

Lec11: Inferencia Causal II: Experimentos & Obs Studies

Isidoro Garcia Urquieta

2023

Agenda

Evaluación de impacto Experimentos

- ▶ ITT
- ▶ Non compliance (IV)
- ▶ Efectos heterogeneos
- ▶ Experimentos de bajo poder: P-values exactos de Fisher
- ▶ Diferencias en Diferencias

Estudios observacionales

- ▶ Dif-in-Dif Doubly Robust
- ▶ Regresión Discontinua

Medición de impacto

La medición de impacto tiene las siguientes directrices:

- ▶ Se puede hacer comparación de medias.
- ▶ SIEMPRE deben incluir en la medición el diseño. Esto es, incluir efectos fijos por estrato del diseño.

Esto genera que los estimadores pierdan varianza. Adicionalmente, genera que sea más sencillo correr una regresión para medir los impactos.

- ▶ Cuando la variable y métrica de interés es dicotómica, se debe correr un **Modelo de Probabilidad Lineal** y no un logit. La razón es no debemos imponer formas funcionales que generen sesgo. Más aún, es muy poco probable que nos encontremos con $P > 1$, $P < 0$.
- ▶ Podemos agregar controles en la medición. Esto no es necesario pero puede ayudar a la eficiencia. Se recomienda agregar variables pre-tratamiento que salieron desbalanceadas (para rebalancear).

Medición de Experimentos

Vamos a cubrir con mejor detalle la medición de los experimentos de control aleatorio.

- ▶ De donde viene la aleatoriedad en los experimentos?
- ▶ Viene del sampling como en las regresiones?
- ▶ No! viene del set de posibilidades de aleatorización.
- ▶ Veamos primero los errores exactos que vienen de la aleatorización

Distribuciones de la asignación de tratamiento

Imaginen que tenemos un experimento donde:

- ▶ $N=20$: Tamaño total de la muestra
- ▶ $N_t = 10$: Tamaño del grupo de tratamiento
- ▶ $N_c = 10$: Tamaño del grupo de control

Vamos a hacer una asignación aleatoria. De cuantas maneras la podemos hacer?

$$\binom{N}{N_t} = \frac{N!}{N_t(N_c)} = \frac{20!}{10!10!} = 184,756$$

Había 184,756 maneras de asignar el experimento. Al medir el mismo, el investigador debería hacer pruebas de hipótesis comparado el estadístico observado vs los 184,755 casos restantes.

Distribuciones de la asignación de tratamiento

Imaginen que tenemos un experimento donde:

- ▶ $N=20$: Tamaño total de la muestra
- ▶ $N_t = 10$: Tamaño del grupo de tratamiento
- ▶ $N_c = 10$: Tamaño del grupo de control

Vamos a hacer una asignación aleatoria. De cuantas maneras la podemos hacer?

$$\binom{N}{N_t} = \frac{N!}{N_t(N_c)} = \frac{20!}{10!10!} = 184,756$$

Había 184,756 maneras de asignar el experimento. Al medir el mismo, el investigador debería hacer pruebas de hipótesis comparado el estadístico observado vs los 184,755 casos restantes.

Distribución de la asignación de tratamiento

Qué pasa si estratificamos el experimento en 2 bloques (Género): 10 mujeres y 10 hombres:

$$\binom{N_b}{N_{tb}} \binom{N_b}{N_{tb}} = \frac{10!}{N_t(N_c)} = \frac{10!}{5!5!} \frac{10!}{5!5!} = 63,504$$

- ▶ Lesson de diseño: Miren como si estratificamos perdemos variabilidad (que luego se traduce en menos varianza en los estimadores!).
- ▶ Lesson 2: Si sobre estratificamos, podemos perder tantas posibilidades de asignaciones aleatorias que deja de ser aleatorio (poder!).
- ▶ Lesson 3: Bajo una hipótesis de no impacto ($\tau = 0$) podemos sacar todas los posibles $\hat{\tau}$ para medir nuestro experimento.

P-values exactos de Fisher

Para hacer una medición con p-values exactos se tiene que hacer lo siguiente:

1. Formular tu hipótesis nula (Típicamente $H_0 : \tau = 0$)
2. Estimar todos los posibles escenarios de asignación
3. Con la hipótesis, “rellenar” los datos contrafactuales a nivel i
4. Calcular el estadístico de prueba
5. Ver qué tan probable es que el estadístico observado se observe en la distribución bajo la hipótesis nula.

Ejemplo

Tomemos un ejemplo de un experimento de 6 personas que les asignaron tomar una medicina como tratamiento

Potential Outcomes				
Unit	Frecuencia de Tos		Observado	
i	$Y_i(0)$	$Y_i(1)$	T_i	Y_{obs}
1	?	3	1	3
2	?	5	1	5
3	?	0	1	0
4	4	?	0	4
5	0	?	0	0
6	1	?	0	1

Si nuestra hipótesis es que no hay impacto

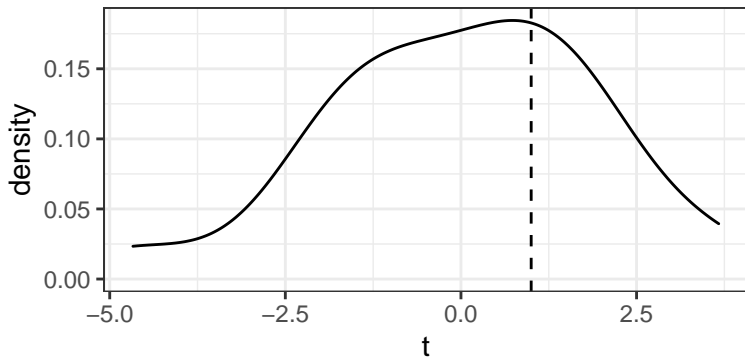
$$H_0 : Y_i(0) = Y_i(1) \quad i = 1, \dots, 6$$

$$\text{Estadístico: } T(T_i, Y_{obs}) = |\bar{Y}_t - \bar{Y}_c|$$

$$\text{Estadístico observado: } T_{obs} = |8/3 - 5/3| = 1$$

Ejemplo (2)

Si ahora vemos los $\binom{6}{3} = 20$ formas de asignar el experimento



16 de 20 casos el estadístico es, en valor absoluto, mayor a 1. Es decir, 80 % de las veces hubieramos observado un valor mayor a 1 bajo la hipótesis nula. No podemos rechazarla con este valor.

Porque no usamos p-values de Fisher siempre?

Uno se podría preguntar por qué no usamos estos errores siempre.




