Profesor: Manuel Ferna´ndez

* David Arboleda
* Camilo Arias
* Valentina Daza
* Douglas Newball
* Santiago Torres.

Taller 2

Econometr´ıa Avanzada, 2022-1 Fecha de Entrega: 1 de abril



# Primer ejercicio

El Ministerio de Agricultura (MinAgricultura) implemento´ un experimento aleatorio controlado a gran escala con el fin de evaluar el impacto de tener buenas pra´cticas de cultivo y cosecha sobre la productividad de los predios agr´ıcolas. El tratamiento asociado al experimento consistio´ en la entrega de dos sesiones informativas sobre las buenas pra´cticas de cultivo y cosecha, uso del fertilizantes, y aprovechamiento de semillas. El tratamiento inclu´ıa la entrega de un folleto que resum´ıa la informacio´n dada en las sesiones. Dada la naturaleza del experimento, el tratamiento fue asignado de forma aleatoria para poder medir su efectividad. Todos los predios que fueron asignados a la capacitacio´n la recibieron efectivamente. No obstante, con el fin de evitar resultados contaminados por sesgos experimentales, y, al mismo tiempo, identificar su presencia, la aleatorizacio´n del tratamiento fue hecha en dos etapas. En una primera etapa, el MinAgricultura dividio´ los municipios en tres grupos: Grupos 1, 2 y 3. En una segunda etapa, al interior de cada grupo, se dividieron los predios de la siguiente forma:

**Grupo 1:** Tratamiento aleatorizado a nivel de municipio. La mitad de los municipios de este grupo fueron asignados al tratamiento y la otra mitad al grupo de control. Todos los predios de los municipios tratados participaron del programa.

**Grupo 2:** Se realizo´ una asignacio´n aleatoria estratificada por municipio: para cada uno de los municipios de este grupo, la mitad de los predios fue asignada al grupo de tratamiento y la otra mitad al grupo de control.

**Grupo 3:** Ninguno de los predios fue asignado al tratamiento, pero fueron informados de que eran parte de un experimento.

*Nota:* salvo los predios del Grupo 3, ningu´n otro predio fue informado de que era parte de un experimento. Suponga tambi´en que no se filtro´ la informacio´n recibida en las capacitaciones entre predios de municipios distintos. Tampoco se filtro´ la informacio´n que algunos predios estaban siendo parte de un experimento.

Ustedes son contratados como consultores externos para realizar la evaluacio´n de impacto del programa. En particular, a sus empleadores les interesa el impacto del programa sobre la productividad de los predios productores, pues se evalu´a la posibilidad de extender las capacitaciones a otros predios que no hicieron parte del experimento. Para lograr este objetivo, a ustedes les entregan la base de datos “producto predios.dta”. Esta base contiene informacio´n a nivel de predio-municipio con las siguientes variables:

*dpto:* Co´digo de Departamento. *mpio:* Co´digo de Municipio. *predio:* Identificador de predio.

*distancia mpio:* Distancia en kilo´metros desde el predio al centro del municipio al que pertenece.

*calidad tierra:* ´Indice estandarizado de calidad de la tierra.

*productividad LB:* Productividad del predio en l´ınea de base (i.e. antes del experimento) medida en kilogramos cosechados por hecta´rea.

*productividad:* Productividad del predio seis meses despu´es de la finalizacio´n del experimento medida en kilogramos cosechados por hecta´rea. La productividad fue medida en una visita sorpresa no anunciada.

*grupo:* Variable catego´rica que captura el grupo experimental al que fue asignado cada municipio.

*T:* Dico´toma que toma el valor de uno si el predio recibio´ el tratamiento y cero de lo contrario.

Respondan las siguientes preguntas teniendo en cuenta el contexto y los datos disponibles.[1](#_bookmark0)

1El experimento descrito y los datos asociados a este son ficticios. Los resultados obtenidos en este taller no proveen informacio´n alguna sobre si capacitar o no a los duen˜os de predios productores incrementa su productividad o no.

