

Data Challenge

Fecha de entrega: Viernes 27 de noviembre, 11:59 P.M.

Uno de los podcasts que escuchamos este semestre hablaba acerca del trabajo de Raj Chetty, quien parte de su investigación la ha enfocado en analizar la movilidad intergeneracional. En este Data Challenge ustedes llevarán a cabo un análisis de movilidad intergeneracional utilizando datos de México.

Challenge:

Deberán escribir un blog que presente de forma gráfica y con regresiones evidencia acerca de la movilidad intergeneracional en México y si la educación es un factor que ayuda a la movilidad hacia arriba. De forma opcional, ustedes pueden decidir enriquecer su análisis dando evidencia de factores (diferentes a educación) que influyan de forma significativa la movilidad hacia arriba. Su análisis forzosamente deberá utilizar modelos probit, logit y/o probit ordenado. Ustedes deberán definir las variables dependientes a utilizar, pero los análisis de movilidad típicamente utilizan variables dummies que identifican si un individuo está en ciertos intercuantiles.¹

1. ¿Cuál es la idea del Data Challenge?

En los “Data Challenge”, a diferencia de las tareas, les presento una pregunta general que deberán resolver en equipo de la mejor manera posible. Ustedes tendrán que decidir qué variables explicativas elegir, qué regresiones y gráficas presentar, etc. Pese a que estoy disponible para resolver dudas vía correo, los motivo a que mas bien con la información que les proporcioné ustedes tomen sus decisiones.

2. Requisitos

- Su blog deberá ser escrito de forma que pueda ser entendido por público en general (excepto tal vez la tabla de regresiones).
- Ustedes deberán decidir si entregan un archivo de Word, pdf o html.
- Al final del blog deberán incluir dos secciones:
 1. **Apéndice estadístico.** Aquí deberán incluir una descripción con lenguaje técnico de las especificaciones que estimaron. Alguien con experiencia en econometría debería poder replicar sus resultados al leer su apéndice . Deberán

¹Definidos como estar entre el cuantil a y b de la distribución.

incluir detalles de cualquier decisión relevante que hayan tenido que tomar con respecto a sus datos (ej. datos faltantes, de donde obtuvieron su información, etc.) y ecuaciones que describan las regresiones que estimaron, tipo de errores que usaron, etc.

2. **Apéndice de contribuciones.** Aquí deberán describir de manera general en qué contribuyó cada miembro del equipo. No es necesario que todos participen en todo. La idea es que busquen la forma de organizarse de manera eficiente, pero si es importante que todos contribuyan y que todos estén perfectamente enterados de forma tal que todos pudieran contestar preguntas del blog.
- Ustedes son libres de elegir la fuente de datos que quieran. Al final del Data Challenge les describo una posible alternativa de fuente de información, pero ustedes son libres de seguirla o no.
 - No hay un mínimo ni un máximo de número de palabras, cantidad de gráficas o regresiones que deberán tener en el blog.
 - Deberán enviarme la base de datos que hayan formado y el archivo (R-script, Do file, etc.) que replique sus resultados.

3. Equipos

1. Daniel Castañón (dcastao2@itam.mx); Coral Hernández (chern140@itam.mx); Jesús Hernández (jherna39@itam.mx); Mauricio López (mlopez73@itam.mx); Jonathan Vázquez (jvazqu82@itam.mx); Diego Mata (dmatamat@itam.mx); Nayeli Castillo (ncastil8@itam.mx);
2. Marisol Nava (mnavaleo@itam.mx); Luis Fierro (lfierros@itam.mx); Nestor Rodriguez (nrodri10@itam.mx); Cynthia Valdivia (cynthia.valdivia@itam.mx); Daniela Pinto (daniela.pinto@itam.mx); Rafael Sandoval (rsandov3@itam.mx);
3. Luis de la Peña (ldelape3@itam.mx); Karen Hazan (khazanme@itam.mx); Marco Ramos (mramosju@itam.mx); Mónica Contreras (mcontre4@itam.mx); Alexa González (agonz251@itam.mx); Marcelo Torres (ftorre10@itam.mx)
4. Sofía Corona (sofia.corona@itam.mx); David Rojas (jrojasav@itam.mx); Samuel Salazar (gsalaza9@itam.mx); Jorge Herrera (jherre13@itam.mx); Adrián Martínez (adrian.martinez@itam.mx); Fidel González (fgonzal4@itam.mx)
5. Carlos López de la Cerda (carlos.lopezdelacerda@itam.mx); Daniel Fernández (dferna15@itam.mx); Héctor Nuñez (hnuezveg@itam.mx); Daniela Brito (dbritoca@itam.mx); Samantha Cervantez (mcervan5@itam.mx); Alfredo Lefranc (alefranc@itam.mx)

4. Aspectos que se califican

La calificación será compartida entre todos los miembros del equipo, a menos que algún miembro del equipo no haya participado en llevar a cabo el Data Challenge según lo que reporten en su apéndice de contribuciones. Ustedes recibirán un documento con los principales comentarios y deducción de puntos de mi parte. Aspectos que serán relevantes para la calificación:

- Presentación clara y descripción de resultados adecuada de sus hallazgos principales (económicamente hablando). En particular, fijarse en la interpretación de resultados, significancia estadística, claridad en sus tablas y gráficas.
- Análisis econométrico correctamente planteado. Es decir, uso adecuado de la variable dependiente, variables explicativas, aplicación adecuada de modelos probit/logit, decisiones adecuadas respecto a observaciones utilizadas u omitidas.
- Asegúrsense que sus tablas y gráficas sean claras y fáciles de leer. Incluyan notas al pie que ayuden a aclarar el contenido de las gráficas y tablas.
- Descripción adecuada en su apéndice estadístico que me permita replicar el análisis que están presentando utilizando la base de datos y el script que me entreguen. Revisaré que su script se ejecute de forma adecuada y produzca los resultados que están reportando.

5. Una posible fuente de datos

ESRU-EMOVI:

El Centro de Estudios Espinoza Yglesias (CEEY) ha dedicado buena parte de su labor al análisis de la movilidad intergeneracional en México y ha recabado bases de datos que sirven para analizarlo. En su página de internet podrán [descargar](#)² las bases de datos de la Encuesta ESRU-EMOVI 2006, 2011 y 2017.

Para analizar la movilidad intergeneracional, los investigadores del CEEY utilizan datos de las características de los hogares donde los individuos habitaban cuando tenían 14 años de edad como un proxy del ingreso de la generación de sus padres. Ustedes pueden seguir esta perspectiva y utilizar ya sea los datos de ingreso de los individuos o las características de los hogares donde actualmente viven para los ingresos de la nueva generación. Ustedes deberán decidir eso. Asimismo ustedes deberán elegir qué año o años utilizar.

²URL: <https://ceey.org.mx/contenido/que-hacemos/emovi-pre/>