



**UNIVERSIDAD NACIONAL AUTÓNOMA DE MÉXICO**  
PROGRAMA DE MAESTRÍA Y DOCTORADO EN CIENCIAS MATEMÁTICAS Y DE LA  
ESPECIALIZACIÓN EN ESTADÍSTICA APLICADA

EL DISEÑO DE REGRESIÓN DISCONTINUA

TESIS  
QUE PARA OPTAR POR EL GRADO DE:  
MAESTRA EN CIENCIAS

PRESENTA:  
MARÍA REBECA SÁMANO ALCÁNTARA

DIRECTOR DE TESIS: DR. EDUARDO ARTURO GUTIÉRREZ PEÑA  
IIMAS, UNAM

MÉXICO, FEBRERO 2016.



Universidad Nacional  
Autónoma de México

Dirección General de Bibliotecas de la UNAM

**Biblioteca Central**



**UNAM – Dirección General de Bibliotecas**  
**Tesis Digitales**  
**Restricciones de uso**

**DERECHOS RESERVADOS ©**  
**PROHIBIDA SU REPRODUCCIÓN TOTAL O PARCIAL**

Todo el material contenido en esta tesis esta protegido por la Ley Federal del Derecho de Autor (LFDA) de los Estados Unidos Mexicanos (México).

El uso de imágenes, fragmentos de videos, y demás material que sea objeto de protección de los derechos de autor, será exclusivamente para fines educativos e informativos y deberá citar la fuente donde la obtuvo mencionando el autor o autores. Cualquier uso distinto como el lucro, reproducción, edición o modificación, será perseguido y sancionado por el respectivo titular de los Derechos de Autor.



# Índice general

<b>Índice general</b>	<b>3</b>
<b>1. Introducción y Preliminares</b>	<b>1</b>
1.1. Diseños cuasi-experimentales . . . . .	2
1.2. Aplicaciones de la Regresión Discontinua . . . . .	3
1.3. El enfoque bayesiano . . . . .	4
1.4. Importancia de los métodos numéricos . . . . .	6
1.5. Estructura de la tesis . . . . .	7
<b>2. El Diseño de Regresión Discontinua</b>	<b>9</b>
2.1. Estructura del Diseño de Regresión Discontinua . . . . .	10
2.2. Ejemplo 1 . . . . .	11
2.2.1. Análisis clásico del modelo del Ejemplo 1 . . . . .	12
2.2.2. Análisis bayesiano del modelo del Ejemplo 1 . . . . .	14
2.3. Paquete rdd . . . . .	16
2.4. Consideraciones y variaciones de un Diseño de Regresión Discontinua . . . . .	19
2.4.1. Puntos de corte múltiples . . . . .	20
2.4.2. Variables de asignación múltiples . . . . .	20
2.4.3. Variantes del tratamiento . . . . .	21
2.4.4. Variantes en las escalas de medición . . . . .	21
2.4.5. Validez interna . . . . .	21
2.4.6. Potencia estadística . . . . .	23
2.5. Análisis de un Diseño de Regresión Discontinua . . . . .	24
2.5.1. Supuestos para el análisis . . . . .	24

<b>3. Inferencia Bayesiana para el DRD</b>	<b>27</b>
3.1. Los DRD <i>sharp</i> y <i>fuzzy</i> . . . . .	28
3.1.1. Diseño <i>sharp</i> . . . . .	28
3.1.2. Diseño <i>fuzzy</i> . . . . .	29
3.2. Inferencia bayesiana para un DRD . . . . .	29
3.3. Supuestos para la inferencia . . . . .	32
3.4. Inferencia bayesiana . . . . .	36
3.5. Aplicación a la evaluación de becas universitarias italianas . . . . .	41
<b>4. El Diseño de Regresión Discontinua Multivariado</b>	<b>43</b>
4.1. Diseño de Regresión Discontinua . . . . .	44
4.1.1. Ejemplo de un DRDM . . . . .	45
4.2. Diseño de Regresión Discontinua Multivariado . . . . .	46
4.3. El problema de estimación para DRDM . . . . .	48
4.4. Estrategias de estimación . . . . .	49
4.4.1. Enfoque frontera . . . . .	50
4.4.2. Enfoque univariado . . . . .	55
4.4.3. Enfoque basado en la distancia . . . . .	56
4.4.4. Enfoque centrado . . . . .	56
4.5. Simulaciones . . . . .	57
4.5.1. Simulación 1: ejemplo de un DRDM . . . . .	57
4.5.2. Simulación 2: enfoque frontera . . . . .	60
4.5.3. Simulación 3: enfoque centrado . . . . .	63
4.5.4. Análisis bayesiano del modelo del enfoque centrado . . . . .	67
4.5.5. Resultados de las simulaciones 2 y 3 . . . . .	71
<b>5. Conclusiones</b>	<b>72</b>
<b>A. Códigos en WinBUGS</b>	<b>74</b>
<b>B. Códigos en R</b>	<b>76</b>
<b>C. Método bootstrap</b>	<b>85</b>





# Capítulo 1

## Introducción y Preliminares

El Diseño de Regresión Discontinua (DRD) es un diseño cuasi-experimental usado para evaluar la efectividad de tratamientos o programas. Se diferencia de los experimentos aleatorizados sobre todo en la asignación de los sujetos de estudio. En esta última clase de experimentos los sujetos son asignados al grupo tratamiento y al grupo control (sujetos a los que no se les aplica el tratamiento) aleatoriamente. En este caso, cualquier efecto detectado se puede atribuir al programa debido a que los efectos de confusión son anulados por la asignación al azar.

Sin embargo, en muchos programas de intervención social y de educación se considera eficaz y ético ofrecer un tratamiento o programa a las personas, con base en una medida relacionada con la necesidad o el mérito. Esto implica asignar los sujetos de estudio a un grupo de tratamiento, de manera que si los sujetos están por debajo o por encima de un determinado punto de corte respecto a dicha medida, entonces pertenecen a este grupo. Los resultados del tratamiento son analizados usando las técnicas de regresión discontinua, asumiendo que el efecto del tratamiento puede ser detectado en el punto de corte a través del análisis de regresión.

Un ejemplo donde se puede aplicar un DRD es el siguiente. Se realiza un estudio de un curso de corrección de escritura para estudiantes universitarios, los cuales son asignados a tal curso cuando su calificación en un determinado examen de evaluación está por debajo de un valor determinado o punto de corte (Lee y Munk, 2008). En este tipo de situaciones no es posible realizar una asignación aleatoria, pues la asignación de los sujetos de estudio está en función del puntaje obtenido. Es decir, la asignación se basa en el valor de una variable puntaje con referencia a un punto de corte. Se pretende utilizar esta asignación en la evaluación del efecto

del programa, asumiendo que la variable respuesta es una función continua de la puntuación asignada. Si el programa tiene un efecto en la variable respuesta, entonces se verá un brinco o salto (es decir, una discontinuidad) en la línea de regresión justo en el punto de corte. El tamaño estimado del salto en esta discontinuidad se utiliza para estudiar la efectividad del programa. Por ello, el DRD es visto como un método útil para determinar la efectividad de un tratamiento o programa.

En este capítulo, se introducen algunos términos y definiciones preliminares que se emplearán en el desarrollo de este trabajo, para el análisis del DRD. Además, se dan a conocer algunas ventajas que tiene este tipo de diseños en comparación con otros diseños experimentales, y se mencionan algunas de las aplicaciones que tienen los DRD en el campo profesional. Posteriormente, se describe el enfoque bayesiano y se discute la importancia de los métodos numéricos en la implementación práctica de los diseños experimentales y cuasi-experimentales.

## 1.1. Diseños cuasi-experimentales

Un diseño cuasi-experimental es un experimento de un estudio en particular, en el que uno tiene poco o ningún control sobre la asignación de los sujetos de estudio al tratamiento, programa u otros factores de estudio ([http://en.wikipedia.org/wiki/Design\\_of\\_quasi-experiments](http://en.wikipedia.org/wiki/Design_of_quasi-experiments)). La diferencia de este enfoque empírico, en comparación con otros diseños experimentales, es la falta de asignación aleatoria. El término cuasi-experimento se refiere a que el diseño tiene semejanzas con el diseño experimental tradicional o con el ensayo clínico controlado aleatorizado (RCT, por sus siglas en inglés; diseño experimental aleatorizado utilizado normalmente en investigaciones médicas), pero no tiene el elemento de la asignación aleatoria.

La asignación al azar permite a los participantes tener la misma probabilidad de ser asignados a una condición de tratamiento dado. Como tal, la asignación aleatoria asegura que el grupo experimental y el grupo control son equivalentes. Sin embargo, si la asignación al azar no es adecuada, entonces las pruebas estadísticas pueden ser sesgadas y por lo tanto inadecuadas. En un diseño cuasi-experimental, la asignación a un tratamiento se basa en algo más que la asignación aleatoria. Dependiendo del tipo de diseño cuasi-experimental, el investigador puede tener cierto control sobre la asignación al tratamiento a través de una condición (por ejemplo, un punto de corte), o puede no tener ningún control sobre la asignación al tratamiento, y los criterios utilizados para la asignación pueden ser desconocidos. Factores como el costo, la viabilidad, y

consideraciones políticas, éticas o de conveniencia pueden influir en cómo los participantes son asignados a un tratamiento, y por lo tanto, los diseños cuasi-experimentales están sujetos a cuestionamientos relativos a la validez interna (es decir, si los resultados del experimento se pueden utilizar para hacer una inferencia causal).

Los diseños cuasi-experimentales se utilizan cuando la asignación al azar es imposible o poco práctica, y por lo general son más fáciles de implementar que los diseños experimentales, en los cuales se requiere la asignación aleatoria de los sujetos de estudio. El lector interesado puede consultar también Gibbons & Herman (1997) y Shadish, Cook & Campbell (2002).

## 1.2. Aplicaciones de la Regresión Discontinua

La Regresión Discontinua (RD), diseñada para evaluar los efectos causales de intervenciones, tratamientos o programas, fue introducido por primera vez por los psicólogos Thistlethwaite y Campbell (1960) en el área de educación. Estos métodos no atrajeron mucha atención en la literatura económica hasta hace poco. A partir de finales de los años 1990s, ha crecido el número de estudios relacionados con el DRD enfocados a economía y ha surgido una amplia aplicación de estos métodos (Cook, 2008). Algunas de estas aplicaciones han sido en las ciencias médicas y sociales (Zuckerman, 2006). A continuación se mencionan algunas de ellas.

La investigación en servicios de salud a menudo implica la evaluación de las intervenciones médicas o programas de servicio fuera de un ambiente controlado. Estas intervenciones o programas pueden ser el resultado de las políticas públicas o el resultado de la difusión de servicios de reciente introducción o de normas. El investigador de servicios de salud se enfrenta al reto de evaluar la eficacia de dichos programas. A pesar de que el RCT puede ser el diseño de estudio ideal, es con frecuencia inviable debido a limitaciones éticas, financieras, logísticas y políticas (ver Hyunshik & Munk, 2008). Los investigadores han ideado muchos diseños alternativos para evaluar el impacto de las intervenciones cuando los ensayos aleatorizados no son posibles. Éstos incluyen las comparaciones de los grupos control y tratamiento, así como series de tiempo interrumpidas. El diseño control-tratamiento es más fácil de implementar y es un método común de evaluación para la revisión de utilización de medicamentos (Gurwitz, Noonan, y Soumerai 1992; Ahluwalia et al. 1996; Farris et al. 1996; Collins et al. 1997; Monane et al. 1998). Este enfoque de dos grupos pre-post es utilizado en el DRD.

En este contexto de la evaluación de los efectos causales de las intervenciones, tratamientos

o programas, el enfoque bayesiano puede jugar un papel muy importante, ya que cada problema es único y tiene su propio contexto, del cual se deriva la información inicial sobre el parámetro (o cualquier otra característica) de interés. Además, los métodos bayesianos de inferencia se desarrollan de manera natural a partir de la Teoría Probabilística y por lo tanto no presentan contradicciones internas. A continuación se describe brevemente el enfoque bayesiano así como los métodos numéricos necesarios para su implementación. El lector interesado puede consultar el libro de Bernardo y Smith (1994).

### 1.3. El enfoque bayesiano

El propósito del enfoque bayesiano es diseñar una Teoría Estadística, basada en una pequeña serie de principios básicos (axiomas), que nos permita estructurar la solución a cualquier problema de inferencia; este tipo de enfoque le da a la Estadística una estructura coherente, ya que con otros enfoques pueden presentar casos en los que no hay una solución razonable o se presentan paradojas. Esto se logra a través de la Teoría de la Decisión.

En el contexto de la Estadística los elementos de un problema de decisión en ambiente de incertidumbre son los siguientes:

1. El espacio de acciones potenciales disponibles:  $A$ .
2. El espacio parametral, que contiene los posibles estados de la naturaleza:  $\Theta$ .
3. El espacio de las consecuencias:  $C = A \times \Theta$ .
4. Una relación de preferencia en  $C : \succeq$ .

Para poder resolver un problema de decisión, es necesario cuantificar tanto la incertidumbre sobre  $\Theta$  como las consecuencias en  $C$ . La forma de cuantificar la incertidumbre es a través de una medida de probabilidad,  $p(\theta)$ , y las consecuencias deben cuantificarse por medio de una función de pérdida,  $L(a, \theta)$ , o equivalentemente, una función de utilidad  $u(a, \theta)$ .

El resultado fundamental de la teoría es que debe elegirse aquella acción que minimice la pérdida esperada

$$L^*(a) = \int_{\Theta} L(a, \theta)p(\theta)d\theta.$$

En términos de la función de utilidad, se debe elegir aquella acción que maximice la utilidad esperada

$$u^*(a) = \int_{\Theta} u(a, \theta) p(\theta) d\theta.$$

En problemas estadísticos se cuenta con información adicional en forma de una muestra  $X_1, \dots, X_n \sim p(x|\theta)$ , la cual se puede incorporar al problema de decisión a través del Teorema de Bayes, pues nos permite combinar las dos fuentes de información,  $p(\theta)$  y  $\mathbf{x} = (x_1, \dots, x_n)$ , y de esta manera producir la distribución final  $p(\theta|\mathbf{x})$ .

El Teorema de Bayes establece que

$$p(\theta|\mathbf{x}) = \frac{p(\theta)p(\mathbf{x}|\theta)}{\int p(\tilde{\theta})p(\mathbf{x}|\tilde{\theta})d\tilde{\theta}}.$$

En este caso, la mejor acción será aquella que minimice la pérdida esperada final

$$L_x^*(a) = \int_{\Theta} L(a, \theta) p(\theta|\mathbf{x}) d\theta.$$

En términos de la función de utilidad, la mejor acción será aquella que maximice la utilidad esperada final

$$u_x^*(a) = \int_{\Theta} u(a, \theta) p(\theta|\mathbf{x}) d\theta.$$

En cierto sentido, la diferencia fundamental respecto al enfoque clásico es que en el enfoque bayesiano el (valor del) parámetro es tratado como aleatorio. Las inferencias se basan en  $p(\theta|\mathbf{x})$  en lugar de  $p(\mathbf{x}|\theta)$ ; es decir, en la distribución de probabilidad (del valor desconocido) del parámetro dados los datos observados, en vez de la distribución de los datos dado el valor del parámetro. Lo anterior da lugar a inferencias más naturales, pero para ello es necesario un ingrediente más: una distribución inicial,  $p(\theta)$ , que describe la información que se tiene sobre el valor del parámetro antes de observar los datos.

El enfoque bayesiano reconoce explícitamente que toda asignación de probabilidades es subjetiva (es decir, dependen del estado de información del individuo que las asigna). No pretende ser un enfoque “objetivo”.

En resumen, los cuatro pasos a seguir dentro del enfoque bayesiano son:

1. Especificación de un modelo (verosimilitud),  $p(\mathbf{x}|\theta)$ .
2. Especificación de una distribución inicial,  $p(\theta)$ .
3. Cálculo de la distribución final,  $p(\theta|\mathbf{x})$ , vía el Teorema de Bayes.

4. Resumen de la información contenida en  $p(\theta|\mathbf{x})$  para hacer inferencias sobre las cantidades de interés (parámetros, observaciones futuras,...).

## 1.4. Importancia de los métodos numéricos

La implementación práctica de las técnicas Bayesianas usualmente requiere de un esfuerzo computacional muy alto, para calcular ciertas características de la distribución final del parámetro de interés que permitan resumir de alguna forma la información contenida en ella. En muchos casos, estos resúmenes inferenciales se reducen a calcular integrales de la forma

$$\int q(\theta) \cdot p(\theta) \cdot p(\mathbf{x}|\theta) d\theta$$

donde la dimensión del espacio parametral de  $\theta$  puede ser muy elevada. En cualquier caso, nos enfrentamos a complicados problemas de integración que han constituido la principal dificultad del análisis bayesiano. Se han desarrollado distintos métodos de integración numérica, mediante aproximaciones determinísticas (ver Bernardo y Smith, 1994; O'Hagan, 1994; Robert y Casella, 1999).

Sin embargo, desde hace poco más de 20 años, los métodos basados en simulación Monte Carlo mediante Cadenas de Markov (MCMC, por sus siglas en inglés), permiten la resolución de problemas que hasta entonces no eran analíticamente tratables y que precisaban distintas aproximaciones numéricas para las integrales implicadas. Estos métodos permiten muestrear la distribución final gracias a la construcción de una cadena de Markov cuya distribución estacionaria es precisamente  $p(\theta|\mathbf{x})$ .

El proyecto BUGS (Bayesian Inference Using Gibbs Sampling) es un software diseñado para el análisis de modelos bayesianos usando MCMC (ver <http://www.mrc-bsu.cam.ac.uk/software/bugs/>).

WinBUGS es un software de uso general para ajustar modelos bayesianos complejos usando métodos Monte Carlo con Cadenas de Markov, los cuales en la práctica han facilitado las aplicaciones de los métodos de inferencia bayesiana. WinBUGS permite al usuario especificar un modelo bayesiano usando un lenguaje con características similares a las del paquete estadístico R (ver <http://www.R-project.org>). En función de esto, el software determina el kernel de transición de una cadena de Markov y genera muestras de la distribución final conjunta de las cantidades desconocidas de interés, especificando el número de cadenas paralelas que se correrán, el número de iteraciones, las cantidades desconocidas que se monitorearán y algunos

chequeos de convergencia. Los resultados finales de las distribuciones marginales univariadas vienen dados en resúmenes tanto numéricos como gráficos.

Otro software que permite la resolución de problemas bayesianos es el paquete estadístico R mencionado anteriormente. Dicho paquete permite realizar tareas estadísticas de todo tipo, ya que posee muchas funciones para análisis estadísticos y gráficos. Entre otras características dispone de almacenamiento y manipulación efectiva de datos, también dispone de una amplia, coherente e integrada colección de herramientas para el análisis de datos, así como gráficas, que funcionan directamente sobre pantalla o impresora, y un lenguaje de programación bien desarrollado, simple y efectivo, que incluye condicionales, ciclos, funciones recursivas y la posibilidad de entradas y salidas. La ventaja que posee este lenguaje es que permite programar, por lo tanto es posible realizar desde los cálculos más básicos hasta los más complejos, lo que a su vez, permite implementar métodos tanto de Estadística Clásica como de Estadística Bayesiana.

## 1.5. Estructura de la tesis

La importancia de los diseños de regresión discontinua radica en que pueden ser considerados métodos transparentes para la asignación de los beneficiarios a un tratamiento o programa de intervención social. Debido a que la asignación de los sujetos de estudio se basa en la necesidad o mérito, se tiene que la mayor ventaja de este tipo de diseños es que resultan ser métodos éticos para la evaluación de programas o tratamientos.

El objetivo de este trabajo es describir el análisis del Diseño de Regresión Discontinua, considerando los supuestos que se deben hacer para su estudio. En primera instancia, se analizará el DRD a través del enfoque bayesiano. Así mismo, se tiene como objetivo analizar el Diseño de Regresión Discontinua Multivariado, que es una variante del DRD simple, y se estudiarán algunos enfoques de estimación para este tipo de diseños. A continuación, se describe la estructura de esta tesis.

En el siguiente capítulo, se da la definición formal del DRD y se ilustra con un ejemplo sencillo. Este ejemplo se analiza mediante métodos clásicos y bayesianos y se comparan los resultados obtenidos. También en este capítulo, se dan a conocer algunas consideraciones y variaciones de los DRD que se deben tomar en cuenta para su análisis. Cabe mencionar que también se describe el paquete estadístico rdd, del software R, diseñado especialmente para el análisis de los DRD.

En el Capítulo 3, se describen los DRD *sharp* y *fuzzy*, y se discute el análisis del DRD *fuzzy* desde el punto de vista bayesiano, incluyendo una reseña de una aplicación del DRD *fuzzy* a la evaluación de becas universitarias italianas.

Posteriormente, en el Capítulo 4 se aborda una variante del DRD, el Diseño de Regresión Discontinua Multivariado (DRDM). En este capítulo, se describe la estructura del DRDM y se analizan algunos de los enfoques de estimación para este tipo de diseños. Por último, se estudian algunos ejemplos simulados para comparar la eficiencia de estos enfoques.

## Capítulo 2

# El Diseño de Regresión Discontinua

El Diseño de Regresión Discontinua (DRD) es un diseño cuasi-experimental que puede utilizarse para evaluar los efectos de un tratamiento, programa o intervención. En este diseño los participantes se asignan a dos grupos, control y tratamiento, con base en una puntuación límite o punto de corte. El término de Regresión Discontinua indica que el efecto del tratamiento se presenta como un “salto” o discontinuidad en la función de regresión que relaciona la variable puntaje (variable de asignación) y la variable respuesta.

La mayor ventaja de los Diseños de Regresión Discontinua es que pueden ser utilizados para estimar los efectos de un tratamiento o programa administrado a los individuos que más lo necesitan o merecen. Los individuos se asignan al tratamiento o programa de acuerdo a su “necesidad” y nada más; la asignación se hace a través de cierto puntaje (punto de corte, valor de corte) en la variable puntaje. Los Diseños de Regresión Discontinua tienen ventajas éticas sobre los diseños experimentales, debido a que no requieren que algunos individuos necesitados sean asignados a un grupo sin tratamiento (grupo control).

El DRD fue introducido por primera vez por los psicólogos Thistlethwaite y Campbell (1960), y fue discutido con detalle por Campbell y Stanley (1963). Se ha reinventado varias veces; por ejemplo, por el economista Goldberger (1972), el estadístico Rubin (1977) y por investigadores de salud pública como Finkelstein (1996). A pesar de su constante reinención, el Diseño de Regresión Discontinua no ha tenido muchas aplicaciones fuera de lo que es la educación compensatoria, donde estudiantes con calificaciones por debajo de un puntaje determinado en una prueba de rendimiento se asignan a cursos de recuperación (Trochim, 1984).

La baja frecuencia en la utilización de los DRD, puede deberse a que son novedosos o a que

los administradores requieren asignar los participantes a condiciones únicamente con base en indicadores cuantitativos, limitando el papel del juicio, la discreción o el favoritismo. Finalmente, la comprensión del diseño depende de una cierta familiaridad con el análisis de regresión, por lo que es difícil comunicar al público no estadístico.

## 2.1. Estructura del Diseño de Regresión Discontinua

En su versión más simple, el Diseño de Regresión Discontinua es un diseño con dos grupos, un antes y un después (pre-post); es decir, se considera que todos los sujetos de estudio o participantes son medidos antes y después de un tratamiento. La asignación a cada uno de estos grupos es a través de un único punto de corte, basado en una variable puntaje. Dos versiones de los grupos del Diseño de Regresión Discontinua pueden ser que el tratamiento esté siendo contrastado con una condición de ausencia del tratamiento (control), o que dos tratamientos alternativos se comparen.

Para describir la idea del diseño, supongamos que la relación entre la variable puntaje  $X$  y la variable respuesta  $Y$ , es la siguiente:

$$Y = \alpha + \beta X + \varepsilon,$$

donde  $\alpha$  y  $\beta$  son los coeficientes de regresión y  $\varepsilon$  un término de error. Después del tratamiento, si los sujetos tratados se ven afectados por un efecto de tratamiento constante  $\beta_0$  sobre la variable respuesta, entonces la regresión puede ser reescrita como:

$$Y_i = \alpha + \beta_0 T_i + \beta_1 X_i + \varepsilon_i, \tag{2.1}$$

donde  $T_i$  es la variable que asigna el valor de 1 si el tratamiento se aplicó al sujeto  $i$  y 0 si no se le aplicó (y por lo tanto se asigna al grupo control).

Debido a la suposición del efecto constante, la pendiente de la recta de regresión no cambia, pero el término intercepto cambia a  $\alpha + \beta_0$  para el grupo tratamiento. El término  $\beta_0$  sigue siendo el efecto constante.

En la Figura 2.1 se muestra un diagrama de un Diseño de Regresión Discontinua, donde los puntos a la izquierda y derecha del punto de corte pertenecen al grupo control y al grupo tratamiento, respectivamente. Así mismo, se muestra el efecto del tratamiento.

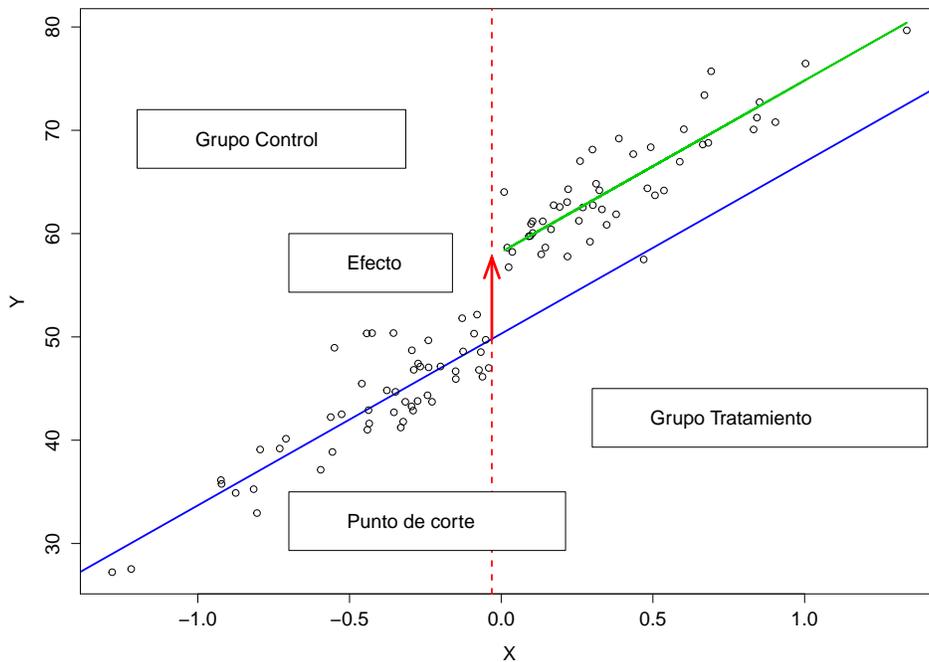


Figura 2.1: Diagrama de un Diseño de Regresión Discontinua. Se muestra la recta de regresión (línea azul), la recta de regresión proyectada en el grupo control (línea verde) y el efecto del tratamiento (flecha roja).