1. Para lograr el objetivo de este estudio, ¿cua´les predios comparar´ıan? Propongan un modelo de regresio´n lineal que les permita recuperar el efecto de inter´es. Asegu´rense de describir cada uno de los t´erminos (variables y para´metros) que componen su ecuacio´n a estimar. Mencionen y discutan los supuestos necesarios para que el estimador por MCO sea consistente e insesgado. ¿Son los supuestos plausibles en este contexto?
2. Antes de realizar la estimacio´n, el MinAgricultura les solicita evidencia que soporte la correcta aleatoriza- cio´n de los predios de los grupos 1 y 2. Para esto, presenten:
   1. Una tabla de balance muestral para el grupo 1.
   2. Una tabla de balance muestral para el grupo 2.

Deben tener en cuenta el tipo de aleatorizacio´n que se hizo (e.g.,si fue estratificada) y si deben ajustar por errores esta´ndar tipo clu´ster. A partir de las tablas, argumenten: ¿hay evidencia de que la aleatorizacio´n se hizo de forma correcta? ¿Para qu´e grupos parece cumplirse el balance en l´ınea base?

*Pista:* Para hacer pruebas de balance muestral condicionales, pueden recurrir al concepto de [*partialling out*](https://www.dropbox.com/s/grsard9s9u8st0t/Partial%20Regression.pdf?dl=0)

o pruebas de diferencias de medias condicionales.

1. Sus empleadores esta´n preocupados por el cumplimiento del supuesto de SUTVA. Por esta razo´n, les solicitan presentar evidencia que les permita abogar a favor o en contra de su validez.
   1. Establezcan una relacio´n matema´tica (en t´erminos de los resultados potenciales) que represente el cumplimiento del supuesto SUTVA en la muestra.
   2. A partir de su respuesta en el inciso I, describan un experimento que les permitir´ıa verificar SUTVA.
   3. Describan una comparacio´n o ecuacio´n a estimar basada en los datos disponibles que les permita verificar el supuesto de SUTVA.
   4. Implementen y presenten en una tabla la comparacio´n/metodolog´ıa planteada en el inciso III.

A partir de los resultados encontrados, ¿parece cumplirse el supuesto de SUTVA? En caso de que no se cumpla, ¿para qu´e grupos parece no cumplirse?

1. Por otra parte, ustedes esta´n preocupados por la presencia de sesgos comportamentales producto del experimento que pueden alterar los resultados de su estimacio´n. En este inciso,
   1. Establezcan una relacio´n matema´tica (en t´erminos de los resultados potenciales) que represente la presencia de efecto Hawthorne en la muestra.
   2. A partir de su respuesta en el inciso I, describan el experimento ideal que les permitir´ıa verificar si se da el efecto Hawthorne.
   3. Argumenten a la luz del contexto y el experimento si es razonable que se d´e o no el efecto Hawthor- ne. En caso de que s´ı, ¿para qu´e grupos se dar´ıa?
2. En este inciso debe presentar evidencia a favor o en contra de la presencia del efecto John Henry. Para esto,
   1. Establezcan una relacio´n matema´tica (en t´erminos de los resultados potenciales) que represente la presencia de efecto John Henry en la muestra.
   2. A partir de su respuesta en el inciso I, describan el experimento ideal que les permitir´ıa verificar si se da el efecto John Henry.
   3. Describan una comparacio´n o ecuacio´n a estimar basada en los datos disponibles que les permita verificar si se da el efecto John Henry o no.
   4. Implementen y presenten en una tabla la comparacio´n/metodolog´ıa planteada en el inciso III.

A partir de los resultados encontrados en este y el anterior inciso, ¿parece cumplirse el supuesto de no sesgos experimentales? En caso de que no se cumpla, ¿para qu´e grupos parece no cumplirse?

