

## Tarea 3

**Fecha de entrega: Miércoles 15 de mayo**

**Nota:** Deberán subir a *Canvas* **dos archivos**. Un **primer archivo** de texto con sus respuestas. Este archivo debe ser auto-suficiente. Esto quiere decir que con revisar este archivo debe ser posible calificar su tarea en su totalidad. Para este archivo pueden utilizar el formato de su preferencia (e.g. pdf, L<sup>A</sup>T<sub>E</sub>X, Word u hojas escritas a mano y escaneadas). El **segundo archivo** es un archivo de soporte que genere todos sus resultados. Típicamente este será un archivo de programación, mismo que al ejecutarse replicaría los resultados que estás reportando en tu primer archivo. Dado que ustedes son libres de elegir el software que utilizarán, este puede ser un R-script, Do-File o similar.

### Seguro Médico de Oregon

Como parte de tu rol en la Organización Mundial de la Salud y trabajando en colaboración con el ITAM, has sido contratado por el estado de Oregon, que enfrenta una crisis significativa en su sistema de salud pública. Oregon tiene actualmente una lista de espera de 90,000 personas de bajos ingresos que buscan acceder a Medicaid, pero sólo dispone de 30,000 cupos disponibles. Adicionalmente, el estado carece de datos concretos para verificar si el acceso a Medicaid realmente tiene un impacto positivo en la salud autoreportada de los beneficiarios.

Ante esta situación, has propuesto utilizar esta oportunidad para realizar un análisis detallado sobre los efectos del acceso al seguro de salud en la salud autoreportada de los individuos. Para abordar de manera efectiva la sobredemanda y la falta de datos, sugieres la implementación de una lotería como método de selección aleatoria para asignar los cupos disponibles de Medicaid. Este enfoque no solo permite manejar la limitación de cupos de manera justa, sino que también crea un diseño experimental natural para evaluar los impactos del acceso a Medicaid.

El estado de Oregon procedió con esta recomendación y llevó a cabo la lotería, resultando en la selección aleatoria de 29,834 individuos de una lista inicial depurada de casi 75,000 solicitantes (los otros 15,000 solicitantes no eran elegibles para el programa).

Para evaluar el impacto del programa, se organizó una serie de encuestas en tres momentos clave: antes de la implementación del programa, a los seis meses y a los doce meses:

- Encuesta inicial (*oregon\_0.csv*): Se realizó antes de comenzar el programa para recoger datos basales sobre la salud autoreportada de los participantes.
- Encuesta de seis meses (*oregon\_6.csv*): Se centró en una submuestra de aproximadamente 5,000 individuos de la encuesta inicial, proporcionando datos preliminares sobre los efectos a corto plazo de la cobertura de Medicaid.

- Encuesta de doce meses (*oregon\_12.csv*): Adoptó un protocolo más intensivo que incluyó incentivos monetarios para mejorar la tasa de respuesta, aplicada sobre la misma muestra inicial, para evaluar los efectos a más largo plazo del programa.

Para tu análisis, ya tienes acceso a las tres bases de datos correspondientes. Utilizarás estos datos para medir y analizar el efecto de tener acceso a aplicar a Medicaid en la salud autoreportada. Las variables de interés son *needmet\_med\_Xm*, *health\_gen\_bin\_Xm*, *baddays\_phys\_Xm* y *baddays\_ment\_Xm*. Este estudio no solo es crucial para el estado de Oregon, sino que también puede proporcionar evidencia valiosa sobre cómo los programas de salud pública impactan en las poblaciones vulnerables, informando decisiones políticas y prácticas en otros contextos similares.

### 1. [Balances]

- (a) Utilizando solo los datos basales, proporciona evidencia de que la asignación al tratamiento o control parece (o no) haber sido aleatoria. Deberás incluir las tablas y pruebas relevantes, indicar en qué nos debemos de enfocar en dicho análisis y qué concluyes.
- (b) Por limitaciones en el presupuesto disponible, la encuesta de seis meses se realizó sobre una submuestra. La empresa encuestadora no te dijo nada acerca de la forma en que se eligieron a los individuos de esta encuesta. Evalúa si hacer un análisis con esta submuestra no afecta el balance de la asignación al tratamiento inicial (el que evaluaste en la pregunta anterior). Repite este mismo ejercicio con la encuesta realizada 12 meses después de la asignación inicial.
- (c) ¿Consideras que el uso de la muestra de los seis meses afecta la **representatividad** de los resultados en comparación con emplear la encuesta de doce meses que siguió un protocolo de seguimiento más intensivo?

