Treatment-effects estimation Number of obs = 1,728

Estimator : inverse-probability weights

Outcome model : weighted mean

Treatment model: logit

------------------------------------------------------------------------------------------

| Robust

bweight | Coefficient std. err. z P>|z| [95% conf. interval]

-------------------------+----------------------------------------------------------------

ATE |

mmarried |

(married vs notmarried) | 103.099 28.54709 3.61 0.000 47.14773 159.0503

-------------------------+----------------------------------------------------------------

POmean |

mmarried |

notmarried | 3191.532 20.32589 157.02 0.000 3151.694 3231.37

------------------------------------------------------------------------------------------

**(h)** *Estimar el efecto promedio de tratamiento utilizando el estimador doblemente robusto.*

Treatment-effects estimation Number of obs = 1,728

Estimator : augmented IPW

Outcome model : linear by ML

Treatment model: logit

----------------------------------------------------------------------------------------

| Robust

bweight | Coefficient std. err. z P>|z| [95% conf. interval]

-----------------------+----------------------------------------------------------------

ATE |

mbsmoke |

(smoker vs nonsmoker) | -203.1215 26.27575 -7.73 0.000 -254.621 -151.622

-----------------------+----------------------------------------------------------------

POmean |

mbsmoke |

nonsmoker | 3343.847 19.01012 175.90 0.000 3306.588 3381.107

----------------------------------------------------------------------------------------

**Ejercicio 4: Estimador de Matching sin Características Observables.**

*Suponer que vale el supuesto de unconfoundedness. En este ejercicio, se propone computar la expresión analítica del estimador de PSM cuando la probabilidad de recibir el tratamiento no depende de características oservables. Utilizar el corolario del Ejercicio 2.*

ATE= E [E [ | X= 1, e (W)] - E [ | X= 0, e (W)]].

= E ( - )

= E [ - ]

= E [ - ]

= E [ - ]

= E [ ]

= E [ ]

= E []

= E (),

donde:

=

= .