1. Finalmente, contraste sus respuestas de los incisos *b*) a *d*) con la comparacio´n propuesta en el inciso *a*). ¿Es su comparacio´n va´lida? En caso de que s´ı, estimen el efecto de inter´es a partir de la compara- cio´n propuesta en el inciso *a*). En caso de que no, propongan una nueva comparacio´n o ecuacio´n a estimar que sea va´lida a pesar de los problemas encontrados en los incisos *b*) a *d*). Estimen el efecto de inter´es a partir de esta nueva comparacio´n. Presenten en una tabla sus resultados e interpr´etenlos. Asegu´rense de que sus resultados sean interpretados en t´erminos de significancia estad´ıstica y econo´mica.

# Segundo ejercicio

La idea de que todos los individuos son iguales ante la ley es uno de los principios fundamentales de los sistemas judiciales modernos en pa´ıses democra´ticos. Diferencias raciales, econo´micas y sociales no deber´ıan tener injerencia al momento de juzgar a un individuo. Sin embargo, estudios sobre este tema sugieren que es comu´n encontrar tratos diferenciados en el manejo de los procesos judiciales asociadas al nivel de ingreso de los individuos. Medir efectivamente dichas desigualdades es una tarea importante para poder intentar corregirlas.

En el pueblo de Angristville, donde recientemente han aumentado las protestas por parte de grupos que cuestionan la imparcialidad del sistema judicial, este tema esta´ en el centro del debate pu´blico. Los grupos que protestan aseguran que las personas ma´s pobres de la sociedad son sistema´ticamente desfavorecidas en las cortes, por lo cual suelen tener mayor probabilidad de recibir penas y dichas penas son, generalmente, ma´s severas que las impuestas contra individuos ma´s ricos. El aumento de las tensiones sociales ha llevado a los dirigentes de Angristville a considerar una reforma al sistema judicial, pero desconocen si las protestas son realmente el reflejo de una situacio´n real o corresponden al estado de opinio´n del momento. Por ello, los han contratado a ustedes, asesores del gobierno, para que estudien el tema.

En Angristville, todos los acusados son asignados aleatoriamente a uno de 8 jueces posibles acorde al d´ıa en el que se acude a una audiencia inicial. Sin embargo, cada juez esta´ disponible para el sorteo en so´lo algunos d´ıas y an˜os. En las audiencias iniciales, se define la fianza que debe pagar el acusado a fin de permanecer en libertad durante el curso del juicio. Segu´n la ley, las fianzas son fijadas segu´n la naturaleza y gravedad del cargo presentado. No obstante, algunos abogados han asegurado que el cara´cter del juez asignado juega un papel central en el resultado de dichas audiencias. Au´n as´ı, la sentencia final, en la que se define la culpabilidad o inocencia del acusado, es decisio´n de un jurado conformado por personas seleccionadas aleatoriamente entre la poblacio´n.

Suponga que el Gobierno les provee la base de datos “sentencias.dta”, la cual contiene informacio´n al nivel de caso sobre las sentencias dictadas, el juez encargado, y algunas caracter´ısticas del individuo en cuestio´n como su raza, sexo, historial criminal y si fue o no encarcelado durante el curso del juicio (i.e. si pago´ o no la fianza). Una descripcio´n de las variables se encuentra en el pdf titulado “diccionario de variables p2”:

1. Para empezar, a partir de estad´ısticas descriptivas resumidas en una tabla, analicen las caracter´ısticas de- mogra´ficas de los acusados. Asegu´rense de solucionar problemas evidentes en los datos y sean expl´ıcitos en co´mo los resolvieron.
2. Un mecanismo por medio del cual el nivel de ingreso puede afectar los resultados de los juicios es a trav´es de las fianzas. ¿Creen ustedes que esta hipo´tesis es cierta? ¿Por qu´e?
3. A ra´ız de lo discutido en el punto anterior, a ustedes les interesa conocer el efecto de ser encarcelado previo al juicio (i.e. no poder pagar la fianza fijada) sobre la probabilidad de ser declarado culpable. Un compan˜ero les sugiere desentenderse pronto de este trabajo y simplemente estimar por MCO una regresio´n que incorpore adecuadamente sus variables de inter´es.
   1. Plantee dicho modelo. Mencione qu´e representa cada variable incorporada. ¿Qu´e variables/grupos de variables de su base de datos podr´ıan servir como controles en una posible regresio´n? Justifique. No es necesario que explique una a una por qu´e incluir´ıa cada variable en su modelo, sino que d´e una idea general de por qu´e ser´ıa deseable controlar por ciertas caracter´ısticas.
   2. Estimen el modelo propuesto y presenten sus resultados. ¿Qu´e sugieren los resultados sobre la existencia de sesgos en la imposicio´n de las sentencias?
   3. ¿La anterior estimacio´n identifica el efecto causal de inter´es? Sean expl´ıcitos en los supuestos que plausiblemente se esta´n (in)cumpliendo.
4. Propongan una estrategia de identificacio´n alternativa que le permita atender el problema descrito en el punto anterior (*Pista: recuerden que la probabilidad de ser encarcelado depende de la severidad del juez asignado*).
   1. Expliquen con claridad co´mo implementar´ıan dicha estrategia (i.e. plantee la(s) ecuaciones a estimar), los supuestos que se deben cumplir y justifiquen su (im)plausibilidad. Por ahora, supongan que el efecto del encarcelamiento es igual para todos los individuos. ¿Qu´e efecto esta´n estimando con su nueva estrategia (i.e. ¿qu´e representa el para´metro de inter´es?).
   2. ¿Cua´les ser´ıan los supuestos que necesitar´ıa si el ser encarcelado previo al juicio tuviera efectos distintos para cada individuo? ¿Qu´e tipo de efecto promedio estar´ıan estimando en este caso?
   3. Argumenten a partir de los datos si tienen evidencia que soporte el (in)cumplimiento de alguno(s) de sus supuestos. ¿Qu´e (problemas) ventajas puede representar esto?
   4. Empleando la estrategia propuesta, estimen el modelo y presente sus resultados. Interprete. ¿Qu´e diferencias encuentran con los resultados obtenidos al efectuar la estimacio´n por MCO? ¿Qu´e sugieren los resultados sobre la existencia de sesgos en la imposicio´n de las sentencias?

# Tercer ejercicio

Entre 2014 y 2018, la administracio´n del expresidente Juan Manuel Santos implemento´ el programa Ser Pilo Paga (SPP). El objetivo principal de SPP fue garantizar que los mejores estudiantes del pa´ıs con me- nos recursos econo´micos accedieran a Instituciones de Educacio´n Superior (IES) de alta calidad. SPP se financio´ con fondos pu´blicos. Para los beneficiarios, cubr´ıa el costo total de la matr´ıcula de un programa de pregrado de cuatro o cinco an˜os en cualquier universidad con acreditacio´n de alta calidad en el pa´ıs. La elegibilidad al programa se definio´, entre otros criterios, con un puntaje ma´ximo en el SISBE´N y un puntaje m´ınimo en el examen de salida de la educacio´n media, Saber 11 (el puntaje m´ınimo para ser elegible era de 310).[2](#_bookmark1) Adicionalmente, para acceder a los beneficios los estudiantes ten´ıan que ser admitidos a un programa universitario en una IES con acreditacio´n de alta calidad. Debido a las caracter´ısticas del programa, ustedes utilizara´n la metodolog´ıa de regresio´n discontinua para determinar el impacto de esta iniciativa estatal.

Ustedes esta´n interesados en estudiar el efecto del programa sobre la matr´ıcula en educacio´n superior y la eleccio´n de universidad de los estudiantes (universidad de alta calidad o de baja calidad, universidad pu´blica o privada). Para esto, disponen u´nicamente informacio´n de la primera cohorte de SPP, en la cual los estudiantes desconoc´ıan la existencia del programa hasta que recibieron los resultados de las pruebas del Saber 11. Estos datos se encuentran en el archivo “SPP Base”, que esta´ acompan˜ado del diccionario “diccionario de variables punto3.pdf”, el cual contiene la descripcio´n de las variables de la base de datos.