- ▶ Ventajas: Noten como no hago ningún supuesto de normalidad de Y . Además estos son los errores exactos.
- ▶ Desventajas: Si el número de N crece, el número de posibles asignaciones crece de manera dramática. Esto hace que sea casi imposible calcular cada estadístico posible para comparar.
- ▶ En general, cuando N es muy grande, la regresión (con sus supuestos de sampling) arroja errores estándar muy similares a los Exactos de Fisher.
- ▶ Los errores de Fisher siguen siendo muy útiles cuando te enfrentas a un experimento con muy pocas observaciones.

Medición con regresión de un experimento

Ya vimos que con la regresión se puede medir un experimento, si N es suficientemente grande:

- ▶ Siempre hay que tomar en cuenta el diseño en la medición. Esto es, incluir efectos fijos de cada estrato en la regresión
- ▶ Esto es equivalente a hacer una comparación de medias dentro de cada estrato y luego promediar (ponderando) estas diferencias para sacar el ATE.
- ▶ Cuando la Y es una dicotómica, se debe usar una Modelo de Probabilidad Lineal. No debemos imponer formas funcionales.
- ▶ Podemos agregar algunos controles para mejorar la eficiencia (\downarrow varianza) del estimador. Especialmente variables que salieron desbalanceadas.

Ejemplo en R

Impact_eval (RCT)
R Documentation

Impact Evaluation of Treatment Effects

Description

Impact Evaluation of Treatment Effects

Usage

```

impact_eval(
  data,
  endogenous_vars,
  treatment,
  heterogeneous_vars,
  cluster_vars = "id",
  fixed_effect_vars = "g",
  control_vars
)

```

Arguments

data	A data frame, tibble or data table
endogenous_vars	Vector of Ys on which treatment effects will be evaluated
treatment	Variable indicating the treatment status
heterogeneous_vars	Vector of variables for which you wish to assess treatment distributions/heterogeneities
cluster_vars	Vector of variables to cluster the standard errors. Default is without clustered std errors
fixed_effect_vars	Vector of variables to add as fixed effects. Default is without fixed effects
control_vars	Vector of variables to control for in the evaluation. Default is without controls

Details

This function carries out the evaluation of treatment effects on endogenous variables. It automatically runs the regressions of all the endogenous_vars supplied & all the combinations of endogenous_vars and heterogeneous_vars. Additionally, the function has the option of include fixed_effects, controls and cluster variables for clustered std errors.

Value

impact_eval() returns a list of regression tables. The names of the list are the same as the endogenous variables. for heterogeneities the names are endogenous_var_heterogeneous_var

Examples

```

data <- data.frame(y_1 = rnorm(n = 100, mean = 100, sd = 15),
  y_2 = rnorm(n = 100, mean = 8, sd = 2),
  treat = rep(c(0,1,2,3), each = 25),
  heterogeneous_var1 = rep(c("X_01", "X_02", "X_03", "X_04"), times = 25),
  cluster_var1 = rep(c(1:5), times = 25),
  fixed_effect_var1 = rep(c(1,2), times = 50),
  control_var1 = rnorm(n = 100, mean = 20, sd = 1))

evaluation<-impact_eval(data = data,
  endogenous_vars = c("y_1", "y_2"),
  treatment = "treat",
  heterogeneous_vars = c("heterogeneous_var1"),
  cluster_vars = c("cluster_var1"), fixed_effect_vars = c("fixed_effect_var1"),
  control_vars = c("control_var1"))

```

Package ACT version 1.4
[Index](#)

Efectos Heterógeneos

Los efectos heterógeneos son una parte muy importante del análisis experimental pues nos permiten ver el **Conditional Average Treatment Effect (CATE)**.

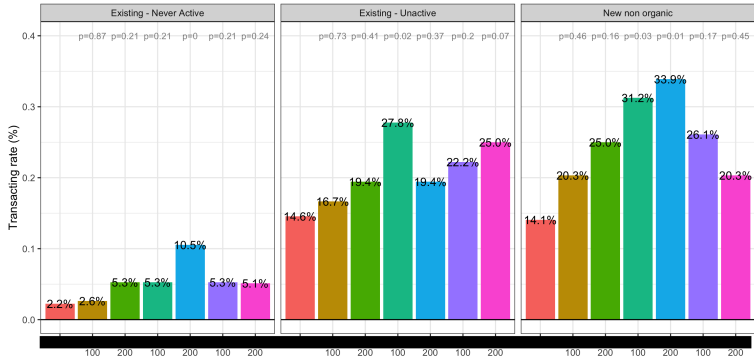
Esto es muy importante pues nos permite evaluar preguntas muy interesantes de policy o business que apuntan a **personalización/targeting**:

- ▶ Debo darle incentivos monetarios a todos los usuarios para que abran la cuenta? A cuáles segmentos?
- ▶ La condicionalidad de llevar a los hijos a chequeos clínico tiene el mismo impacto para jefes del hogar hombres o mujeres?

Noten como el targeting apunta a mejorar el impacto de la intervención: Si focalizo mi intervención, puedo excluir grupos con impactos de tratamiento bajos y sólo enfocarme en grupos donde el impacto es alto. Esto genera un concepto muy útil llamado **Gain**.

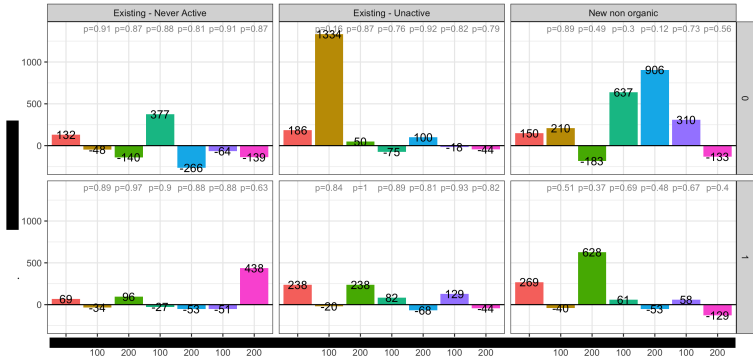
Ejemplo: HTE (1 sola interacción)

Qué segmento deposita o tradea cuando recibe efectivo?



Ejemplo: HTE (doble interaccion)

Si vemos por segmento y por genero?



Variables a analizar HTE

En general, las variables sobre las que se quiere evaluar impactos heterógeneos deben ser declaradas previo al experimento. Es decir, deben ser variables que usamos para estratificar la asignación del mismo.

- Evaluar sobre variables a posteriori puede ser algo fishy. No está prohibido; pero se debe corroborar antes el balance en pre-treatment variables e limitarlo a preguntas muy interesantes.

Si podemos hacer dobles interacciones o incluso triples, pero empezamos a caer en dos problemas conocidos: Falta de poder y evaluación de pruebas múltiples! Hay que hacerlas de manera seleccionada.