A continuación se muestra un ejemplo de un DRD, donde se plantean los elementos de este tipo de diseños.

## 2.2. Ejemplo 1

Para ilustrar el Diseño de Regresión Discontinua, la variable puntaje  $X$  se simuló de una distribución normal con media 0 y varianza 0.5. El punto de corte fue la media muestral de la variable  $X$ , es decir  $-0.03$ . La variable respuesta  $Y$  se modeló como sigue:

$$Y_i = 49.8 + 9.89 \cdot T_i + 14.8 \cdot X_i + \varepsilon_i, \quad i = 1, 2, \dots, n, \quad (2.2)$$

donde  $\varepsilon_i$  se distribuye como una normal de media 0 y varianza 6.63, y la variable  $T_i$  toma el valor de 1 si  $X_i > -0.03$  y toma el valor de 0 si  $X_i \leq -0.03$ . El tamaño de la muestra fue de  $n = 100$ .

En la Figura 2.2 se muestra la gráfica de la variable puntaje  $X$  contra la variable respuesta  $Y$ . En este diagrama de dispersión podemos apreciar que hay un “salto” o “discontinuidad” en  $x \approx 0$ . En la Figura 2.1 se muestra la gráfica bivariada de los datos y la recta de regresión ajustada para cada grupo del diseño. En ella se puede ver cómo el “tratamiento” tuvo un efecto “positivo” en promedio.

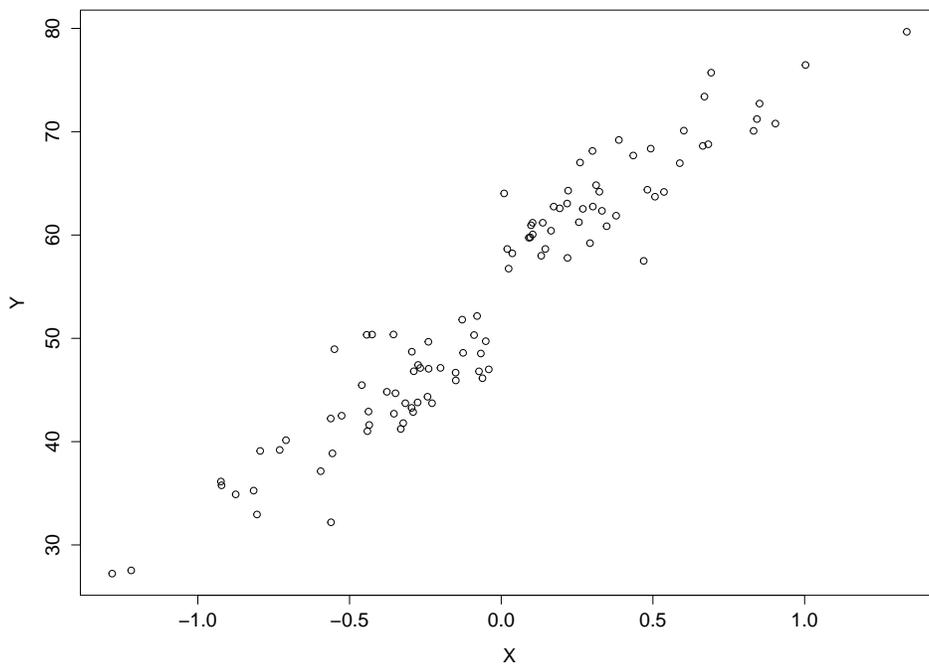


Figura 2.2: Gráfica de la variable puntaje  $X$  contra la variable respuesta  $Y$ , de una muestra de tamaño  $n = 100$ .

### 2.2.1. Análisis clásico del modelo del Ejemplo 1

A continuación se hace un análisis clásico del modelo del Ejemplo 1, descrito en la sección anterior. El análisis se realizó ejecutando un algoritmo en el paquete estadístico R, empleando la función  $lm$ , y los resultados obtenidos fueron los siguientes. Los coeficientes estimados son  $\hat{\alpha} = 50.31$  ( $ee = 0.5855$ ),  $\hat{\beta}_0 = 7.90$  ( $ee = 1.0096$ ) y  $\hat{\beta}_1 = 16.61$  ( $ee = 1.0013$ ); los valores en paréntesis son los errores estándar ( $ee$ ) correspondientes. Por ello, el modelo ajustado es:

$$\widehat{Y}_i = 50.31 + 7.90 \cdot T_i + 16.61 \cdot X_i \quad i = 1, 2, \dots, n. \quad (2.3)$$

El intervalo de confianza del 95 % para  $\beta_0$  es  $[5.919469, 9.894056]$ . De aquí, podemos decir que el “tratamiento” tiene un efecto significativo en los sujetos a los cuales se les aplicó, ya que el verdadero valor del efecto del tratamiento, 9.89, está contenido en este intervalo de confianza.

La gráfica de residuales estandarizados contra los valores ajustados, se muestra en la Figura 2.3. Esta gráfica sugiere que tanto para el grupo tratamiento como para el grupo control, el ajuste es adecuado.

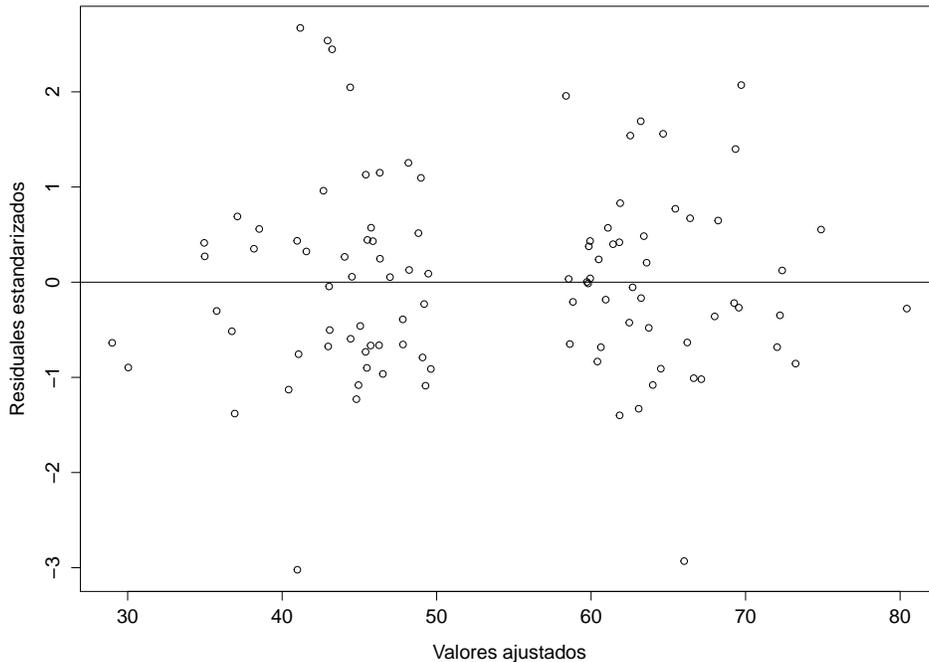


Figura 2.3: Gráfica de residuales estandarizados contra valores ajustados del modelo (2.3).

El valor del coeficiente de determinación múltiple ajustado es 0.9406811, lo cual indica que un 94.06 % de la variabilidad de  $Y$  es explicada por las variables  $X$  y  $T$ .

En la siguiente sección se hace un análisis bayesiano considerando el mismo modelo del Ejemplo 1.

## 2.2.2. Análisis bayesiano del modelo del Ejemplo 1

### Análisis no informativo

Para efectos de comparación, en este análisis bayesiano primero se utilizó una distribución inicial no informativa para los parámetros del modelo. La distribución inicial no informativa que se utilizó se aproxima a la del análisis de referencia haciendo uso del análisis conjugado (ver Gutiérrez-Peña, 1998); los detalles de la especificación de la distribución inicial se pueden ver en el Apéndice A. Aunque, se trata de un análisis de tipo conjugado se utilizó el paquete estadístico WinBUGS para fines ilustrativos del mismo.

El análisis se hizo ejecutando un algoritmo en el WinBUGS (ver Apéndice A). Después de una fase de calentamiento de 1000 iteraciones y de monitorear los nodos  $\alpha$ ,  $\beta_0$  y  $\beta_1$  por 10000 iteraciones más, se obtuvieron los siguientes resultados. Los coeficientes estimados fueron  $E(\alpha|\underline{y})=50.03$  ( $de = 0.5779$ ),  $E(\beta_0|\underline{y})=9.968$  ( $de = 0.9445$ ) y  $E(\beta_1|\underline{y})= 14.44$  ( $de = 0.873$ ), los valores en paréntesis son las desviaciones estándar ( $de$ ) de las distribuciones finales correspondientes. Por ello, el modelo ajustado es:

$$\widehat{Y}_i = 50.03 + 9.968 \cdot T_i + 14.44 \cdot X_i \quad i = 1, 2, \dots, n.$$

El intervalo de probabilidad del 95 % para  $\beta_0$  es [8.131, 11.85], el cual contiene el verdadero valor de  $\beta_0$ .

En la Figura 2.4 se muestran las gráficas para el diagnóstico de convergencia de los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ ; estas gráficas sugieren que el algoritmo converge. En la Figura 2.5 se muestran las gráficas de la densidad, para cada uno de los coeficientes del modelo. Éstas también sugieren un ajuste significativo.

### Análisis informativo

En este análisis bayesiano se utilizó una distribución inicial informativa para  $\beta_0$ , una normal de media 7 y varianza 3. De igual forma que en el análisis no informativo, se ejecutó un algoritmo en el paquete estadístico WinBUGS. Después de una fase de calentamiento de 1000 iteraciones y de monitorear los nodos  $\alpha$ ,  $\beta_0$  y  $\beta_1$  por 10000 iteraciones más, se obtuvieron los siguientes resultados. Los coeficientes estimados fueron  $E(\alpha|\underline{y})=50.39$  ( $de = 0.5268$ ),  $E(\beta_0|\underline{y})=9.29$  ( $de = 0.8323$ ) y  $E(\beta_1|\underline{y})= 14.94$  ( $de = 0.8077$ ), los valores en paréntesis son las desviaciones estándar ( $de$ ) de las distribuciones finales correspondientes. Por ello, el modelo ajustado es:

$$\widehat{Y}_i = 50.39 + 9.29 \cdot T_i + 14.94 \cdot X_i \quad i = 1, 2, \dots, n.$$

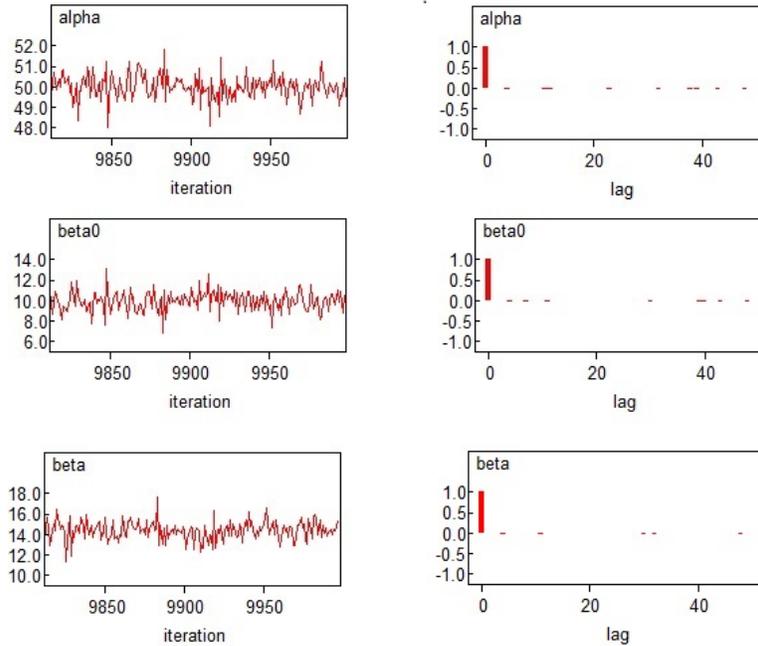


Figura 2.4: Gráficas de diagnóstico de convergencia. A la izquierda las gráficas de traza y a la derecha las gráficas de autocorrelación correspondientes a los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ , respectivamente.

El intervalo de probabilidad del 95 % para  $\beta_0$  es  $[7.651, 9.303]$ , el cual contiene prácticamente el valor verdadero de  $\beta_0$ , 9.89.

En la Figura 2.6 se muestran las gráficas para el diagnóstico de convergencia de los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ ; estas gráficas sugieren que el algoritmo converge.

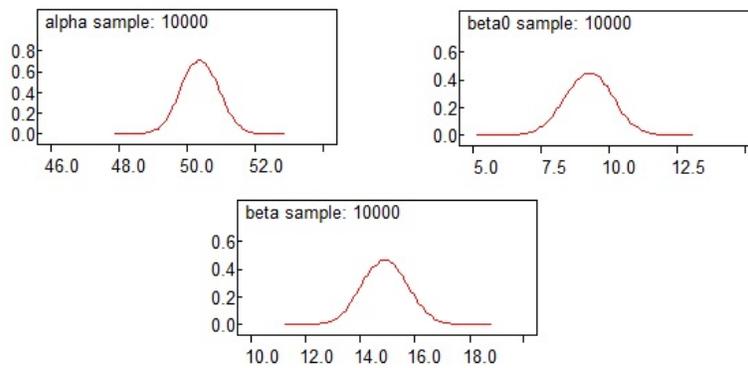


Figura 2.7: Gráficas de las densidades de los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ , respectivamente.

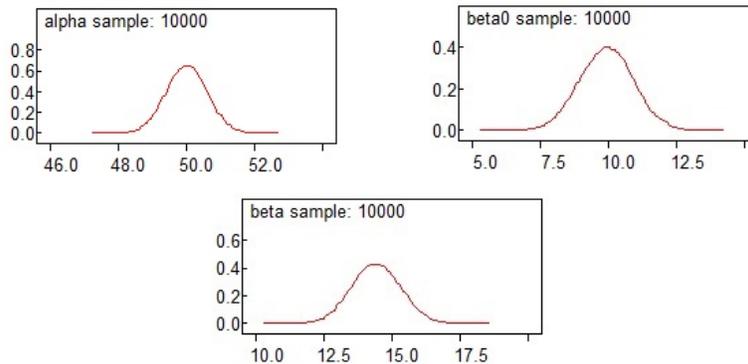


Figura 2.5: Gráficas de las densidades de los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ , respectivamente.

En la Figura 2.7 se muestran las gráficas de la densidad final, para cada uno de los coeficientes del modelo. Éstas también sugieren un ajuste significativo.

Comparando los estimadores del efecto del tratamiento obtenidos en el análisis clásico (7.90, efecto estimado del tratamiento) y en el análisis bayesiano, se tiene que el efecto estimado más cercano al valor verdadero de efecto del tratamiento, 9.89, fue el obtenido en el análisis bayesiano no informativo, 9.968. Además, el verdadero valor del tratamiento está contenido prácticamente en el intervalo del 95 % de confianza (análisis clásico) y en el intervalo del 95 % de probabilidad del análisis bayesiano informativo, y está contenido en el intervalo del 95 % de probabilidad del análisis bayesiano no informativo.

Para efectos comparativos, a continuación se describe el paquete `rdd`, implementado en R y diseñado para el análisis estadístico del DRD.

## 2.3. Paquete `rdd`

Este paquete proporciona herramientas para llevar a cabo la estimación de los diferentes tipos de diseños de regresión discontinua (<http://cran.r-project.org/web/packages/rdd/rdd.pdf>). La estimación se lleva a cabo utilizando regresión lineal local, por ambos lados del punto de corte, empleando el cálculo de ancho de banda óptimo Imbens-Kalyanaraman, denotado por I-K (Imbens & Kalyanaraman, 2009); cabe mencionar que este ancho de banda opera en la variable  $X$ .

El paquete `rdd` proporciona la función `RDestimate`, la cual sirve para la estimación de un DRD. Esta función también abarca tanto el diseño *fuzzy* como el diseño *sharp* discutidos

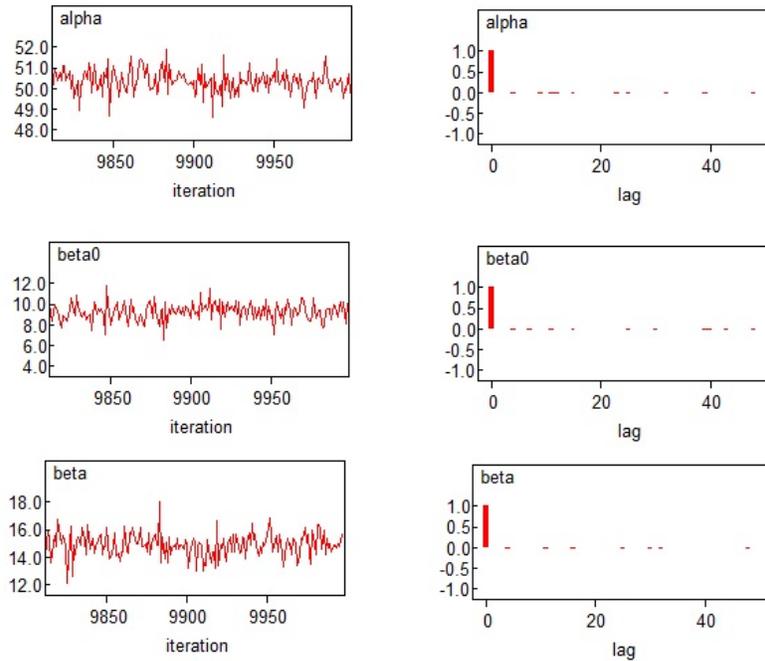


Figura 2.6: Gráficas de diagnóstico de convergencia. A la izquierda las gráficas de traza y a la derecha las gráficas de autocorrelación correspondientes a los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ , respectivamente.

en el capítulo siguiente. Algunos de los argumentos que requiere son: la fórmula del DRD, la especificación del punto de corte (si ésta se omite, el punto de corte se considera como 0), un vector numérico especificando el ancho de banda para la estimación del DRD (si ésta se omite, se considera el ancho de banda óptimo Imbens-Kalyanaraman (I-K), y se considera la mitad del ancho de banda  $((I-K)/2)$  y el doble del ancho de banda  $2(I-K)$  para la evaluación del diseño). Los métodos tradicionales para la selección del ancho de banda óptimo generalmente tratan de minimizar algún tipo error cuadrático medio, integrado a lo largo de todo el rango de los datos, por medio de técnicas de validación cruzada. Este procedimiento no es relevante en el caso del diseño de regresión discontinua, donde lo que interesa es el comportamiento en una vecindad del punto de corte. El error cuadrático medio propuesto por Imbens & Kalyanaraman (2009) depende de la función a estimar solamente a través sus dos valores en el punto de corte y de las estimaciones correspondientes, que a su vez son función de ancho de banda. Imbens & Kalyanaraman (2009) desarrollan una aproximación asintótica de este criterio, la cual es usada en la práctica para elegir el ancho de banda óptimo.

A continuación se analiza el Ejemplo 1, empleando el paquete estadístico rdd.

En la implementación de este paquete para el análisis del Ejemplo 1, se consideraron las variables  $X$  y  $Y$  descritas en el Ejemplo 1 y el punto de corte  $-0.03$ . Los resultados que se obtuvieron se resumen en el Cuadro 2.1.

	Ancho de banda	Estimador	Observaciones	Error estándar	z value
I-K	0.3547	9.369	56	1.826	5.1131***
(I-K)/2	0.1774	11.635	24	3.111	3.740***
2(I-K)	0.7094	9.329	84	1.172	7.958***
Nivel de significancia:	0***				

Cuadro 2.1: Ajuste del modelo (2.2) con el paquete rdd.

En la Figura 2.8 se muestra la gráfica obtenida empleando la función  $RDestimate$ , para el análisis del modelo descrito en el Ejemplo 1.

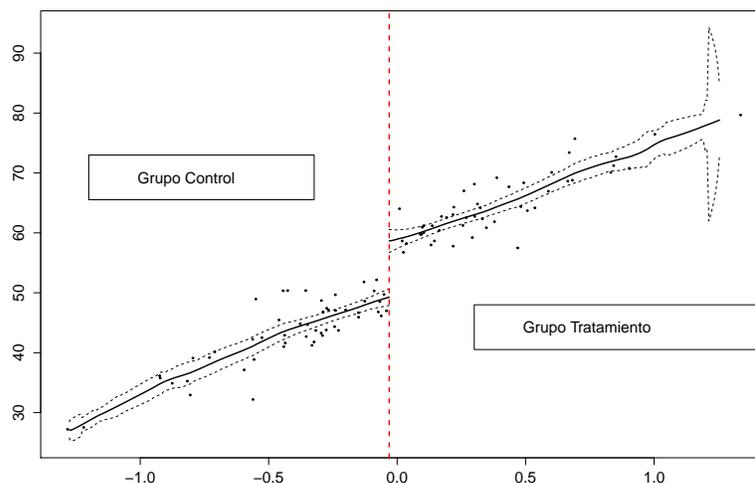


Figura 2.8: Gráfica de la variable puntaje  $X$  contra la variable respuesta  $Y$  del Ejemplo 1, utilizando el paquete rdd. Las líneas punteadas delimitan la banda de confianza al 95 %, y la línea continua muestra el modelo de regresión ajustado por la función  $RDestimate$ .

Por otro lado, en el Cuadro 2.2, se resumen los resultados del análisis del DRD descrito en el Ejemplo 1, correspondientes al análisis clásico, análisis bayesiano y los resultados obtenidos empleando el paquete rdd.

	Análisis clásico	Análisis bayesiano	Paquete rdd
Estimador del efecto	7.90	7.93	9.369
Error estándar	1.0096	1.003	1.826

Cuadro 2.2: Resultados de los ajuste para el modelo (2.2).

El efecto del tratamiento del Ejemplo 1 es 9.89. Por lo anterior, considerando el Cuadro 2.2, se tiene que el estimador más cercano a este valor es el que se obtuvo con el paquete estadístico rdd. Sin embargo, este estimador es el que tiene mayor error estándar debido a que ocupa menos observaciones del modelo para hacer el ajuste.

## 2.4. Consideraciones y variaciones de un Diseño de Regresión Discontinua

Esta sección está basada en el artículo Trochim (1990). Aquí se describen algunas consideraciones y variaciones de un DRD, las cuales se requieren para su aplicación.

1. La asignación al grupo tratamiento y control se determina, totalmente o parcialmente, a través de la variable puntaje  $X$ , por medio de un punto de corte.
2. La variable puntaje y los resultados potenciales están asociados a través de una relación suave. Cualquier discontinuidad o salto en la distribución condicional (o esperanza condicional) de los resultados como función de la variable puntaje, en el punto de corte, se interpreta como evidencia de un efecto causal del tratamiento.
3. La elección del punto de corte se puede hacer únicamente con base en los recursos disponibles o de acuerdo a argumentos circunstanciales. Aunque el uso de un punto de corte distingue el Diseño de Regresión Discontinua, a menudo es difícil de implementar. En muchas situaciones, no parece dejar espacio para el juicio profesional o la discrecionalidad.
4. Los sujetos de estudio son asignados al grupo tratamiento si su puntaje está por arriba (o abajo) del punto de corte y son asignados al grupo control si su puntaje está por abajo (o arriba) del punto de corte. Si la asignación se hace únicamente a través de la variable puntaje, el diseño es llamado “*sharp*”. Si la asignación es por medio de la variable

puntaje condicionada a una covariable o si hay “cruces” de tratamiento, el diseño es llamado “*fuzzy*” (ver Capítulo 3).

5. Por otro lado, se considera que no hay otro factor que cause la discontinuidad en el punto de corte, de modo que cualquier discontinuidad en el punto de corte se atribuye al tratamiento.
6. Otro supuesto es el principio de “no interferencias”, donde el resultado obtenido de la aplicación del tratamiento a un individuo no interfiere en el resultado del tratamiento en otros individuos.

Algunas variaciones del Diseño de Regresión Discontinua descrito hasta ahora se plantean a continuación.

### **2.4.1. Puntos de corte múltiples**

Los DRD no se limitan a un único punto de corte. El caso más sencillo de puntos de corte múltiples es el que consta de dos puntos de corte. En esta variante, los individuos con valores en la variable puntaje mayores al punto de corte superior son asignados a un tratamiento, y los individuos con valores en la variable puntaje menores al punto de corte inferior son asignados a otro tratamiento. Finalmente, los individuos con puntuación en el intervalo determinado por estos dos puntos pertenecen al grupo control.

El DRD involucra puntos de corte múltiples cuando la asignación al tratamiento se lleva a cabo por etapas, quizá comenzando con los más necesitados antes de que sucesivamente se den a los menos necesitados. En la primer etapa, se utiliza un único punto de corte, y se asignan los más necesitados al tratamiento y posteriormente se mide el efecto del tratamiento. En la segunda etapa, el valor de corte se traslada para incluir el siguiente grupo de más necesitados, con respecto a la medida de asignación. Con el tiempo, todos los participantes pueden recibir el tratamiento, pero en grupos ordenados de acuerdo a la necesidad y la etapa.

### **2.4.2. Variables de asignación múltiples**

La exigencia del cumplimiento estricto de los criterios de corte implica un indicador cuantitativo único. El uso de variables de asignación múltiples puede ayudar a incorporar más información en una estrategia de asignación cuantitativa.

La variación de un Diseño de Regresión Discontinua aquí descrita da lugar al Diseño de Regresión Discontinua Multivariado (DRDM). En el Capítulo 3 se discute este tipo de diseño.

### **2.4.3. Variantes del tratamiento**

Cuando las comparaciones de tratamientos son absolutas, el grupo control no recibe ningún tratamiento formal o, en todo caso, recibe un tratamiento placebo. Cuando son relativas, el grupo control recibe un tratamiento alternativo. Muy a menudo la intervención que recibe el grupo tratamiento se considera novedoso, y por ello debe ser comparado con los tratamientos establecidos como exitosos.