### 2. [Efecto de tratamiento simple]. Empezaremos haciendo un análisis simple en el cual no explotaremos la característica panel de la encuesta. En el análisis de esta parte no debes preocuparte si es que el balance se perdió al usar los datos de la encuesta realizada 12 meses después de la asignación inicial.

- (a) Dada la descripción inicial. Define conceptualmente cuál es el efecto promedio de tratamiento (ATE) y el efecto promedio en los tratados (TOT) en este contexto. Aquí no debes hacer ningún análisis con los datos, sino simplemente describir estos parámetros en el contexto que se describe en este problema. De estos dos parámetros (ATE y TOT), ¿cuál consideras que es más relevante para un policy-maker?
- (b) Utilizando el *estimador de Neyman*, evalúa si la asignación al tratamiento (ganar la lotería de Medicaid) afecta la salud general de los beneficiarios doce meses

después de la asignación inicial (*health\_gen\_bin\_12m*). En esta pregunta deberás aclarar cuál es la prueba de hipótesis que empleas y describe si el resultado es significativo (y a qué nivel).

- (c) Replica el resultado anterior realizando una estimación de OLS simple (sin controles) con errores heterocedásticos.
  - (d) Realiza una estimación del efecto de tratamiento con un OLS agregando los controles que consideres relevantes y errores heterocedásticos. Debes justificar qué controles añadiste.
  - (e) Repite este último cálculo utilizando como variable dependiente el efecto seis meses después de la asignación inicial. Ojo: solo lleva a cabo el estimador con controles y errores heterocedásticos utilizando los mismos controles que en el inciso anterior.
  - (f) Reporta tus resultados en una tabla con el formato que hemos empleado durante el semestre. Describe con palabras los resultados encontrados en los incisos anteriores. Argumenta de alguna manera si el efecto es grande o pequeño indicando claramente cómo le haces para justificar el tamaño del efecto encontrado.
3. **[Primeras diferencias. Dos periodos de tiempo]**. A partir de este inciso empezaremos a utilizar los datos como un panel. Es importante que a partir de esta pregunta deberán de poder cruzar las bases de datos para lo cual utilizaran la variable *person\_id*. Para llevar a cabo el análisis de primeras diferencias tendrás que crear una nueva base de datos partiendo de la base original, como lo vimos en clase.
- (a) Utilizando solo los datos basales y doce meses después, utiliza un modelo de primeras diferencias para estimar el impacto del tratamiento sobre la salud general autoreportada (*health\_gen\_bin*). Incluye una breve interpretación de tus hallazgos. [Tip: en esta pregunta utilizarás solo observaciones de individuos que están tanto en la base de datos basal como en la de doce meses después.]
  - (b) Ahora, utilizando tus datos basales y doce meses después lleva a cabo una tabla de 2x2 para calcular un estimador de diferencias en diferencias simple. Lleva a cabo la tabla y calcula el valor estimado del tratamiento. Notarás que si utilizas todos los datos disponibles, tu valor estimado será distinto al del inciso anterior. ¿A qué se debe eso? [Tip: a diferencia del inciso anterior, en esta pregunta se usan todas las observaciones de los datos basales y de doce meses después de la asignación.]
  - (c) Replica tu resultado del inciso anterior utilizando MCO. Deberás indicar la ecuación que estas estimando y asociar las variables que estás creando con variables que generes en tu código. [Tip: aquí deberás obtener exactamente el mismo resultado del inciso anterior].