Variaciones del ejemplo base

Muchos periodos

Podemos usar Dif in Dif si el experimento corrió varios periodos y tenemos los datos pre-tratamiento. En principio, las diferencias previas entre grupos de tratamiento y control debería ser cero. Cómo podemos *asegurar*lo?

- ▶ El Dif-Dif (En su TWFE form) te permite evaluar impactos dinámicos
- ▶ Incluir la variable endógena para varios periodos como estratificador es super recomendable.

Participación Parcial

Algo muy común en los experimentos es la participación parcial. Esto es, personas asignadas al tratamiento no cumplan con su asignación inicial.

- ▶ Las Variables instrumentales son el go-to teórico. No obstante, veremos que tiene limitantes serias en la práctica.
- ▶ Análisis del estimador Treatment on the Treated (ToT) suelen ser mejor alternativa.

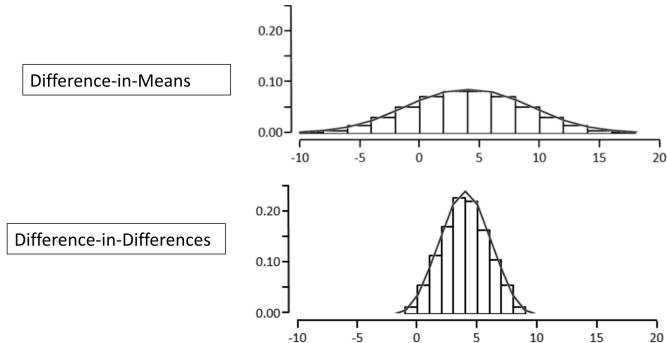
Muchos periodos: Dif-in-Dif

El método de diferencias en diferencias puede ser utilizado en experimentación.

- ▶ Si un experimento duró varios periodos y tenemos observaciones previas al tratamiento, podemos usar diferencias en diferencias para mejorar (\downarrow varianza) la estimación de ATE.
- ▶ Piensen en lo siguiente: Como los grupos fueron asignados aleatoriamente, a “fuerzas” tienen tendencias paralelas en los periodos previos a la intervención.
- ▶ Más aún, el estimador va a restar cualquier desbalance (aunque no significativo) de las observaciones previas.
- ▶ En la práctica, esto cambia ligeramente los estimadores de impacto. Pero no debería de ser un número significativo para el negocio.

Muchos periodos: Dif-in-Dif II

En general, el dif in dif no debería de generar cambios en el estimador de impacto. Su principal ganancia es la disminución de la varianza de los estimadores:



Es importante tener la variable endógena como estratificador para garantizar esto! (Control for what you can, randomize what you can't).

Dif-in-Dif: Estrategia de identificación

El método de diferencias es muy popular. Esta es su estrategia de identificación causal:

- ▶ No asumimos nada de la función de asignación $P(treat)$ (Ojo, si vemos que se cumpla SUTVA)
- ▶ Encontramos un grupo, que intentamos sea lo más parecido al grupo de tratamiento, que no haya recibido el tratamiento y que se haya evolucionado de la misma manera en el tiempo. (Overlap!)
- ▶ Esto es, que tenga **tendencia paralela**. (Unconfoundness!)
- ▶ Con esto, podemos inferir el $ATE = [\bar{Y}_t - \bar{Y}_c | durante = 1] - [\bar{Y}_t - \bar{Y}_c | durante = 0]$. Es decir, la diferencia post menos las diferencias previas.
- ▶ Típicamente, esto se estima con:

$$y = \beta_0 + X\beta + \delta * T + \beta_d * durante + \delta_{dif} durante * T$$

- ▶ Donde δ_{dif} es el estimador de diferencias en diferencias

Diferencias en Diferencias (2)

Cómo evaluamos si hay tendencias paralelas en las observaciones previas a la intervención:

- ▶ Típicamente se corre una prueba placebo.

Imaginense que observan $t = 1, 2, 3, 4, 5, 6$. En $t = 4$ empezó la intervención.

- ▶ La prueba placebo consiste en correr el estimador de diferencias en diferencias ($y = \beta_0 + X\beta + \delta * T + \beta_d * durante + \delta_{dif} durante * T$) para los periodos previos 1, 2, 3 asumiendo que cada uno de ellos es donde sucedió la intervención.

$$durante_j = 1\{t \geq j\} \quad j = 1, 2$$

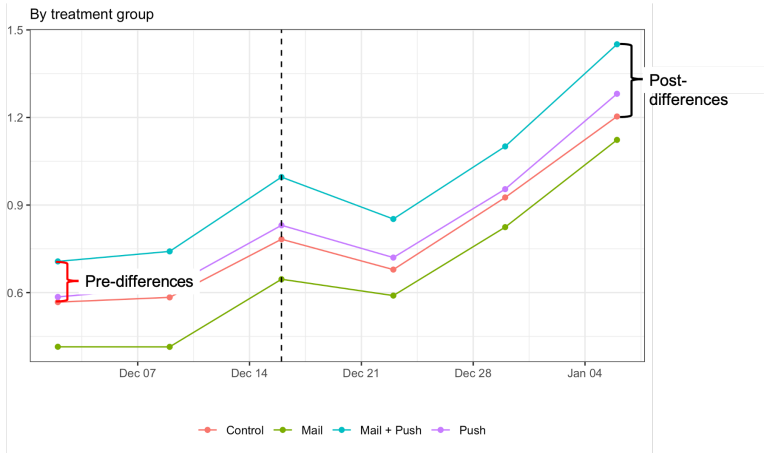
- ▶ Si hay tendencias paralelas, el estimador de diferencias en diferencias en periodos previos debería ser pequeño y no significativo.