El DRD es flexible para examinar las variantes de un tratamiento. Por ejemplo, supongamos que se tienen tres variantes de tratamiento: la estándar y dos más riesgosas. Considerando dos puntos de corte podemos asignar la variante más riesgosa del tratamiento a los individuos más necesitados, los individuos menos necesitados se les asigna la variante estándar, y a los individuos comprendidos entre los dos puntos de corte, se les asigna la variante de moderado riesgo. En caso de no contar con información inicial que determine qué variante del tratamiento es más riesgosa en comparación con la otra, una alternativa es considerar un sólo punto de corte, y se asignan los individuos menos necesitados al tratamiento estándar; posteriormente los individuos restantes se asignan aleatoriamente a las variantes más riesgosas.

### **2.4.4. Variantes en las escalas de medición**

En los DRD, la variable puntaje y la variable respuesta no necesariamente deben estar en las mismas medidas o unidades. Por ejemplo, se podría asignar a los individuos a un tratamiento de educación para la salud en base a los ingresos del hogar, pero luego examinar sus efectos sobre las actitudes hacia la salud. En este caso la variable puntaje es un indicador económico y la variable respuesta es un indicador cognitivo. Estos diseños no se limitan a un solo resultado, y para cada variable respuesta se puede analizar un DRD por separado. Esto es útil para crear un resultado agregado.

### **2.4.5. Validez interna**

La validez es un aspecto importante del diseño de investigación. La validez se refiere a la pertinencia de la medida. Una medida es válida, en general, cuando mide aquello para lo cual

está diseñada. La validez interna y la validez externa son los dos tipos principales de validez; cada una de ellas refleja aspectos diferentes del diseño y de los resultados obtenidos.

La validez interna, también considerada como validez causal, se refiere al grado en que una inferencia causal es razonable. La validez interna es relevante en estudios que tratan de establecer una relación causal, no necesariamente en la mayoría de estudios observacionales o descriptivos. La pregunta clave en este caso es si los cambios observados pueden atribuirse al tratamiento aplicado y no a otras causas posibles. Otros factores que causalmente pueden inducir una discontinuidad en la relación de la variable puntaje y la variable respuesta en el punto de corte, se consideran que son amenazas a la validez interna. Por ejemplo, en un estudio de dos grupos control-tratamiento, un aumento en la puntuación del grupo control al grupo tratamiento puede ser atribuible al tratamiento o a otros posibles factores, tales como acontecimientos históricos que ocurren en el tiempo.

Muchas de estas amenazas se pueden descartar con la inclusión de un grupo control. Suponiendo que el grupo control es equivalente al grupo tratamiento, previos al estudio, el grupo control proporciona evidencia de que el cambio se debe atribuir a factores del tratamiento. Por lo tanto, los diseños experimentales aleatorizados son fuertes con respecto a la validez interna, debido a la equivalencia previa entre los grupos de comparación, la cual es resultado de la asignación aleatoria. Esto asegura que el grupo control proporcionará un reflejo legítimo de todos los factores que podrían afectar los resultados.

En los diseños que no utilizan la asignación aleatoria, la preocupación central gira en torno a su validez interna, ya que existe la posibilidad de que los grupos de comparación no sean equivalentes antes de aplicar el tratamiento.

Se emplea el término “sesgo de selección” cuando el procedimiento de selección de los grupos control y tratamiento determina sistemáticamente que existan diferencias en una o más características de estudio entre los distintos grupos, entre sí, o entre ellos y la población. Cualquier factor que se presente diferenciando los grupos de comparación (control y tratamiento) puede causar un sesgo de selección y una amenaza a la validez interna.

En los Diseños de Regresión Discontinua, debido a la deliberada diferencia entre los grupos, hay varias amenazas para la validez interna, lo cual a primera vista puede parecer un problema. Por ejemplo, una posible amenaza es el efecto de “maduración”, que se refiere a todos aquellos posibles procesos biológicos o psicológicos que varían de modo más o menos sistemático en función del tiempo, independientemente de ciertos acontecimientos externos, como puede ser

el propio experimento. El aumento de estatura o edad, así como la fatiga de los individuos de estudio, son ejemplos de este efecto. Este tipo de efecto es difícil de controlar, sobre todo si el periodo de tiempo transcurrido en el experimento es demasiado largo, ya que los cambios ocurridos durante el mismo en los individuos de estudio pueden deberse a los efectos de maduración y no a la variable independiente. Otra amenaza a la validez interna es el fenómeno de “regresión a la media” que surge cuando se considera una muestra heterogénea de la población. Este efecto se refiere al hecho de que la media muestral del grupo tratamiento es más cercana (en unidades estandarizadas) a la media de la población, en comparación con la media muestral del grupo control que es más lejana de la media muestral de la población. En otras palabras, del grupo control al grupo tratamiento la media muestral parece regresar a la media poblacional. En los diseños de regresión discontinua, se crean deliberadamente muestras heterogéneas y por lo tanto se espera una regresión a la media en ambos grupos. El término regresión fue utilizado originalmente para referirse al hecho de que una línea de regresión describe la regresión a la media.

Aunque inicialmente el DRD puede parecer susceptible a los sesgos de selección, en realidad no lo es. En principio, el diseño de regresión discontinua es tan fuerte en la validez interna como sus alternativas experimentales aleatorizadas, debido a que el proceso de selección de los tratamientos se conoce totalmente y sólo aquellos factores que causalmente podrían inducir una discontinuidad en la relación *pre-post* en el punto de corte, pueden ser considerados una amenaza.

Sin embargo, en la práctica la validez también depende de cómo el analista puede modelar la relación *pre-post* para que una posible relación no lineal subyacente no se disfrace como discontinuidad en el punto de corte.

#### **2.4.6. Potencia estadística**

La discusión anterior argumenta que el DRD es fuerte con respecto a la validez interna, sin duda más fuerte que el diseño de grupos no equivalentes y quizás tan fuerte como los experimentos aleatorizados. Sin embargo, se sabe que los diseños de regresión discontinua no son tan potentes estadísticamente como los experimentos aleatorizados (Trochim, 1990). Específicamente, se ha visto que, con el fin de alcanzar el mismo nivel de precisión estadística, un DRD necesita 2.75 veces más participantes que un experimento aleatorizado (Goldberger, 1972). Por ejemplo, si un experimento aleatorizado necesita de 100 participantes para alcanzar un

cierto nivel de “potencia estadística”, el DRD podría llegar a necesitar hasta 275 participantes. El atractivo real de los Diseños de Regresión Discontinua, es que permite asignar el tratamiento o programa a los sujetos de estudio que más lo necesitan o merecen. Por lo tanto, el verdadero atractivo del DRD es ético, con lo cual no se tiene que negar el programa o el tratamiento a los participantes que pudieran necesitarlo, como lo hacemos en estudios aleatorizados.

## 2.5. Análisis de un Diseño de Regresión Discontinua

Para el análisis de un DRD, se necesita un modelo estadístico que incluya un término que represente al grupo control, uno para el grupo tratamiento y una variable *dummy* que representa si el programa se aplicó o no. A continuación se discuten los supuestos principales para el análisis de este diseño. Esta sección también se basa en el artículo de Trochim (1990).

### 2.5.1. Supuestos para el análisis

Es importante que antes de discutir el análisis del modelo de regresión discontinua se especifiquen los supuestos que deben cumplirse. Considerando el diseño básico de RD, descrito anteriormente, hay cinco supuestos centrales que deben cumplirse para el análisis del modelo, cada uno de los cuales se discuten a continuación:

1. *El criterio de corte.* Este criterio debe seguirse sin excepción. Cuando hay una mala asignación en relación con el punto de corte (a menos que se conozca que la asignación es aleatoria) surge una amenaza a la validez interna del diseño; por ello es probable que las estimaciones del efecto del programa sean sesgadas.

2. *La relación de la variable puntaje y la variable respuesta.* Se supone que la relación entre estas variables se puede describir como una función polinomial. Si la relación verdadera es logarítmica, exponencial o alguna otra función, el modelo estará mal especificado, y es probable que las estimaciones del efecto del programa sean sesgadas. Sin embargo, si los datos se pueden transformar antes del análisis para crear una relación polinomial, el modelo puede ser adecuado, aunque es posible que sea más difícil de interpretar.

3. *Varianza de la variable puntaje en el grupo de control.* Debe existir un número suficiente de valores de la variable puntaje en el grupo control que permita la estimación adecuada de la línea de regresión pre-post para ese grupo. Por lo general, es conveniente conocer la variabilidad en el grupo tratamiento, aunque no es estrictamente necesario, pues se puede proyectar la línea

de regresión del grupo de control, incluso a un solo punto del grupo tratamiento.

4. *Continuidad de la distribución del grupo control.* Los grupos control y tratamiento deben provenir de una distribución del p continua, con sólo una discontinuidad entre los grupos determinada por el punto de corte. En algunos casos se pueden encontrar grupos homogéneos (por ejemplo, dos grupos de pacientes de dos áreas geográficas diferentes) que casualmente se dividen en alguna medida con el fin de implicar algún punto de corte. Tales grupos, naturalmente discontinuos, deben utilizarse con precaución debido a que se tiene mayor probabilidad de que si difieren en el punto de corte antes del tratamiento, tal diferencia podría reflejar un sesgo de selección que podría introducir discontinuidades entre los grupos control y tratamiento en ese punto.

5. *Programa de ejecución.* Se supone que el programa o tratamiento se entrega de manera uniforme a todos los destinatarios, es decir, que todos reciben la misma dosis, la misma duración de estancia, igual intensidad de entrenamiento, etc. Si este no es el caso, es necesario modelar explícitamente el tratamiento tal como se aplica, lo que complica el análisis.

Con el fin de interpretar los resultados obtenidos en el análisis, se debe conocer la naturaleza de las variables. Sin esta información, no hay un patrón de resultados que indique directamente si un efecto de tratamiento obtenido es positivo o negativo. Tiene sentido intuitivamente que la responsabilidad de un tratamiento o programa dependa en gran medida de la transparencia de la asignación de los beneficiarios al tratamiento.

Los tres principales diseños de grupos pre y post son los experimentos aleatorizados, el diseño de regresión discontinua y el diseño de grupos no equivalentes. Estos son análogos a los tres tipos de esquemas de asignación de tratamientos que los administradores pueden elegir. Los experimentos aleatorizados son análogos a la utilización de un sorteo para la asignación del tratamiento. Los diseños de regresión discontinua pueden ser considerados métodos transparentes de la asignación al tratamiento en base a la necesidad o mérito. Por último, los diseños no equivalentes puede ser considerados como un tipo de asignación política, ya que permiten el uso de una asignación no verificable, subjetiva o de motivación política. La mayoría de los tratamientos o programas sociales son asignados políticamente. Incluso cuando los tratamientos se asignan principalmente en función de la necesidad o el mérito, el organismo regulador por lo general se reserva una cierta capacidad discrecional para decidir quién recibe el programa. Sin discutir la necesidad de tal discrecionalidad, es claro que se debe alentar a los administradores para que utilicen diseños tales como el de regresión discontinua. Conforme los administradores

puedan estar convencidos de avanzar hacia la elección de criterios de asignación más explícitos, la utilidad potencial del DRD se incrementará.

# Capítulo 3

## Inferencia Bayesiana para el DRD

En los capítulos anteriores se mencionó que el Diseño de Regresión Discontinua surge de la necesidad de evaluar la efectividad de tratamientos o programas de intervención social o de educación, donde la participación de los sujetos de estudio a estos tratamientos o programas está sujeta a decisiones administrativas. Además, la asignación de incentivos para participar en los tratamientos o programas son limitados debido a la escasez de recursos y al criterio de los administradores.

Existen dos tipos de Diseños de Regresión Discontinua (DRD), el diseño *sharp* y el diseño *fuzzy*. En el diseño *sharp*, el estatus de tratamiento se elige por medio de una función determinista llamada variable puntaje. Todos los individuos con un valor de la variable puntaje mayor o menor a un determinado punto de corte, se asignan a un estatus de tratamiento. En los diseños *fuzzy*, el valor de la variable puntaje no es lo único que determina la recepción del tratamiento, aunque puede actuar como un estímulo o incentivo para participar en el tratamiento. En este caso, la recepción del tratamiento también depende de elecciones individuales, que plantean problemas de autoselección.

En este capítulo se reseña el artículo de Li, Mattei & Mealli (2014) cuyo objetivo es estudiar el DRD *fuzzy* a través del enfoque bayesiano. En la primer sección se describen de manera breve los diseños *sharp* y *fuzzy*; posteriormente se describe a detalle el DRD *fuzzy*, visto desde el enfoque bayesiano. Esta descripción se hace a través de la aplicación del DRD *fuzzy* a la evaluación de becas universitarias italianas.

### 3.1. Los DRD *sharp* y *fuzzy*

Como ya se ha mencionado, en los DRD la asociación de la variable puntaje  $X$  con la variable respuesta  $Y$  se supone suave, por lo cual cualquier discontinuidad en la distribución condicional de la variable  $Y$  dado  $X$ , en el punto de corte, se interpreta como evidencia del efecto causal del tratamiento.

Asumimos que, para cada individuo de estudio  $i$ , existen dos resultados posibles, la observación obtenida de la variable respuesta dado que no se aplicó el tratamiento y la observación obtenida de la variable respuesta dado que se aplicó el tratamiento,  $Y_i(0)$  y  $Y_i(1)$ , respectivamente. El efecto causal se define por la diferencia  $Y_i(1) - Y_i(0)$ .

A continuación se describen brevemente los diseños *sharp* y *fuzzy*. El lector interesado puede consultar Imbens & Lemieux (2008) y Wong, Steiner & Cook (2012).

#### 3.1.1. Diseño *sharp*

Este tipo de diseño se estudió en el Capítulo 2. En el diseño *sharp* la variable indicadora  $T$  es una función determinista de la variable puntaje  $X$ , es decir

$$T_i = T_i(x_0) = \mathbf{1}\{X_i \geq x_0\} \text{ ó } T_i = \mathbf{1}\{X_i \leq x_0\},$$

donde  $\mathbf{1}\{\}$  denota a la función indicadora. El individuo  $i$  se asigna al tratamiento si el valor de la variable puntaje  $X_i$  es mayor o igual al punto de corte  $x_0$ , y se asigna al grupo control si  $X_i$  es menor que  $x_0$ . Alternativamente, el individuo  $i$  se asigna al tratamiento si el valor de la variable puntaje  $X_i$  es menor o igual a  $x_0$ , y se asigna al grupo control si  $X_i$  es mayor que  $x_0$ . En este tipo de diseños se asume que la probabilidad de recibir el tratamiento cambia de 0 a 1, es decir,

$$\lim_{x \uparrow x_0} E[Y_i|X_i = x] - \lim_{x \downarrow x_0} E[Y_i|X_i = x] = 1.$$

Considerando lo anterior, el parámetro de interés está dado por:

$$\tau_S = E[Y_i(1) - Y_i(0)|X_i = x_0]. \tag{3.1}$$

Asumiendo que la variable respuesta es continua en el punto de corte, entonces  $\tau_S$  se interpreta como el efecto causal promedio del tratamiento.

### 3.1.2. Diseño *fuzzy*

En el diseño *fuzzy*, la probabilidad de recibir el tratamiento no necesariamente cambia de 0 a 1. Sin embargo se debe cumplir que:

$$0 < \lim_{x \uparrow x_0} E[Y_i | X_i = x] - \lim_{x \downarrow x_0} E[Y_i | X_i = x] < 1.$$

El efecto causal promedio del diseño *fuzzy* es distinto a (3.1). En el diseño *fuzzy*, además de la hipótesis de continuidad de la variable respuesta, se requiere que no existan individuos “desafiantes”, cuyo valor en  $X$  sea cercano al punto de corte. Un individuo se considera “desafiante” si participa en el tratamiento cuando fue asignado al grupo control, o es parte del grupo control cuando fue asignado al grupo tratamiento. Si las dos hipótesis anteriores se cumplen, entonces se identifica el estimador causal del efecto del tratamiento en el punto de corte para una subpoblación de individuos “cumplidores”. Un individuo  $i$  se considera cumplidor (C), si

$$\lim_{x \uparrow x_i} T_i(x) = 1 \text{ y } \lim_{x \downarrow x_i} T_i(x) = 0,$$

es decir, si el individuo se asigna al grupo tratamiento entonces toma el tratamiento, o si se asigna al grupo control entonces permanece en el grupo control. El efecto causal promedio para el diseño *fuzzy* está dado por:

$$\tau_F = E[Y_i(1) - Y_i(0) | X_i = x_0, i \in C].$$

A continuación se discute el DRD *fuzzy*, considerando el enfoque bayesiano, a través una aplicación a la evaluación de becas universitarias italianas.

## 3.2. Inferencia bayesiana para un DRD

Algunas de las razones por las cuales es importante adoptar el enfoque bayesiano en el estudio de los DRD son las siguientes. En primer lugar, la inferencia causal en los DRD generalmente implica datos observacionales complejos, con múltiples fuentes de incertidumbre, incluidos datos faltantes. En segundo lugar, el análisis de regresión discontinua por lo general se basa en una muestra de individuos con valores en la variable puntaje cercanos a un solo punto, y el tamaño de dicha muestra puede ser pequeño; los métodos bayesianos no dependen de aproximaciones asintóticas, y por ello son atractivos cuando se consideran muestras pequeñas. En tercer lugar, en el paradigma bayesiano los resultados potenciales faltantes (no observables) son tratados

como variables aleatorias y todas las inferencias se basan en las distribuciones finales. Así, la inferencia sobre muestras finitas y poblaciones grandes se puede obtener usando los mismos procedimientos de inferencia. Finalmente, las covariables pueden ser fácilmente incorporadas en el enfoque bayesiano y pueden mejorar la eficiencia del análisis; es decir, reducir la variabilidad de la distribución final.

El principal objetivo del estudio descrito en este capítulo es investigar los efectos de las becas universitarias italianas para dar igualdad de oportunidades y poder bajar la tasa de deserción escolar en la educación superior, evitando así que estudiantes pertenecientes a familias de bajos ingresos abandonen la educación superior. En el sistema universitario italiano, la asignación de una beca se basa en los criterios de elegibilidad y una solicitud voluntaria. Sólo los estudiantes que cumplen los criterios de elegibilidad y solicitan una beca puede recibir tal ayuda. El estatus de elegibilidad depende de una medición económica de los ingresos y bienes (variable puntaje) de la familia del estudiante en relación a un determinado punto de corte (en este caso, el punto de corte es 15 000 euros). Esta norma define el DRD *fuzzy*, en el sentido de que no todos los estudiantes elegibles, con valor en la variable puntaje menor al punto de corte, consiguen una beca. Los estudiantes elegibles deben solicitarla para recibir tal ayuda. Los estudiantes no elegibles suelen solicitarla, ya sea porque no son conscientes de su condición de elegibilidad, o porque esperan que su solicitud se tome en cuenta quizás a causa de fondos adicionales u otras consideraciones.

A continuación se proporciona una definición formal del DRD *fuzzy*, visto como un diseño experimental aleatorizado local, basándose en una descripción del mecanismo de asignación, es decir, en el proceso que describe el por qué los individuos quedan asignados a los diferentes tratamientos. Primero se definen los tratamientos y los posibles resultados. Posteriormente, se presenta una formulación probabilística del mecanismo de asignación de la configuración general del DRD. Para ello, se asume que existe una pequeña subpoblación alrededor del punto de corte en donde puede suponerse explícitamente que hay un traslape local, reemplazando el supuesto usual de continuidad por el supuesto de aleatorización local dentro de dicha subpoblación. Por último, se introducen algunos supuestos estructurales que mejoran la inferencia.

A continuación se introduce la notación que se va a emplear en esta sección.

**Notación:**

$X$ : variable puntaje. (Medida combinada de ingresos, ajustada por el tamaño de la familia).

$x_0 = 15\ 000$  euros; punto de corte.

$T$  : función determinista de  $X$ ,  $T_i = 1(X_i \leq x_0)$ ;  $t \in \{0, 1\}$ ,  $t = 1$  indica que el estudiante es elegible y  $t = 0$  indica que el estudiante no es elegible.

$$A_i(t) = \begin{cases} 1 & \text{si el estudiante } i \text{ solicita la beca} \\ 0 & \text{el estudiante } i \text{ no solicita la beca} \end{cases}$$

$$W_i(t) = \begin{cases} 1 & \text{si el estudiante } i \text{ recibe la beca} \\ 0 & \text{el estudiante } i \text{ no recibe la beca} \end{cases}$$

$Y_i(t)$ : resultado (es decir, deserción del estudiante  $i$ ).

$A_i^{obs} = A_i(T_i)$  : estatus observado de solicitud.

$W_i^{obs} = W_i(T_i)$  : estatus observado de recepción de la beca.

$Y_i^{obs} = Y_i(T_i)$ : resultado observado.

$G_i = (A_i(0), A_i(1))$  : valores conjuntos del estatus de solicitud.

$\mathbf{R}$  : matriz de covariables de dimensión  $n \times p$ .

En esta matriz el  $i$ -ésimo renglón será denotado por  $\mathbf{R}_i$ .

En el contexto de evaluación de becas, para definir los parámetros causales de mayor interés y su conexión con el efecto de recibir una beca, los individuos se pueden clasificar en cuatro estratos principales: 1)  $G_i = (0, 1) = c$ , los estudiantes que no son elegibles no solicitan beca, pero que sí son elegibles si solicitan beca ; 2)  $G_i = (1, 1) = a$ , los estudiantes que sí solicitan beca independientemente de su estado de elegibilidad; 3)  $G_i = (0, 0) = n$ , los estudiantes que no solicitan beca independientemente de su elegibilidad y 4)  $G_i = (1, 0) = d$ , los estudiantes que no son elegibles si solicitan beca, y los que si son elegibles pero que no solicitan beca.

La estratificación de los estudiantes con respecto al estatus de solicitud de beca, muestra que los solicitantes elegibles y no elegibles observados son poblaciones heterogéneas: la primera es una mezcla de 1) y 2), y la otra es una mezcla de 2) y 4).

Como resultado, un análisis simple que compara directamente los resultados observados entre los solicitantes elegibles y no elegibles por lo general lleva a estimaciones sesgadas de los efectos causales. La estratificación también deja claro que sólo en los estratos 1) y 2), podemos observar el resultado para algunos estudiantes elegibles que reciben una beca y el resultado para otros estudiantes no elegibles que no reciben una beca. Los estudiantes pertenecientes a los otros dos estratos, por el contrario, nunca reciben una beca, por lo que podemos sólo observar sus resultados en ausencia de una beca. Por lo tanto, sacar conclusiones para el efecto causal de la subvención sólo es posible en los estratos 1) y 2). En este caso, el conjunto de parámetros de principal interés son los efectos causales en la población promedio.

El efecto de interés considerando el contexto anterior está dado por:

$$\tau = E[Y_i(1) - Y_i(0)|G_i = \{a, c\}].$$

Sean  $\tau_a = E[Y_i(1) - Y_i(0)|G_i = a]$ ,  $\tau_c = E[Y_i(1) - Y_i(0)|G_i = c]$ ,  $\pi_a = \Pr(G_i = a)$  y  $\pi_c = \Pr(G_i = c)$ , donde  $g \in \{a, c\}$ . Entonces,

$$\tau = \frac{\tau_a \pi_a + \tau_c \pi_c}{\pi_a + \pi_c}.$$

En el análisis de estudios observacionales, un paso preliminar importante es evaluar cuidadosamente la posibilidad de conceptualizar que los datos observados surgieron a partir de un experimento aleatorizado complejo, donde el mecanismo de asignación al tratamiento se ha perdido y debe ser reconstruido.

A continuación se introduce formalmente el mecanismo de asignación probabilística para el DRD que se está considerando. La clave de esta propuesta radica en la reconstrucción del experimento hipotético bajo el DRD, viendo la variable puntaje como una variable aleatoria con una distribución de probabilidad.

### 3.3. Supuestos para la inferencia

En esta sección se plantean los supuestos a considerar en este enfoque, que garantizan la existencia de la subpoblación  $U_{x_0}$ , donde se asume el supuesto de aleatorización local. Así mismo, se define el mecanismo de asignación probabilística para el DRD en cuestión.

El supuesto 1, garantiza que se puede ver a la variable puntaje como una variable aleatoria con una distribución de probabilidad.

**Suposición 1.** (*Stable Unit Treatment Value Assumption*). Consideremos dos estados de elegibilidad  $T'_i = \mathbf{1}(X'_i \leq x_0)$  y  $T''_i = \mathbf{1}(X''_i \leq x_0)$ , donde  $X'_i \neq X''_i$ . Si  $T'_i = T''_i$ , es decir, si  $X'_i \leq x_0$  y  $X''_i \leq x_0$ , ó  $X'_i > x_0$  y  $X''_i > x_0$ , entonces  $A_i(\mathbf{T}') = A_i(\mathbf{T}'')$ ,  $W_i(\mathbf{T}') = W_i(\mathbf{T}'')$  y  $Y_i(\mathbf{T}') = Y_i(\mathbf{T}'')$ .

Este supuesto puede ser controversial para toda la población de estudio, pues depende del máximo de  $X$  para cada individuo  $i$ . Sin embargo, parece ser válido para una subpoblación de individuos que tienen mayor probabilidad de que los valores realizados de  $X$  caigan en una vecindad alrededor de  $x_0$ , la cual se define más adelante.

El mecanismo de asignación, es una función que asigna probabilidades a todos los  $2^n$  posibles vectores de asignación  $\mathbf{T}$  de dimensión  $n$ , como una función que asigna probabilidades a todos

los posibles vectores de dimensión  $n$  de las realizaciones de la variable puntaje  $X$ , mayores o menores al punto de corte  $x_0$ . Es decir,

$$\begin{aligned} & \Pr(\mathbf{T}|\mathbf{A}(0), \mathbf{A}(1), \mathbf{W}(0), \mathbf{W}(1), \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{R}) \\ = & \Pr(\mathbf{X} \in \mathbf{\Lambda}|\mathbf{A}(0), \mathbf{A}(1), \mathbf{W}(0), \mathbf{W}(1), \mathbf{Y}(0), \mathbf{Y}(1), \mathbf{R}), \end{aligned}$$

donde  $\mathbf{T} \in \{0, 1\}^n$  y

$$\mathbf{\Lambda} \in \left\{ \begin{array}{l} (-\infty, x_0]^n, (-\infty, x_0]^{n-1} \times (x_0, \infty), (x_0, \infty) \times (-\infty, x_0]^{n-1}, \\ \dots, (-\infty, x_0] \times (x_0, \infty)^{n-1}, (x_0, \infty)^{n-1} \times (-\infty, x_0], (x_0, \infty) \end{array} \right\}.$$

Se asume que este mecanismo es localmente independiente, es decir, el mecanismo de asignación es separable en probabilidades de asignación de individuos, y no depende de las covariables o resultados potenciales de otros individuos. En principio, para cada estudiante, cada valor de la variable puntaje es posible *a priori*, es decir, antes de que ocurra una realización de la variable puntaje. Por ello es razonable asumir que el mecanismo de asignación antes de la realización de la variable puntaje es probabilístico, lo cual implica que para cada estudiante  $i$ , ambos eventos  $X_i \leq x_0$  y  $X_i > x_0$  tienen *a priori* una probabilidad no cero de ocurrencia.