- (d) Ahora emplea un modelo de efectos fijos para mejorar tu estimación del inciso anterior. Utiliza efectos fijos a nivel individuo. ¿A qué valor estimado de los anteriores se parece más tu resultado y por qué crees que sucede esto? [Tip: de aquí en adelante tu deberás de decidir qué observaciones utilizar.]
  - (e) En el caso del inciso anterior compara el uso de errores heterocedásticos con el uso de errores cluster a nivel individuo. ¿Por qué crees que sucede esto?
4. **[Primeras diferencias. Tres periodos de tiempo]**. Ahora emplearemos los datos como un panel con tres periodos de tiempo. Ya no es necesario que hagas una nueva base de datos para los estimadores de primeras diferencias (si quieres hacerlo adelante), puedes utilizar un comando de R o Stata para llevar a cabo esto.
- (a) Utilizando todos los periodos de tiempo, lleva a cabo una estimación de primeras diferencias.
  - (b) Ahora haz una tabla de 3x2 utilizando todas las observaciones disponibles para los tres periodos de tiempo. Solo empleando los valores de la tabla argumenta: ¿cuál sería el estimador de TOT relevante?
  - (c) Replica los resultados del inciso anterior utilizando MCO.
  - (d) Compara el tamaño de los efectos estimados en los incisos anteriores con los que calculaste en la sección de primeras diferencias con dos periodos de tiempo. Argumenta: ¿qué tan comparables son estos resultados? ¿Cuál es la manera adecuada de comparar el tamaño de los efectos. Haz un párrafo de menos de 200 palabras.
  - (e) Por último, empleando todas las observaciones disponibles en las tres bases de datos, lleva a cabo una estimación utilizando efectos fijos a nivel individuo.
  - (f) Contrasta el uso de errores heterocedásticos con errores cluster a nivel individuo.
5. **[Reporte de la OMS]**
- (a) Utilizando una tabla con el formato que hemos empleado, reporta los resultados que consideres mas adecuados basado en todos los que estimaste en las secciones anteriores. Tu tabla deberá de tener a lo mas cuatro columnas y deberás considerar si agregas controles o no y qué tipo de errores empleas. Esta tabla será parte de un reporte profesional, así que debes asegurarte que sea presentado de forma que una persona que sepa econometría la pueda entender.
  - (b) Considerando que la OMS es una organización internacional, debes reportar tus resultados de forma muy profesional. Acompaña tu tabla anterior de un breve reporte de menos de 500 palabras donde describas los principales hallazgos obtenidos en tu análisis.

Tabla 1: Descripción de variables  $X_m$ .

Variable	Descripción
<i>person_id</i>	Identificador único para cada persona
<i>treatment</i>	Dummy que indica si el individuo fue seleccionado para el tratamiento. Dummy (0, 1)
<i>birthyear_list</i>	Año en el que nació el individuo
<i>female_list</i>	Dummy que indica si el individuo es mujer por nacimiento
<i>hhinc_cat_Xm</i>	Categoría de ingresos del hogar a los 6 meses. Categórica (1-16) 1 siendo \$0 y 16 \$35001-\$37500
<i>race_hisp_Xm</i>	Identificación como hispano. Dummy (0, 1)
<i>race_white_Xm</i>	Identificación como blanco. Dummy (0, 1)
<i>race_hisp_Xm</i>	Identificación como hispano. Dummy (0, 1)
<i>race_black_Xm</i>	Identificación como negro. Dummy (0, 1)
<i>race_amerindian_Xm</i>	Identificación como indio-americano. Dummy (0, 1)
<i>race_asian_Xm</i>	Identificación como asiático. Dummy (0, 1)
<i>race_pacific_Xm</i>	Identificación como hawaiano nativo o de las islas del pacífico. Dummy (0, 1)
<i>race_pacific_Xm</i>	Identificación como cualquier otra raza no mencionada previamente. Dummy (0, 1)
<i>edu_Xm</i>	Nivel educativo alcanzado 1 = menos que la escuela secundaria, 2 = diploma de escuela secundaria, 3 = títulos profesionales o 2 años, 4= título de 4 años
<i>ins_any_Xm</i>	Si posee algún tipo de seguro a los 6 meses. Dummy (0, 1). 1 = tiene algún seguro 0 = NO tiene algún seguro.
<i>needmet_med_Xm</i>	Es igual a uno si la persona obtuvo todos los servicios médicos necesarios o no necesitó de servicios médicos. Es igual a cero si necesitó servicios médicos pero no los recibió. Dummy (0, 1)
<i>health_gen_bin_Xm</i>	toma un valor de 0 si el encuestado indicó que su salud era excelente, muy buena o buena, y toma un valor de 1 si así lo indicó que la salud era regular o mala.
<i>baddays_phys_Xm</i>	Días con mala salud física en los últimos X meses (0-30)
<i>baddays_ment_Xm</i>	Días con mala salud mental en los últimos X meses (0-30)

El término  $X_m$  representa 0m, 6m y 12m según corresponda a la encuesta inicial, de seis o doce meses después de la intervención.