Para cada inciso, adema´s de contestar las preguntas respectivas, incluya el co´digo utilizado para llegar a la solucio´n.

1. Importe la base de datos. Presente una tabla de estad´ısticas descriptivas para las variables *puntaje saber11, edad, estrato, miembros familia*, donde se reporte el nu´mero de observaciones, el promedio de la variable, su desviacio´n esta´ndar y el m´ınimo y ma´ximo observados. Reporte toda la informacio´n con dos decimales. Esta tabla debe ser generada completemente en Stata/R para su exportacio´n en Word o en Tex.
2. Presente dos gra´ficos que le permitan analizar si hay o no discontinuidad en la probabilidad de (i) ser elegible para el programa y (ii) acceder efectivamente al programa. Para la gra´fica (ii) utilice un polinomio de grado 2 diferenciado a ambos lados del punto de corte. Interprete sus resultados y discuta brevemente la relevancia de encontrar una discontinuidad en el gra´fico (ii).

2La prueba Saber 11 es un examen estandarizado de salida de la educacio´n media. Este examen lo presentan todos los estudiantes que se gradu´an de la educacio´n media independientes de si acceden o no a la educacio´n superior.

1. Con base en el inciso *b*), estime, presente en una sola tabla e interprete los efectos de la elegibilidad de SPP y de ser beneficiario de dicho programa sobre la matr´ıcula en educacio´n superior y la eleccio´n de universidad de los estudiantes (universidad de alta o baja calidad, universidad pu´blica o privada). Deben estimar estas cantidades usando m´etodos no-param´etricos. Para ello, usen el comando rdrobust (quiza requieran instalarlo usando el comando “ssc install rdrobust” en Stata o “install.packages(‘rdrobust’)” en R). Adema´s, utilicen un polinomios de grado 1 y 2, y como ponderador un kernel triangular. Incluyan en la tabla el nu´mero de observaciones disponibles y el taman˜o del ancho de banda usado en cada estimacio´n.

Una estructura sugerida para la tabla es la siguiente:

**Polinomio de grado 1 Polinomio de grado 2**

Matr´ıcula IES alta calidad IES privada Matr´ıcula IES alta calidad IES privada (1) (2) (3) (4) (5) (6)

*Panel A: Elegibilidad*

Elegible para SPP [1=S´ı]

Observaciones Ancho de Banda

Observaciones en banda

*Panel B: Beneficiario del programa*

Beneficiario de SPP [1=S´ı]

Observaciones Ancho de Banda

Observaciones en banda

Luego discutan

* 1. ¿Qu´e tipo efecto de tratamiento se identifica en el panel A y en el Panel B? ¿Cua´l es ma´s grande?¿Por qu´e?
  2. ¿Son los valores encontrados en la columna 1 estad´ısticamente significativos? ¿Son ´estos econo´mi- camente significativos?

**Nota:** Utilicen la opcio´n “all” del comando rdrobust y reporten los coeficientes cuyo sesgo ha sido corregido y cuyos errores esta´ndar son robustos a esta correccio´n.

Para que las metodolog´ıas de regresio´n discontinua identifique el efecto causal del tratamiento se necesitan ciertos supuestos. Si bien ´estos son imposibles de probar, podemos proveer evidencia emp´ırica de su posible cumplimiento. El primero de ellos es el de continuidad local, esto es, que las variables no afectadas por el tratamiento no tienen una discontinuidad en el umbral.

1. Para evaluar el cumplimiento del supuesto, evaluaremos si existe una discontinuidad en el umbral de otras variables que posiblemente determinan el tratamiento. Para ello, presente en una u´nica gra´fica (no tabla) los valores (con sus respectivos intervalos de confianza) resultantes de estimar de manera no-param´etrica una posible discontinuidad en las variables *mujer, edad, minoria, estrato, miembros familia y col privado*.