El mecanismo de asignación localmente independiente de los resultados potenciales define los experimentos aleatorizados clásicos. Las reglas de asignación fundamentales de los diseños de RD sugieren que estos supuestos son más razonables para una subpoblación de individuos que tienen relativamente una mayor probabilidad de que sus valores realizados de la variable puntaje caigan en una vecindad del punto de corte  $x_0$ . Para esta subpoblación, se supone que la distribución de la variable puntaje no está relacionada con las características observables y no observables de los estudiantes. La siguiente suposición garantiza que dicha subpoblación exista.

**Suposición 2.** (*Local overlap*) Sea  $U$  una muestra aleatoria (o población) de individuos de estudio. Existe un subconjunto de individuos,  $U_{x_0}$ , tal que para cada  $i \in U_{x_0}$ ,  $\Pr(X_i \leq x_0) > \epsilon$  y  $\Pr(X_i > x_0) > \epsilon$  para  $\epsilon > 0$  suficientemente grande.

En esta suposición se asume que existe una subpoblación de individuos, cada uno de los cuales tiene una probabilidad distinta de cero de ser asignado a cualquiera de los estatus de tratamiento. La siguiente suposición ayuda en la práctica a la construcción de tal subpoblación; además indica cuáles estudiantes pertenecientes a  $U_{x_0}$  tienen un valor de la variable puntaje  $X$  que cae en una vecindad de  $x_0$ .

**Suposición 3.** Existe  $h > 0$  tal que para cada  $\epsilon > 0$ ,  $\Pr(x_0 - h \leq X_i \leq x_0 + h) > 1 - \epsilon$ , para cada  $i \in U_{x_0}$ .

La Suposición 4, presentada a continuación, formaliza el concepto del DRD visto como un experimento aleatorizado local; es decir, que en una vecindad del punto de corte  $x_0$ , la variable puntaje no depende de cualquiera de los resultados potenciales o de covariables.

**Suposición 4** (*Local randomization*). Para cada  $i \in U_{x_0}$ ,

$$\Pr(X_i | A_i(0), A_i(1), W_i(0), W_i(1), Y_i(0), Y_i(1), \mathbf{R}_i) = \Pr(X_i).$$

Esta suposición implica que los valores de la variable puntaje mayores o menores que el punto de corte  $x_0$ , y por tanto los estados de elegibilidad, se asignan al azar en una pequeña vecindad,  $U_{x_0}$ , alrededor de  $x_0$ .

Así mismo, la aleatorización local implica que para cada  $i \in U_{x_0}$ , los valores de la distribución de probabilidad de los resultados potenciales, condicionada en la variable puntaje  $X$ , son constantes, y por ello la distribución es continua en  $x$ . Sin embargo, la suposición de continuidad estándar en la literatura de los diseños RD es un supuesto diferente y es menos estricto que la aleatorización local, que implica la independencia del mecanismo de asignación y los resultados posibles sólo en el punto de corte. Dado el supuesto de continuidad, los métodos existentes suelen centrarse en los parámetros causales definidos en un único punto de discontinuidad ( $x_0$ ) y acoplan el supuesto de continuidad con los procedimientos de estimación específicos tales como polinomios locales, con el objetivo de encontrar un equilibrio óptimo entre la precisión y el sesgo. Por el contrario, el método propuesto no está ligado a ningún procedimiento de estimación en particular, y la inferencia se basa por completo en los supuestos estructurales del 1 al 4. Los supuestos 5 y 6, que se mencionan a continuación, en la práctica facilitan la identificación paramétrica de los efectos causales.

**Supuesto 5.**  $A_i(1) \geq A_i(0)$  para todo  $i \in U_{x_0}$ .

**Supuesto 6.**  $\Pr(Y_i(1) | G_i = n, i \in U_{x_0}) = \Pr(Y_i(0) | G_i = n, i \in U_{x_0})$ .

Si se conociera la subpoblación  $U_{x_0}$ , de acuerdo con los supuestos mencionados anteriormente, el efecto causal correspondiente al estrato a) sería identificable y estimable puntualmente de manera no paramétrica a través del método de variables instrumentales. Sin embargo, los efectos causales para los estratos 1) y 2) sólo son parcialmente identificables desde el punto de vista no paramétrico bajo los supuestos dados anteriormente. La inferencia de estos estimadores se basa en aproximaciones para muestras grandes, por ello pueden no ser confiables en algunos

casos debido a que la muestra es pequeña en  $U_{x_0}$ . Por otra parte, en aplicaciones reales, la subpoblación  $U_{x_0}$  es generalmente desconocida y debe ser seleccionada por el investigador.

A pesar de que las covariables no son necesarias para la estimación de los efectos del tratamiento, incluirlas en el análisis mejorarán la precisión de la inferencia causal y, con ello, los resultados serán menos sensibles a la elección de  $U_{x_0}$ .

Para la selección de  $U_{x_0}$  se propone un método de selección basado en las covariables (ver Cattaneo, Frandsen & Titiunik, 2015). Este método puede basarse tanto en covariables predeterminadas (determinadas antes de asignar el tratamiento, por lo tanto no se ven afectadas por el mismo) como en covariables placebo (determinadas después de la asignación del tratamiento; sin embargo, se espera que se vean afectadas por el tratamiento dado el conocimiento teórico previo de como opera el mismo). En la mayoría de las aplicaciones empíricas del DRD, los investigadores tienen acceso a covariables predeterminadas y como ya se menciono anteriormente estas son utilizadas para reducir la variabilidad de la distribución final.

En la práctica, para seleccionar la subpoblación  $U_{x_0}$  es necesario elegir el tamaño del intervalo,  $h$ , en torno al punto de corte. Para seleccionar  $h$ , se comienza eligiendo un valor pequeño alrededor de  $x_0$ . Se considera un test de aleatorización y se hace el contraste de hipótesis; se considera como hipótesis nula que no hay efecto del tratamiento, esto significa que los resultados observados son fijos e independientes de la asignación a  $U_{x_0}$ . Posteriormente se aumenta el valor de  $h$ , hasta que se rechaza la hipótesis nula, lo cual puede interpretarse como evidencia contra la Suposición 4. El valor más grande de  $h$ , para el cual la hipótesis nula no fue rechazada es el óptimo. Una característica importante de este enfoque es que se interesa por la posibilidad de no rechazar la hipótesis nula cuando esta es falsa (hipótesis de error tipo II). En este contexto, la hipótesis nula se utiliza como prueba de que la aleatorización local no se cumple, y el objetivo final es saber si los datos apoyan la existencia de una vecindad alrededor del punto de corte donde la hipótesis nula no se puede rechazar.

Para el contraste de hipótesis se pueden emplear diferentes pruebas estadísticas tales como el estadístico diferencia de medias, el de suma rangos de Wilcoxon o el estadístico Kolmogorov-Smirnov, esto debido a que la distribución nula de estas pruebas es conocida.

En la práctica es poco probable que el ancho de banda elegido coincida exactamente con el valor verdadero. Por ello, la sensibilidad de los resultados con respecto a la elección de  $h$  debe ser evaluada. Para más detalles en cuanto a la elección de  $h$ , el lector puede consultar Cattaneo, Frandsen & Titiunik (2015).

En la siguiente sección se desarrolla la inferencia bayesiana para estimar los efectos causales que tienen las becas universitarias italianas.

### 3.4. Inferencia bayesiana

El desarrollo del enfoque bayesiano que se aborda en esta sección está basado en los trabajos de Rubin (1978) y Imbens & Rubin (1997).

En inferencia bayesiana, para estimar el efecto causal del tratamiento en cuestión, los valores observados se consideran realizaciones de variables aleatorias y los valores no observados, al ser desconocidos, se modelan como variables latentes, también aleatorias. Las cantidades involucradas en esta estimación son las siguientes:  $Y_i(0)$ ,  $Y_i(1)$ ,  $A_i(0)$ ,  $A_i(1)$ ,  $\mathbf{R}_i$ ,  $X_i$  de las cuales  $\mathbf{R}_i$ ,  $X_i$ ,  $A_i^{obs} = A_i(T_i)$ , y  $Y_i^{obs} = Y_i(T_i)$  son observables, mientras que  $A_i^{mis} = A_i(1 - T_i)$  y  $Y_i^{mis} = Y_i(1 - T_i)$  no lo son.

Sea  $\Pr(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{A}(0), \mathbf{A}(1), \mathbf{R}, \mathbf{X})$  la distribución de probabilidad conjunta de las variables  $\mathbf{Y}(0)$ ,  $\mathbf{Y}(1)$ ,  $\mathbf{A}(0)$ ,  $\mathbf{A}(1)$ ,  $\mathbf{R}$  y  $\mathbf{X}$ . Se asume que esta distribución es intercambiable, es decir, invariante bajo permutaciones de los índices. Por el teorema de De Finetti, existe un vector de parámetros desconocido  $\theta$ , que es a su vez una variable aleatoria con distribución inicial conocida,  $p(\theta)$ , y que cumple lo siguiente:

$$\Pr(\mathbf{Y}(0), \mathbf{Y}(1), \mathbf{A}(0), \mathbf{A}(1), \mathbf{R}, \mathbf{X}) = \int \prod_{i=1}^n \Pr(Y_i(0), Y_i(1), A_i(0), A_i(1), \mathbf{R}_i, X_i | \theta) p(\theta) d\theta.$$

Considerando los datos observados, la distribución final de  $\theta$  se puede ver de la forma siguiente:

$$\begin{aligned} & \Pr(\theta | \mathbf{Y}^{obs}, \mathbf{A}^{obs}, \mathbf{R}, \mathbf{X}) & (3.2) \\ & \propto p(\theta) \times \int \int \prod_{i=1}^n \Pr(Y_i(0), Y_i(1), A_i(0), A_i(1), \mathbf{R}_i, X_i | \theta) dY_i^{mis} dA_i^{mis} \\ & = p(\theta) \times \int \int \prod_{i=1}^n \Pr(Y_i(0), Y_i(1), G_i, \mathbf{R}_i, X_i | \theta) dY_i^{mis} dA_i^{mis} \\ & = p(\theta) \times \prod_{i=1}^n \int \int [\Pr(X_i | Y_i(0), Y_i(1), G_i, \mathbf{R}_i; \theta_{X|Y,G,R}) \\ & \quad \times \Pr(Y_i(0), Y_i(1), G_i | \mathbf{R}_i; \theta_{Y,G|R}) \times \Pr(\mathbf{R}_i | \theta_R)] dY_i^{mis} dA_i^{mis}, \end{aligned}$$

donde  $\theta_{X|Y,G,R}$ ,  $\theta_{Y,G|R}$  y  $\theta_R$  son parámetros desconocidos correspondientes a las variables  $(X|Y, G, R)$ ,  $(Y, G|R)$  y  $R$ , respectivamente.

Considerando ahora que dentro de la subpoblación  $U_{x_0}$  el mecanismo de asignación no depende de los resultados potenciales, ni tampoco de las covariables, se tiene que

$$\begin{aligned} \Pr(\theta|\mathbf{Y}^{obs}, \mathbf{A}^{obs}, \mathbf{R}, \mathbf{X}; U_{x_0}) &\propto p(\theta) \times \\ &\prod_{i \in U_{x_0}} \left[ \int \int \Pr(X_i|\theta_X) \times \Pr(\mathbf{R}_i|\theta_R) \right. \\ &\quad \left. \times \Pr(Y_i(0), Y_i(1), G_i|\mathbf{R}_i; \theta_{Y,G|R}) dY_i^{mis} dA_i^{mis} \right]. \end{aligned} \quad (3.3)$$

Los estimadores causales son funciones de  $Y_i(z)$ 's y  $A_i(z)$ 's, por lo cual el vector de parámetros  $\theta$  queda determinado por el vector  $\theta_{Y,G|R}$ . Considerando que los vectores de parámetros  $\theta_X, \theta_R$  y  $\theta_{Y,G|R}$  son independientes entre sí (Rubin, 1978), entonces las distribuciones  $\Pr(X_i|\theta_X)$  y  $\Pr(\mathbf{R}_i|\theta_R)$  pueden ser ignoradas en la distribución final  $\theta_{Y,G|R}$ , por lo cual se tiene que:

$$\begin{aligned} \Pr(\theta_{Y,G|R}|\mathbf{Y}^{obs}, \mathbf{A}^{obs}, \mathbf{R}, \mathbf{X}; U_{x_0}) &\propto p(\theta_{Y,G|R}) \times \\ &\prod_{i \in U_{x_0}} \left[ \int \int \Pr(Y_i(0), Y_i(1), G_i|\mathbf{R}_i; \theta_{Y,G|R}) dY_i^{mis} dA_i^{mis} \right] \\ &= p(\theta_{Y|G,R}) \times p(\theta_{G|R}) \times \\ &\prod_{i \in U_{x_0}} \int \int \left[ \Pr(Y_i(0), Y_i(1)|G_i, \mathbf{R}_i; \theta_{Y|G,R}) \right. \\ &\quad \left. \Pr(G_i|\mathbf{R}_i; \theta_{G|R}) dY_i^{mis} dA_i^{mis} \right] \end{aligned} \quad (3.4)$$

donde  $\theta_{Y,G|R} = (\theta_{Y|G,R}, \theta_{G|R})$ .

La descomposición anterior indica que para encontrar la distribución final es necesario especificar los dos modelos siguientes: (1) el modelo de resultados potenciales condicionados en los estatus y las covariables,  $\Pr(Y_i(0), Y_i(1)|G_i, \mathbf{R}_i; \theta_{Y|G,R})$ , y (2) el modelo de los estatus condicionados en las covariables,  $\Pr(G_i|\mathbf{R}_i; \theta_{G|R})$ .

Sea  $\pi_{i,g} = \Pr(G_i = g|\mathbf{R}_i; \theta_{G|R})$  y  $f_{i,gt} = \Pr(Y_i(t)|G_i = g, \mathbf{R}_i; \theta_{Y|G,R})$ , donde  $g = a, c, n$  y  $t = \mathbf{1}\{x \leq x_0\}$ . Entonces la distribución (3.4), considerando los datos observados, se puede ver de la siguiente forma:

$$\begin{aligned}
\Pr(\theta_{Y,G|R}|\mathbf{Y}^{obs}, \mathbf{A}^{obs}, \mathbf{R}, \mathbf{X}; U_{x_0}) &\propto p(\theta_{Y,G|R}) \\
&\times \prod_{\{i \in U_{x_0} : X_i > x_0, A_i^{obs} = 0\}} (\pi_{i,c} f_{i,c0} + \pi_{i,n} f_{i,n}) \\
&\times \prod_{\{i \in U_{x_0} : X_i > x_0, A_i^{obs} = 1\}} \pi_{i,a} f_{i,a0} \times \prod_{\{i \in U_{x_0} : X_i \leq x_0, A_i^{obs} = 0\}} \pi_{i,n} f_{i,n} \\
&\times \prod_{\{i \in U_{x_0} : X_i \leq x_0, A_i^{obs} = 1\}} (\pi_{i,a} f_{i,a1} + \pi_{i,c} f_{i,c1}),
\end{aligned} \tag{3.5}$$

donde  $f_{i,n} = f_{i,n0} = f_{i,n1}$ , por el supuesto de que  $\Pr(Y_i(1)|G_i = n, i \in U_{x_0}) = \Pr(Y_i(0)|G_i = n, i \in U_{x_0})$ , el cual implica que los estudiantes que no solicitan la beca no recibirán tal apoyo, independientemente de su estatus de elegibilidad. La ecuación (3.5) no depende de  $Y_i(0)$  y  $Y_i(1)$ , entonces esta distribución final es igual a su distribución inicial siempre y cuando los parámetros de asociación sean *a priori* independientes.

En el contexto de evaluación de becas, los efectos causales promedio para la muestra se obtienen bajo el supuesto de que para cada individuo  $i$ , los resultados potenciales,  $Y_i(0)$  y  $Y_i(1)$ , son independientes condicionados en  $\mathbf{R}_i$  y  $\theta$ .

El cálculo de la distribución final se puede llevar acabo especificando  $\pi_{i,g}$ ,  $f_{i,gt}$  y las distribuciones *a priori*. Sin embargo, el cálculo se complica debido a que la distribución de probabilidad final implica mezclas de distribuciones.

Por esta razón se utiliza el muestreo de Gibbs con un paso del algoritmo de aumento de datos para imputar los datos faltantes  $A_i^{mis}$ , y por consiguiente  $G_i$ . El algoritmo procede simulando iterativamente de las dos distribuciones finales

$$\Pr(\theta_{Y,G|R}|\mathbf{Y}^{obs}, \mathbf{A}^{obs}, \mathbf{R}, \mathbf{X}; U_{x_0}) \text{ y } \Pr(\mathbf{A}^{mis}|\mathbf{Y}^{obs}, \mathbf{A}^{obs}, \mathbf{R}, \mathbf{X}, \theta_{Y,X|R}; U_{x_0}).$$

Tomando en cuenta los estatus de solicitud se tiene que:

$$\begin{aligned}
\Pr(\theta_{Y,G|R}|\mathbf{Y}^{obs}, \mathbf{A}^{obs}, \mathbf{A}^{mis}, \mathbf{R}, \mathbf{X}; U_{x_0}) &\propto p(\theta_{Y,G|R}) \\
&\times \prod_{\{i \in U_{x_0} : X_i > x_0, G_i = c\}} \pi_{i,c} f_{i,c0} \times \prod_{\{i \in U_{x_0} : X_i > x_0, G_i = a\}} \pi_{i,a} f_{i,a0} \\
&\times \prod_{\{i \in U_{x_0} : G_i = n\}} \pi_{i,n} f_{i,n} \times \prod_{\{i \in U_{x_0} : X_i \leq x_0, G_i = a\}} \pi_{i,c} f_{i,a1} \\
&\times \prod_{\{i \in U_{x_0} : X_i \leq x_0, G_i = c\}} \pi_{i,c} f_{i,c1}.
\end{aligned}$$

La especificación de  $\pi_{i,g}$ ,  $f_{i,gt}$  y las distribuciones iniciales dependen de cada aplicación específica. Como norma general, se sugiere especificar a  $\pi_{i,g}$ , y  $f_{i,gt}$  condicionando en las covariables  $\mathbf{R}$  y la variable puntaje  $X$ . Si se conociera el ancho del intervalo  $h$ , entonces se conocería  $U_{x_0}$ , y en teoría no sería necesario ajustar para  $X$  condicionando en  $h$ , ya que la aleatorización local garantiza que para cada individuo en  $U_{x_0}$  el valor de la variable puntaje está por arriba o debajo del punto de corte, independientemente de los resultados potenciales. Sin embargo, en la práctica  $h$  es desconocido, y una vez seleccionado un  $h^*$  óptimo pueden haber algunos individuos con valor en la variable puntaje  $x_0 - h^*$  y  $x_0 + h^*$ , que no pertenecen a  $U_{x_0}$ . Para estas unidades puede haber una relación entre los resultados y el valor de la variable puntaje, y estas dependencias necesitan ser modeladas. Específicamente, las dependencias entre las estudiantes elegibles y no elegibles, las cuales pueden afectar a la inferencia de la presencia de los estudiantes que no pertenecen a  $U_{x_0}$ .

La inclusión del vector de covariables  $\mathbf{R}$  en el análisis mejora la precisión de la inferencia causal, y hace que los resultados sean menos sensibles a la elección de  $U_{x_0}$ . A continuación se describen los modelos empleados para la estimación del DRD, aplicado a la evaluación de becas italianas universitarias.

Para cada individuo  $i$  de la subpoblación  $U_{x_0}$ , se asume un modelo paramétrico para las variables ( $f_{gt}$ ) y ( $\pi_g$ ), donde

$$\begin{aligned} f_{i,gt} &= \Pr(Y_i(t)|G_i = g, \mathbf{R}_i; \theta_{Y|G,R}), \text{ y} \\ \pi_{i,g} &= \Pr(G_i = g|\mathbf{R}_i; \theta_{G|R}). \end{aligned}$$

El modelo para  $\pi_g$ , el estrato principal de solicitud, consta de dos modelos probit condicionales definidos mediante las variables indicadoras  $G_i^*(a)$  y  $G_i^*(n)$ , para el individuo  $i$ , el cual siempre solicita una beca o para el individuo  $i$  que nunca solicita una beca, es decir:

$$\begin{aligned} \Pr(G_i = a) &= \Pr(G_i^*(a) \leq 0) \\ \Pr(G_i = n) &= \Pr(G_i^*(a) > 0) \text{ y } \Pr(G_i^*(n) \leq 0) \\ \Pr(G_i = c) &= 1 - \Pr(G_i = a) - \Pr(G_i = n), \end{aligned}$$

donde

$$G_i^*(a) = \alpha_{a0} + \alpha_a^{(X)} X_i^* + \mathbf{R}_i' \alpha_a^{(R)} + \epsilon_{ai}, \quad G_i^*(n) = \beta_{n0} + \beta_n^{(X)} X_i^* + \mathbf{R}_i' \beta_n^{(R)} + \epsilon_{ni},$$

con  $\epsilon_{ai} \sim N(0, 1)$ ,  $\epsilon_{ni} \sim N(0, 1)$  independientes, y  $X_i^* = (X_i - x_0)/1000$ . Considerando que la variable respuesta, la deserción escolar, es una variable binaria entonces se asume un modelo lineal generalizado con función liga probit, es decir:

$$\Pr(Y_i(t) = 1 | G_i = g, X_i, \mathbf{R}_i) = \Phi(\beta_{0,g,t} + \beta_{g,t}^{(X)} X_i^* + \mathbf{R}_i' \beta_{g,t}^{(R)}),$$

donde  $\Phi$  es la función de densidad de una normal estándar. Para calcular los estimadores de los promedios muestrales, es decir:

$$\tau_g^X \equiv \sum_{i:G_i=g} [Y_i(1) - Y_i(0)]/n_g,$$

donde  $g = a, c, \{a, c\}$  y  $n_g$  es el número de unidades en el estrato  $g$ , se asume que la correlación entre  $Y_i(0)$  y  $Y_i(1)$  es 0.

Definiendo  $\alpha_g = [\alpha_{g0}, \alpha_g^{(X)}, \alpha_g^{(R)}]'$ ,  $g = a, n$  y  $\beta_{g,z} = [\beta_{0,g,t}, \beta_{g,t}^{(X)}, \beta_{g,t}^{(R)}]$ ,  $g = a, c, n$ ;  $t = 0, 1$ , entonces por la Suposición 6 mencionada en la sección anterior, se tiene que  $\beta_{n,0} = \beta_{n,1}$ . Por otro lado, se asume que los parámetros son inicialmente independientes y tienen distribuciones normales multivariadas, es decir:

$$\alpha_g \sim N(\mu_{\alpha_g}; \sigma_{\alpha_g}^2 \mathbf{I}), \quad \beta_{g,t} \sim N(\mu_{\beta_{g,t}}; \sigma_{\beta_{g,t}}^2 \mathbf{I}), \quad \beta^{(R)} \sim N(\mu_{\beta}; \sigma_{\beta}^2 \mathbf{I})$$

donde  $\mathbf{I}$  es la matriz identidad. Por otro lado, se asume una distribución inicial vaga poco informativa. Específicamente, se asume que los parámetros iniciales son los siguientes:  $\mu_{\alpha_g}, \mu_{\beta_{g,t}}, \mu_{\beta}$  son vectores nulos, y  $\sigma_{\alpha_g}^2 = \sigma_{\beta_{g,t}}^2 = \sigma_{\beta}^2 = 10$  para  $g = a, c, n$  y  $t = 0, 1$ .

Se construyó una cadena de Markov (MCMC) que tiene, como distribución de equilibrio, la distribución final de los parámetros del modelo. El algoritmo de la cadena de Markov que se adoptó, emplea el método de Aumento de Datos para imputar los indicadores faltantes y obtener una distribución final que emplea datos completos para actualizar la distribución de parámetros. Tras obtener muestras finales de los parámetros, se calcularon los efectos poblacionales promedio ( $\tau_a, \tau_c$ ) en  $U_{x_0}$ , los efectos poblacionales promedio en  $x_0$  y el efecto muestral promedio,  $\tau_g^X$ , en  $U_{x_0}$ . Los efectos poblacionales promedio en  $U_{x_0}$ , se calcularon promediando el modelo basado en las proporciones de deserción escolar, entre la distribución empírica de las covariables y la variable puntaje, es decir:

$$\frac{\sum_{i \in U_{x_0}} \pi_{i,g} \Phi(\beta_{0,g,1} + \beta_{g,1}^{(X)} X_i^* + \mathbf{R}_i' \beta_{g,1}^{(R)})}{\sum_{i \in U_{x_0}} \pi_{i,g}} - \frac{\sum_{i \in U_{x_0}} \pi_{i,g} \Phi(\beta_{0,g,0} + \beta_{g,0}^{(X)} X_i^* + \mathbf{R}_i' \beta_{g,0}^{(R)})}{\sum_{i \in U_{x_0}} \pi_{i,g}},$$

$g = \{a, c\}$ . Los efectos promedio poblacionales en  $x_0$ , se calcularon de forma similar considerando ahora que  $X_i^* = 0$  (es decir,  $X_i = x_0$ ). Para estimar  $\tau_g^X$  en  $U_{x_0}$ , se obtuvieron las distribuciones finales de la variable respuesta para cada  $i$  en  $U_{x_0}$ , y basados en esto se obtuvo la estimación del efecto deseado.