Ejemplo

Noten como las diferencias previas son no significativas (experimento bien asignado). Aún así, el dif-dif limpia -3 pp y -2pp de diferencias previas.

term	estimate	std.error	statistic	p.value
factor(treat_final)Mail	-0.03	0.04	-0.78	0.43
factor(treat_final)Mail + Push	0.00	0.02	0.01	1.00
factor(treat_final)Push	-0.02	0.02	-1.04	0.30
durante	0.32	0.02	17.70	0.00
durante:treat_finalMail	0.06	0.05	1.31	0.19
durante:treat_finalMail + Push	0.06	0.02	2.92	0.00
durante:treat_finalPush	0.03	0.02	1.48	0.14

Ejemplo II



Dif in Dif con RCT

Para incluir dif in dif en la librería de RCT, sólo tenemos que hacer:

```
library(RCT)
dif_dif<-impact_eval(data = data,
                      endogenous_vars = "y_1",
                      treatment = "treat",
                      control_vars = c('durante',
                                      "durante:treat"),
                      fixed_effect_vars = 'strata')
```

One-sided Non-compliance en Experimentos

El experimento pudo tener participación parcial de un lado (Non-defiers):

- ▶ Imaginen que lanzan un experimento donde envían SMS a los usuarios para ahorrar de manera aleatoria
- ▶ El vector que define que mensaje recibía el usuario es T_i (asignado aleatoriamente)
- ▶ El vector que define quien vio el mensaje cuando lo recibió es G_i

Cómo medimos el impacto del experimento? Tenemos dos opciones:

- ▶ **Intend to Treat (ITT):** Medimos $Y_i = \alpha + \tau T_i$. Esto nos arrojará el impacto de la intención de tratar sobre la Y_i de manera insesgada (dado que sigue la asignación aleatoria).
- ▶ **Treatment on the Treated:** El impacto de los que efectivamente vieron el mensaje sobre Y_i

Intend to Treat (ITT)

Este estimador es insesgado por definición: Mide el impacto de la asignación inicial (aleatoria). Sin embargo, puede no ser el estimador relevante.

- ▶ Imaginemos que mandamos el mensaje a 5 personas y sólo 2 lo vieron. Guardamos 5 personas en el grupo de control.
- ▶ Tenemos que $T_i = (0, 0, 0, 0, 0, 1, 1, 1, 1, 1)$ pero $G_i = (0, 0, 0, 0, 0, 1, 0, 0, 0, 1)$
- ▶ Imaginemos que las personas que ven el mensaje incrementan Y_i en 5 unidades $Y_{i|G_i=0} = 17$. Los que no se quedan con el valor del grupo control $Y_{i|G_i=0} = 12$
- ▶ Cuando midamos el impacto con $\bar{Y}_{T_i=1} - \bar{Y}_{T_i=0} = 14 - 12 = 2$.
- ▶ Noten como el ITT nos dice que el impacto fue 2 cuando en realidad es 5 (para los tratados). Estoy promediando impactos de $(5, 0, 0, 0, 5)$. Así, subestimo el impacto el experimento sobre los tratados.

Intend to Treat (ITT)

Para correr el ITT estimator, lo único que tengo que hacer es correr la regresión de los outcomes vs la asignación de tratamiento:

$$Y_i = \alpha_s + \tau T_i + \epsilon_i$$

Donde τ es el ITT estimator.

Treatment on the Treated (ToT)

La otra alternativa es intentar estimar el impacto de los usuarios que fueron efectivamente tratados.

Tenemos 2 opciones:

- ▶ Usar Variables Instrumentales para estimar el Local Average Treatment Effect (LATE) sobre los compliers.
- ▶ Usar G_i (effective treatment) en lugar de T_i (As per treated analysis)

Veamos las ventajas y desventajas de uno y otro

Variables Instrumentales

Las variables instrumentales son otro método popular de inferencia causal:

- ▶ Buscas estimar $Y_i = \beta_0 + \tau T_i + \epsilon_i$
- ▶ Existe una variable G_i (instrumento) que está **MUY** correlacionada con el tratamiento de interés T_i (Relevancia).
- ▶ Esta variable además **no está correlacionada** con la métrica de interés Y_i . (Exogeneidad/Unconfoundness). Formalmente esto implica $Cov(U_i, Z_i)$ que el instrumento no tiene correlación con ninguna otra variable no incluida en la regresión.
- ▶ La varianza de T_i explicada por G_i es causal!

$$T_i = \gamma_0 + \gamma_1 G_i + v_i \quad 1st \quad Stage$$

$$Y_i = \eta_0 + \eta_1 G_i + \omega_i \quad 2nd \quad Stage$$

$$\hat{\tau} = \frac{\hat{\eta}_1}{\hat{\gamma}_1}$$

Variables Instrumentales II

La anterior muestra una forma de estimar IV que se llama Two-Stage Least Squares:

- ▶ Estimamos el First Stage
- ▶ Predicimos los scores del first stage (toda la varianza exógena de T_i)
- ▶ Usamos esta predicción \hat{T}_i y la metemos a la regresión $Y_i = \beta_0 + \tau \hat{T}_i + \epsilon_i$
- ▶ $\hat{\tau}$ es el estimador IV 2sls. Noten como es lo mismo que dividir los dos coeficientes del first stage y second stage

Formalmente, usamos una corrección de los errores de libertad para obtener los errores estándar exactos.

Variables Instrumentales III

Probemos que lo anterior es correcto:

$$Y_i = \beta_0 + \tau T_i + \epsilon_i$$

$$\text{cov}(Y_i, G_i) = \text{cov}(\beta_0 + \tau T_i + \epsilon_i, G_i)$$

$$\text{cov}(Y_i, G_i) = \text{cov}(\beta_0, G_i) + \text{cov}(\tau T_i, G_i) + \text{cov}(u_i, G_i)$$

$$\text{cov}(Y_i, G_i) = \tau \text{cov}(T_i, G_i)$$

$$\tau = \frac{\text{cov}(Y_i, G_i)}{\text{cov}(T_i, G_i)} = \frac{\eta_1}{\gamma_1}$$

Notas en IV en Experimentos

El impacto encontrado es para los “compliers”.

- ▶ Los compliers son los usuarios que de recibir T_i , se tratan efectivamente
- ▶ Este impacto excluye a los Never takers: Usuarios que no tienen el cambio en el comportamiento sin importar si reciben el tratamiento o no.
- ▶ Igualmente excluye a los Always takers: Usuarios que siempre observan el cambio en su comportamiento
- ▶ Por definición se asume que no habrá defiers

Como se les ocurre traducir esto a experimentos digitales en negocios?

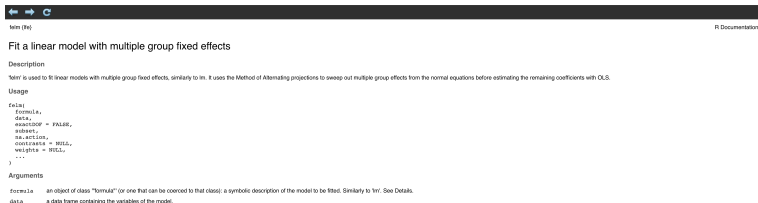
Notas en IV en Experimentos II: Supuestos

Recordemos que necesitamos que el instrumento sea **Relevante** y **Exógeno**:

- ▶ La exogeneidad es muy fácil de justificar. Dado que el tratamiento fue asignado aleatoriamente, es natural que sea independiente a todas las variables
- ▶ Necesitas un instrumento **fuerte** en la primera etapa para que el estimador sea bueno (i.e. La primera etapa que tenga un $F - statistic \geq 10$). Más aún, necesitas que la participación sea alta (i.e. $P(G_i) \sim T_i \rightarrow 1$); de lo contrario el instrumento será débil.
- ▶ Esto último hace complicado usar IV en experimentos digitales. Como la participación parcial es alta, el IV estimator tiende a tener muchísima varianza.