Una gra´fica sugerida viene del realizar los siguientes pasos:

* 1. Estandarice las variables mencionadas para que los resultados sean comparables.
  2. Utilice el comando *rdrobust* usando como variable de focalizacio´n el puntaje de la Prueba Saber y como variable dependiente cada uno de los confounders. Utilice un polinomio local de grado 1 y un kernel triangular como ponderador.
  3. Guarde en una matriz los coeficientes y los errores esta´ndar de cada estimacio´n.
  4. Muestre en una figura cada uno de los valores estimados junto con sus intervalos de confianza a un nivel de significancia del 95 %. Para ello, puede ser u´til el comando *coefplot*.
  5. Interprete los resultados.

Otro posible test es evaluar que no es posible manipular la regla de asignacio´n. A continuacio´n vamos a implementar dos tipos de tests diferentes para evaluar la veracidad de esta afirmacio´n.

1. Una posible nocio´n de no manipulacio´n es pensar que para puntajes lo suficientemente cerca del umbral, estar a un lado o al otro es casi que aleatorio. Una manera de modelar esta intuicio´n matema´ticamente es pensar que

*Ei* = 1[{Puntaje*i* ≥ 310}]|Puntaje*i* ∈ [310 − *λ,* 310 + *λ*] ∼ Bernoulli(0*.*5)

donde *Ei* es una dummy que indica si un individuo es elegible o no. Esto es, bajo este modelo, caer a un lado u otro del umbral es equivalente al resultado de lanzar una moneda para aquellas observaciones que esta´n a lo sumo a *λ >* 0 puntos del umbral.

Vamos a implementar un test de Monte-Carlo que nos permita poner a prueba esta idea:

* 1. Fije un *λ >* 0.
  2. Para cada *i*, genere *l* = 1*,* · · · *,* 10000 variables de asignacio´n ficticias *D*(*l*) ∼ Bernoulli(0*.*5)

*i*

* 1. Para cada *l*, calcule *p*(*l*):

(*l*)

#{*i* : *D* = 1 & Puntaje ∈ [310*,* 310 + *λ*]}

*p*(*l*) = *i i*

#{*i* : Puntaje*i* ∈ [310 − *λ,* 310 + *λ*]}

* 1. Calcule *p*∗ para la muestra original.

*p*∗ = #{*i* : *Ei* = 1 & Puntaje*i* ∈ [310*,* 310 + *λ*]}

#{*i* : Puntaje*i* ∈ [310 − *λ,* 310 + *λ*]}

* 1. Utilice la distribucio´n aproximada de los *p*(*l*) para estimar cua´n extremo es *p*∗ bajo este modelo. Para ello calcule:

#{*l* : |*p*(*l*) − 0*.*5| *>* |*p*∗ − 0*.*5|}

*χ* =

10000

Calcule *χ* para *λ* = 3. Concluya qu´e significa este resultado a la luz del supuesto de no-manipulacio´n.

1. Otra posibilidad para descartar manipulacio´n, es pensar que no existe una discontinuidad en la funcio´n de densidad de la variable de focalizacio´n en el umbral. Presente un figura en donde estime usando polinomios locales la funcio´n de densidad a cada lado del umbral, y reporte el estad´ıstico *T* de la discontinuidad estimada as´ı como su respectivo p-valor. Concluya qu´e dice esto sobre la validez del supuesto.

**Nota:** Una manera sencilla de hacer eso es usar el comando “rddensity”.

# Punto doctorado

En no ma´s de tres pa´ginas, realice un *referee report* del art´ıculo: [Card, D., (2001) Immigrant Inflows, Native](https://davidcard.berkeley.edu/papers/immig-inflows.pdf) [Outflows, and the Local Market Impacts of Higher Immigration, Journal of Labor Economics, Vol 19, No 1](https://davidcard.berkeley.edu/papers/immig-inflows.pdf) [(Enero., 2001)](https://davidcard.berkeley.edu/papers/immig-inflows.pdf). Sin importar que el art´ıculo se encuentre publicado, debe ser cr´ıtico a la hora de evaluar su contenido, identificando de manera clara los puntos d´ebiles y fuertes de la estrategia emp´ırica.