### 3.5. Aplicación a la evaluación de becas universitarias italianas

La deserción escolar en la universidad es un fenómeno relevante en Italia. Para estudiar este fenómeno se utilizaron datos de estudiantes inscritos al primer año de las universidades de Florencia y Pisa. Posteriormente, se aplicó el marco conceptual propuesto anteriormente, para evaluar los efectos que tienen las becas universitarias italianas. Esto se hizo con el fin de prevenir la deserción escolar en estudiantes que pertenecen a familias de bajos recursos.

En la práctica, las cuestiones importantes a considerar dentro del marco conceptual fueron la elección de la subpoblación  $U_{x_0}$  (es decir, la elección del ancho de banda  $h$ ) y la especificación de la relación que hay entre la variable respuesta  $Y$  y la variable puntaje  $X$ .

En la aplicación del marco conceptual, en este contexto, se consideró una regresión polinomial en  $X$ . Para el análisis se tomaron en cuenta varios ancho de banda y se consideraron distintos modelos de regresión polinomial con diferentes órdenes.

Los resultados que se observaron fueron los siguientes. Las distribuciones finales de los efectos causales de interés no cambiaron substancialmente con el orden del polinomio; por ello, sólo se presentan los resultados obtenidos considerando sólo un polinomio de primer orden en  $X$ .

h	$\pi_{a,x_0}$		$\pi_{c,x_0}$		$\pi_{n,x_0}$		$\tau_{a,x_0}$		$\tau_{c,x_0}$	
	Media	SD	Media	SD	Media	SD	Media	SD	Media	SD
500	0.324	0.015	0.060	0.018	0.616	0.021	-0.149	0.075	0.044	0.187
1000	0.315	0.012	0.073	0.018	0.612	0.017	-0.152	0.054	0.061	0.234
1500	0.331	0.009	0.046	0.006	0.623	0.010	-0.179	0.056	0.112	0.297
2000	0.343	0.010	0.048	0.014	0.609	0.014	-0.135	0.046	0.308	0.297
2500	0.315	0.008	0.067	0.011	0.617	0.011	-0.137	0.041	0.089	0.204

Cuadro 3.1: Estadísticas de las distribuciones finales, para un polinomio de regresión en  $X$  de primer orden y distintos ancho de banda.

En el Cuadro 3.1 se muestran las medias de las distribuciones finales y las desviaciones estándar de los parámetros causales de interés, considerando distintos anchos de banda, donde las distribuciones finales se simularon de una sola cadena de 20 000 iteraciones después de un periodo de calentamiento de 10 000 iteraciones. Las distribuciones finales del efecto causal para los estudiantes que siempre solicitan la beca,  $\tau_{a,x_0}$ , están centradas en valores negativos y tienen desviaciones estándar relativamente pequeñas, independientemente de la elección del ancho de banda. El estimador  $\tau_{a,x_0}$ , da una reducción en la tasa de deserción escolar de estudiantes de bajos ingresos que siempre solicitan una beca; esta reducción va desde el 13.5 % al 17.9 %.

Como resultado también se observó que la precisión de los estimadores aumenta conforme el ancho de la banda aumenta. Sorprendentemente, también se encontró que el acceso a una beca incrementó la tasa de deserción para estudiantes que no solicitan una beca cuando no son elegibles, aunque los efectos se estiman de manera muy imprecisa. En realidad, los datos no parecen contener mucha información sobre los efectos causales en los estudiantes que no solicitan una beca siendo no elegibles (se estima que la proporción de estos estudiantes es muy pequeña). Cabe señalar que los resultados obtenidos para los estudiantes que no solicitan beca siendo no elegibles son sensibles a la elección del ancho de la banda, lo que genera duda sobre la validez de la asignación al azar (Suposición 4) para anchos de banda superiores a los 1000 euros.

Estos resultados sugieren que las becas universitarias italianas actuales son eficaces en reducir la tasa de deserción escolar de los estudiantes que pertenecen a familias cuya medida económica anual es de aproximadamente 15 000 euros. El análisis hecho anteriormente, también revela que los estratos de los estudiantes que siempre solicitan o solicitan cuando no son elegibles son heterogéneos con respecto al efecto de las becas. Por ello, desde una perspectiva económica, parece más benéfico para los administradores educativos reducir los criterios de elegibilidad (disminuyendo el punto de corte) para permitir a más estudiantes obtener una beca, que aumentar el monto de las becas a los estudiantes beneficiados.

## Capítulo 4

# El Diseño de Regresión Discontinua Multivariado

En un Diseño de Regresión Discontinua simple se presenta una discontinuidad, a lo largo de una medida continua, debido a la asignación de los individuos de estudio a un tratamiento o programa. Para medir el efecto del tratamiento se estima la diferencia de los límites del promedio de los resultados observados en ambos lados de la discontinuidad.

Los Diseños de Regresión Discontinua Multivariados (denominados así por considerar variables puntaje múltiples) se distinguen de los diseños de regresión discontinua simples, pues más de una variable puntaje puede determinar la asignación a más de dos tratamientos. Además, en lugar de proporcionar estimaciones de un único efecto del tratamiento para una sola población (el efecto del tratamiento para las personas con puntuaciones cerca del punto de corte), los Diseños de Regresión Discontinua Multivariados (en adelante DRDM) pueden proporcionar varias estimaciones de efectos, que corresponden a los posibles contrastes de los diversos tratamientos, y a los diferentes subgrupos de población en donde se estiman los efectos. Lo anterior lleva a que el investigador se enfrente a una gama más amplia de estrategias para la estimación de los efectos de los tratamientos; la elección entre las diferentes estrategias de estimación tiene implicaciones importantes para la precisión, el sesgo, y la generalización de las estimaciones.

Este capítulo está basado principalmente en los artículos Wong, Steiner y Cook (2012) y Reardon y Robinson (2012). En las siguientes secciones se describen el DRD y el DRDM; así mismo se presenta un ejemplo de un DRDM. Finalmente, se describen las estrategias de

estimación del efecto promedio de tratamiento y, a través de una serie de simulaciones, se examina el desempeño de dos de los enfoques que se mencionan.

## 4.1. Diseño de Regresión Discontinua

Recordemos que, para describir el Diseño de Regresión Discontinua simple, se considera sólo una variable puntaje, sea  $X$ , con punto de corte  $x_c$ , (sin pérdida de generalidad, supongamos  $x_c = 0$ ), y dos posibles tratamientos,  $a$  y  $b$ . Si  $X_i \geq 0$  el individuo  $i$  se asigna al tratamiento  $a$ , y si  $X_i < 0$  se asigna al tratamiento  $b$ .

Cada individuo  $i$  tiene dos posibles resultados: se asigna al tratamiento  $a$  (denotamos esta asignación como  $Y_i^a$ ) o se asigna al tratamiento  $b$  (denotamos esta asignación como  $Y_i^b$ ). El resultado esperado bajo el tratamiento  $a$  para individuos con  $X = x$  es denotado por  $\bar{Y}^a(x) = E[Y|T = a, X = x]$  y el resultado esperado bajo el tratamiento  $b$  para individuos con  $X = x$  es denotado por  $\bar{Y}^b(x) = E[Y|T = b, X = x]$ . Suponiendo que  $\bar{Y}^a(x)$  y  $\bar{Y}^b(x)$  son funciones continuas de  $x$  en  $x = 0$ , el efecto promedio del tratamiento  $a$  relativo a  $b$  en  $x = 0$ , está descrito por:

$$\delta(0) = \delta(x)|(x = 0) = \lim_{x \rightarrow 0^+} \bar{Y}^a(x) - \lim_{x \rightarrow 0^-} \bar{Y}^b(x). \quad (4.1)$$

Bajo la suposición de continuidad, ambos límites a la derecha de la ecuación (4.1) se pueden estimar a partir de los datos observados. La diferencia de los límites estimados es el estimador de regresión discontinua.

Con el fin de obtener una estimación insesgada de esta diferencia, sólo se necesitan estimaciones insesgadas de cada límite. Estas estimaciones se pueden obtener de un modelo de regresión paramétrica, bajo el supuesto de que se conoce la forma funcional del modelo especificado correctamente.

Una versión paramétrica general del modelo es:

$$Y_i = f(x_i) + \delta(T_i) + \epsilon_i,$$

donde la variable  $x$  está centrada en el punto de corte,  $f$  es una función continua en  $x = 0$ , y  $T$  es una variable indicadora de que  $x \geq 0$  o no. Podemos también estimar  $\delta(x)|(x = 0)$  de forma no paramétrica, usando un estimador de suavizamiento para cada uno de los límites de la ecuación (4.1).

La definición de un estimador requiere de: 1) los resultados de interés  $Y$ , 2) la población de interés  $P$ , y 3) el contraste de tratamientos ( $a$  relativo a  $b$ ).

A continuación se presenta un ejemplo del DRDM. Aquí, se considera solamente el caso en el que hay dos variables puntaje y cuatro posibles tratamientos.

#### 4.1.1. Ejemplo de un DRDM

Un grupo de estudiantes es asignado a cursos de verano con base en sus calificaciones en dos pruebas, una de matemáticas y la otra de inglés. En este caso, los estudiantes que obtienen una puntuación menor a un determinado punto de corte en el examen de matemáticas, se asignan a los cursos de verano de matemáticas, mientras que los estudiantes que obtienen una puntuación menor a un determinado punto de corte en la prueba de inglés, se asignan a los cursos de verano de inglés. Esto da lugar a cuatro posibles “condiciones de tratamiento”: no asisten a ningún curso, asisten al curso de matemáticas, asisten a ambos cursos, y asisten al curso de inglés.

En este ejemplo se consideran dos variables  $X^1$  y  $X^2$ , que indican el curso de matemáticas y el curso de inglés, respectivamente. Los tratamientos  $a$ ,  $b$ ,  $c$ , y  $d$  corresponden a las cuatro posibles condiciones de tratamiento, respectivamente. Aquí, las posibles estimaciones del efecto de un tratamiento relativo a otro se ilustran en la Figura 4.1 y se describen a continuación:

1. El efecto promedio de  $a$  (no asisten a ningún curso) contra  $b$  (asisten al curso de matemáticas) en la región  $X^{A|B}$  (región determinada por la población que es asignada a los tratamientos  $a$  y  $b$ );
2. El efecto promedio de  $b$  (asisten al curso de matemáticas) contra  $c$  (asisten a ambos cursos) en la región  $X^{B|C}$ ;
3. El efecto promedio de  $c$  (asisten a ambos cursos) contra  $d$  (asisten al curso de inglés) en la región  $X^{C|D}$ ;
4. El efecto promedio de  $a$  (no asisten a ningún curso) contra  $d$  (asisten al curso de inglés) en la región  $X^{A|D}$ ;
5. El efecto promedio de  $a$  (no asisten a ningún curso) contra  $c$  (asisten a ambos cursos) en el origen (intersección de las cuatro regiones) ;
6. El efecto promedio de  $b$  (asisten al curso de matemáticas) contra  $d$  (asisten al curso de inglés) en el origen.

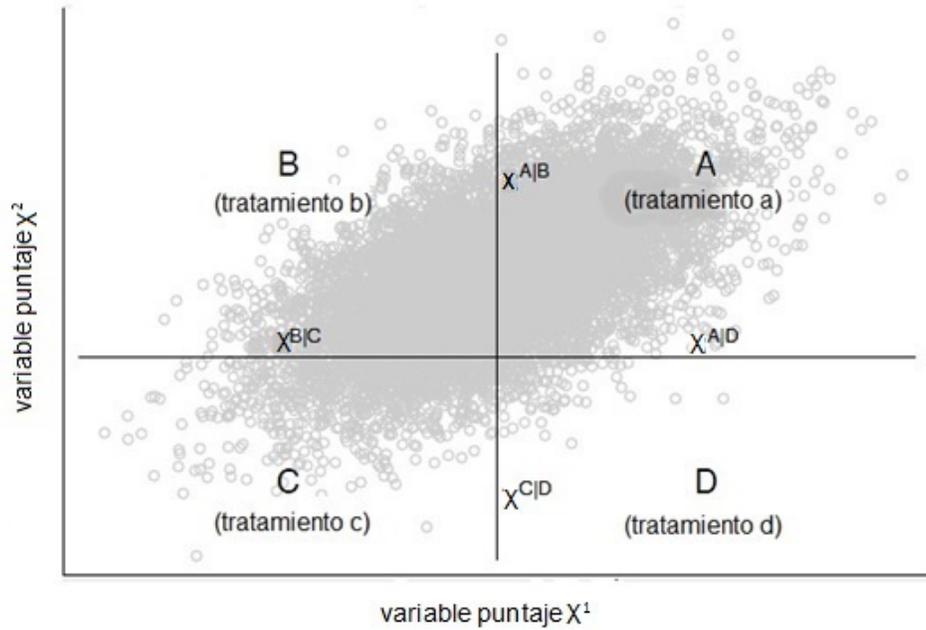


Figura 4.1: Ilustración de una regresión discontinua con dos variables puntaje  $X^1$  y  $X^2$ , y cuatro tratamientos.

## 4.2. Diseño de Regresión Discontinua Multivariado

Para describir el Diseño de Regresión Discontinua con variables puntaje múltiples, por simplicidad consideramos sólo dos variables puntaje  $X^1$  y  $X^2$ , con puntos de corte  $x^1$  y  $x^2$ , respectivamente. Sea  $p$  un punto en el plano definido por  $X^1$  y  $X^2$ .

Sin pérdida de generalidad, supongamos que  $x^1 = 0$  y  $x^2 = 0$ , y consideremos el caso donde hay cuatro tratamientos distintos  $a$ ,  $b$ ,  $c$ , y  $d$ . La asignación a los tratamientos es la siguiente: individuos con valores en las variables  $X^1$  y  $X^2$  mayores a  $x^1$  y  $x^2$  se asignan al tratamiento  $a$ , con valores menores a  $x^1$  y mayores a  $x^2$  se asignan al tratamiento  $b$ , valores mayores que  $x^1$  y menores a  $x^2$  se asignan al tratamiento  $d$ , y finalmente individuos con valores en las variables puntaje menores a  $x^1$  y  $x^2$  se asignan al tratamiento  $c$ , (ver Figura 4.1). Dado que hay cuatro tratamientos, cada individuo tiene cuatro posibles resultados, los cuales denotamos por  $Y_i^a$ ,  $Y_i^b$ ,  $Y_i^c$ , y  $Y_i^d$ . Denotemos por  $\bar{Y}^t(p)$  al resultado esperado bajo el tratamiento  $t$  en el punto  $p$ .

El efecto promedio de un tratamiento (supongamos  $a$ ) relativo a otro tratamiento

(supongamos  $b$ ) en el punto  $p$ , está dado por  $\delta^{ab}(p) = \bar{Y}^a(p) - \bar{Y}^b(p)$ . Considerando lo anterior, tenemos que el efecto promedio de  $a$  relativo  $b$  en la población  $P$ , es:

$$\delta_P^{ab}(p) = \int_{p \in \mathbf{R}} \delta^{ab}(p) \rho(p) dp,$$

donde  $\rho(p)$  es la densidad de la población en el punto  $p$  y  $\mathbf{R}$  es el espacio bidimensional definido por  $X_1$  y  $X_2$ .

En un experimento aleatorizado, se pueden obtener estimaciones de  $\delta_P^{ab}$  (así como de  $\delta_P^{ac}$ ,  $\delta_P^{ad}$ , y así sucesivamente). Sin embargo, en un Diseño de Regresión Discontinua, no se obtienen estimaciones que generalicen a toda la población  $P$ . Por el contrario, se obtienen estimaciones insesgadas del efecto promedio del contraste de dos tratamientos, en este caso de  $a$  en relación con  $b$  para el subconjunto  $P^{AB}$  (población determinada por la asignación de los tratamientos  $a$  y  $b$ ) de  $P$ .

En un Diseño de Regresión Discontinua Multivariado se deben asumir dos supuestos para obtener estimaciones insesgadas de los efectos promedio de los tratamientos. En primer lugar, se supone que cada resultado promedio de la superficie potencial  $\bar{Y}^t(p)$ , es una función continua de  $p$  (es decir, se trata de una función continua de  $X^1$  y  $X^2$ ) en el límite (o límites) en el que se está estimando el impacto del tratamiento. En segundo lugar, se asume que los puntos de corte de  $X^1$  y  $X^2$  se determinan de manera exógena. Denotamos por  $d(p, q)$  la distancia euclídeana entre los puntos  $p$  y  $q$  en  $X^2$ . Entonces, para dos condiciones de tratamiento  $t$  y  $u$ , tenemos lo siguiente:

$$\begin{aligned} \delta^{tu}(p) &= \bar{Y}^t(p) - \bar{Y}^u(p) \\ &= \lim_{s \rightarrow 0^+} \bar{Y}^t(q) | (d(p, q) = s) - \lim_{s \rightarrow 0^+} \bar{Y}^u(q) | (d(p, q) = s) \\ &= \lim_{s \rightarrow 0^+} \bar{Y}^t(q) | (d(p, q) = s, T = t) - \lim_{s \rightarrow 0^+} \bar{Y}^u(q) | (d(p, q) = s, T = u). \end{aligned} \tag{4.2}$$

Los límites de la última línea de la ecuación (4.2) pueden estimarse a partir de los datos observados, siempre y cuando los datos observados de los casos asignados al tratamiento  $t$  estén arbitrariamente cerca del punto  $p$ ; de igual forma, los datos observados asignados al tratamiento  $u$  estén arbitrariamente cerca del punto de  $p$ . Por otra parte, suponiendo que la forma funcional del estimador utilizado para obtener estos límites es apropiada, y que la estimación de los límites es insesgada, entonces se tiene una estimación insesgada del efecto promedio del tratamiento, en el punto  $p$ . De este modo, podemos obtener estimaciones insesgadas del efecto promedio del tratamiento  $t$  en comparación con el tratamiento  $u$  sólo para los puntos de la frontera que

determina la asignación a las condiciones de tratamiento  $t$  y  $u$ . Sin embargo, en lugar de estimar  $\delta^{tu}(p)$  en algún punto o varios puntos específicos a lo largo de esta frontera, se calcula el valor medio de  $\delta^{tu}$  sobre dicha frontera.

En la siguiente sección se describe el problema de estimación de un DRDM.

### 4.3. El problema de estimación para DRDM

El marco conceptual de una estimación analítica en un modelo de regresión discontinua simple o multivariado es el siguiente: se usan los datos observados para estimar los límites del promedio de las funciones potenciales de los resultados en la frontera de dos regiones de tratamiento asignado, y se toma la diferencia de estos límites estimados. La complicación radica en que estos límites deben ser estimados en la frontera de los datos observados para cada condición de tratamiento; lo cual requiere ajustar un modelo de regresión. Tales modelos de regresión son de la forma:

$$Y_i = f(X_i^1, X_i^2, \dots, X_i^J) + \sum_{k=1}^K \delta_k T_i^k + \mathbf{R}_i \mathbf{B} + \epsilon_i, \quad (4.3)$$

donde  $J$  es el número de variables puntaje y  $K$  es el número de las distintas condiciones de tratamiento,  $\{X_i^1, X_i^2, \dots, X_i^J\} \in D \subset \mathbb{R}$ ; cada  $X_i^j$  son las diferentes variables puntaje que determinan la asignación de tratamientos, con  $j \in \{1, 2, \dots, J\}$ ; y  $T_i^k$  denota las variables binarias que indica si un individuo  $i$  es asignado al tratamiento  $k$ .

Siguiendo este marco conceptual, los estimadores difieren en dos aspectos importantes: 1) la especificación de la función  $f$ , y 2) el dominio  $D$  de observaciones utilizadas en la estimación del modelo. La inclusión de un vector de covariables  $\mathbf{R}_i$ , en el modelo puede aumentar la precisión de las estimaciones.

La elección de la forma funcional de  $f$  utilizada para modelar la superficie media de los resultados es importante, sobre todo cuando los datos son relativamente escasos en la región cercana a la frontera. Lo ideal es considerar una amplia cantidad de información cerca de una frontera. En este caso, se utilizan sólo aquellas observaciones cercanas de uno y otro lado de la frontera para el cálculo de los límites. Esta es una estrategia que minimiza la necesidad de hacer supuestos fuertes acerca de la forma funcional de  $f$ .

Bajo el supuesto de continuidad, las observaciones a ambos lados de una frontera son similares entre sí, en promedio, excepto por el tipo de tratamiento. Por otro lado, si limitamos nuestro análisis sólo a los datos cercanos a la frontera, podemos analizar los datos como si hubieran sido

producidos por un experimento de ruptura (experimento que tiene por objetivo examinar las reacciones de la gente con respecto a las violaciones de normas sociales comúnmente aceptadas). Sin embargo, en ausencia de grandes cantidades de datos, el análisis se restringe a puntos muy cercanos a los puntos de corte, dando así estimaciones insesgadas de los límites, haciendo necesario el uso de información concerniente a los puntos de corte y a los supuestos sobre la forma funcional de la superficie media de respuesta.

Idealmente, se espera que las estimaciones sean a la vez objetivas y precisas, pero estas metas están un poco en desacuerdo unas con otras, pues se puede mejorar la precisión mediante la inclusión de nuevas observaciones cercanas a la frontera y asumiendo una forma funcional para estimar los límites de las funciones promedio de resultados, pero esto aumenta el sesgo de los límites estimados. Por otra parte, se puede reducir el sesgo al reducir el ancho de banda, pero generalmente a costa de la precisión.

En principio, podría decirse que para estimar los efectos de tratamiento se puede utilizar la misma lógica de estimación de los diseños de regresión discontinua simple; sin embargo, para estimar estos efectos existe mayor complejidad. En particular, se presentan varias estimaciones posibles y varias estrategias de estimación. En la siguiente sección se discuten cuatro diferentes enfoques de estimación de tratamiento para un DRDM.

## 4.4. Estrategias de estimación

Algunos posibles enfoques para estimar el(los) efecto(s) del(los) tratamiento(s), en un Diseño de Regresión Discontinua con variables puntaje múltiples son los siguientes:

1. Enfoque frontera. En este enfoque se ajusta un modelo paramétrico a la superficie de respuesta completa, y el efecto del tratamiento se considera como aquellas discontinuidades de esta superficie en los límites de asignación del tratamiento.
2. El segundo enfoque, es llamado univariado. Se considera un subconjunto de individuos para estimar el efecto de un par de tratamientos, ajustando sólo una variable puntaje en el modelo de regresión discontinua, de un conjunto de datos adyacentes en una región de  $\mathbf{R}$ ; es decir, si se consideran individuos sólo en dos regiones  $A$  y  $B$ , se estima el efecto del tratamiento  $a$  relativo a  $b$ .
3. Enfoque basado en la distancia. Este enfoque utiliza la distancia a un punto (por ejemplo

el origen) como la variable puntaje, y así se comparan los contrastes de tratamiento disponibles.

4. Finalmente, el enfoque centrado. Aquí se construye una sola variable unidimensional a partir de las variables puntaje múltiples, la cual determina la asignación al tratamiento a través de un sólo punto de corte.

Cabe aclarar que estos enfoques difieren en sus estimaciones y su potencia estadística, así como en la medida en que los supuestos sobre los que se basan resultan fáciles de evaluar. Además, cada método tiene sus ventajas y deficiencias.

Para determinar qué enfoque se va a utilizar en el análisis del DRDM, lo primero que se debe considerar es el estimador de interés. Posteriormente, dependiendo del número de tratamientos o condiciones del tratamiento, de la cantidad de variables puntaje, y de la subpoblación de mayor interés, el investigador puede estimar diferentes parámetros. Dependiendo de la elección del estimador, uno o más enfoques diferentes pueden ser útiles. La elección entre estos enfoques debe hacerse considerando su potencia estadística y la facilidad con la cual los supuestos del DRDM son evaluados.

Independientemente de la estrategia utilizada, se debe evaluar la información que proporcionan los datos en cuanto a la determinación de las variables puntaje y los puntos de corte; esta determinación debe ser exógena. También debe tomarse en cuenta la suposición de que la variable respuesta es continua y el supuesto de que la forma funcional del modelo es correcta.

En la siguiente sección se describe a detalle cada uno de los enfoques ya mencionados.

#### **4.4.1. Enfoque frontera**

Tal vez la manera más obvia para estimar los efectos de un(os) tratamiento(s) con respecto a otro(s), es modelar los efectos del tratamiento como los desplazamientos de una superficie multidimensional (ver Robinson, 2008). Es decir, simplemente ajustar un modelo de la forma mostrada en la ecuación (4.3), donde  $f$  es una función continua que describe la forma de la superficie promedio de los datos observados y los coeficientes  $\delta$ 's indican los efectos promedio del tratamiento.

Cuando se utilizan variables puntaje múltiples para asignar los individuos a uno de dos tratamientos posibles (tratamiento y control), la superficie de respuesta de un modelo de

regresión discontinua es

$$Y_i = f(X_i^1, X_i^2, \dots, X_i^J) + \delta T_i + \epsilon_i.$$

En este modelo,  $f$  es una función continua y  $T_i$  indica el tipo de tratamiento asignado con base en  $\{X_i^1, X_i^2, \dots, X_i^J\}$ . Notemos que el modelo asume que el efecto de tratamiento no está relacionado con las variables puntaje. Cuando se utilizan las variables puntaje  $J \geq 2$ , para asignar los individuos a uno de los  $K$  posibles tratamientos (donde,  $2 \leq K \leq 2^J$ ), la superficie de respuesta de un modelo de regresión discontinua es:

$$Y_i = f(X_i^1, X_i^2, \dots, X_i^J) + \sum_{k=1}^K \delta_k T_{ik} + \epsilon_i,$$

donde  $f$  es una función continua y  $T_k$  indica la asignación al tratamiento  $k$ .

Bajo el supuesto de que  $f$  tiene la forma funcional correcta, se tienen estimaciones insesgadas de los efectos promedio de tratamiento. Además, se debe tomar en cuenta que el enfoque de superficie de respuesta estima cada contraste de tratamientos posible, y puede ser utilizado independientemente de la cantidad de puntos de corte o de tratamientos.

Para ilustrar el procedimiento que sigue este enfoque, se considera un modelo de regresión discontinua con dos variables puntaje. Por simplicidad de notación se supone en esta sección que  $X^1=X$  y  $X^2=M$ . Para cada una de estas variables se consideran los puntos de corte  $x_c$  y  $m_c$ , respectivamente. Dentro de los diseños de regresión discontinua, aquí se consideran sólo los llamados “*sharp*”. En este caso, se asigna el tratamiento si se “alcanzan” los puntos de corte  $x_c$ ,  $m_c$  o ambos; es decir, los individuos son asignados a la condición control  $C$ , si  $X_i \geq x_c$  y  $M_i \geq m_c$ ; y son asignados al grupo tratamiento, si  $X_i < x_c$  o  $M_i < m_c$ . Así, el conjunto de individuos está particionado en tres subespacios: el subespacio  $T_1$ , donde las unidades sólo son menores que  $x_c$  y mayores que  $m_c$ , el subespacio  $T_2$ , las unidades son menores que  $x_c$  y  $m_c$ , y el subespacio  $T_3$ , las unidades son menores que  $m_c$  y mayores que  $x_c$ . Se consideran tres subespacios del grupo tratamiento, pero se asume que todos los individuos reciben el mismo tratamiento. La Figura 4.2 ilustra la descripción de la asignación al tratamiento.