Instrumentos en R

La mejor librería para esto y varias cosas de econometría tipo panel es lfe



lfe (R)

R Documentation

Fit a linear model with multiple group fixed effects

Description

'lfe' is used to fit linear models with multiple group fixed effects, similarly to lm. It uses the Method of Alternating projections to sweep out multiple group effects from the normal equations before estimating the remaining coefficients with OLS.

Usage

```
lfe(lm,
    formula,
    data,
    exactDCP = FALSE,
    subset,
    na.action,
    contrasts = NULL,
    weights = NULL,
    ...)
```

Arguments

formula an object of class "formula" (or one that can be coerced to that class): a symbolic description of the model to be fitted. Similarly to 'lm'. See Details.

data a data frame containing the variables of the model.

Instrumentos en R (2)

La mejor librería para esto y varias cosas de econometría tipo panel es lfe



Details

This formula is intended for use with large datasets with multiple group effects of large cardinality. If dummy-encoding the group effects results in a manageable number of coefficients, you are probably better off by using [lm](#).

The formula specification is a response variable followed by a four part formula. The first part consists of ordinary covariates, the second part consists of factors to be projected out. The third part is an IV-specification. The fourth part is a cluster specification for the standard errors. I.e. something like $y \sim x1 + x2 \mid z1 + z2 \mid (Q1W - x3W4) \mid c1u1 + c1u2$ where y is the response, $x1, x2$ are ordinary covariates, $z1, z2$ are factors to be projected out, $Q1$ and W are covariates which are instrumented by $c1$ and $x4$, and $c1u1, c1u2$ are factors to be used for computing cluster robust standard errors. Parts that are not used should be specified as 0, except if it's at the end of the formula, where they can be omitted. The parentheses are needed in the third part since $|$ has higher precedence than \sim . Multiple left hand sides like $y \mid w1x - x1 + x2 \mid z1 + z2 \dots$ are allowed.

Interactions between a covariate and a factor (as can be projected out with the syntax $x \mid x:f$) are not possible with things like $y \sim x1 + x \mid x:f$, rather one would specify $y \sim x1 + x \mid x:f + f$. Note that $f \sim x$ also works, since R's parser does not keep the order. This means that in interactions, the factor must be on the right, whereas a non-instrumented factor will be coerced to a factor. I.e. in $y \sim x1 \mid x:f1 + f2$, the $f1$ must be a factor, whereas it will work as expected if $f2$ is an integer vector.

In older versions of the formula you would specify $y \sim x1 + x2 + Q1(z1 - x3W4) \mid c1u1 + c1u2$. This syntax still works, but yields a warning. Users are strongly encouraged to change to the new multi-part formula syntax. The old syntax will be removed at a later time.

The standard errors are adjusted for the reduced degrees of freedom coming from the dummies which are implicitly present. (An exception occurs in the case of clustered standard errors and, specifically, where clusters are nested within fixed effects; see [here](#).) In the case of two factors, the exact number of implicit dummies is easy to compute. If there are more factors, the number of dummies is estimated by assuming there's one reference-level for each factor, this may be a slight over-estimation, leading to slightly too large standard errors. Setting `exactDOF="on"` computes the exact degrees of freedom with `exactlyDOF()` in package `Matrix`.

For the W -part of the formula, it is only necessary to include the instruments on the right hand side. The other explanatory covariates, from the first and second part of `formula`, are added automatically in the first stage regression. See the examples.

The `clusters` argument is similar to the one in `lm()`, it is used for factors in the first part of the formula. The factors in the second part are analyzed as part of a possible subsequent `getFe()` call.

The `clustered` argument may affect the clustered covariance matrix (and thus regressor standard errors), either directly or via adjustments to a degrees of freedom scaling factor. In particular, Cameron, Gelbach and Miller (CGMM11, sec. 2.3) describe two possible small cluster corrections that are relevant in the case of multilevel clustering.

- The first approach adjusts each component of the cluster-robust variance estimator (CRVE) by its own c adjustment factor. For example, the first component (with G clusters) is adjusted by $c_1 = G/(G-1)/(N-1)/(N-K)$, the second component (with H clusters) is adjusted by $c_2 = H/(H-1)/(N-K)$, etc.
- The second approach applies the same adjustment to all CRVE components: $c = G/(G-1)/(H-1)/(N-K)$ in the case of two-way clustering, for example.

Any differences resulting from these two approaches are likely to be minor, and they will obviously yield exactly the same results when there is only one cluster dimension. SB, CGMM11 adopt the former approach in their own paper and simulations. This is also the default method that `fe2se` uses (i.e. `method="cgmm"`). However, the latter approach has since been adopted by several other packages that allow for robust inference with multilevel clustering. This includes the popular `Stata` package `xtspline`, as well as the `TwoStageModelKit` implementation in Julia. To match results from these packages exactly, use `method="cgmm2"` (or its alias, `method="correlated"`). It is possible that some residual differences may still remain; see discussion [here](#).

The old syntax with a single part formula with the `Q()` syntax for the factors to transform away is still supported, as well as the `clustervec` and `iv` arguments, but users are encouraged to move to the new multi-part formulas as described here. The `clustervec` and `iv` arguments have been moved to the `...` argument list. They will be removed in some future update.

Value

`fe2se` returns an object of class `"fe2se"`. It is quite similar to an `"lm"` object, but not entirely compatible.

The generic `summary()` method will yield a summary which may be printed. The object has some resemblance to an `"lm"` object, and some postprocessing methods designed for `lm` may happen to work. It may however be necessary to coerce the object to succeed with this.