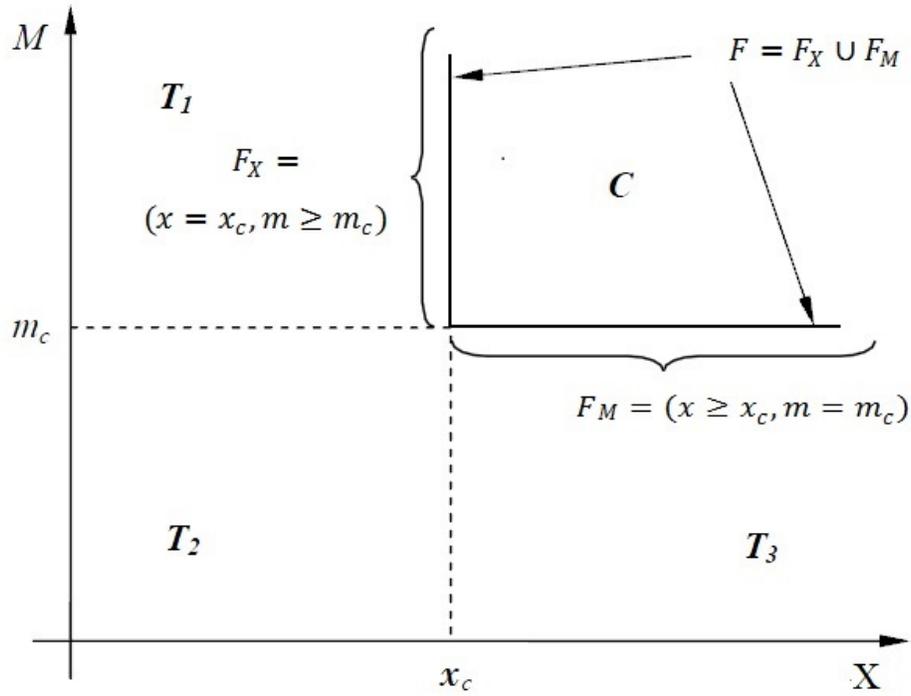


Figura 4.2: Regresión Discontinua con dos variables puntuaje  $X$  y  $M$ , y dos puntos de corte.

Basándose en un diseño de regresión discontinua “*sharp*”, es de interés estimar el efecto promedio del tratamiento, de aquí en adelante denotado por  $\tau_{RDM}$ , en la frontera de corte  $F = \{(x, m) : (x \geq x_c, m = m_c) \cup (x = x_c, m \geq m_c)\}$ . Sea  $T$  la variable binaria que indica si un individuo se asigna o no al tratamiento; es decir,  $T_i = 1$  si el individuo  $i$  pertenece al grupo tratamiento y  $T_i = 0$  si pertenece al grupo control. Denotemos por  $Y_i(0)$  y  $Y_i(1)$ , los resultados posibles, con  $Y_i(0)$  como el resultado que se observa si el individuo  $i$  no está expuesto al tratamiento ( $T_i = 0$ ) y  $Y_i(1)$  el resultado que se observa si el individuo  $i$  está expuesto al tratamiento ( $T_i = 1$ ). En la práctica, se observa sólo el posible resultado de tratamiento o el resultado de control para cada individuo en función del tratamiento recibido, pero no ambos. Considerando lo anterior tenemos que el valor de la variable respuesta está dado por  $Y_i = (1 - T_i) \cdot Y_i(0) + T_i \cdot Y_i(1)$ .

Considerando lo expuesto en la Secciones 4.2 y 4.3 de este capítulo, tenemos que el efecto promedio del tratamiento se define como:

$$\tau_{RDM} = E[Y_i(1) - Y_i(0) | (X_i, M_i) \in F]. \quad (4.4)$$

Sea  $F_X = \{(x, m) : (x = x_c, m \geq m_c)\}$ ,  $F_M = \{(x, m) : (x \geq x_c, m = m_c)\}$ ,  $G_i = Y_i(1) - Y_i(0)$ , y  $f(x, m)$  la función de densidad conjunta de las variables  $X$  y  $M$ . El efecto promedio del tratamiento  $\tau_{RDM}$  se puede descomponer en una media ponderada de efectos del tratamiento en la frontera  $F$ . Esto es, se descompone como el promedio ponderado de esperanzas condicionales en cada una de las fronteras  $F_X$  y  $F_M$

$$\begin{aligned} \tau_{RDM} &= E[G_i | (X_i, M_i) \in F] = w_X E[G_i | X_i \in F] + w_M E[G_i | M_i \in F] \\ &= w_X \tau_X + w_M \tau_M, \end{aligned}$$

donde

$$\begin{aligned} w_X &= \frac{\int_{m \geq m_c} f(x = x_c, m) dm}{\int_{m \geq m_c} f(x = x_c, m) dm + \int_{x \geq x_c} f(x, m = m_c) dx}, \\ w_M &= \frac{\int_{x \geq x_c} f(x, m = m_c) dx}{\int_{m \geq m_c} f(x = x_c, m) dm + \int_{x \geq x_c} f(x, m = m_c) dx}. \end{aligned} \quad (4.5)$$

Las esperanzas condicionales representan el promedio de los efectos del tratamiento  $\tau_X$  y  $\tau_M$  en las dos fronteras  $F_X$  y  $F_M$ , es decir

$$\tau_X = E[G_i | X_i \in F_X] = \frac{\int_{m \geq m_c} g(x, m) f(x = x_c, m) dm}{\int_{m \geq m_c} f(x = x_c, m) dm} \quad (4.6)$$

y

$$\tau_M = E[G_i | M_i \in F_M] = \frac{\int_{x \geq x_c} f(x, m = m_c) dx}{\int_{x \geq x_c} f(x, m = m_c) dx},$$

donde  $g(x, m) = y_1(x, m) - y_0(x, m)$  es la diferencia en los resultados potenciales.

Teniendo en cuenta que  $\tau_{RDM}$  puede ser descompuesto en una media ponderada de los efectos en una frontera específica, todos los supuestos requeridos para una regresión discontinua univariada se deben cumplir para cada frontera  $F_X$  y  $F_M$ . El diseño requiere discontinuidad en las probabilidades de tratamiento sobre  $F_X$  y  $F_M$ . Además, las esperanzas de los resultados deben ser continuas en  $F_X$  y  $F_M$ , es decir

$$\lim_{x \downarrow x_c} E[Y_i(1) | X_i = x, M_i \geq m_c] = \lim_{x \uparrow x_c} E[Y_i(1) | X_i = x, M_i \geq m_c]$$

y

$$\lim_{m \downarrow m_c} E[Y_i(1)|X_i \geq x_c, M_i = m] = \lim_{m \uparrow m_c} E[Y_i(1)|X_i \geq x_c, M_i = m].$$

Una igualdad similar se cumple para  $Y(0)$ . Notemos que no se requiere la continuidad de  $M < m_c$  en  $X = x_c$  y  $X < x_c$  en  $M = m_c$ , porque estas fronteras no pertenecen a la frontera de discontinuidad,  $F$ . Además, el supuesto de continuidad no implica la continuidad de  $E[Y(1)]$  y  $E[Y(0)]$  a lo largo de la frontera  $F$ .

Este procedimiento estima el tamaño de la discontinuidad a lo largo de ambas fronteras al mismo tiempo, y posteriormente pondera lo antes mencionado para obtener el efecto promedio  $\tau_{RDM}$ . A continuación se describe el procedimiento que sigue este enfoque.

En primer lugar, se estima la superficie de respuesta discontinua  $\hat{g}(x, m)$ , utilizando un método de regresión paramétrico, semi-paramétrico o no-paramétrico. Dado que el estimador de interés es el efecto promedio del tratamiento en la frontera de corte  $F$ , se estima la función  $\hat{g}(x, m|F)$  donde  $g(x, m|F) = y_1(x, m|F) - y_0(x, m|F)$ . Posteriormente, se estima la función de densidad conjunta  $\hat{f}(x, m)$ , usando un estimador de densidad de kernel bivariado. Por último, se sustituye  $\hat{g}(x, m|F)$  y  $\hat{f}(x, m)$  en las ecuaciones (4.5) y (4.6) y a través de integración numérica se obtienen  $\tau_X$  y  $\tau_M$ , y los pesos  $w_M$  y  $w_X$ , para la estimación de  $\tau_{RDM}$ .

La estimación precisa de la densidad bivariada  $\hat{f}(x, m)$ , requiere recursos computacionales intensos y de un gran número de observaciones. Por ello, es preferible una estimación no paramétrica de las densidades univariadas condicionales y marginales  $f_{X|M}$ ,  $f_{M|X}$ ,  $f_X$  y  $f_M$ . Por lo tanto, utilizando integración numérica y sustituyendo las densidades univariadas estimadas y la función  $\hat{g}(x, m|F)$  en las ecuaciones anteriores se tiene que:

$$\begin{aligned} w_X &= \frac{\int_{m \geq m_c} f_{M|X}(m|x = x_c) dm \cdot f_X(x = x_c)}{\int_{m \geq m_c} f_{M|X}(m|x = x_c) dm \cdot f_X(x = x_c) + \int_{x \geq x_c} f_{X|M}(x|m = m_c) dx \cdot f_M(m = m_c)} \quad (4.7) \\ w_M &= \frac{\int_{x \geq x_c} f_{X|M}(x|m = m_c) dx \cdot f_M(m = m_c)}{\int_{m \geq m_c} f_{M|X}(m|x = x_c) dm \cdot f_X(x = x_c) + \int_{x \geq x_c} f_{X|M}(x|m = m_c) dx \cdot f_M(m = m_c)} \end{aligned}$$

y

$$\begin{aligned} E[G|F_X] &= \frac{\int_{m \geq m_c} g(x = x_c, m) \cdot f_{M|X}(m|x = x_c) dm}{\int_{m \geq m_c} f_{M|X}(m|x = x_c) dm}, \\ E[G|F_M] &= \frac{\int_{x \geq x_c} f_{X|M}(x|m = m_c) dx \cdot g(x, m = m_c)}{\int_{x \geq x_c} f_{X|M}(x|m = m_c) dx} \end{aligned} \quad (4.8)$$

y por tanto, se obtienen los efectos de frontera  $\tau_X$  y  $\tau_M$ , los pesos  $w_M$  y  $w_X$ , y el efecto promedio de tratamiento  $\tau_{RDM}$ . Esta estrategia se puede aplicar con más de dos variables de asignación.

La gran ventaja de este enfoque es que utiliza todos los datos disponibles de una manera relativamente parsimoniosa, produciendo estimaciones relativamente precisas del efecto promedio del tratamiento. La debilidad del método es que requiere de supuestos fuertes sobre la forma funcional para estimar el efecto promedio del tratamiento, y toma en cuenta los datos que están lejos de los puntos de corte. Si la forma funcional del modelo está mal especificada, las estimaciones pueden ser sesgadas.

#### 4.4.2. Enfoque univariado

Un método más simple que el de modelar una superficie de respuesta, es tomar un subconjunto de datos por estatus (es decir, mayores o menores al punto de corte) en todas menos en una de las variables puntaje, y luego modelar la discontinuidad a lo largo de la variable puntaje restante usando métodos de regresión discontinua simple.

Por ejemplo, para estimar el efecto del tratamiento de  $a$  en comparación con  $b$  (ver Figura 4.1) se limita la muestra a aquellos individuos para los cuales  $X^2 \geq 0$  y se usa regresión discontinua simple para estimar el efecto promedio de  $a$  en relación con  $b$ , a lo largo de la región  $X^{A|B}$ .

El efecto promedio de tratamiento local de  $a$  en relación con  $b$  (ver Figura 4.1) puede ser estimado ajustando el modelo de regresión

$$Y_i = f(X_i) + \delta^{a|b}T_{ia} + \epsilon_i, \quad (4.9)$$

donde la muestra cumple con  $X^2 \geq 0$ , dentro de un dominio especificado alrededor del punto de corte. El parámetro  $\delta^{a|b}$  indica el efecto de  $a$  relativo a  $b$  en la frontera definida por  $X^{A|B}$ .

Si se incluye una función de otras variables puntaje (por ejemplo,  $X^2$ ) y de covariables en el modelo, se puede mejorar la precisión de las estimaciones de  $\delta^{a|b}$  :

$$Y_i = f(X_i) + \delta^{a|b}T_{ia} + g(X_i^2) + \mathbf{R}_i\mathbf{B} + \epsilon_i.$$

De una manera similar, se puede estimar el efecto promedio de  $b$  contra  $c$ ,  $c$  contra  $d$  y  $d$  contra  $a$ . Es importante señalar que estas estimaciones del efecto no son necesariamente comparables, ya que cada uno se aplica a diferentes subpoblaciones. Es decir, no podemos estimar el efecto de  $a$  contra  $c$  simplemente con la adición de los efectos promedio estimados de  $a$  contra  $b$  en  $X^{A|B}$  y  $b$  contra  $c$  en  $X^{B|C}$ , a menos que se asuma que los efectos son homogéneos en toda la

población de interés (o por lo menos entre las poblaciones de estas dos subregiones). Si sólo hay dos tratamientos, entonces el enfoque univariado puede proporcionar alguna evidencia con respecto a la heterogeneidad de los efectos, mediante la comparación de los efectos promedio del tratamiento entre los individuos a lo largo de la frontera  $X^{A|B}$  y aquellos a lo largo de frontera  $X^{B|C}$ .

### 4.4.3. Enfoque basado en la distancia

Estimar el contraste del tratamiento  $a$  con  $c$  (o  $b$  con  $d$ ), puede ser difícil con muestras pequeñas, porque se comparan observaciones en un punto con las de otro punto (en lugar de a lo largo de una línea, como se haría si se examina el tratamiento  $a$  contra  $b$ , por ejemplo). Una de las estrategias para el estudio de estos contrastes es la construcción de una nueva variable, que es la distancia euclidiana desde el punto  $p_i$  al origen. Si se tienen dos variables puntaje, se construye una nueva variable puntaje,  $d_i = \sqrt{(X_i^1)^2 + (X_i^2)^2}$ , usando sólo observaciones expuestas a uno de los dos tratamientos de interés (por ejemplo, sólo de las regiones  $A$  y  $C$ ). Posteriormente se ajusta el modelo

$$Y_i = f(d_i) + \delta^{a|c}T_{ia} + \epsilon_i.$$

Este método se puede utilizar para estimar el efecto en cualquier punto  $(X^{1*}, X^{2*}, \dots, X^{J*})$ , donde la distancia a ese punto es  $d_i = \sqrt{\sum_{j=1}^J (X_i^j - X^{j*})^2}$ . Sin embargo, la mayoría de las aplicaciones de este método tienen una baja potencia estadística para estimar el efecto en un sólo punto.

### 4.4.4. Enfoque centrado

Este enfoque se basa en la construcción de una variable puntaje  $Z = g(X_i^1, \dots, X_i^J)$ , donde  $g$  es una función monótona en cada  $X^j$ , tal que  $g(0, 0, \dots, 0) = 0$ . La pregunta aquí es cuál función  $g$  elegir. Una de las estrategias es considerar a  $g$  como el mínimo de las variables puntaje estandarizadas, basándose en el argumento de que esto hace que la distancia desde el punto de corte sea comparable para cada una de las variables puntaje. Sin embargo, cualquier función  $g$  proporcionará una estimación insesgada del efecto del tratamiento (suponiendo que se puede identificar una forma funcional  $f$  correcta para el modelo de relación  $Y$  con  $Z$ ); aunque también la elección de  $g$  puede afectar el modelo potencial de regresión discontinua, porque se ve afectada la correlación entre  $Z$  y  $Y$ , condicionada en  $T$ , lo cual afecta la potencia del estimador del modelo

de regresión discontinua. Aún existe poca orientación, empírica o teórica, sobre cómo elegir una función  $f$  que maximice la precisión del enfoque centrado.

Para aumentar la precisión de las estimaciones, puede ser útil incluir individualmente las variables puntaje en el modelo como covariables. En este caso, el modelo se convierte en

$$\begin{aligned} Y_i &= f(Z_i) + h(X_i^1, \dots, X_i^J) + \delta T_i + \epsilon_i \\ &= f(g(X_i^1, \dots, X_i^J)) + h(X_i^1, \dots, X_i^J) + \delta T_i + \epsilon_i \\ &= f(X_i^1, \dots, X_i^J) + \delta T_i + \epsilon_i. \end{aligned}$$

Este enfoque es atractivo porque se puede usar toda la información disponible para estimar el efecto del tratamiento, pero lo hace de manera simple transformando el vector multidimensional de variables puntaje en un vector unidimensional que determina la asignación al tratamiento.

En la siguiente sección se ilustra una simulación de un ejemplo de un diseño de regresión discontinua multivariado. De igual forma, a través de una serie de simulaciones se examina el desempeño de los enfoques frontera y centrado. Se consideran sólo estos dos enfoques debido a que son estrategias generales (no dependen del número de variables puntaje, del número de puntos de corte y del número de tratamientos) que se pueden aplicar para estimar el efecto promedio del tratamiento de un DRDM.

## 4.5. Simulaciones

Para ilustrar el análisis de un DRDM, así como el desempeño de los enfoques frontera y centrado descritos en la sección anterior, se llevaron a cabo una serie de simulaciones, las cuales se ejecutaron en el paquete estadístico R (ver Apéndice B).

### 4.5.1. Simulación 1: ejemplo de un DRDM

Se consideró un diseño con dos variables puntaje  $R$  y  $M$ , con puntos de corte  $r_c$  y  $m_c$ , respectivamente. Se asume que hay tres distintos tratamientos  $a$ ,  $b$ ,  $c$  y un grupo control. Los individuos se asignan al grupo control  $C$ , si  $R_i \geq r_c$  y  $M_i \geq m_c$ ; y son asignados a un tratamiento si  $R_i < r_c$  o  $M_i < m_c$  (ver Figura 4.2).

Para ilustrar este diseño, se simuló una muestra de tamaño  $n = 1000$  de una distribución normal multivariada, con vector de medias  $\mu = (0, 0)$  y matriz de covarianza  $\sigma^2 = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}$ .

Se consideró a las variables puntaje  $R$  y  $M$ , como los vectores transversales del vector bivariado que se generó anteriormente.

Los puntos de corte fueron  $r_c = 0.007781275$  y  $m_c = 0.03267760$ , que son las medias muestrales de  $R$  y  $M$ , respectivamente. Finalmente, la variable respuesta que se generó fue:

$$Y_i = 4 + 0.5 \cdot R_i + M_i + 1.2 \cdot T_{i1} + 2 \cdot T_{i2} + 3.4 \cdot T_{i3} + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.10)$$

donde  $\varepsilon$  se distribuye como una normal de media 0 y desviación estándar 2, y los términos  $T_1$ ,  $T_2$  y  $T_3$  son variables indicadoras. La variable  $T_1$  toma el valor de 1 si  $R_i < r_c$  y  $M_i > m_c$ ,  $T_2$  es 1 si  $R_i < r_c$  y  $M_i < m_c$  y  $T_3$  es 1 cuando  $R_i > r_c$  y  $M_i < m_c$ .

En la Figura 4.3 se muestra la superficie de respuesta con discontinuidades, y en la Figura 4.4 se muestra una gráfica bidimensional, donde los contrastes de color indican los saltos de discontinuidad, así como las regiones en el plano determinadas por los individuos a los que se le aplicó cada tratamiento.

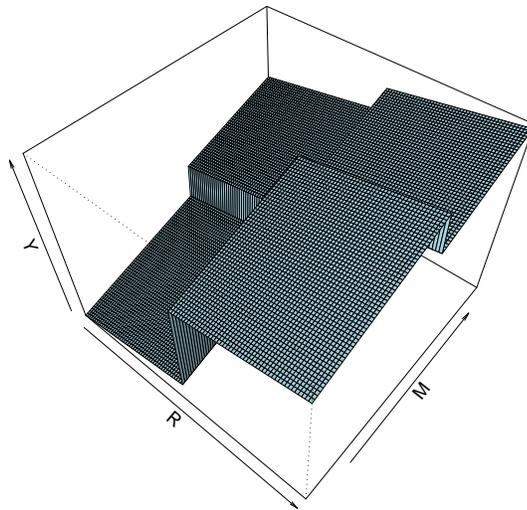


Figura 4.3: Superficie de respuesta de los datos observados para la Simulación 1.

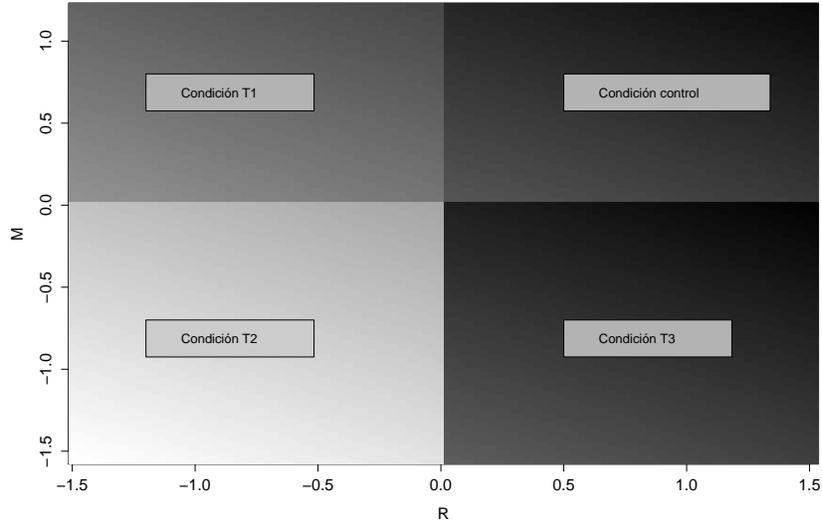


Figura 4.4: Contrastes de color que ilustran los saltos de discontinuidad de la superficie de respuesta, de los datos observados para la Simulación 1.

La superficie de respuesta que se ajustó bajo el enfoque clásico fue:

$$\hat{Y}_i = 3.9034 + 0.5722 \cdot R_i + 0.7315 \cdot M_i + 1.4784 \cdot T_1 + 2.1803 \cdot T_2 + 3.3547 \cdot T_3, \quad i = 1, \dots, n. \quad (4.11)$$

En el Cuadro 4.1, se muestran los resultados obtenidos en el ajuste de la superficie (4.11).

	Coefficiente	Error estándar	p-value
(Intercepto)	3.7337	0.1836	2e-16 ***
R	0.8114	0.2347	0.000569 ***
M	1.0411	0.2440	2.17e-05 ***
T1	1.5082	0.2518	2.95e-09 ***
T2	2.5287	0.3119	1.50e-15 ***
T3	3.6027	0.2454	2e-16 ***
Nivel de significancia: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1			

Cuadro 4.1: Ajuste del modelo (4.11).

Del cuadro anterior podemos notar que los tres tratamientos aplicados resultaron ser significativos. Considerando el error estándar del estimador del tratamiento  $T_3$ , éste es menor que los errores estándar de los estimadores de los tratamientos  $T_1$  y  $T_2$ , lo cual implica que

el tratamiento  $T_3$  tuvo un efecto promedio más significativo en comparación con los demás tratamientos y en comparación con el grupo control. Esto, a su vez se ve reflejado en la superficie de respuesta de los datos observados (Figura 4.3) y en la Figura 4.4, ya que en ambas figuras se notan los saltos de discontinuidad de los datos observados, siendo el salto de discontinuidad mayor el determinado por el tratamiento  $T_3$ .

### 4.5.2. Simulación 2: enfoque frontera

Para ilustrar el enfoque frontera se llevó a cabo la siguiente simulación: se utilizó una muestra de tamaño  $n = 65$  de una distribución normal bivariada, con vector de medias  $\mu = (0, 0)$  y matriz de covarianza  $\sigma^2 = \begin{pmatrix} 0.2 & 0 \\ 0 & 0.2 \end{pmatrix}$ . Se toman como variables puntaje  $R$  y  $M$ , a los vectores transversales generados en la simulación de esta distribución bivariada, y como puntos de corte las medias muestrales de  $R$  y  $M$ , es decir  $r_c = 0.02640$  y  $m_c = -0.00924$ .

Se supone que todos los individuos reciben el mismo tratamiento. Los individuos se asignan al grupo control  $C$ , si  $R_i \geq r_c$  y  $M_i \geq m_c$ ; y son asignados al tratamiento si  $R_i < r_c$  ó  $M_i < m_c$ .

Finalmente la variable respuesta que se generó fue:

$$Y_i = 4 + 0.5 \cdot R_i + M_i + 2 \cdot T_{i1} + 2 \cdot T_{i2} + 2 \cdot T_{i3} + \varepsilon_i, \quad i = 1, \dots, n, \quad (4.12)$$

donde  $\varepsilon$  se distribuye como una normal de media 0 y desviación estándar 1, y los términos  $T_1$ ,  $T_2$  y  $T_3$  son variables indicadoras. La variable  $T_1$  toma el valor de 1 si  $R_i < r_c$  y  $M_i > m_c$ ,  $T_2$  es 1 si  $R_i < r_c$  y  $M_i < m_c$  y  $T_3$  es 1 cuando  $R_i > r_c$  y  $M_i < m_c$ .

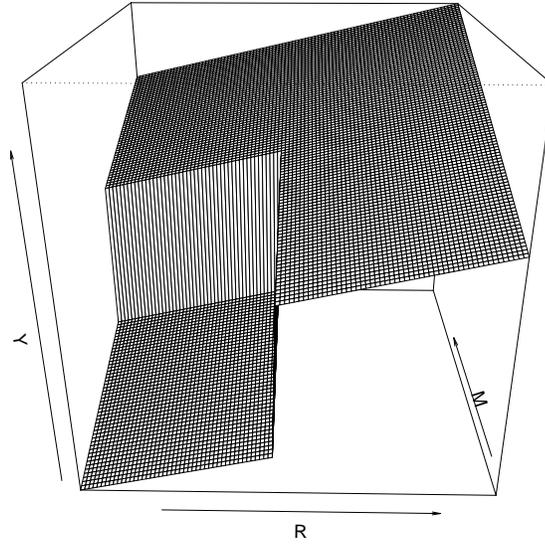


Figura 4.5: Superficie de respuesta de los datos observados para la Simulación 2. Notemos que en esta gráfica sólo hay un salto de discontinuidad, debido a que se asume que hay un único tratamiento.