The `"fe2se"` object is a list containing the following fields:

<code>coefficients</code>	a numerical vector. The estimated coefficients.
<code>n</code>	an integer. The number of observations
<code>p</code>	an integer. The total number of coefficients, including those projected out.
<code>response</code>	a numerical vector. The response vector.
<code>fitted.values</code>	a numerical vector. The fitted values.
<code>residuals</code>	a numerical vector. The residuals of the full system, with dummies. For IV-estimations, this is the residuals when the original endogenous variables are used, not their predictions from the 1st stage.
<code>r.residuals</code>	a numerical vector. Reduced residuals, i.e. the residuals resulting from predicting without the dummies.
<code>iv.residuals</code>	numerical vector. When using instrumental variables, residuals from 2. stage, i.e. when predicting with the predicted endogenous variables from the 1st stage.
<code>weights</code>	numeric. The square root of the argument <code>weights</code> .
<code>efactor</code>	factor of length N . The factor describing the correlated components of the two first terms in the second part of the model formula.
<code>vcv</code>	a matrix. The variance-covariance matrix.
<code>fe</code>	list of factors. A list of the terms in the second part of the model formula.
<code>stage1</code>	The <code>"fe2se"</code> objects for the IV 1st stage, if used. The 1st stage has multiple left hand sides if there are more than one instrumented variable.
<code>ivstage1st</code>	list of numerical vectors. For IV 1st stage. F -value for excluded instruments, the number of parameters in restricted model and in the unrestricted model.
<code>x</code>	matrix. The expanded data matrix, i.e. from the first part of the formula. To save memory with large datasets, it is only included if <code>fe2se\$keepCV=TRUE</code> is specified. Must be included if <code>bootstrap</code> or <code>permutation</code> is to be used for correcting limited mobility bias.
<code>ex</code>	matrix. The expanded data matrix. Only included if <code>fe2se\$keepCV=TRUE</code> .

As per Treated Analysis

Que hacemos si tenemos un experimento donde la participación parcial fue alta?

- ▶ El ITT diluye demasiado el impacto; generando dificultad en observar impactos
- ▶ La asignación de tratamiento es un instrumento débil, lo que genera que el IV estimator tenga mucha varianza y números sin sentido (γ_1 es muy pequeño. $\tau = \eta_1/\gamma_1$ se hace gigante)

Idealmente, lo mejor sería tener un **placebo** en el grupo de control. Esto es, darles un tratamiento inocuo

- ▶ En medicina, darle mentas en lugar de medicamento
- ▶ En business, un mensaje simple sin CTA

Placebo identified analysis

Con el grupo placebo, uno puede identificar a los compliers de ambos grupos.

1. Filtrar a los controles que no se tomaron el placebo/ no recibieron el mensaje/etc.
 2. Filtrar los usuarios tratados que se tomaron el medicamento/ no recibieron el mensaje/etc.
 3. Realizar la evaluación de impacto comparando ambos compliers
- Previo corroborar buen Unconfoundness de participación en ambos grupos!

As Per Treated Analysis

La otra opción final es, sino tengo placebo group, es filtrar a los usuarios tratados del análisis. Esto puede ser problemático y se debe hacer con cuidado:

- ▶ Típicamente la no participación es NO aleatoria; lo que indica cierto sesgo de selección.

Por ende, es recomendable optar por

- ▶ Un análisis de regresión con controles en desbalance, y/o
- ▶ Un diferencias en diferencias. En la práctica, estos grupos tienden a tener tendencias paralelas. El Dif-in-Dif nos ayuda a eliminar las diferencias de nivel

De cualquier manera, conforme se va bajando en la calidad de los análisis, el Data Scientist debe reconocer la certeza/robustez de los resultados

Métodos cuasiexperimentales

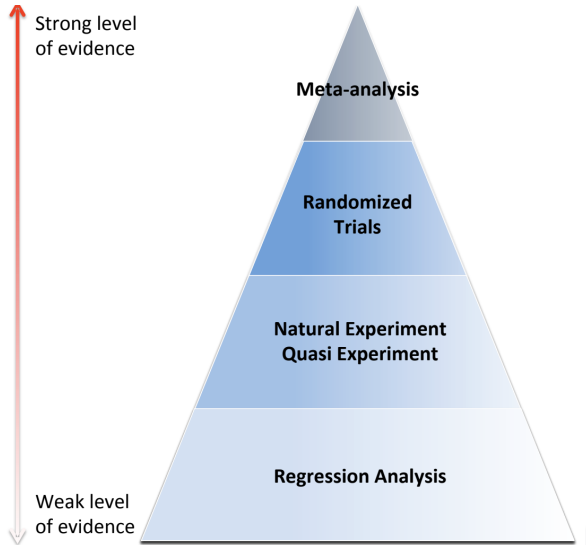
Los métodos cuasiexperimentales son aquellos en los que el Data Scientist no controla la función de asignación de tratamientos PERO tiene mucho conocimiento de ella.

- ▶ Todos los métodos cuasiexperimentales usan este conocimiento para llegar a comparativas (as good as random) en las que se pueda inferir causalidad entre las variables

Hoy vamos a ver 3:

- ▶ Ajuste por covariables
- ▶ Diferencias en Diferencias Doubly Robust y Event Studies
- ▶ Regresión Discontinua

Métodos cuasiexperimentales II



Métodos cuasiexperimentales III

Recordemos las 3 características de asignación de tratamiento y cómo cambian en cuasiexperimentación:

1. **Asignación Individual (SUTVA):** Se mantiene.
2. **Asignación Probabilística (Overlap):** Las distribuciones de las covariables (X) son lo suficientemente cercanas entre los grupos de tratamiento.
 - Cuando las distribuciones difieren mucho, no podemos usar métodos de econometría clásicos. Debemos intentar inferir causalidad mediante extrapolación para las observaciones extremas y excluirlas del diseño cuasiexperimental.
3. **Unconfoundness (Conditional Ignorability):** Vamos a intentar construir mecanismos que nos lleven a esta comparaciones as good as random. *No es comprobable*

Ajuste por covariables

En principio, si tenemos SUTVA, podemos ajustar regresores causales con regresión:

$$y_i = X\beta + \epsilon_i$$

- ▶ SUTVA implica que la β que nos interesa depende sólo de las observaciones de cada i
- ▶ Overlap se puede alcanzar controlando por muchos covariables X
- ▶ En este caso el problema es el unconfoundness (exogeneidad). Sólo llegaremos a un estimador causal si y sólo si observamos cada variable que este correlacionada con la variable de interés y con la métrica y_i . Esto sería un modelo de tratamientos estructural
- ▶ En un mundo de Big Data (siguiente clase!), nos estamos acercando mucho más a la posibilidad de estimar estos modelos.

Doubly Robust Dif-in-Dif

Si bien les explique el estimador de Dif-in-Dif más sencillo, en la actualidad hay estimadores mucho más sofisticados. Estos generalizan al dif-in-dif tradicional en varias dimensiones:

- ▶ El tratamiento sucedió en varios periodos: Tenemos muchos puntos de before and after.
- ▶ El tratamiento fue lanzado en distintos periodos (staggering)
- ▶ Evalúan de manera más robusta las tendencias paralelas para grupos “not treated yet” o “never treated”.