En la Figura 4.5 se muestra la superficie de respuesta de los datos observados. La superficie de respuesta que se ajustó mediante métodos paramétricos clásicos fue:

$$\hat{Y}_i = 3.7903 + 0.5542 \cdot R_i + 0.5972 \cdot M_i + 2.4062 \cdot T_1 + 2.2979 \cdot T_2 + 2.1255 \cdot T_3, \quad i = 1, \dots, n. \quad (4.13)$$

	Coefficiente	Error estándar	p-value
(Intercepto)	3.7903	0.3616	4.28e-15 ***
R	0.5542	0.5175	0.288586
M	0.5972	0.4176	0.157978
T1	2.4062	0.5050	1.27e-05 ***
T2	2.2979	0.6863	0.001421 **
T3	2.1255	0.5652	0.00039 ***
Nivel de significancia: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1			

Cuadro 4.2: Ajuste del modelo (4.13).

En el Cuadro 4.2 se muestran los resultados obtenidos del ajuste de la superficie (4.13).

El estimador del efecto promedio utilizando el enfoque frontera fue de 2.249831. Para estimar el error estándar se utilizaron los métodos *bootstrap* paramétrico y no paramétrico (ver Apéndice C), y se generó para ambos métodos una muestra de tamaño  $n = 65$  de los estimadores del efecto promedio del tratamiento. Las estimaciones obtenidas de las muestras *bootstrap* se resumen en el Cuadro 4.3. Para comparar que tan confiables son las estimaciones obtenidas por el método *bootstrap* se cálculo el error estándar real (esto se logra debido a que se conoce el modelo verdadero de los datos). Para ello, se simulo un gran número de muestras del modelo verdadero (4.12) del mismo tamaño que el modelo verdadero. Las estimaciones obtenidas se presentan en el Cuadro 4.3.

Estimaciones	Bootstrap paramétrico	Bootstrap no paramétrico	Error estándar real
Media	2.1476	1.4215	2.0036
Error estándar	0.4415	0.4768	0.3797
Intervalo de confianza al 95 %	[2.038243, 2.257075]	[1.303434, 1.539727]	[1.9662, 2.0111]

Cuadro 4.3: Estimaciones del efecto promedio del DRDM de la Simulación 2.

En las Figura 4.6 y 4.7 se muestran las densidades y los histogramas, obtenidos de las muestras de los estimadores del efecto promedio generadas a través de los métodos *bootstrap* paramétrico y no *paramétrico*, respectivamente.

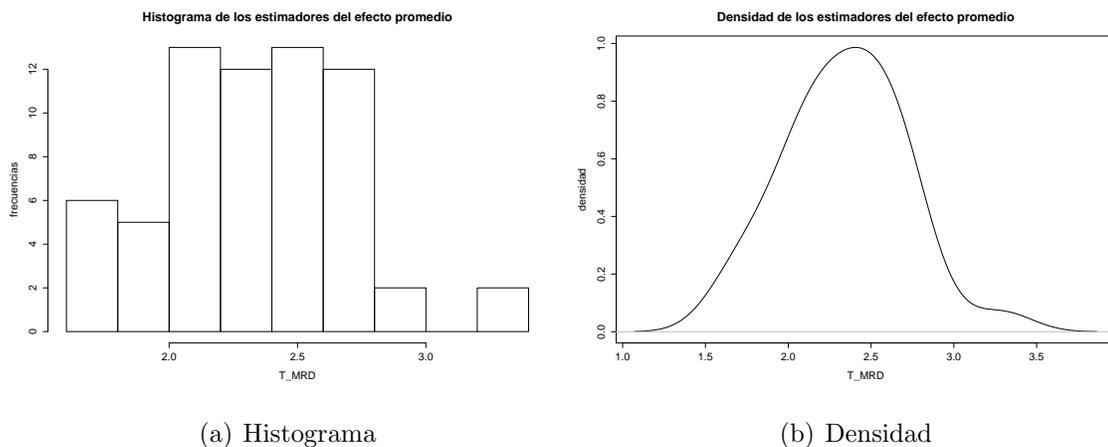


Figura 4.6: Gráficas de la muestra de los estimadores del efecto promedio generados por el método *bootstrap* paramétrico.

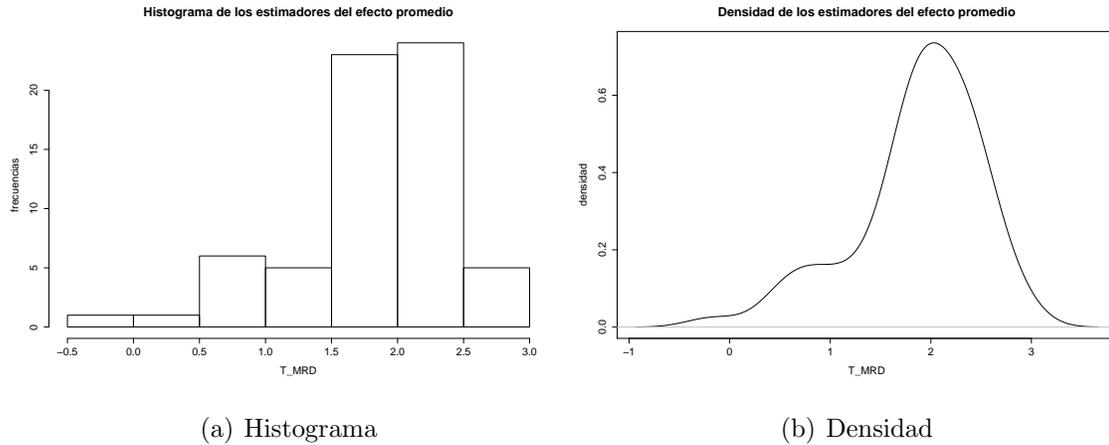


Figura 4.7: Gráficas de la muestra de los estimadores del efecto promedio generados por el método *bootstrap* no paramétrico.

De los resultados obtenidos anteriormente se concluye lo siguiente: de la superficie de respuesta que se ajusta a los datos observados se puede observar que efectivamente existe una discontinuidad en el límite de asignación entre el grupo control y el grupo tratamiento. Además el Cuadro 4.2 muestra que los coeficientes de las condición tratamiento son significativas, es decir, el efecto promedio del tratamiento resulta ser significativo, debido a que estas condiciones hacen referencia a un solo tratamiento. Por otro lado, el efecto promedio del tratamiento obtenido a través del enfoque frontera, 2.249831, es muy cercano al efecto promedio verdadero que es 2; además el error estándar obtenido es más cercano al error estándar real en comparación con el obtenido por el método *bootstrap* no paramétrico. Considerando el Cuadro 4.3 podemos notar que se obtuvieron estimaciones más precisas con el método *bootstrap* paramétrico que con el no paramétrico; ya que el primer método tiene menor error estándar y el intervalo de confianza obtenido por este método prácticamente contiene al valor del efecto promedio del tratamiento obtenido a través del enfoque frontera. En contraste, el intervalo de confianza obtenido a través del método no paramétrico ni siquiera contiene al valor verdadero del efecto del tratamiento 2.

En la siguiente simulación se ilustra el enfoque centrado.

### 4.5.3. Simulación 3: enfoque centrado

En esta simulación, el Diseño de Regresión Discontinua Multivariado que se utilizó es el mismo que se describió en la Simulación 2; su objetivo es ilustrar el desempeño del enfoque centrado.

Como se mencionó en la Sección 4.4.4, el enfoque centrado se basa en la construcción de una nueva variable puntaje  $Z$ , definida como  $Z = g(X_i^1, \dots, X_i^J)$ , donde  $g$  es una función monótona en cada  $X^J$ , tal que  $g(0, \dots, 0) = 0$ . En nuestro caso,  $Z$  se define como  $Z_i = \min\{R_i - r_c, M_i - m_c\}$  para cada  $i = 1, \dots, n$ . Se considera como punto de corte a  $z_c = 0$  y como variable de asignación  $T$ ; por ello  $T_i = 1$  si  $Z_i \geq 0$ . Es decir el individuo  $i$  se asigna al grupo tratamiento, de lo contrario  $T_i = 0$  y se asigna al grupo control. Ahora, el DRD a analizar es:

$$Y_i = \alpha + \beta_0 \cdot T_i + \beta_1 \cdot Z_i + \epsilon_i, \quad i = 1, \dots, n. \quad (4.14)$$

El modelo ajustado fue:

$$\hat{Y}_i = 3.8367 + 2.0011 \cdot T_i + 0.9021 \cdot Z_i, \quad i = 1, \dots, n. \quad (4.15)$$

El coeficiente de la variable de asignación  $\beta_0$  es considerado como la magnitud del efecto del tratamiento.

Empleando este enfoque se obtuvo que el estimador del efecto promedio es de 2.0011 con error estándar igual a 0.5400. El intervalo de confianza del 95 % para  $\beta_0$  es  $[0.92159, 3.08051]$ . En la Figura 4.8 se muestra el diagrama del diseño de regresión discontinua simple (4.14), y en el Cuadro 4.4, se muestran los resultados obtenidos del análisis de este diseño.

	Coficiente	Error estándar	p-value
(Intercepto)	3.8367	0.3329 < 2e-16 ***	
$\beta_0$	0.9021	0.441	0.045110*
$\beta_1$	2.0011	0.5400	0.000452***
Nivel de significancia:	0 *** 0.001 **	0.01 * 0.05 .	0.1 1

Cuadro 4.4: Ajuste del modelo (4.15).

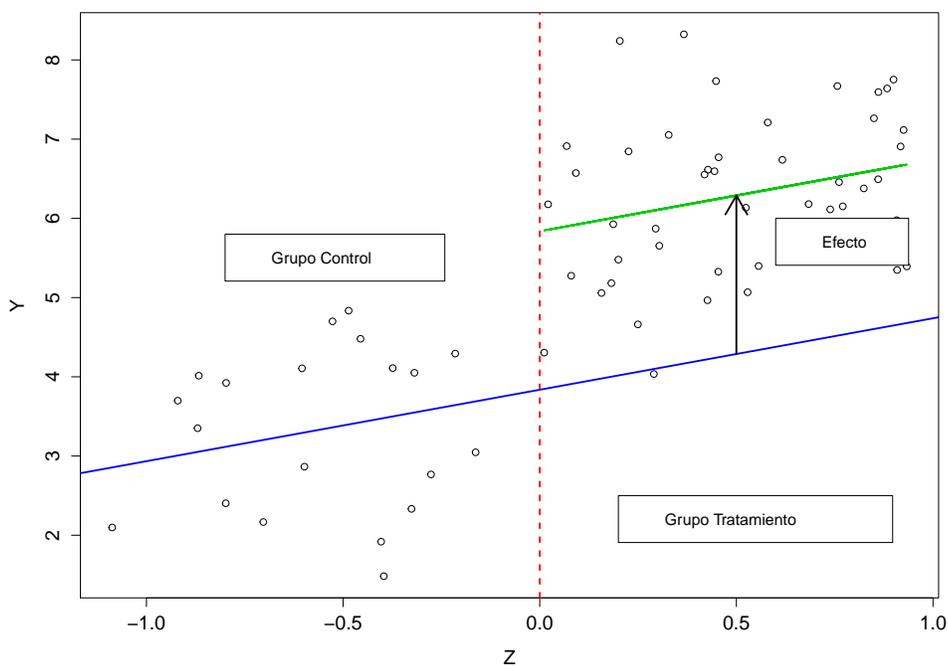


Figura 4.8: Diagrama del Diseño de Regresión Discontinua, donde se considera como variable puntaje a  $Z$ . Se muestra la recta de regresión (línea azul), la recta de regresión proyectada en el grupo control (línea verde) y el efecto del tratamiento (flecha).

A continuación se emplea el paquete estadístico rdd, descrito en el Capítulo 2, para compararlo con los resultados obtenidos en esta sección.

En la implementación del paquete rdd, para el análisis de la Simulación 3, se consideraron las variables  $Y$  y  $Z$ , descritas en la sección anterior y el punto de corte 0. Los resultados que se obtuvieron se resumen en el siguiente cuadro:

	Ancho de banda	Estimador	Observaciones	Error estándar	z value
I-K	0.6778	2.647	42	0.8102	3.267**
(I-K)/2	0.3389	2.334	21	1.2267	1.903.
2(I-K)	1.3555	2.347	65	0.5789	4.054***
Nivel de significancia:	0***	0 0.001**	0.01*	0.05.	

Cuadro 4.5: Ajuste del modelo (4.14) con el paquete rdd.

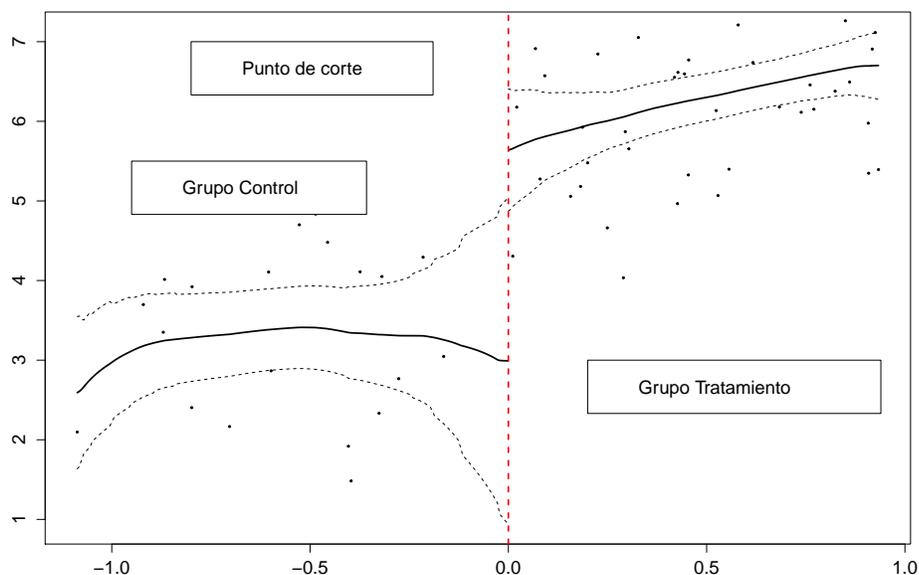


Figura 4.9: Gráfica de la variable puntaje  $Z$  contra la variable respuesta  $Y$  de la Simulación 3, utilizando el paquete estadístico rdd. Las líneas punteadas delimitan la banda de confianza al 95 %, y la línea continúa muestra la recta de regresión ajustada por la función  $RDestimate$ .

En la Figura 4.9 se muestra la gráfica obtenida empleando la función  $RDestimate$ , para el análisis del modelo descrito en la Simulación 3. En el Cuadro 4.6 se resumen los resultados correspondientes al enfoque centrado y a los resultados obtenidos con el paquete rdd.

	Enfoque centrado	Paquete rdd
Estimador del efecto	2.0011	2.647
Error estándar	0.5400	0.8102

Cuadro 4.6: Resultados del ajuste para el modelo descrito en la Simulación 3. El estimador considerado empleando el paquete rdd fue el estimador óptimo de I-K.

El estimador del efecto del tratamiento que se obtuvo con el enfoque centrado, es 2.011. Por lo tanto, considerando el Cuadro 4.6, se tiene que el estimador más cercano al valor del efecto del tratamiento, cuyo valor es 2, es el que se obtuvo con el enfoque centrado; a su vez este estimador es el de menor error estándar.

#### 4.5.4. Análisis bayesiano del modelo del enfoque centrado

##### Análisis no informativo

A través de métodos bayesianos y utilizando una distribución inicial no informativa, se analiza el modelo unidimensional descrito en el enfoque centrado. La distribución inicial no informativa que se utilizó se aproxima a la del análisis de referencia haciendo uso del análisis conjugado (ver Gutiérrez-Peña, 1998); los detalles de la especificación de la distribución inicial se pueden ver en el Apéndice A. Aunque, se trata de un análisis de tipo conjugado se utilizó el paquete estadístico WinBUGS para fines ilustrativos del mismo. El análisis se hizo ejecutando un algoritmo en el paquete estadístico WinBUGS. Después de una fase de calentamiento de 1000 iteraciones y de monitorear los nodos  $\alpha$ ,  $\beta_0$  y  $\beta_1$  por 10000 iteraciones más, se obtuvieron los siguientes resultados. Los coeficientes estimados son  $E(\alpha|\underline{y})=3.839$  ( $de = 0.3341$ ),  $E(\beta_0|\underline{y})=1.997$  ( $de =0.5452$ ) y  $E(\beta_1|\underline{y})=0.9022$  ( $de =0.4447$ ), donde los valores en paréntesis ( $de$ ) son las desviaciones estándar de las correspondientes distribuciones finales. Por ello, el modelo ajustado es:

$$\widehat{Y}_i = 3.839 + 1.997 \cdot T_i + 0.9022 \cdot Z_i, \quad i = 1, 2, \dots, n.$$

El intervalo de probabilidad del 95 % para  $\beta_0$  es  $[0.9288, 3.074]$ . Comparando con los coeficientes del modelo (4.15) y considerando los errores estándar, podemos decir que los coeficientes estimados son significativos.

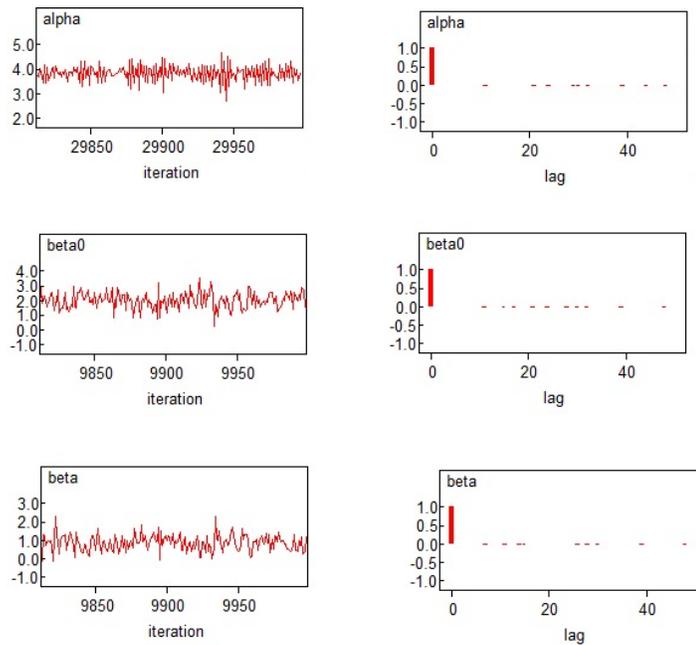


Figura 4.10: Gráficas de Diagnóstico de convergencia. A la izquierda las gráficas de traza y a la derecha las gráficas de autocorrelación correspondientes a los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ , respectivamente.

En la Figura 4.10 se muestran las gráficas para el diagnóstico de convergencia de los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ ; estas gráficas sugieren que el algoritmo converge.

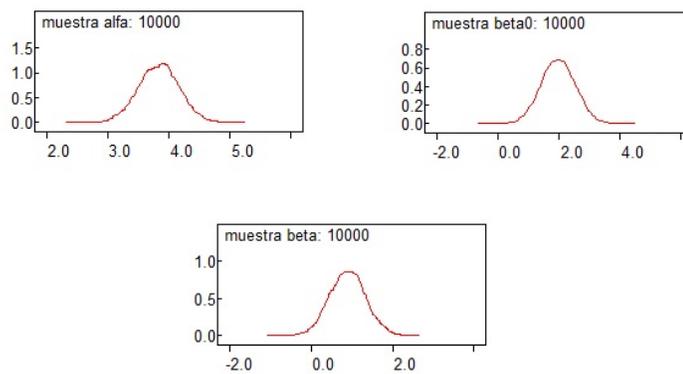


Figura 4.11: Gráficas de las densidades de los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ , respectivamente.

En la Figura 4.11 se muestran las gráficas de la densidad para cada uno de los coeficientes

estimados. Éstas sugieren un ajuste significativo.

A continuación se analiza el mismo modelo considerado aquí, pero ahora tomando en cuenta una distribución inicial informativa para  $\beta_0$ .

### **Análisis informativo**

A través de métodos bayesianos y utilizando como distribución inicial informativa para  $\beta_0$  una distribución normal de media 2 y desviación estándar 5, se analiza el modelo unidimensional descrito en el enfoque centrado. El análisis se hizo ejecutando un algoritmo en el paquete estadístico WinBUGS. Después de una fase de calentamiento de 1000 iteraciones y de monitorear los nodos  $\alpha$ ,  $\beta_0$  y  $\beta_1$  por 10000 iteraciones más, se obtuvieron los siguientes resultados. Los coeficientes estimados son  $E(\alpha|\underline{y})=3.839$  ( $de = 0.3266$ ),  $E(\beta_0|\underline{y})=1.996$  ( $de =0.5302$ ) y  $E(\beta_1|\underline{y})=0.9012$  ( $de =0.4345$ ), donde los valores en paréntesis son las desviaciones estándar de las correspondientes distribuciones finales. Por ello, el modelo ajustado es:

$$\hat{Y}_i = 3.839 + 1.996 \cdot T_i + 0.9012 \cdot Z_i, \quad i = 1, 2, \dots, n.$$

El intervalo de probabilidad del 95 % para  $\beta_0$  es [0.9615, 3.041]. Comparando con los coeficientes del modelo (4.15) y considerando los errores estándar, podemos decir que los coeficientes estimados son significativos.

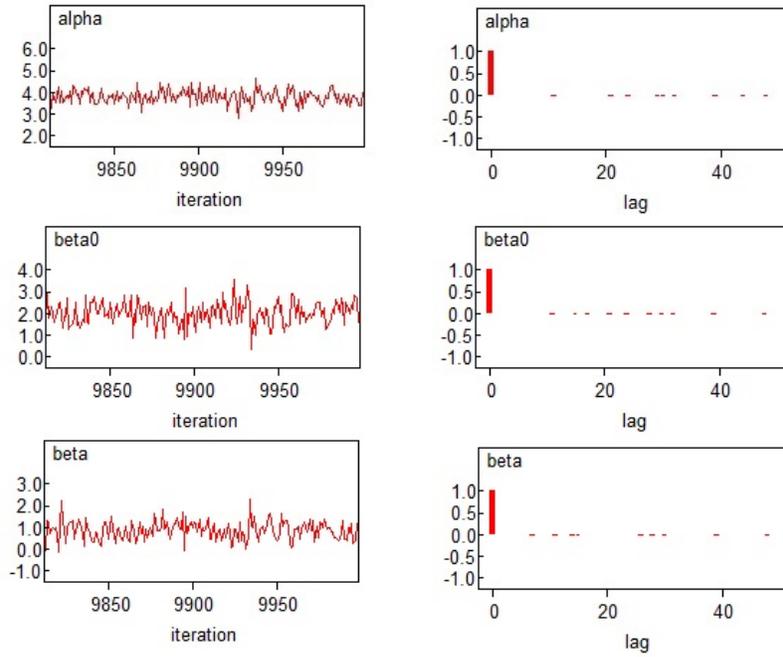


Figura 4.12: Gráficas de Diagnóstico de convergencia. A la izquierda las gráficas de traza y a la derecha las gráficas de autocorrelación correspondientes a los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ , respectivamente.

En la Figura 4.12 se muestran las gráficas para el diagnóstico de convergencia de los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ ; estas gráficas sugieren que el algoritmo converge.

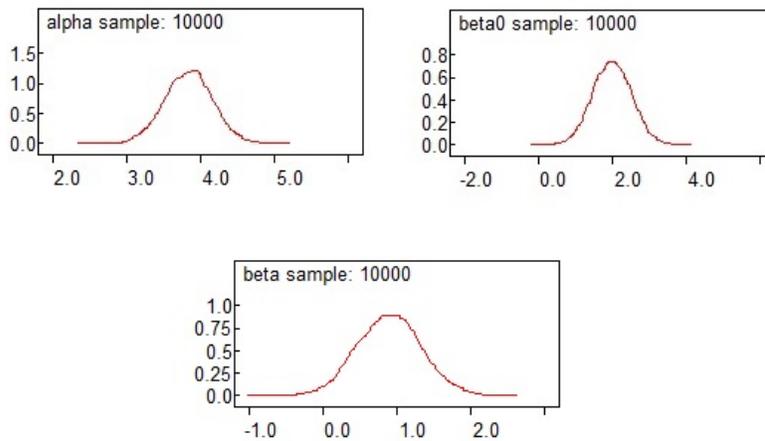


Figura 4.13: Gráficas de las densidades de los coeficientes  $\alpha$ ,  $\beta_0$  y  $\beta_1$ , respectivamente.

En la Figura 4.13 se muestran las gráficas de la densidad para cada uno de los coeficientes estimados. Éstas sugieren un ajuste significativo.

Los resultados obtenidos por el enfoque centrado y el análisis bayesiano (informativo y no informativo) son similares. Tanto el estimador del efecto de tratamiento obtenido por el enfoque centrado como el estimador obtenido a través de métodos bayesianos son muy próximos al valor verdadero del efecto del tratamiento, y sus errores estándar son muy cercanos entre sí. Además el valor verdadero del efecto del tratamiento, 2, está dentro del intervalo de confianza del 95 % y dentro de los intervalos de probabilidad del 95 %, por lo cual se concluye que los resultados obtenidos del enfoque centrado son significativos. En la siguiente sección se resumen los resultados obtenidos de las simulaciones.

#### **4.5.5. Resultados de las simulaciones 2 y 3**

De acuerdo al DRDM empleado para la ilustración de los enfoques frontera y centrado, en la Simulación 2 y 3, se tiene que el efecto del tratamiento acorde al modelo verdadero es de 2.0, ya que es la magnitud de la discontinuidad entre el grupo control y el grupo tratamiento.

El estimador del efecto promedio empleando el enfoque frontera es de 2.2498, mientras el estimador del efecto promedio empleando el enfoque centrado fue de 2.0011. Comparando estos estimadores con el efecto promedio verdadero, se puede decir que son relativamente cercanos a 2. Sin embargo, analizando los errores estándar se tienen los siguientes resultados:

- Empleando el enfoque frontera, el error estándar que se obtuvo de una muestra generada por el método bootstrap paramétrico de los estimadores del efecto promedio del tratamiento fue 0.4415, y de la muestra generada por el método bootstrap no paramétrico, el error estándar fue 0.4768.
- Finalmente, empleando el enfoque centrado se obtuvo un error estándar de 0.5400, el cual surge del ajuste de la nueva variable a considerar en este enfoque.