Doubly Robust Dif-in-Dif

Assumption 1: El tratamiento es irreversible:

$$D_{t-1} = 1 \rightarrow D_t = 1 \text{ for } t \in \{t, \infty\}$$

Assumption 2: Los grupos G se definen por el periodo en el cuál la persona fue expuesta al tratamiento por primera vez

$$G = 1\{\text{group treated time } t\}$$

La Probabilidad de ser tratado (Overlap!) en el periodo t para el grupo t está definida como:

$$p_{g,t}(X) = p(G_g = 1|X)$$

Doubly Robust Dif-in-Dif

Los potential outcomes están definidos como:

$$Y_{i,t} = Y_{i,t}(0) + \sum_{g=2}^{\tau} (Y_{i,t}(g) - Y_{i,t}(0)) \cdot G_{i,g}$$

Donde:

- ▶ $Y_{i,t}(0)$ es el outcome en los periodos donde la persona no fue expuesta al tratamiento
- ▶ $Y_{i,t}(g)$ es el outcome cuando la persona empezó a ser expuesta

En un setup con 2 periodos (before-after), el ATT es (y para los tratados - y para los tratados en ese mismo periodo):

$$ATT = E[Y_2(2) - Y_2(0) | G_2 = 1]$$

Generalizando para muchos periodos:

$$ATT(g, t) = E[Y_t(g) - Y_t(0) | G_t = 1]$$

Doubly Robust Dif-in-Dif

Hay dos grupos de comparación posibles:

Never treated group (dif-dif “clásicos”)

- ▶ Tiene que tener tendencias paralelas

$$E[Y_t(0) - Y_{t-1}(0)|X, G_g = 1] = E[Y_t(0) - Y_{t-1}(0)|X, G_g = 0, C = 0]$$

- ▶ Tiene que ser suficientemente similar (OVERLAP!)

Eventually treated (Staggered Dif-in-Diff/ Event Studies)

- ▶ Tiene que tener tendencias paralelas $E[Y_t(0) - Y_{t-1}(0)|X, G_g = 1] = E[Y_t(0) - Y_{t-1}(0)|X, G_g = 0, D_s = 0]$

En ambas, debe haber overlap $0 < p(G_g = 1|X) < 1$

Doubly Robust Dif-in-Dif

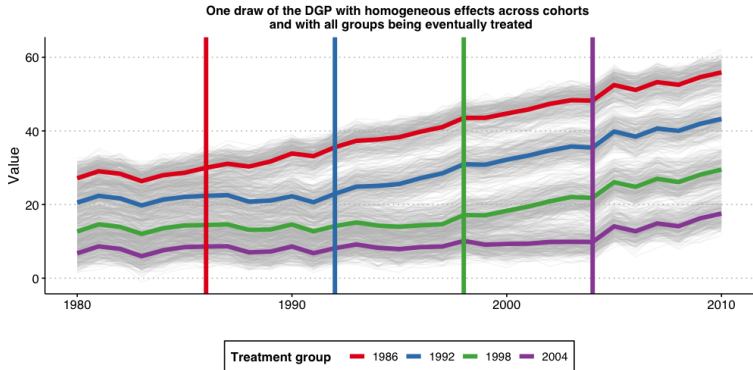
El estimador doble robusto es:

$$ATT_{DR}(g, t; \delta) =$$

$$E \left[\left(\frac{G_g}{E[G_g]} - \frac{\frac{p_{g,t+\delta}(X)(1-D_{t+\delta}(1-G_g))}{1-p_{g,t+\delta}(X)}}{E \left[\frac{p_{g,t+\delta}(X)(1-D_{t+\delta}(1-G_g))}{1-p_{g,t+\delta}(X)} \right]} \right) (Y_t - Y_{g-\delta-1} - m_{g,t,\delta}) \right]$$

DR Dif-in-Dif Ejemplo

Veamos un ejemplo de los autores:

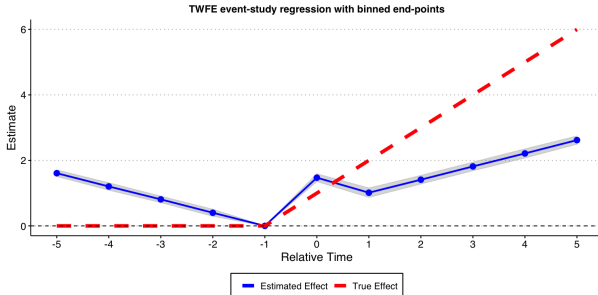


DR Dif-in-Dif Ejemplo

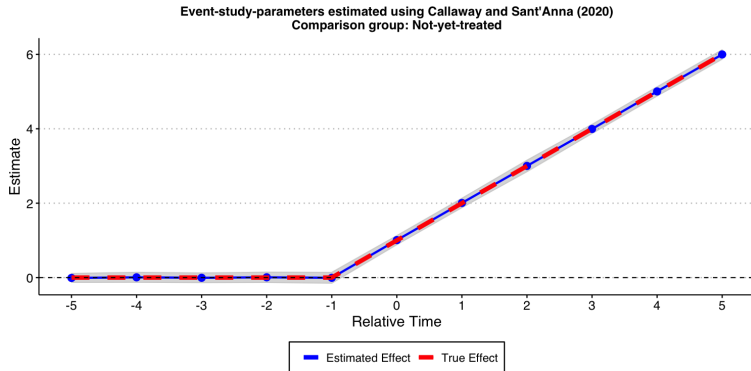
Podríamos usar un Event Study clásico para identificar los efectos de un tratamiento escalonado.

$$t = \text{tiempo} - t_{\text{tratamiento}}$$

$$y_{it} = \alpha_t + X\beta + \sum_{j=-\infty}^{\infty} \tau_j \mathbf{1}\{t = j\}$$



DR Dif-in-Dif Ejemplo



DR Dif-in-Dif Ejemplo R

```
#####  
# Sin controles  
#####  
event_robust<-map(endogenas,  
                   ~att_gt(yname = .,  
                           tname = 'semana_num',  
                           idname = 'user_id',  
                           gname = 'semana_treated',  
                           control_group = 'notyettreated',  
                           allow_unbalanced_panel = T,  
                           data = panel %>% filter(t>=-4)))  
  
save(event_robust, file = 'Modelos/event_winsor_semana_nivel.Rdat')  
  
# Explorando la tabla  
names(event_robust)<-endogenas
```

DR Dif-in-Dif Ejemplo R

```
# Como event study
att_event<-map(event_robust,
               ~aggte(., type = 'dynamic', na.rm = T))

att_event<-map_dfr(att_event, ~tidy())
att_event$variable<-rep(endogenas, each = 8)
att_event<-att_event %>% select(variable, everything()) %>% sele
save(att_event, file = 'Tablas/att_event_winsor_nivel.Rdata')
write.xlsx(att_event, file = 'Tablas/att_event_winsor_nivel.xlsx')
```