De los resultados obtenidos de la examinación del enfoque frontera y centrado podemos notar que para este DRDM, el enfoque frontera produce estimaciones más precisas, ya que se obtuvo un error estándar de los estimadores del efecto promedio menor, en comparación con el obtenido con el enfoque centrado. Este error fue obtenido a partir de la muestra generada por el método de bootstrap paramétrico.

# Capítulo 5

## Conclusiones

En este trabajo se definió y analizó el DRD, el cual es un diseño eficaz para la evaluación de programas y tratamientos de intervención social, ya que la mayor ventaja de los DRD es que pueden ser utilizados para estimar los efectos de un tratamiento administrado a los individuos que más lo necesitan o merecen. Como base para el análisis, se discutieron las consideraciones y supuestos que se deben cumplir para implementar un DRD. Apoyados en estos supuestos y consideraciones, se llegó a la conclusión de que, con el fin de interpretar los resultados obtenidos en el análisis de un DRD, se debe conocer la naturaleza de las variables. Sin esta información, no hay un patrón de resultados que indique directamente si un efecto de tratamiento obtenido es positivo o negativo. Por ello, el enfoque bayesiano puede jugar un papel muy importante, ya que cada problema tiene su propio contexto, del cual se deriva la información inicial sobre el parámetro (o cualquier otra característica) de interés.

Cuando se trata de analizar un diseño *fuzzy*, resulta más conveniente analizarlo a través del enfoque bayesiano, debido a las diversas ventajas que tiene el enfoque bayesiano sobre otros enfoques. En este trabajo se discutió cómo aplicar el enfoque bayesiano al análisis de un DRD *fuzzy*, dando los supuestos que se deben cumplir para poder analizar el diseño. Como ilustración, se consideró un DRD *fuzzy* aplicado a la evaluación de becas universitarias italianas, y los resultados sugieren que las becas universitarias italianas actuales son eficaces en reducir la tasa de deserción escolar de los estudiantes que pertenecen a familias cuya medida económica anual es de aproximadamente 15 000 euros.

Finalmente, se discutieron los enfoques de estimación para un Diseño de Regresión Discontinua Multivariado, por medio del cual se pueden proporcionar estimaciones de efectos

múltiples (que corresponde a los posibles contrastes de los tratamientos múltiples, y a los diferentes subgrupos de población en donde se estiman los efectos). Lo anterior lleva a que el investigador se enfrente a una gama más amplia de estrategias para la estimación de los efectos de los tratamientos, tomando en cuenta que la elección entre las diferentes estrategias de estimación tiene implicaciones importantes para la precisión, el sesgo y la generalización. Por ello, es importante conocer la naturaleza del diseño a evaluar, para así elegir una estrategia de estimación adecuada al problema. Los enfoques frontera y centrado resultan ser estrategias de estimación que se pueden generalizar a cualquier DRDM; es decir, el análisis no depende del número de variables puntaje y sirven para calcular cualquier estimador del tratamiento. Para elegir alguna de estas estrategias sólo se debe tomar en cuenta la potencia estadística que tiene cada una. Sin embargo, el enfoque univariado y el enfoque basado en la distancia, son estrategias que resultan eficaces cuando se pretende contrastar dos tratamientos.

Comparando el enfoque frontera con el centrado se obtuvo que el enfoque frontera resultó ser más eficiente, ya que produce estimaciones más precisas. Sin embargo, este enfoque emplea mayor tiempo de cómputo en el software estadístico R.

Cabe mencionar que aún existen varios problemas por resolver en el estudio de los diseños de Regresión Discontinua y que son de gran importancia, pues facilitarían el estudio de diversos programas y tratamientos de intervención social. Uno de estos problemas es la estimación de un DRDM con puntos de corte múltiples, para el cual en la literatura aún no se han especificado estrategias de estimación.

Como consideración final, debe exhortarse a los administradores de programas de intervención social (que desean que sus programas sean responsables de hacer explícitos sus criterios de elegibilidad) a que utilicen diseños cuasi-experimentales tales como el DRD.

# Apéndice A

## Códigos en WinBUGS

- El siguiente código corresponde al análisis bayesiano del modelo del Ejemplo 1.

Ejemplo: analisis bayesiano de un DRD

1. Especificacion del modelo en el lenguaje de BUGS:

```
model;
{
  for( i in 1 : n ) {
    y[i] ~dnorm(mu[i],tau)
  }
  for( i in 1 : n ) {
    mu[i] <- alpha + beta0*group[i]+beta * x[i]
  }
  alpha ~dnorm( 0.0,1.0E-6)
  beta0 ~dnorm( 0.0,1.0E-6)
  beta ~dnorm( 0.0,1.0E-6)
  tau ~dgamma(0.01,0.01)
  sigma2 <- 1 / tau
  for( i in 1 : n ) {
# y.pred[i] ~dnorm(mu[i],tau)
  }
}
```

4.1. Datos

```

list(n=100)
# Los datos estan en RDD.txt
Valores iniciales
list(alpha=0,beta0=0,beta=0,tau=1)

```

- Código correspondiente al análisis bayesiano informativo de la Sección 4.5.4.

Ejemplo: analisis bayesiano informativo de un DRD

1. Especificacion del modelo en el lenguaje de BUGS:

```

model;
{
  for( i in 1 : n ) {
    y[i] ~dnorm(mu[i],tau)
  }
  for( i in 1 : n ) {
    mu[i] <- alpha + beta0*group[i]+beta * x[i]
  }
  alpha ~dnorm( 0.0,1.0E-6)
  beta0 ~dnorm(2,0.2)
  beta ~dnorm( 0.0,1.0E-6)
  tau ~dgamma(0.01,0.01)
  sigma2 <- 1 / tau
  for( i in 1 : n ) {
# y.pred[i] ~dnorm(mu[i],tau)
  }
}

```

4.1. Datos

```

list(n=65)
# Los datos estan en datosEC1.txt
Valores iniciales
list(alpha=0,beta0=0,beta=0,tau=1)

```

# Apéndice B

## Códigos en R

A continuación se presentan algunos códigos del Capítulo 4 en R .

- Código de la Simulación 1

```
###Datos del pretest: x#####
set.seed(5)
  library(mvtnorm)
  N<-1000
  mp<-c(0,0)
  vcp<-matrix(c(0.2,0,0,0.2),2,2)
  R_M<-rmvnorm(N,mp,vcp)
####Puntos de corte: cut#####
rc=mean(R_M[,1])
mc=mean(R_M[,2])
###Variables grupo: group#####
t1<-function(x,cut1,cut2){
g=which(x[,1]<cut1)
g1=which(x[,2][g]>cut2)
  group=array(0,N)
  group[g][g1]=1
  return(group) }
```

```

t3<-function(x,cut1,cut2){
g=which(x[,2]<cut2)
g1=which(x[,1][g]>cut1)
  group=array(0,N)
  group[g][g1]=1
  return(group) }

t2<-function(x,cut1,cut2){
g=which(x[,1]<cut1)
g1=which(x[,2][g]<cut2)
  group=array(0,N)
  group[g][g1]=1
  return(group) }
T1<-t1(R_M,rc,mc)
T2<-t2(R_M,rc,mc)
T3<-t3(R_M,rc,mc)
### Error: e #####
  e=rnorm(N,0,2)
### Variable posttest: y ###
R=R_M[,1]
M=R_M[,2]

Y=4+0.5*R+M+1.2*T1+2*T2+3.4*T3+e
ajuste=lm(Y~R+M+T1+T2+T3)
### Grafica de la superficie de respuesta de los datos observados
yhat = function(R,M,rc,mc){
  if(R < rc & M < mc) Y = 4 + 0.5*R + M+2
# print("Region T2")
  if(R < rc & M > mc) Y = 4 + 0.5*R + M + 1.2
# print("Region T1")
  if(R >= rc & M >= mc) Y = 4 + 0.5*R + M

```

```

# print("Region C")
  if(R > rc & M < mc) Y = 4 + 0.5*R + M + 3.4
# print("Region T3")
  Y
}

Rgrid = seq(min(R),max(R),length=100)
Mgrid = seq(min(M),max(M),length=100)

Ym = matrix(0,100,100)
for(i in 1:100){
  for(j in 1:100){
    Ym[i,j] = yhat(Rgrid[i],Mgrid[j],rc,mc)
  }
}

image(Rgrid,Mgrid,Ym,col=rainbow(50,start=0.9,end=0),xlab="R",ylab="M")
image(Rgrid,Mgrid,Ym,col=gray(100:1 / 100),xlab="R",ylab="M")
legend(0.5,0.8,"tratamiento a",cex=0.8,merge = TRUE, bg = 'gray70')
legend(-1.2,0.8,"tratamiento b",cex=0.8,merge = TRUE, bg = 'gray70')
legend(-1.2,-0.7,"tratamiento c",cex=0.8,merge = TRUE, bg = 'gray80')
legend(0.5,-0.7,"tratamiento d",cex=0.8,merge = TRUE, bg = 'gray70')
persp(Rgrid,Mgrid,Ym,xlab="R",ylab="M",zlab="Y",
theta = 40, phi = 48, r = sqrt(5), d = 1)

```

- Código para generar los datos que se emplean en la Simulación 2 y 3

```

# 1. Datos
set.seed(2)
###Datos del pretest: x#####
library(mvtnorm)
N<-65
mp<-c(0,0)
vc<-matrix(c(0.2,0,0,0.2),2,2)

```

```

###Variables grupo: group#####
t1<-function(x,cut1,cut2){
  g=which(x[,1]<cut1)
  g1=which(x[,2][g]>cut2)
  group=array(0,N)
  group[g][g1]=1
  return(group) }

t3<-function(x,cut1,cut2){
  g=which(x[,2]<cut2)
  g1=which(x[,1][g]>cut1)
  group=array(0,N)
  group[g][g1]=1
  return(group) }

t2<-function(x,cut1,cut2){
  g=which(x[,1]<cut1)
  g1=which(x[,2][g]<cut2)
  group=array(0,N)
  group[g][g1]=1
  return(group) }
R_M<-rmvnorm(N,mp,vcp)
####Puntos de corte: cut#####
rc=mean(R_M[,1])
mc=mean(R_M[,2])
T1<-t1(R_M,rc,mc)
T2<-t2(R_M,rc,mc)
T3<-t3(R_M,rc,mc)
T=cbind(T1,T2,T3)
##### Error: e #####
e=rnorm(N,0,1)

```

```

###Variable posttest: y ###
R=R_M[,1]
M=R_M[,2]
Y = 4 + 0.5*R + 1.0*M + 2.0*T1 + 2.0*T2 + 2.0*T3 + e

```

- Código de la Simulación 2

```

#2. ENFOQUE FRONTERA #####
### Estimacion de la superficie de respuesta:
    efectos de tratamiento constante###
ajuste.1<-lm(Y~R+M+T1+T2+T3)
Y_hat=function(r,m,t1,t2,t3){
ajuste.1$coefficient[1]+ajuste.1$coefficient[2]*r
+ajuste.1$coefficient[3]*m+ajuste.1$coefficient[4]*t1
+ajuste.1$coefficient[5]*t2+ajuste.1$coefficient[6]*t3}
#### Estimacion de la densidad de R y M #####
library(ks)
H<-Hpi(x=cbind(R,M))

##### Pesos #####
a=array()
b=array()
a1=array()
b1=array()

##### se calcula la regla del trapecio para calcular la integral###
d=abs((mc-(max(M)+0.5))/(100-1))
m=seq(mc,max(M)+0.5,length.out=100)
for(i in 1:length(m)){
a[i]=kde(x=cbind(R,M), H=H, eval.points=c(rc,m[i]))$estimate
}
for(i in 1:(length(m)-1))b[i]=d*(a[i+1]+a[i])/2
int=sum(b)

```

```

d1=abs((rc-(max(R)+0.5))/(100-1))
r=seq(rc,max(R)+0.5,length.out=100)
for(i in 1:length(r)){
a1[i]=kde(x=cbind(R,M), H=H, eval.points=c(r[i],mc))$estimate
}
for(i in 1:(length(r)-1)) b1[i]=d1*(a1[i+1]+a1[i])/2
int1=sum(b1)

##### Pesos: wR y wM #####
wR.1=int/(int+int1)
wM.1=int1/(int+int1)

##### Efectos TR y TM #####
A=array()
A1=array()
gr_hat=function(m){
  #evaluando g en la frontera FR
  if(m >= mc) g=Y_hat(rc,m,1,0,0)-Y_hat(rc,m,0,0,0)
  return(g) }
gm_hat=function(r){
  #evaluando g en la frontera FM
  if(r >= rc) g=Y_hat(r,mc,0,0,1)-Y_hat(r,mc,0,0,0)
  return(g) }

for(i in 1:(length(m)-1)){
A[i]=gr_hat(m[i])}
int_p=sum(A*b)
for(i in 1:(length(r)-1)){
A1[i]=gm_hat(r[i])}
int1_p=sum(b1*A1)
TR.1=int_p/int    #se genera una muestra de
                  #tamano N=65 para los efectos estimados del tratamiento

```

```

TM.1=int1_p/int1
print(j)
#### EFECTO DEL TRATAMIENTO UTILIZANDO EL ENFOQUE FRONTERA####
T_MRD.1=wR.1*TR.1+wM.1*TM.1

## RESULTADOS DEL ENFOQUE FRONTERA
T_MRD.1
summary(ajuste.1)

```

- Código de la Simulación 3

```

#1. ENFOQUE CENTRADO ##
Rz=R-rc
Mz=M-mc
T=cbind(T1,T2,T3)
r=array()
for(i in 1:N)
  r[i]=sum(T[i,])
reg_cont=which(r==0)
reg_trat=which(r==1)
### Variable Puntaje:Z ###
z1=array()
z2=array()
Z=array()
for(i in 1:N){
  z1[i]=min(Rz[i],Mz[i])
  z2[i]=max(Rz[i],Mz[i])
Z[reg_cont]=z1[reg_cont]
Z[reg_trat]=z2[reg_trat]
### Variable grupo: group ###
g=which(Z>0)
group=array(0,N)
group[g]=1

```

```

x1=Z[which(group==0)]
x2=Z[g]
y1=Y[which(group==0)]
y2=Y[g]

##### Ajuste #####
ajus<-lm(Y~Z+group)
summary(ajus)
par(mfrow = c(1, 1))

##### Grafica #####
plot(Z,Y,abline(reg=ajus,lwd=2,col=4),xlab="pre-test",ylab="post-test")
points(x2,ajus$coeff[1]+ajus$coeff[2]*x2+ajus$coeff[3],type="l",lwd=2,col=3)
abline(v=0,lty=2,lwd=1.9,col=2)
legend(-0.6,8,"Punto de corte",cex=0.8)
legend(0.2,2.5,"Grupo Tratamiento",cex=0.8)
legend(-0.8,5.8,"Grupo Control",cex=0.8)
arrows(0.5,ajus$coeff[1]+ajus$coeff[2]*0.5,0.5,
ajus$coeff[1]+ajus$coeff[2]*0.5+ajus$coeff[3],
angle=30,code=2,lwd=2)
legend(0.1,7,"Discontinuidad",cex=0.8)

## Ajuste del modelo con el paquete rdd##
library(rdd)
rd=RDestimate(Y~Z)
summary(rd)
plot(rd, xlab="Z",Ylab="Y")
abline(v=0,lty=2,lwd=1.9,col=2)
legend(-0.6,6,"Punto de corte",cex=0.8)
legend(0.2,3,"Grupo Tratamiento",cex=0.8)
legend(-0.8,2,"Grupo Control",cex=0.8)

```

```
##### se guardan datos #####  
datos=cbind(Y,group,Z)  
write(t(datos),"datos_EC1.txt",nc=3)  
# FIN: ENFOQUE CENTRADO #
```

# Apéndice C

## Método bootstrap

El enfoque *bootstrap* propuesto por Bradley Efron (1979), proporciona estimaciones del error estadístico, imponiendo escasas restricciones sobre las variables aleatorias analizadas y se establece como un procedimiento de carácter general, independientemente del estadístico considerado. Dicho enfoque ve a la muestra observada como una población. La función de distribución para dicha población es la función de distribución empírica de los datos, y la estimación de los parámetros basados en la muestra observada son tratados como los parámetros del modelo actual. Aunque los resultados simulados basados en el proceso *bootstrap* no representan las propiedades exactas de los estimadores en muestras finitas, sí proveen una aproximación que mejora conforme el tamaño de la muestra observada aumenta.

En años recientes, el procedimiento *bootstrap* ha ganado aceptación en la investigación aplicada por diversas razones. Primero, el *bootstrap* evita la mayoría de las fuertes suposiciones distribucionales requeridas en los tradicionales procedimientos Monte Carlo. Segundo, como en la simulación Monte Carlo, los procesos *bootstrap* pueden ser usados para resolver estimaciones intratables por medios analíticos y problemas de inferencia más allá de la dependencia de las soluciones asintóticas, las cuales podrían llegar a ser muy complicadas en problemas no estándar. Tercero, las aproximaciones *bootstrap* son frecuentemente equivalentes a los resultados asintóticos de primer orden en muestras grandes y en algunos casos podrían dominarlos.

Existen diversos procedimientos para llevar a cabo una simulación *bootstrap*, pero todas ellas pertenecen a dos grandes rubros, los procedimientos paramétricos y los no paramétricos.

A continuación se detallan las ideas básicas de cada uno de estos rubros.

- *No paramétrico*: Sean  $X_i$  variables aleatorias, para  $i = 1, \dots, n$ . Supongamos que

$X_1, \dots, X_n \sim F(x)$  y que deseamos estimar a  $\theta = H(F)$ .

La idea del método es la siguiente:

Para  $r = 1, \dots, B$ .

1. Generar una muestra de variables aleatorias de  $X_1^{(r)}, \dots, X_n^{(r)} \stackrel{i.i.d.}{\sim} F_n(x)$ , donde

$$F_n(x) = \frac{1}{n} \sum_{i=1}^n I(x_i \leq x)$$

denota la función de distribución empírica de la muestra  $x_1, \dots, x_n$ .

2. Calcular  $\hat{\theta}^{(r)} = H(F_n^{(r)})$ , donde  $F_n^{(r)}$  denota la función de distribución empírica de la muestra  $x_1^{(r)}, \dots, x_n^{(r)}$ .

La idea de este método es utilizar la muestra  $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(B)}$  para aproximar la distribución del estimador  $\hat{\theta}$  y de esta manera hacer inferencia sobre  $\theta$ . La suposición básica de este método es que  $F(x) = F_n(x)$ .

- *Paramétrico*: Supongamos ahora que  $X_1, \dots, X_n \sim p(x|\theta)$ .

La idea del método es la siguiente:

Para  $r = 1, \dots, B$ .

1. Generar una muestra de variables aleatorias  $X_1^{(r)}, \dots, X_n^{(r)} \stackrel{i.i.d.}{\sim} p(x|\hat{\theta})$ , donde  $\hat{\theta} = \hat{\theta}(x_1, \dots, x_n)$  es un estimador consistente de  $\theta$ .
2. Calcular  $\hat{\theta}^{(r)} = \hat{\theta}(x_1^{(r)}, \dots, x_n^{(r)})$ .

Como en el caso del *bootstrap* no paramétrico, la muestra  $\hat{\theta}^{(1)}, \dots, \hat{\theta}^{(B)}$  es utilizada para aproximar la distribución del estimador  $\hat{\theta}$  y de esta forma hacer inferencias de  $\theta$ . Este método esencialmente supone que  $\theta = \hat{\theta}$ .

# Bibliografía

- [1] Ahluwalia J., Weisenberger M., Bernard A. & McNagny S. (1996). Changing Physician Prescribing Behavior: A Low-Cost Administrative Policy That Reduced the Use of Brand Name Nonsteroidal Anti-Inflammatory Drugs. *Preventive Medicine*. **25(6)**, 668-672.
- [2] Bernardo J. M. & Smith A. (1994). F. M. *Bayesian Theory*. Chichester: Wiley.
- [3] Berk R. & Rauma D. (1983). Capitalizing on Nonrandom Assignment to Treatments: A Regression Discontinuity Evaluation of a Crime Control Program. *Journal of the American Statistical Association*. **78(381)**, 21-27.
- [4] Campbell D. & Stanley J. (1963). Experimental and quasi-experimental designs for research on teaching. In N. L. Gage (Ed.), “Handbook of research on teaching“. Chicago, IL: Rand McNally. 171-246.
- [5] Cattaneo M. D., Frandsen B., & R. Titiunik. (2015). Randomization Inference in the Regression Discontinuity Design: An Application to the Study of Party Advantages in the U.S. Senate. *Journal of Causal Inference*, forthcoming.
- [6] Collins T., Mott D., Bigelow W. & Zimmerman D. (1997). A Controlled Letter Intervention to Change Prescribing Behavior: Results of a Dual-Targeted Approach. *Health Services Research*. **32(4)**, 471-489.
- [7] Cook T. (2008). “Waiting for Life to Arrive”: A history of the regression-discontinuity design in Psychology, Statistics and Economics. *Journal of Econometrics* . **142**, 636-654.
- [8] Efron B. (1979). Bootstrap Methods: Another Look at the Jackknife. *Ann. Statist.*, **7**, 1-26.

- [9] Farris K., Kirking D., Shimp L. & Opdycke R. (1996). Design and Results of a Group Counter-Detailing DUR Educational Program. *Pharmaceutical Research*. **13(10)**, 1445-1452.
- [10] Finkelstein M., Levin B. & Robbins H. (1996). Clinical and prophylactic trials with assured new treatment for those at greater risk: II. Examples. *American Journal of Public Health*. **86(5)**, 696-705.
- [11] Gribbons B. & Herman J. (1997). True and quasi-experimental designs. *Practical Assessment, Research & Evaluation*, *5(14)*. Retrieved June 14, 2011 from <http://PAREonline.net/getvn.asp?v=5&n=14>.
- [12] Goldberger A. (1972). *Selection Bias in Evaluating Treatment Effects: Some Formal Illustrations*. Discussion Paper No. 123, University of Wisconsin, Madison, Institute for Research on Poverty.
- [13] Gurwitz J., Noonan J. & Soumerai S. (1992). Reducing the Use of H<sub>2</sub>-Receptor Antagonists in the Long-Term-Care Setting. *Journal of the American Geriatrics Society*. **40(4)**, 359-364.
- [14] Gutiérrez-Pea, E. (1998). Análisis Bayesiano de Modelos Jerárquicos Lineales. *Serie Monografías*, Vol. 7, No. 16. IIMAS, UNAM. **3**
- [15] Hyunshik L. & Munk T. (2008). *Using Regression Discontinuity Design for Program Evaluation*. Section on Survey Research methods-JSM.
- [16] Imbens G. & Rubin D. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*. **25**, 305-327.
- [17] Imbens G. & Lemieux T. (2008). Regression discontinuity design: A guide to practice. *Journal Econometrics*. **142(2)**, 615-635.
- [18] Imbens G. & Kalyanaraman K. (2009). Optimal Bandwidth Choice for the regression discontinuity estimator. *NBER Working Paper Series 14726*.
- [19] Lee, H. & Munk, T. (2008). Using Regression Discontinuity Design in Program Evaluation. *Survey Research Methods*. **511**, 1675-1682.

- [20] Li F., Mattei A. & Mealli F. (2014, under review). Evaluating the Effect of University Grants on Student Dropout: Evidence from a Regression Discontinuity Design Using Bayesian Principal Stratification Analysis. *Duke University Department of Statistical Science Discussion Paper*.
- [21] Monane M., Matthias D., Nagle B. & Kelly M. (1998). Improving Prescribing Patterns for the Elderly through an Online Drug Utilization Review Intervention: A System Linking the Physician, Pharmacist, and Computer. *Journal of the American Medical Association*. **280**(14), 1249-1252.
- [22] O'Hagan A. (1994). *Kendall's Advanced Theory of Statistics, Volume 2B, Bayesian Inference*. Edward Arnold, London.
- [23] R Core Team. (2014). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL <http://www.R-project.org>.
- [24] Reardon S. & Robinson J. (2012). Regression discontinuity designs with multiple rating-score variables. *Journal of Research on Educational Effectiveness*. **5**, 83-104.
- [25] Robert C. & Casella G. (1999). *Monte Carlo Statistical Methods*. New York: Springer-Verlag.
- [26] Robinson J. (2008). *Essays on the effectiveness of policies and practices for reducing cognitive gaps between linguistic groups and socioeconomic groups*. Stanford University doctoral dissertation.
- [27] Rubin D. (1977). Assignment to treatment group on the basis of a covariate. *Journal of Educational Statistics*. **2**(1), 1-26.
- [28] Rubin D. (1978). Bayesian inference for causal effects. *The Annals of Statistics*, **6**, 34-58.
- [29] Shadish W., Cook D. & Campbell D. (2002). *Experimental and Quasi-Experimental Designs for Generalized Causal Inference*. New York: Houghton Mifflin Company.

- [30] Thistlethwaite, D. L., & Campbell D. (1960). Regression-Discontinuity Analysis: An Alternative to the Ex Post Facto Experiment. *Journal of Educational Psychology*. **51(6)**, 309-317.
- [31] Trochim W. (1984). *Research design for program evaluation: The regression-discontinuity design*. Beverly Hills, CA: Sage Publications.
- [32] Trochim W. (1990). Regression-discontinuity design in health evaluation. In: Sechrest L, Perrin E, Bunker J (eds.) *Research Methodology: Strengthening Causal Interpretations of Nonexperimental Data*. US Dept. of HHS, Agency for Health Care Policy and Research, Rockville, MD. 119-139.
- [33] Trochim W. (1994). The Regression-Discontinuity Design: An Introduction. *Research Methods Paper Series, Number 1, Thresholds National Research and Training Center on Rehabilitation and Mental Illness, Chicago, IL*.
- [34] Wong V., Steiner P., Cook T. (2012). Analyzing regression-discontinuity designs with multiple assignment variables: A comparative study of four estimation methods. *Journal of Educational and Behavioral Statistics*.
- [35] Zuckerman I., Lee E., Wutoh A., Xue Z., & Stuart B. (2006). Application of Regression-Discontinuity Analysis in Pharmaceutical Health Services Research. *Health Serv Res*. **41(2)**, 550-563