Regresión Discontinua

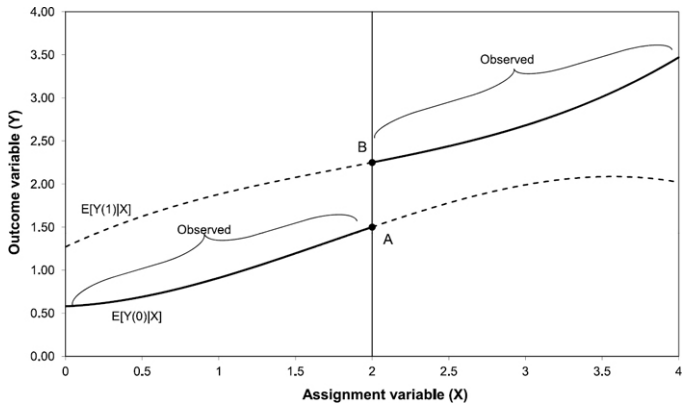
Este es otro diseño cuasi-experimental muy utilizado en policy y en business.

- ▶ La asignación al tratamiento es definida por una **forcing variable** que define la pertenencia al tratamiento.

$$T_i = 1\{\textit{forcing} \geq k\}$$

- ▶ Ejemplos: Score de pobreza en Prospera, Score de elasticidad para asignar descuentos en tasa, reglas de edad para programas infantiles, etc.
- ▶ El mecanismo para llegar a unconfoundness es: Para las unidades muy cercanas al corte en ambos lados, estar de un lado o el otro fue casi aleatorio. Por ende las comparaciones entre ellas se pueden interpretar como causales.

Regresión Discontinua II



Regresión Discontinua: Supuestos

Para estimar la RDD se necesita:

1. Unconfoundness condicional en la forcing variable.
2. Continuidad: Si la Y_i es continua en el corte, entonces el valor observado del otro grupo sirve como contrafactual.
3. No manipulación de la forcing variable: Debemos examinar la densidad de la forcing variable para comprobar que no hay acumulación (bunching) de un lado del corte.

Como se cumplen SUTVA y Overlap aquí?

Estimación de RDD

Estimar los impactos causales con RDD es posible via MCO:

$$y_i = \beta_0 + \beta_1(\textit{forcing} - k) + \tau T_i + \beta_3(\textit{forcing} - k)T_i$$

Para observaciones “cercanas” $|\textit{forcing} - k| \leq k$

- ▶ El primer término es el escalar
- ▶ El segundo término denota la forma funcional de $y(\textit{forcing})$. Este puede ser polinómico.
- ▶ El tercer término es nuestro término de impacto de tratamiento.
- ▶ El cuarto término nos permite identificar si la forma funcional $y(\textit{forcing})$ cambia entre los lados de k

Adicionalmente, típicamente se corren las regresiones con pesos triangulares $w = \max\{0, h - |\textit{forcing}|\}$

Estimación RDD II

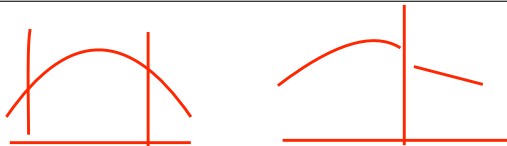
La parte “difícil” en el RDD es seleccionar:

- ▶ h (que tanto es tantito?): La forma óptima de elegirlo es el bandwidth de Imbens-Kalyanaraman. Este intenta encontrar el bandwidth que minimice (en promedio) el sesgo.
- ▶ La forma funcional de $y(\textit{forcing})$: Típicamente se empieza por una regresión lineal. Esto se fundamenta en la aproximación de Taylor. Cualquier $f(\cdot)$ es localmente \sim lineal.

De cualquier manera, casi siempre se corren pruebas de robustez donde:

- ▶ Se incrementa y disminuye h : La idea es ver que tan sensible es tu estimador a personas no tan distintas.
- ▶ Formas polinómicas de $y(\textit{forcing})$: Se intenta ver robustez a polinomios de 2,3 o cuarto grado.

Otras pruebas



Típicamente se corren otras pruebas:

- Densidad en el corte: Se verifica que no haya bunching en la forcing variable. Se estima el McCrary test
- Pruebas placebo: Si corremos la RDD en otro corte k^* , no deberíamos de encontrar impactos en Y . Esto también sirve para evaluar la continuidad
- Pruebas placebo II: Evaluar el RDD sobre variables que no deberían ser afectadas por el tratamiento. No deberías encontrar impacto

Fuzzy RDD

La regresión discontinua tipo Fuzzy también asume que una forcing variable gobierna la asignación de tratamiento, sólo que esta no es Sharp:

- ▶ El punto del forcing variable marca la elegibilidad a recibir el tratamiento, pero este no es obligatorio.
- ▶ Noten como esto es el setting de variables instrumentales. La regla del corte de la forcing variable instrumenta la participación en el tratamiento.
- ▶ Lo único que le pedimos a la Fuzzy RDD es monotonicidad: No defiers. Es decir, que la probabilidad de participar aumente si cruzas al lado elegible. No hay personas que participen si no son elegibles.

RDD en R

Ya hay librerías que te ayudan a hacer la prueba de McCrary y calcular el h de IK. En principio con el resto, podrías hacer la estimación con `lm`



The screenshot shows the CRAN page for the 'rdd-package (rdd)'. The title is 'Regression Discontinuity Estimation Package'. The description states it is a 'Regression discontinuity estimation package'. The details section mentions that 'rdd' supports both sharp and fuzzy RDD utilizing the 'AER' package for 2SLS regression under the fuzzy design. The author is Drew Dimmery, with email 'ddim@nyu.edu'. The 'See Also' section lists related packages: 'RDestimates', 'DCemaisiv', 'fbandwidth', 'summary.RDcint.RD', and 'hcrmlsda'. The footer indicates the package version is 0.57 and provides a link to the index.

rdd-package (rdd) R Documentation

Regression Discontinuity Estimation Package

Description
Regression discontinuity estimation package

Details
rdd supports both sharp and fuzzy RDD utilizing the **AER** package for 2SLS regression under the fuzzy design. Local linear regressions are performed to either side of the cutpoint using the Imbens-Kalyanasaraman optimal bandwidth calculation, [fbandwidth](#).

Author(s)
Drew Dimmery ddim@nyu.edu

See Also
[RDestimates](#), [DCemaisiv](#), [fbandwidth](#), [summary.RDcint.RD](#), [hcrmlsda](#)

[Package rdd version 0.57 [Index](#)]